Constellation Optimization for the Non-Linear Fiber Channel

by

Amirhosein Soleimanzade

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Communications

Department of Electrical and Computer Engineering

University of Alberta

Abstract

Due to the exponential growth of Internet usage and inefficiency of the traditional modulations, the capacity of the current optical fiber networks is not sufficient for future Internet demands. As a result, increasing the capacity of the fiber networks is of great importance. Since most of the available bandwidth of fiber systems has been occupied, improving the spectral efficiency (SE) of the fiber channel is an appropriate approach to increase the capacity. Constellation optimization (also known as constellation shaping or constellation design) is an efficient SE enhancement technique, and has two important categories: (1) probabilistic shaping and (2) geometric shaping. Probabilistic shaping changes the uniform distribution of the constellation points into a non-uniform distribution. Geometric shaping relocates the position of the equiprobable constellation points such that the achievable rate increases. In this thesis, we propose two constellation optimization methods. The first method is a geometric shaping method that maximizes mutual information (MI) of the amplitude-phase shift keying (APSK) constellations. We optimize APSK constellations for the additive white Gaussian noise (AWGN) channel and non-linear fiber channel. For the fiber channel, the optimization is performed at the maximum modified signal-to-noise ratio (SNR) of the optical system. By doing so, our optimization algorithm maximizes the MI rate while the impacts of shaping on the non-linear interference noise (NLIN) power are considered. The second shaping method is a hybrid method that combines probabilistic shaping and geometric shaping. Our hybrid method maximizes

the generalized mutual information by considering the impacts of shaping on non-linear interference noise. We show that our hybrid method outperforms both geometrically and probabilistically-shaped constellations.

Preface

The results of Chapter 3 have been submitted to the IEEE Journal of Lightwave Technology under the title "EGN-Based Optimization of the APSK Constellations for the Non-Linear Fiber Channel Based on the Symbol-Wise Mutual Information," and the paper has been accepted for publication.

This work is dedicated to my parents for their love.

Acknowledgements

Firstly, I would like to express my thanks to my supervisor, Dr. Masoud Ardakani, for his continuous support throughout this research project. His guidance, support, and mentorship helped me in all the time of research and writing of this thesis. It was a pleasure that I could research under his supervision. Secondly, I would like to express my gratitude to the committee members, Dr. Yindi Jing, Dr. Ivan Fair, and Dr. Mrinal Mandal, for their insightful comments and for dedicating their time to this thesis.

Contents

List of Tables

List of Figures

List of Abbreviations

AIR	achievable information rate
AWGN	additive white Gaussian noise
ASE	amplified spontaneous emission
APSK	amplitude-phase shift keying
APS	arbitrary probabilistic shaping
BSA	binary switching algorithm
BICM	bit-interleaved coded modulation
BMD	bit-metric decoding
B2S	bit-to-symbol
CUT	channel-under-test
c.c.	complex conjugate
DM	distribution matcher
EGN	enhanced Gaussian-noise
EDFA	erbium-doped fiber amplifier
FEC	forward error correction
FWM	four-wave mixing
GN	Gaussian-noise
GMI	generalized mutual information
GS	geometric shaping
GVD	group velocity dispersion
HPGS	hybrid probabilistic-geometric shaping
IGN-model	incoherent Gaussian-noise model
i.i.d	independent and identically distributed
ISI	inter-symbol interference
MB	Maxwell–Boltzmann
MI	mutual information

NLI	non-linear interference
NLIN	non-linear interference noise
NLSE	non-linear Schrödinger equation
OSNR	optical signal-to-noise ratio
PO	pairwise optimization
PSO	particle swarm optimization
PSK	phase shift keying
PM	polarization multiplexing
PSD	power spectral density
PAS	probabilistic amplitude shaping
PFS	probabilistic fold shaping
PS	particle swarm optimization
PMF	probability mass function
PAM	pulse-amplitude modulation
QAM	quadrature amplitude modulations
RRC	root-raised-cosine
SNR	signal-to-noise ratio
SE	spectral efficiency
SSFM	split-step Fourier method
SMD	symbol-metric decoding
S2B	symbol-to-bit
WDM	wavelength division multiplexing

Chapter 1

Introduction

1.1 Motivation

The Internet has a significant, undeniable impact on the life of humankind. This technology is the most rapid means of communication, enabling lightning-speed-level communication. Moreover, in recent years, it has led to the creation or improvement of a wide variety of industries such as E-learning, digital marketing, media broadcasting, etc. The benefits provided by the Internet would not be possible to achieve without ultra-fast and reliable communication channels.

Fiber optic has attracted much attention in recent years, owing to its capability of handling high-rate communication. Fiber optics has several significant advantages over other communication channels, including ultra-high bandwidth (as high as several THz), low signal loss (as low as 0.2 dB/km), and low electromagnetic interference, making this channel a powerful asset to handle high data rates. Additionally, by employing multiplexing techniques such as polarization multiplexing (PM) and wavelength division multiplexing (WDM), a huge increase in the capacity of the optical fiber networks is achieved [1]. Although the mentioned characteristics make fiber optics an ideal channel for communication, it is prone to major impairments, which degrade the optical system performance.

The main impairment of the fiber channel is its non-linearity behavior, which is directly proportional to the instantaneous power of the optical signal [2]. The non-linear effects are so significant that in non-linear regimes of

fiber, increasing the average launch power decreases the performance rather than improving it. Consequently, in most scenarios, increasing the launch power cannot be considered as a proper solution to the fiber channel capacity enhancement. Moreover, considering the exponential growth of Internet usage, the capacity of the current optical fiber networks would not be sufficient for future demands [3]. Hence, these challenges necessitate applying modern modifications to the optical systems in order to catch up with the growth of Internet usage.

The constellation used in a communication system has a significant impact on the transmission rate. The well-known constellations such as pulse-amplitude modulation (PAM), quadrature amplitude modulations (QAM), and phase shift-keying (PSK) are not capable of providing transmission rates as high as the channel capacity [4], meaning that the traditional constellations are inefficient. Hence, more efficient constellations must be designed, which we discuss in the following.

1.2 Inefficiency of PAM Constellations on the Additive White Gaussian Noise Channel

To measure inefficiency of a constellation, a metric must be employed. The maximum rate of information that can be transferred over a communication channel is one of the main parameters of a communication link, and to compute this parameter, we need to define a metric to measure the information. The mutual information (MI) is one of the well-known tools to measure the statistical dependence between the transmitted information and the received information. Suppose that random variable X is transferred over a noisy channel, and random variable Y is the received information about X, i.e. Y is a noisy observation of X. The mutual information between random variables X and Y is defined as:

$$\mathbb{I}(X;Y) = \mathbb{H}(X) - \mathbb{H}(X|Y), \tag{1.1}$$

where $\mathbb{H}(\cdot)$ is the entropy.

The maximum information rate or capacity of a channel is the maximum

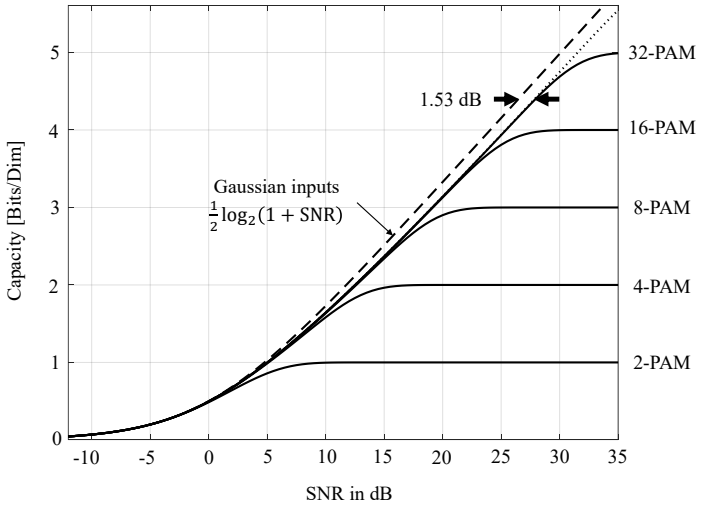

Figure 1.1: Capacity of the ideal AWGN channel with Gaussian inputs and with equiprobable M-PAM inputs.

mutual information between the channel input and output. In the case of an ideal discrete-time additive white Gaussian noise (AWGN) channel, Shannon in [5], [6] shows that the capacity is obtained with a Gaussian distribution over the channel input, and the corresponding capacity in bits per dimension is given by:

$$\mathrm{C} = \frac{1}{2}\log_2(1+\mathrm{SNR})\Big[\frac{\mathrm{Bits}}{\mathrm{Dim}}\Big], \tag{1.2}$$

where SNR is the signal-to-noise ratio.

In practical communication systems, the distribution of the channel input is not Gaussian, which leads to creating a gap between the channel capacity defined in (1.2) and the actual information rate.

M-PAM family of constellations is one of the well-known and widely used constellations in the communication systems, where modulation symbols X_i are chosen from set $\frac{d_0}{2}\{-M+1, -M+3, \cdots, M-1\}$, where d_0 is the minimum distance between symbols. Fig. 1.1 shows the capacity and the mutual information achieved with equiprobable M-PAM, for $M = 2, 4, \cdots, 32$.

According to Fig. 1.1, for low values of SNR, an equiprobable 2-PAM is almost optimal; for high values of SNR, on the other hand, the capacity of

equiprobable M-PAM constellations, before getting saturated, approaches a straight line parallel to the capacity of the AWGN channel. The gap between the capacity of the AWGN channel and the mentioned straight line is $\pi e/6$ (1.53dB) [7], and it is due to not using a Gaussian distribution over the modulation symbols. This deficiency is also true for QAM constellations which are the two-dimensional version of PAMs. To compensate for the 1.53 dB loss, we must optimize constellations to produce a Gaussian-like distribution over the transmitted symbols, which we discuss next.

1.3 Constellation Optimization

Constellation optimization (or constellation shaping) is to find energy-efficient constellations that achieve better performance compared to the traditional QAMs [8]. To optimize a constellation, different objective functions or optimization criteria can be considered. For instance, we can select the achievable information rate (AIR) of a communication system as the optimization criterion or we can select its error rate (such as symbol error rate or bit error rate). In most cases, as it is in this study, AIR is chosen as the objective function, and by maximizing AIR, error-free communication can be achieved at higher transmission rates.

Constellation shaping approaches can be divided into two categories: geometric shaping (GS) and probabilistic shaping (PS). Both GS and PS try to mimic a sampled Gaussian distribution; however, they follow different ideas. The idea behind PS is to change the uniform distribution of the constellation points into a non-uniform distribution. In GS, the position of the equiprobable constellation points is relocated such that the achievable rate increases and approaches the channel capacity [9]. The idea of PS and GS are described in Fig. 1.2. As Fig. 1.2 (b) demonstrates, a Gaussian distribution can be estimated by assuming non-uniform distribution over symbols. Also, in Fig. 1.2 (c), again, the Gaussian distribution is approximated by changing the position of the constellation points.

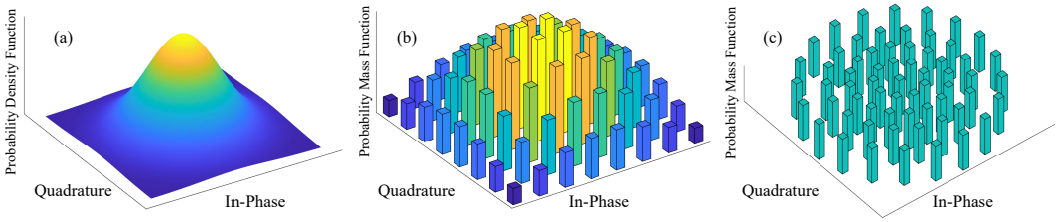

Figure 1.2: Two-dimensional representation of (a) a continuous Gaussian source, (b) a 64-point probabilistic shaped QAM, and (c) a 64-point geometric shaped constellation.

1.4 Constellation Optimization for the Fiber Channel

PS and GS increase the AIR of the AWGN channels significantly; however, in the case of non-linear fiber channel, more considerations are required during the shaping probability or geometry of a constellation. This is because non-linearity of the fiber is a modulation-dependent interference [10]. In this thesis, we introduce two shaping methods for the non-linear fiber channel. The first shaping method is a GS method that maximizes the AIR of the fiber channel. We compare our method with the state-of-the-art GS methods and show that our optimized constellations are capable of outperforming other methods in almost all of the scenarios. Also, we obtain an equation demonstrating that our optimized constellations introduce much lower non-linear interference noise to the system than AWGN-based shaped constellations. The second shaping approach is a hybrid method. In other words, we combine GS and PS to increase the AIR of the fiber channel even further.

1.5 Organization of Thesis

The remainder of this thesis is organized as follows. In Chapter 2, we provide the essential preliminaries. We talk about impairments of the optical fiber communication systems and models of the fiber channel. In addition, we define and formulate constellation shaping approaches. In Chapter 3, we propose and explain our GS method. We optimize the geometry of constellations for

both the AWGN and fiber channel and analyze the results. Also, we compare our method with other GS methods in terms of AIR and the modified SNR of the fiber channel. Our results indicate that our geometric-shaped constellations result in both higher AIR and lower modulation-dependent noise compared to the state-of-the-art GS methods. In Chapter 4, we propose a hybrid probabilistic-geometric shaping method for the non-linear fiber channel. We optimize the position and probability of the points simultaneously to maximize the AIR of the optical system. Our reported AIRs outperform that of both QAMs and probabilistic shaped QAMs in all scenarios. Finally, in Chapter 5, we conclude the thesis and provide some future research directions for our studies.

Chapter 2

Background

2.1 Optical Fiber Communication Systems

In this section, we briefly discuss the characteristics and structure of optical fiber communication systems. First, we explain the structure of the system used in this thesis. After that, we review equations that govern the propagation of optical signals inside the fiber and discuss impairments that result in performance degradation. Finally, we review two well-known discreet memoryless models of the fiber channel, the Gaussian-noise (GN) model and the enhanced Gaussian-noise (EGN) model.

The structure of the fiber link studied in this thesis is provided in Fig. 2.1. We assume that the fiber link between the transmitter and receiver consists of N_s spans whose length is L_s. The link consists of identical spans (the homogenous link assumption). At the end of each span, the loss is exactly compensated for by optical amplification (the transparent link assumption). In the following, we discuss the system of Fig. 2.1 in detail.

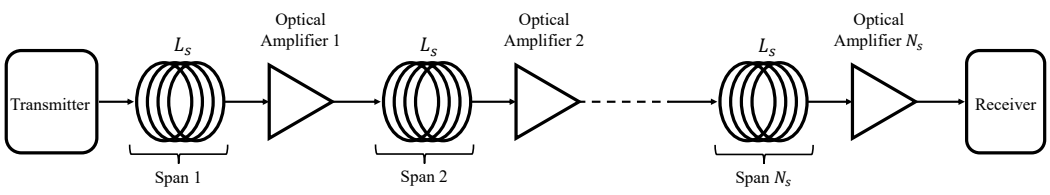

Figure 2.1: Structure of the fiber link

2.1.1 Propagation of Signals Inside Optical Fiber

Since the optical fiber medium behaves similar to a waveguide, propagation of light waves, similar to all electromagnetic fields, is governed by Maxwell's equations [2]:

$$\begin{aligned} \bigtriangledown \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\ \bigtriangledown \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \\ \bigtriangledown \cdot \mathbf{D} &= \rho_f, \\ \bigtriangledown \cdot \mathbf{B} &= 0, \end{aligned} \tag{2.1}$$

where $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$ and $\mathbf{B}$ are the electric field, the magnetic field, the electric flux density, and the magnetic flux density, respectively. The current density vector $\mathbf{J}$ and the charge density ρ_f represent the sources for the electromagnetic field. Since optical fiber is a free-charge medium, we have $\mathbf{J} = \mathbf{0}$ and $\rho_f = 0$.

The flux density $\mathbf{D}$ in (2.1) is related to electric field $\mathbf{E}$ as follows [2]:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \tag{2.2}$$

where ϵ_0 is the vacuum permittivity, and $\mathbf{P}$ is the induced electric polarization. Since optical fiber is a non-linear medium, the induced electric polarization $\mathbf{P}$ can be divided into linear and non-linear terms as follows:

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{P}_{\mathrm{L}}(\mathbf{r}, t) + \mathbf{P}_{\mathrm{NL}}(\mathbf{r}, t), \tag{2.3}$$

where $\mathbf{P}_{\mathrm{L}}(\mathbf{r}, t)$ and $\mathbf{P}_{\mathrm{NL}}(\mathbf{r}, t)$ are linear and non-linear parts, respectively. By substituting (2.3) and (2.2) in (2.1), the pulse propagation equation is achieved. But before that, by considering that the electric field consists of a fast and a slow varying terms (slowly varying envelope approximation), a significant simplifications can be made in the propagation equation. Assuming slowly varying envelope approximation, the electric field can be written as follows:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2}\hat{x}[E(\mathbf{r}, t) \exp(-i\omega_0 t) + \text{c.c.}], \tag{2.4}$$

where $\hat{x}$ is the polarization unit vector, $E(\mathbf{r}, t)$ is a slowly varying function of time, ω_0 is the carrier angular frequency, and c.c. stands for complex conjugate.

In communication systems, we are highly interested in the slow varying term to investigate the impacts of fiber on the pulse shape. Using the separation of variables method, the wave equation for slowly varying component of electric field is obtained as follows (the detailed derivation is provided in [2], Chapter 2):

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + \frac{\alpha}{2}A = i\gamma|A|^2 A, \tag{2.5}$$

where $A(z,t)$ is the slowly varying pulse envelope, β_1 determines the group velocity, β_2 is the group velocity dispersion (GVD) parameter, α is the attenuation constant such that the signal power is attenuated as $\exp(-\alpha z)$, and γ is the Kerr non-linearity coefficient.

Equation (2.5) can be more simplified by defining a new time reference as $T = t - \beta_1 z$. In this new time reference, a frame of reference is moving with the pulse at the group velocity $v_g = \frac{1}{\beta_1}$. By substituting the new time reference in (2.5), we have:

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial T} \times \frac{\partial T}{\partial t} + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial T^2} + \frac{\alpha}{2}A = i\gamma|A|^2 A, \tag{2.6}$$

where

$$\frac{\partial T}{\partial t} = \frac{\partial(t - \beta_1 z)}{\partial t} = 1 - \frac{\beta_1}{\beta_1} = 0. \tag{2.7}$$

Therefore, (2.5) in the new time reference is equal to:

$$\frac{\partial A}{\partial z} + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial T^2} + \frac{\alpha}{2}A = i\gamma|A|^2 A. \tag{2.8}$$

Equation (2.8) is known as the non-linear Schrödinger equation (NLSE) that governs propagation of light inside optical fiber. Now, based on (2.8), we explain GVD and Kerr non-linearity as the impairments of the fiber channel and examine their impacts on the transmitted signal.

Group Velocity Dispersion

Here, to consider impacts of GVD, we assume that the fiber is lossless ($\alpha \approx 0$), and the fiber non-linearity is negligible ($\gamma \approx 0$). Also, we define the normalized time reference τ and normalized amplitude $U(z,\tau)$ as follows:

$$\tau = \frac{T}{T_0}, \tag{2.9}$$

$$A(z,\tau) = \sqrt{P_0}\exp(-\frac{\alpha}{2}z)U(z,\tau), \tag{2.10}$$

where T_0 is the pulse width, and P_0 is the peak power of the pulse. By substituting (2.9) and (2.10) in (2.8) and assuming both γ and α are almost zero, we have:

$$\frac{\partial U}{\partial z} + \frac{i\beta_2}{2T_0^2}\frac{\partial^2 U}{\partial \tau^2} \approx 0. \tag{2.11}$$

Since (2.11) is linear, it is more convenient to solve (2.11) in the frequency domain. In the frequency domain, (2.11) is as follows:

$$\frac{\partial \tilde{U}}{\partial z} - \frac{i\beta_2}{2}\omega^2\tilde{U} = 0, \tag{2.12}$$

where $\tilde{U}$ is the Fourier transform of U. The solution of (2.12) is:

$$\tilde{U}(z,\omega) = \tilde{U}(0,\omega)\exp(\frac{i\beta_2}{2}\omega^2 z). \tag{2.13}$$

As we can see, GVD changes the phase of each frequency component, and the amount phase change depends on both the frequency ω and the distance z. In other words, the more the optical pulse propagates inside the fiber link, the more the phase of the optical pulse changes. Moreover, GVD does not affect the amplitude of the propagating pulse, and new frequency components are not generated during the propagation of the optical pulse. Also, according to (2.13), different frequency components of the optical pulse travel with different velocities inside the optical fiber. Meaning that the more the optical pulse propagates inside the fiber, the more the pulse broadens, which results in considerable inter-symbol interference (ISI). As an example, the impacts of GVD on a root-raised-cosine (RRC) pulse are provided in Fig. 2.2. According to Fig. 2.2, as the fiber length increases, the RRC pulse broadens further.

Kerr Effect

Non-linearity of the optical fiber is one of the most destructive impairments of the fiber. To examine the impacts of this non-linearity on an optical signal, in (2.8), we assume that $\beta_2 \approx 0$. Considering normalized time reference (2.9) and normalized amplitude (2.10), we have:

$$\frac{\partial U}{\partial z} \approx i\gamma P_0 \exp(-\alpha z)|U|^2 U. \tag{2.14}$$

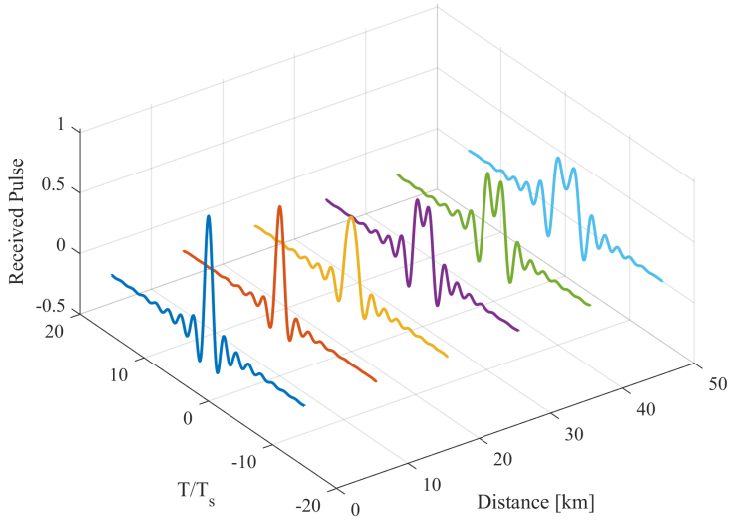

Figure 2.2: Pulse broadening of an RRC pulse due to GVD, note that T_s is symbol time and is equal to 31.25 picoseconds.

Equation (2.14) is a nonlinear differential equation, and it is more convenient to solve in time-domain. To solve (2.14), we assume that:

$$U = V \cdot \exp\left(i\varphi_{\text{NL}}\right). \tag{2.15}$$

By substituting (2.15) in the differential equation of (2.14), we have:

$$\frac{\partial V}{\partial z} = 0, \tag{2.16}$$

$$\frac{\partial \varphi_{\text{NL}}}{\partial z} = \gamma P_0 \exp\left(-\alpha z\right) V^2. \tag{2.17}$$

By solving differential equations of (2.16) and (2.17), $U(z,T)$ is obtained as follows:

$$\begin{gathered} U(z,T) = U(0,T) \cdot \exp\left(i\varphi_{\text{NL}}(z,T)\right), \\ \varphi_{\text{NL}}(z,T) = \gamma P_0 |U(0,T)|^2 \times L_{\text{eff}} \quad , \quad L_{\text{eff}} = \frac{1-\exp\left(-\alpha z\right)}{\alpha}. \end{gathered} \tag{2.18}$$

According to (2.18), the amount of phase shift is intensity-dependent. In other words, the more power is launched to the fiber link, the more phase shift occurs. Also, φ_{NL} increases as the fiber length increases. In Equation (2.18), L_{eff} is an effective length, which is smaller than the fiber length z. The concept

behind the L_{eff} is the loss of power when an optical pulse propagates inside the optical fiber. In other words, since there exists the loss of power when an optical pulse propagates inside the optical fiber, the pulse intensity decreases, which leads to a lower phase shift than the phase shift of the non-attenuated optical pulse.

Another conclusion that can be drawn based on (2.18) is that the amplitude of the optical pulse is not affected by the nonlinear effect. However, due to the time dependence of φ_{NL}, the frequency-domain signal of (2.18) gets broadened as the pulse propagates inside the optical fiber, and new frequency components are generated during the propagation. This is the opposite of what happens in the GVD scenario, i.e. in GVD, the amplitude of the frequency-domain signal is not affected by GVD but the time-domain signal gets broadened; however, when the non-linearity is considered, the time-domain signal does not get broaden but the frequency domain signal does. As an example, Fig. 2.3 indicates the effects of non-linearity on an RRC pulse in time-domain. As seen, the more the RRC pulse propagates inside the fiber, the more phase shift it experiences. Moreover, the low-power points of the RRC pulse are much less affected by the non-linearity than high-power points.

Amplified Spontaneous Emission Noise

As shown in Fig. 2.1, due to the fiber loss, N_s optical amplifiers are placed in the fiber link to compensate for the loss. Similar to other amplifiers, optical amplifiers are not ideal, and they introduce amplification noise to the system, which is called amplified spontaneous emission (ASE) noise. There are several types of optical amplifiers for the fiber channel. In this thesis, the erbium-doped fiber amplifier (EDFA) is employed to compensate for the fiber loss. The ASE noise of EDFA is a circularly symmetric Gaussian noise whose power spectral density (PSD) is as follows [11]:

$$G_{\text{ASE}} = Fh\frac{c}{\lambda}(A_s - 1), \tag{2.19}$$

where F is the EDFA noise figure, h is the Plank's constant, c is the speed of the light, λ is the wavelength, and A_s is the span loss.

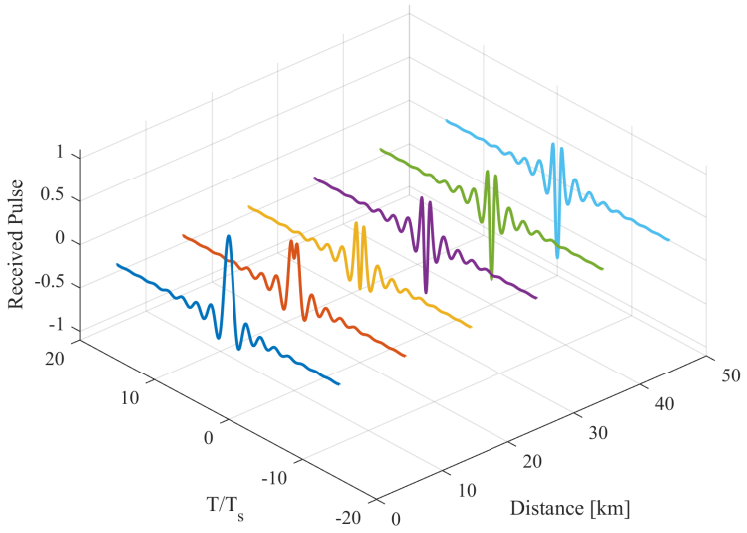

Figure 2.3: Impacts of the fiber non-linearity on an RRC pulse, note that T_s is symbol time and is equal to 31.25 picoseconds.

Multiplexing Techniques

In the previous sections, we discussed propagation of one single pulse (which can be simply extended to a pulse train) inside the optical fiber. However, by employing multiplexing techniques, significant increase in the capacity of the fiber channel can be achieved. Here, we briefly review two of the most widely-used multiplexing techniques in the fiber, wavelength division multiplexing (WDM) and polarization multiplexing (PM).

As mentioned earlier, the bandwidth of the optical fiber is as high as several THz. To efficiently use the whole fiber bandwidth, it is divided into several channels, and within each channel, independent information is transmitted. This multiplexing technique is called WDM. As an example, Fig. 2.4 demonstrates the base-band PSD of a WDM signaling, which consists of nine WDM channels, and withing each WDM channel, RRC pulses with symbol rate of 32 GBaud are transmitted. Also, note that the WDM spacing, the space between adjacent channels, is set to 40 GHz.

As mentioned earlier, since the fiber channel is a non-linear system, new

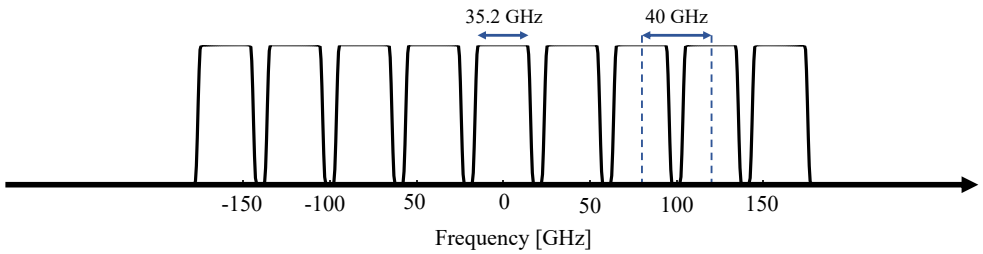

Figure 2.4: PSD of a WDM signaling with RRC pulses whose roll-off is 0.1. The symbol rate is 32 GBaud, and the WDM spacing is 40 GHz.

frequency components are generated as the pulse propagates inside the optical fiber. Consequently, different WDM channels affect each other during the propagation. To analyze the impacts of channels on each other, coupled NLSE is employed [2]. Let A_i be the slowly varying pulse envelope of the i-th WDM channel where $i \in \{1, 2, 3, \cdots, N_{\text{ch}}\}$, and N_{ch} is the total number of WDM channels. According to the coupled NLSE, the propagation equation for A_i becomes [2]:

$$\frac{\partial A_i}{\partial z} + \frac{i\beta_2}{2}\frac{\partial^2 A_i}{\partial T^2} + \frac{\alpha}{2}A_i = i\gamma\Big(\sum_{\substack{1<j<N_{\text{ch}} \\ j\neq i}} 2|A_j|^2 + |A_i|^2\Big)A_i. \tag{2.20}$$

Based on (2.20), the destructive impacts of the other $N_{\text{ch}} - 1$ channels on channel i are two times stronger than the destructive impact of channel i on itself.

In addition to WDM, different polarization of the fiber can carry information. The $\hat{x}$ and $\hat{y}$ components of electric field are orthogonal and can be used for transmitting information simultaneously, which doubles the spectral efficiency of each WDM channel. This multiplexing technique is called PM. In dual-polarized systems, the Manakov equation [12] is used to simulate propagation of optical pulses instead of (2.5):

$$\begin{aligned}
\frac{\partial A_x}{\partial z} + \frac{i\beta_2}{2}\frac{\partial^2 A_x}{\partial T^2} + \frac{\alpha}{2}A_x &= i\frac{8}{9}\gamma(|A_x|^2 + |A_y|^2)A_x, \\
\frac{\partial A_y}{\partial z} + \frac{i\beta_2}{2}\frac{\partial^2 A_y}{\partial T^2} + \frac{\alpha}{2}A_y &= i\frac{8}{9}\gamma(|A_x|^2 + |A_y|^2)A_y,
\end{aligned} \tag{2.21}$$

where A_x and A_y are the slowly varying pulse envelope of the x-polarization and y-polarization, respectively.

Numerical Solution to NLSE

A closed-form solution to both NLSE and Manakov equation does not exist, meaning that numerical methods must be employed to solve (2.5) and (2.21). Up to now, the best approach known for solving these partial differential equations is split-step Fourier method (SSFM) [2], [13]. In this method, to solve (2.5) (or (2.21)), the fiber span is broken into small distances, and for each small step, we follow two steps:

- first, we ignore the non-linearity and apply the dispersion operator to the signal using (2.13).
- Then, we ignore dispersion and apply the non-linearity operator using Equation (2.18).

These steps are iterated until we reach to the end of the span. At the end of the span, the signal is amplified to compensate for the fiber loss, and then, the ASE noise is added to the signal. For the next span, the same procedure is repeated.

SSFM is an accurate method for simulating propagation of pulse inside optical fibers. In this thesis, SSFM is employed to validate accuracy of the EGN-model of fiber as well as the performance of the optimized constellations.

2.1.2 EGN-model of Fiber

As we discussed, a closed form solution for (2.5) and (2.21) does not exist, which makes optimizing constellations for the fiber extremely hard. Therefore, modeling of non-linear interference is of great importance. Due to the large values of accumulated dispersion, as electric field components propagate inside the fiber, they take on identical, statistically independent, zero-mean Gaussian distributions. This phenomenon turns the signal into a noise-like source of non-linear disturbance [14]. Hence, we can model non-linear interference as additive Gaussian noise. In the rest of this section, we address two well-known models, Gaussian-noise (GN) and enhanced Gaussian-noise (EGN) models, which consider non-linear interference as Gaussian noise.

To model non-linear interference (NLI) of the fiber optics, GN-model assumes that NLI is approximately Gaussian and additive. This is one of the main pillars of the GN-model. In other words, in the GN-model, the signal disturbance caused by the non-linearity manifests itself as additive Gaussian noise. By finding the power of NLI, we can predict system performance. PSD of NLI must be computed in order to find the power of NLI. In the derivation of the NLI PSD, we consider three assumptions:

- The non-linearity is relatively small versus the useful signal.
- The transmitted signal statistically behaves as stationary Gaussian noise.
- NLI manifests itself as additive Gaussian noise.

The signal-Gaussianity is the key assumption in the derivation of the NLI PSD. This assumption is not valid at the transmitter; however, as the signal gets dispersed, the distribution of the signal approaches a Gaussian distribution [15]. Although the PSD of NLI can be found for systems with multiple fiber types and different span lengths, here, for simplicity, we consider identical-span systems. If we consider the transparent link assumption and the transmitted signals are dual-polarization, the NLI PSD is equal to [15]:

$$G_{\mathrm{NLI}}(f) = \frac{16}{27}\gamma^2 L_{\mathrm{eff}}^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_{\mathrm{WDM}}(f_1)G_{\mathrm{WDM}}(f_2)G_{\mathrm{WDM}}(f_1+f_2-f) \times \rho(f_1,f_2,f)\chi(f_1,f_2,f)\mathrm{d}f_1\,\mathrm{d}f_2, \tag{2.22}$$

where $G_{\mathrm{WDM}}(f)$ is the PSD of the overall WDM transmitted signal, γ is the fiber non-linearity coefficient, and $\chi(f_1, f_2, f)$ has the role of accounting for NLI accumulation in multi-span links. Also, L_{eff} is the span effective length, and $\rho(f_1, f_2, f)$ is the normalized four-wave mixing (FWM) efficiency, and both L_{eff} and $\rho(f_1, f_2, f)$ depend on the type of amplification. For instance, for EDFA, we have:

$$\rho(f_1,f_2,f) = \left|\frac{1-\exp\left(-2\alpha L_s\right)\exp\left(j4\pi^2\beta_2(f_1-f)(f_2-f)L_s\right)}{2\alpha - j4\pi^2\beta_2(f_1-f)(f_2-f)}\right|^2 \times L_{\mathrm{eff}}^{-2}, \tag{2.23}$$

and L_{eff} is defined in (2.18). In (2.23), L_s is the span length, β_2 is the dispersion coefficient in $[s^2.m^{-1}]$, and α is the fiber **field** (not power) loss coefficient in $[m^{-1}]$, such that the signal power is attenuated as $\exp(-2\alpha z)$.

The interpretation of (2.22) is that the NLI PSD generated at frequency f is the integrated result of the product of G_{WDM} at three frequencies f_1, f_2, and $f_3 = f_1 + f_2 - f$. Also, the efficiency of the beating (or interaction) of these three frequencies is determined by the FWM efficiency $\rho(f_1, f_2, f)$.

As stated, factor $\chi(f_1, f_2, f)$ governs NLI accumulation along the link. If $\chi(f_1, f_2, f)$ considers the produced NLI of each span coherently, it becomes:

$$\chi(f_1, f_2, f) = \frac{\sin^2(2N_s\pi^2(f_1 - f)(f_2 - f)\beta_2 L_s)}{\sin^2(2\pi^2(f_1 - f)(f_2 - f)\beta_2 L_s)}, \tag{2.24}$$

where N_s is the total number of spans. As seen, for $N_s = 1$, $\chi = 1$. One can make the further approximation of completely neglecting coherent interference among NLI generated in different spans. Assuming this approximation, (2.24) becomes:

$$\chi(f_1, f_2, f) = N_s. \tag{2.25}$$

This model is called the incoherent GN-model (IGN-model).

Now that we defined all factors of (2.22), we can calculate NLI power, P_{NLI}, by integrating $G_{\text{NLI}}(f)$. Since we assumed that NLI is Gaussian and additive, the equivalent channel model of fiber according to GN-model becomes:

$$Y = X + N_{\text{NLI}} + N_{\text{ASE}}, \tag{2.26}$$

where N_{NLI} and N_{ASE} are circularly symmetric Gaussian noise whose PSD is determined by (2.22) and (2.19), respectively. According to (2.26), the system performance is governed by the modified optical signal-to-noise ratio (OSNR):

$$\text{OSNR}_{\text{NL}} = \frac{P_{\text{ch}}}{P_{\text{ASE}} + P_{\text{NLI}}}, \tag{2.27}$$

where P_{ch} is the average power per WDM channel, P_{ASE} is the power of the ASE noise which falls within a conventional optical noise bandwidth B_N, and P_{NLI} is the NLI power of the channel of interest. For matched receiver filtering, OSNR_{NL} can be converted to the well-known signal-to-noise ratio (SNR) by

the following equation [14]:

$$\mathrm{SNR}_{\mathrm{NL}} = \frac{B_N}{R_s}\mathrm{OSNR}_{\mathrm{NL}}, \tag{2.28}$$

where R_s is the symbol rate. In this thesis, we assume that $B_N = R_s$. Based on (2.28), we can calculate system performance such as bit error-rate and channel capacity.

The results of the classic GN-model are valid assuming that the transmitted signal statistically behaves as stationary Gaussian noise, i.e., P_{NLI} is independent of the modulation format. However, as shown in [10], [16], the NLI power does depend on the modulation format, especially in the first spans. To resolve this problem, [10] removes the signal-Gaussianity assumption and modifies the classic GN-model by adding some correction terms to the NLI power of the GN-model and calls the modified model enhanced Gaussian noise (EGN) model. According to EGN-model, NLI power can be written as [10]:

$$P_{\mathrm{NLI}} = P_{\mathrm{NLI}}^{\mathrm{GN}} + P_{\mathrm{NLI}}^{\mathrm{corr}}, \tag{2.29}$$

where $P_{\mathrm{NLI}}^{\mathrm{GN}}$ is the NLI power obtained from GN-model, and $P_{\mathrm{NLI}}^{\mathrm{corr}}$ is the correction term of the EGN-model, which contains the modulation-dependent terms of the NLI power. The EGN-model accuracy is discussed more in-depth in [10], and comparison of the results of the EGN-model with the split-step Fourier method (SSFM) simulations shows that the EGN-model is an accurate tool for predicting the fiber optics system performance. Hence, in this thesis, we use the EGN-model.

Assuming all WDM channels are spaced equally and all with the same modulation format, equal symbol rate, and launch power, we can write (2.29) as follows [10]:

$$P_{\mathrm{NLI}} = \eta \times P_{\mathrm{ch}}^3, \tag{2.30}$$

where

$$\eta = \eta_{\mathrm{GN}} + \Phi \cdot \eta_{\Phi} + \Psi \cdot \eta_{\Psi}. \tag{2.31}$$

In Equation (2.31), Φ and Ψ are the modulation-dependent terms and are given for a $X \in \chi$, where $\chi = \{x_1, x_2, \cdots, x_M\}$ is the set of constellation

points as follows:

$$\Phi = \frac{\mathbb{E}\{|X|^4\}}{\mathbb{E}^2\{|X|^2\}} - 2 \quad , \quad \Psi = \frac{\mathbb{E}\{|X|^6\}}{\mathbb{E}^3\{|X|^2\}} - 9 \cdot \frac{\mathbb{E}\{|X|^4\}}{\mathbb{E}^2\{|X|^2\}} + 12, \tag{2.32}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator. Note that complete derivation of η_Φ and η_Ψ can be found in [10]. By substituting (2.30) in (2.28), one can conclude that there is an optimal launch power $P_{\text{ch}}^{\text{opt}}$ at which the modified SNR of (2.28) becomes maximum [15]:

$$P_{\text{ch}}^{\text{opt}} = \sqrt[3]{\frac{P_{\text{ASE}}}{2 \cdot \eta}}. \tag{2.33}$$

Equation (2.33) indicates that unlike other commutation channels that increasing the average launch power results in higher SNRs and lower error rates, in the fiber channel, by increasing the average launch power, we may achieve lower SNRs. This behavior of the fiber channel is arisen from the fact that the non-linearity of fiber is directly proportional to the launch power (see (2.5)). To better demonstrate this phenomenon, we provide Fig. 2.5. Fig. 2.5 shows SNR_{NL} performance of the different types of polarization-multiplexed QAMs (PM-QAM) and Gaussian distribution. Polarization-multiplexed 4-QAM (PM-4QAM) introduce the least amount of NLI noise to the system. Higher order QAMs have higher NLI than PM-4QAM, PM-16QAM and PM-32QAM have similar performance, and PM-64QAM has slightly higher SNR_{NL} than PM-8QAM. Also, the Gaussian distribution introduces the lowest SNR_{NL} among all PM-QAMs. Moreover, for launch powers higher than 1 dBm, SNR_{NL} is monotonically decreasing in P_{ch}. Results indicate that the fiber channel is a modulation-dependent channel and to optimize constellations, constellations must be shaped specifically for the fiber channel.

2.2 Constellation Shaping

As mentioned in Section 1.3, constellation shaping is the process of designing constellations so that they approximate a Gaussian source more accurately. The shaping process can be: (i) probabilistic shaping (PS); (ii) geometric shaping (GS), and (iii) hybrid probabilistic-geometric shaping (HPGS). To

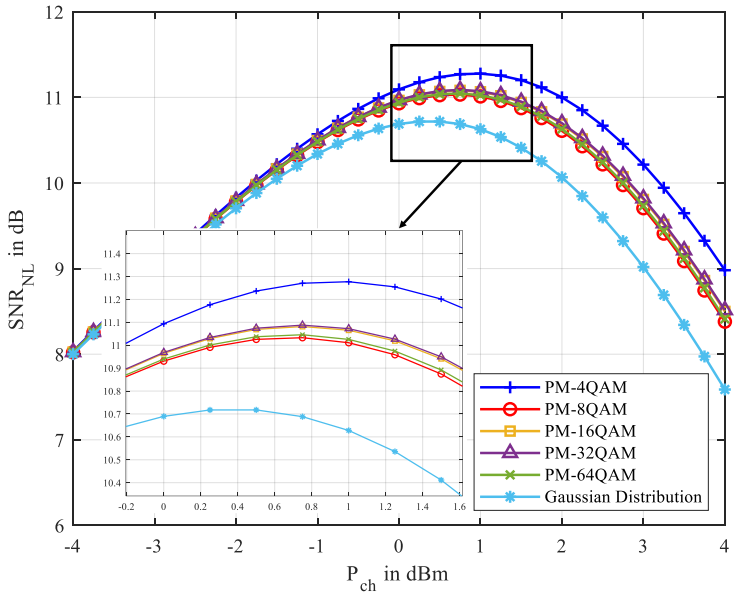

Figure 2.5: SNR_{NL} performance of the PM-QAMs and Gaussian distribution. The simulation is based on the EGN-model of fiber and is done at the distance of 3000 km.

measure how accurate a constellation is approximating a Gaussian source, a metric must be defined. In the following, first, we review the fundamentals of constellation shaping and define our tools for measuring the performance of the shaped constellations. After that, we formulate the mentioned shaping methods based on the defined measures. Finally, we briefly review recent studies on constellation shaping in the literature and our proposed shaping methods.

2.2.1 Fundamentals of Constellation Shaping

On a communication channel, in order to measure the quality of communication for a certain constellation, a metric must be defined. In this thesis, mutual information (MI) and generalized mutual information (GMI) are used as two metrics of the achievable information rate (AIR).

Mutual Information Analysis

MI is utilized as a measure to mimic the symbol-wise AIR [17], [18], [19]. MI depends on the probability and position of each constellation point, but does

not depend on the binary labeling of the constellation points.

Assume that the channel input and output vectors are denoted by $\mathbf{X} = [X_1, X_2, \cdots, X_n]$ and $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_n]$, respectively, where X_is and Y_is are the i-th input and output respectively. It is assumed that X_is are independent and identically distributed (i.i.d). Also, Y_i only depends on X_i. Each $X_i \in \chi$, where $\chi = \{x_1, x_2, \cdots, x_M\}$ is the set of constellation points. The MI between the channel input vector $\mathbf{X}$ and channel output vector $\mathbf{Y}$ is defined as:

$$\frac{1}{n}\mathbb{I}(\mathbf{X};\mathbf{Y}) = \frac{1}{n}\mathbb{E}\left\{ \log_2 \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{Y})} \right\}, \tag{2.34}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator, $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ is the channel transition probability, and $f_{\mathbf{Y}}(\mathbf{y})$ is the probability distribution function of $\mathbf{Y}$.

In some communication channels such as fiber optics, a closed-form equation for $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ does not exist. However, by taking advantage of the mismatched decoding technique [20], [21], we can evaluate the MI rate using the auxiliary channel transition probability $q_{\mathbf{Y}|\mathbf{X}}$ instead of the unknown $f_{\mathbf{Y}|\mathbf{X}}$. In the case of fiber optics channels, by ignoring the correlations over the polarization and time, we have:

$$q_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} q_{Y|X}(y_i|x_i), \tag{2.35}$$

where $q_{Y|X}(y|x)$ is the fixed channel transition probability. Assuming an AWGN auxiliary channel, we have:

$$q_{Y|X}(y|x) = \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|y-x|^2}{\sigma^2}}, \tag{2.36}$$

where σ^2 is the noise variance of the auxiliary channel, $x \in \chi$, and y is a complex number such that $y = y_{\mathrm{Re}} + j y_{\mathrm{Im}}$. Assuming the auxiliary channel transition probability of (2.36), a lower bound for the MI rate can be computed by substituting (2.35) in (2.34) [22]:

$$\frac{1}{n}\mathbb{I}(\mathbf{X};\mathbf{Y}) \geq \mathbb{E}\left\{ \log_2 \frac{q_{Y|X}(Y|X)}{q_Y(Y)} \right\} \triangleq R_{\mathrm{SMD}}, \tag{2.37}$$

where R_{SMD} stands for the MI rate that can be achieved with symbol-metric decoding (SMD) and can be computed as:

$$R_{\mathrm{SMD}} = \sum_{j=1}^{M} \int_{-\infty}^{+\infty} \log_2 \left(\frac{q_{Y|X}(y|x_j)}{q_Y(y)} \right) q_{X,Y}(x_j, y) \mathrm{d}y, \tag{2.38}$$

where

$$q_Y(y) = \sum_{i=1}^{M} q_{Y|X}(y|x_i)P_X(x_i) = \sum_{i=1}^{M} \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|y-x_i|^2}{\sigma^2}} P_X(x_i), \tag{2.39}$$

and

$$q_{X,Y}(x,y) = q_{Y|X}(y|x)P_X(x) = \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|y-x|^2}{\sigma^2}} P_X(x). \tag{2.40}$$

In (2.39) and (2.40), $P_X(x_i)$ is the probability that the random variable X takes the value x_i. Note that the integration over y in (2.38) is a two-dimensional integration over y_{Re} and y_{Im}. Substituting (2.39) and (2.40) in (2.38) yields:

$$\begin{aligned} R_{\mathrm{SMD}} = \sum_{j=1}^{M} \int_{-\infty}^{+\infty} \Bigg(- \frac{|y-x_j|^2}{\sigma^2} \log_2 \mathrm{e} - \log_2 \Bigg(\sum_{i=1}^{M} \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|y-x_i|^2}{\sigma^2}} P_X(x_i) \Bigg) \Bigg) \\ \times \frac{1}{\pi\sigma^2} \mathrm{e}^{-\frac{|y-x_j|^2}{\sigma^2}} P_X(x_j) \mathrm{d}y. \end{aligned} \tag{2.41}$$

Note that although (2.41) is obtained using the assumption of an AWGN auxiliary channel, it can be used for the non-linear fiber channels by making the noise variance in (2.41) dependent on the launch power and the set of constellation points χ [23], [24]. In addition to (2.41), R_{SMD} can be approximated from Monte Carlo simulations. For N transmitted points x_ks and received points y_ks, we have:

$$R_{\mathrm{SMD}} \approx \frac{1}{N} \sum_{k=1}^{N} \log_2 \frac{q_{Y|X}(y_k|x_k)}{q_Y(y_k)} = \frac{1}{N} \sum_{k=1}^{N} \frac{\mathrm{e}^{-\frac{|y_k-x_k|^2}{\sigma^2}}}{\sum_{i=1}^{M} \mathrm{e}^{-\frac{|y_k-x_i|^2}{\sigma^2}} P_X(x_i)}. \tag{2.42}$$

Also, note that (2.41) and (2.42) are valid for any arbitrary probability mass function $P_X(x)$.

Generalized Mutual Information Analysis

MI provides the symbol-wise AIR of a communication system and it does not provide any insight into how efficiently the labels of the constellation points have been assigned. Consequently, MI is not appropriate for bit-interleaved coded modulation (BICM) systems [25]. For BICM systems, bit-metric decoding (BMD) rate, R_{BMD}, is a proper AIR [24], [26].

To compute R_{BMD}, we assume that symbol X consists of m bit levels $\mathbf{B} = [B_1, B_2, \cdots, B_m]$, where $m = \log_2(M)$. Considering $\mathbf{B}$, we have[24]:

$$R_{\text{BMD}} = \mathbb{H}(\mathbf{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y), \tag{2.43}$$

where $\mathbb{H}(\cdot)$ is the entropy. For independent bit-levels (which is the case in this thesis), R_{BMD} becomes the generalized mutual information (GMI) rate [24], [27]. Considering the independent bit-levels assumption, we have:

$$R_{\text{BMD}} = \mathbb{H}(\mathbf{B}) + \sum_{i=1}^{m} \mathbb{E}_{B_i,Y}\left\{ \log_2\Big(\frac{q_{B_i,Y}(B,Y)}{q_Y(Y)}\Big)\right\}, \tag{2.44}$$

where

$$\mathbb{H}(\mathbf{B}) = \mathbb{H}(X) = -\sum_{i=1}^{M} P_X(x_i)\log_2(P_X(x_i)), \tag{2.45}$$

and

$$\mathbb{E}_{B_i,Y}\left\{ \log_2\Big(\frac{q_{B_i,Y}(B,Y)}{q_Y(Y)}\Big)\right\} = \sum_{b=0}^{1}\int_{-\infty}^{+\infty} \log_2\Big(\frac{q_{B_i,Y}(b,y)}{q_Y(y)}\Big) q_{B_i,Y}(b,y)\mathrm{d}y. \tag{2.46}$$

In (2.46), $q_Y(y)$ can be computed from (2.39), and $q_{B_i,Y}(b,y)$ is calculated as follows:

$$q_{B_i,Y}(b,y) = \sum_{x\in\chi_b^i} q_{X,Y}(x,y), \tag{2.47}$$

where χ_b^i is the set of symbols whose i-th bit are equal to b.

By substituting (2.47) in (2.46), we have:

$$\begin{aligned}\mathbb{H}(B_i|Y) = -\sum_{b=0}^{1}\int_{-\infty}^{+\infty} \sum_{x'\in\chi_b^i} \log_2\Big(&\frac{\sum_{x\in\chi_b^i} \frac{\exp\left(-\frac{|y-x|^2}{\sigma^2}\right)}{\pi\sigma^2} P_X(x)}{\sum_{x\in\chi} \frac{\exp\left(-\frac{|y-x|^2}{\sigma^2}\right)}{\pi\sigma^2} P_X(x)}\Big) \\ &\times \frac{\exp\left(-\frac{|y-x'|^2}{\sigma^2}\right)}{\pi\sigma^2} P_X(x')\mathrm{d}y.\end{aligned} \tag{2.48}$$

Now, by substituting (2.48) and (2.45) in (2.43), the GMI rate is competed. Also, R_{BMD} can be estimated using Monte Carlo simulations. For N transmitted bit labels $\mathbf{b}_k = [b_k^1, b_k^2, \cdots, b_k^m]$s and received points y_ks, we have::

$$R_{\text{BMD}} \approx \frac{-1}{N}\sum_{k=1}^{N}\log_2(P_X(x_k)) - \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{m}\log_2\Big(\frac{\sum_{x\in\chi_{b_k^i}^i} \frac{\exp\left(-\frac{|y_k-x|^2}{\sigma^2}\right)}{\pi\sigma^2} P_X(x)}{\sum_{x\in\chi} \frac{\exp\left(-\frac{|y_k-x|^2}{\sigma^2}\right)}{\pi\sigma^2} P_X(x)}\Big). \tag{2.49}$$

2.2.2 Probabilistic Shaping

PS is to change the uniform distribution of a constellation (usually PAMs or QAMs) to a non-uniform distribution such that AIR increases. To find the optimal non-uniform distribution, consider the following AWGN channel:

$$Y = \Delta X + N, \tag{2.50}$$

where Δ is a positive scalar that adjusts the average symbol energy, and N is a circularly symmetric Gaussian noise with zero mean and unit variance so that $\mathbb{E}\{|\Delta \cdot X|^2\}$ represents the average SNR at the receiver, i.e.,

$$\mathbb{E}\{|\Delta \cdot X|^2\} = \text{SNR}. \tag{2.51}$$

[28] shows that distributions from the family of the Maxwell–Boltzmann (MB) distribution are capable of maximizing entropy subject to power constraint (2.51). MB distributions have the following form:

$$P_X(x_i) = \frac{\exp\left(-\lambda |x_i|^2\right)}{\sum_{j=1}^{M} \exp\left(-\lambda |x_j|^2\right)}, \tag{2.52}$$

where λ is the parameter of the MB distribution and controls the entropy rate.

For a fixed λ, we always compute Δ such that power constraint (2.51) holds. By making Δ dependent on λ, we formulate the symbol-wise PS problem as follows: for a set of constellation points χ and *shaping* SNR γ (the SNR value in which we want to maximize AIR), we solve

$$\begin{aligned} &\lambda^* = \underset{\lambda}{\text{argmax}}\, R_{\text{SMD}}(\chi, P_X, \gamma_{(\chi,P_X)}), \\ &\text{subject to: } \sum_{x\in\chi} P_X(x) \times |\Delta \cdot x|^2 = \gamma_{(\chi,P_X)}. \end{aligned} \tag{2.53}$$

Note that depending on the communication channel, shaping SNR γ can depend on the modulation format (in Section 2.1, we discussed that this fact holds for the fiber channel). Also, the bit-wise PS problem is formulated as follows: for a set of constellation points χ, binary labeling $\mathcal{L}$, and shaping SNR γ, we solve

$$\begin{aligned} &\lambda^* = \underset{\lambda}{\text{argmax}}\, R_{\text{BMD}}(\chi, \mathcal{L}, P_X, \gamma_{(\chi,P_X)}), \\ &\text{subject to: } \sum_{x\in\chi} P_X(x) \times |\Delta \cdot x|^2 = \gamma_{(\chi,P_X)}. \end{aligned} \tag{2.54}$$

Note that in most of the PS problems, square QAMs are selected as the constellation χ is (2.53) and (2.54). Since the optimal binary labels of QAMs are already known (Gray labels for square QAMs exist), the main focus is on solving (2.54) rather than (2.53). To achieve the whole gain that (2.53) suggests, non-binary labels are required which makes implantation of symbol-wise-based PS systems extremely complicated. Hence, in this thesis, we concentrate on (2.54) for solving the PS problem.

2.2.3 Probabilistic Amplitude Shaping (PAS): An Implementation of PS

One of the main challenges of PS is to design a transmitter that both shapes the probability of the uniform data and applies a forward error correction (FEC) code on the transmitting data. In order to ideally implement PS and an FEC algorithm, PS and FEC must be applied on data simultaneously, otherwise, either the optimal P_X changes to another probability mass function (PMF) or error bursts happen after PS decoder [29]. Authors in [30] solve this problem by proposing the idea of probabilistic amplitude shaping (PAS). PAS originally was proposed for PAM constellations. Hence, to use PAS for square QAMs, the PAS architecture must be employed for the in-phase and quadrature components of the square QAMs independently.

The PAS architecture is based on two observations:

- Amplitude-Sign Factorization: Since the MB distribution depends on the amplitude of the points and the PAM points are symmetric around zero, we can rewrite (2.52) as follows:

$$P_X(x_i) = P_A(|x_i|) \times P_S(\text{sign}(x_i)), \tag{2.55}$$

  where $P_A(\cdot)$ and $P_S(\cdot)$ are PMF of the amplitude and sign of the PAM points, respectively. Since points are symmetric around zero, the amplitude and sign of a point are stochastically independent, and the sign of the points is uniformly distributed:

$$P_S(+1) = P_S(-1) = \frac{1}{2}. \tag{2.56}$$

- Uniform Check Bit Assumption: The second observation corresponds to the distribution of the redundant bits of a binary code. Assuming that the systematic generator matrix of an (n, k) binary code has the following form:

$$\mathbf{G} = [\mathbf{I}_k|\mathbf{P}], \tag{2.57}$$

one can show that the distribution of the $n-k$ redundant bits is uniform since they are modulo-two sums of k data bits [30]. Note that in (2.57), $\mathbf{I}_k$ is a $k \times k$ identity matrix, $\mathbf{P}$ is the parity matrix, which is a $k \times (n-k)$ matrix.

Based on these two observations, the encoding procedure of PAS is as follows: consider the transmission of n_c symbols of a 2^m-PAM. Since the amplitude and sign of the PAM points are independent, we can independently label the 2^{m-1} amplitudes and the 2 signs by binary strings as follows:

$$\begin{aligned} A &\mapsto \mathbf{b}(A) \in \{0,1\}^{m-1}, \\ S &\mapsto \mathrm{b}(S) \in \{0,1\}. \end{aligned} \tag{2.58}$$

If we use a systematic $\big(n = mn_c, k = (m-1)n_c\big)$ binary code, we can shape distribution of data as follows (the encoding procedure is provided in Fig. 2.6):

1. The discrete memoryless source A generates i.i.d amplitudes $A_1, \cdots, A_{n_c}$ whose PMF is equal to $P_A(\cdot)$.

2. The $(m-1)n_c$ bits of the generated amplitudes are determined based on $\mathbf{b}(A_i)$.

3. The n_c redundant bits of the $(m-1)n_c$ bits are computed based on the parity matrix $\mathbf{P}$.

4. Signs of the n_c redundant bits are determined using the inverse of $\mathrm{b}(S_i)$.

5. The generated amplitudes A_is are multiplied by the corresponding signs S_is, and then, they are scaled by Δ and transmitted.

According to the uniform check bit assumption, the probability of signal x_i in Fig. 2.6 is equal to $P_A(a_i) \times \frac{1}{2}$, which is equal to $P_X(x_i)$ in (2.55). As a result, the distribution of the transmitted data is MB.

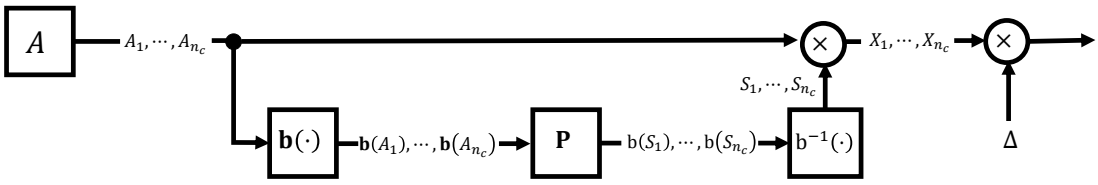

Figure 2.6: The PAS architecture for probabilistic shaping [30]

In general, PAS-based PS provides much higher achievable information rates compared to GS [9], however, implementation of PS is challenging. PS is typically implemented by distribution matchers (DM) (the discrete memoryless source A in Fig. 2.6) which require long block lengths to provide the expected AIR. Unfortunately, the hardware implementation of DM for long block lengths is a significant challenge [31]. Consequently, the imperfect implementation of DM decreases the performance of the PS systems [32]. In addition, other drawbacks of PS are as follows: (i) they experience error propagation after the distribution de-matcher [29]; and (ii) they incur high complexity due to applying the multiple bit-to-symbol (B2S) and symbol-to-bit (S2B) mappings at the transceiver [30].

2.2.4 Geometric Shaping

GS is a technique that increases AIR by changing the position of the equiprobable constellation points, i.e., $P_X(x) = \frac{1}{M}$. As a result, the GS problem can be formulated as follows: for a set of constellation points χ, binary labeling $\mathcal{L}$, and shaping SNR γ, we try to solve:

$$\begin{aligned} \chi^* &= \underset{\chi}{\operatorname{argmax}}\, R_{\mathrm{SMD}}(\chi, \gamma_{(\chi)}), \\ \text{subject to: } & \frac{1}{M}\sum_{x\in\chi} |x|^2 = 1, \end{aligned} \tag{2.59}$$

for the symbol-wise problem, and

$$\begin{aligned} \chi^*, \mathcal{L}^* &= \underset{\chi,\mathcal{L}}{\operatorname{argmax}}\, R_{\mathrm{BMD}}(\chi, \mathcal{L}, \gamma_{(\chi)}), \\ \text{subject to: } & \frac{1}{M}\sum_{x\in\chi} |x|^2 = 1, \end{aligned} \tag{2.60}$$

for the bit-wise problem. In (2.59), there is no need to find labels of the points because R_{SMD} informs us of the maximum transmission rate at which error-free communication is possible. Meaning that R_{SMD} assumes the ideal detection rule.

To solve (2.59), for a D-dimensional M-point constellation, we need to search in a D-dimensional space and pick the best set of M points in terms of MI. Consequently, there are $D \times M$ unknown variables to be determined. Considering the fact that (2.59) is non-convex [9], solving this optimization problem becomes extremely complicated, especially for large values of M. Moreover, to solve (2.60), in addition to the position of the points, the corresponding binary labels must be found, which results in $D \times (M+1)$ unknown variables.

Compared to PS, GS needs high-resolution digital-to-analog and analog-to-digital converters. Also, in GS, there is a gap between the mutual information and generalized mutual information rate due to the non-Gray-mapping [23], [27]. However, GS is more capable of mitigating fiber non-linear effects than PS [23], [33], [34]. Hence, GS has attracted a lot of interest in recent years [23], [26], [27], [33], [35]–[37].

2.2.5 Hybrid Probabilistic-Geometric Shaping

As mentioned in the previous sections, in GS, there is a gap between the MI and GMI rate due to the non-Gray-mapping. In PS transmission systems, which are based on the PAS architecture, there is a rate loss at the distribution matcher block due to using finite blocklengths. To compensate for these losses, one can employ both PS and GS, which is known as hybrid probabilistic-geometric shaping (HPGS). Hence, HPGS can be formulated for bit-wise shaping as follows: for a set of constellation points χ, binary labeling $\mathcal{L}$, PMF equals to MB distribution with parameter λ, and shaping SNR γ, we try to solve:

$$\begin{aligned} &\chi^*, \mathcal{L}^*, \lambda^* = \underset{\chi,\mathcal{L},\lambda}{\operatorname{argmax}}\, R_{\mathrm{BMD}}(\chi, \mathcal{L}, P_X, \gamma_{(\chi,P_X)}), \\ &\text{subject to: } \sum_{x\in\chi} P_X(x) \times |\Delta \cdot x|^2 = \gamma_{(\chi,P_X)}. \end{aligned} \tag{2.61}$$

Note that since PS is used in HPGS, same as PS, we select R_{BMD} as the AIR of the communication system.

2.2.6 Related Works

Here, we briefly review some of the recent studies on constellation shaping in the literature.

PS is employed in the fiber channel in [24], [29]. Both studies find the optimized λ for the AWGN channel, and then, the obtained PMF is studied over the fiber channel. Optimization results of [24] for the 64-QAM show that the probabilistic-shaped 64-QAM is capable of providing GMI gains as high as 0.35 bits/4D-symbol and system reach increases as high as 400 km over 64-QAM.

In [23], a class of Gaussian-like constellation points is introduced and studied over a 100-km fiber-optics transmission system. [23] proposes an iterative algorithm for maximizing MI of M-QAM modulations. In this algorithm, the location of the M constellation points is initialized with the position of the regular M-QAM constellation points, then, a symbol sequence following the Gaussian distribution is generated (in ASE noise limited optical fiber channels, the optimal source distribution is a two-dimensional Gaussian distribution), and the generated sequence is distributed into M clusters based on the Euclidean distance from the current constellation points. After that, the new positions are equal to the average central positions of the labeled symbols. These steps are repeated until the algorithm converges. MI performance as a function of optical fiber length was calculated for 8-QAM, 16-QAM, and 32-QAM. In the case of 8-QAM, at the MI value of 2.5 bits/symbol, the optimized 8-QAM provides 700 km reach improvement over regular 8-QAM. In the simulation of 16-QAM, the optimized 16-QAM achieved a gain of 300 km compared with the regular 16-QAM when the MI is equal to 3.2 bits/symbol. Also, in the 32-QAM scenario, when the MI value is 4 bits/symbol, the reach improvement of 100 km was achieved by the optimized 32-QAM.

The pairwise optimization (PO) algorithm [38] is one of the ways to solve the GS problem. Authors in [38] show that if a zero mean and an average power

constraint is concerned, it is not possible to solve the GS problem by adjusting the position of only a single constellation point, but by taking any pair of points and moving them simultaneously, we can find the optimum position of the two constellation points, while the mentioned constraints are satisfied. Moreover, to satisfy the zero mean and average power constraints, the selected pair of points must be located on two different circles [38]. Hence, the problem of searching over a continuous two-dimensional plane is reduced to searching over a circle. Therefore, the steps of the PO algorithm are to 1) configure an initial constellation satisfying the zero mean and average constraint, 2) randomly select a pair of points, 3) calculate the constrained circles, and 4) find the optimum positions on the constrained circles. Steps 2-4 are repeated until the constellation stabilizes. In [27] and [26], PO is used to maximize mutual information of 64 and 32-point constellations, respectively. Simulations of [26] for an optical fiber link show that at an SNR value of 15 dB, the MI-optimized and GMI-optimized GS modulations achieve gains of 0.33 bits/symbol and 0.2 bits/symbol compared with 64-QAM, respectively. Also, in [27], by taking advantage of the pairwise optimization algorithm and a quasi-Gray bits mapping, the GMI optimization problem was solved for 32-point constellations over a nonlinear optical fiber link, and a gain of 0.15 bits/symbol compared with 32-QAM was achieved. Also, [35] employs PO to minimize the bit error rate of 8 and 32-point constellations. In [39], PO is combined with the binary switching algorithm (BSA)[40] to maximize generalized mutual information of 16, 64, and 256-point constellations.

In [41], probabilistic fold shaping (PFS) [42] is used to design a probabilistic-geometric-shaped 32-QAM. PFS is originally used for shaping PMF of non-square QAM constellations. However, the authors in [41] employ PFS to shape both geometry and probability of 32-ary constellations. In a PFS-based N-fold rotationally symmetrical QAM, the $\log_2(N)$ bits determining the fold index yield the uniform distribution, which can be used to carry the parity-check bits after FEC encoding. The authors in [41] assume the 4-fold condition on the position of the constellation points. As a result, only the position of the constellation points located on the first quadrant is required to be opti-

mized. After finding the optimal location of the first quadrant points, they are mirrored with respect to the x and y-axis to obtain the complete constellation. After that, the PS scheme can be applied to achieve an optimized probabilistic-geometric-shaped 32-QAM (the MB distribution is assumed on the constellation). Results of [41] show that the optimized constellation outperforms 32-QAM by approximately 0.35 bits/4D-symbol. Also, the optimized constellation is capable of extending the reach of 32-QAM by more than 500km.

In [42], arbitrary probabilistic shaping (APS) scheme is introduced. APS is suitable for any geometric-shaped constellation. However, the main drawback of APS is that the symbols generated by APS do not yield an MB distribution. The results of studying the APS scheme over the fiber channel show that the probabilistic-geometric-shaped 32QAM outperforms 32-QAM in terms of bit-error rate by 0.9 dB. Also, results of [42] show that APS cannot outperform probabilistically-shaped QAM constellations in low-transmission-rate regimes [42].

The main drawback of most of these algorithms is that they optimize the location of the constellation points for the AWGN channels. Then, the obtained constellations are studied over fiber-optics channels. Since the optimization is for the AWGN channel, the shaped constellations may introduce much higher NLI noise (NLIN) to the system than the constellations that are shaped specifically for the fiber channel [43]. Consequently, in most cases, the best possible shaping gain is not achieved. Moreover, in these algorithms, since there is no constraint on the position of constellation points (each point moves freely during the optimization process), analyzing the pattern of the location of the obtained points remains a challenge. By assuming and utilizing a model for the position of the constellation points, one can simplify the process of finding the pattern of the optimal points and decrease the complexity of the optimization problem as well.

2.2.7 Overview of Proposed Solutions to Constellation Shaping

Amplitude-phase shift keying (APSK) modulations have a good potential for GS. The reason is that a well-designed APSK can properly approximate a Gaussian source [44]. However, properly designing APSK constellations is a challenge as APSKs have several parameters. In [45], using the method of non-uniformly assigning the constellation points [46], an equation for the optimal radii of APSK rings for the AWGN channels is obtained. This equation, however, does not provide the other optimal parameters of APSK points.

In this thesis, unlike previous work which optimize the constellations for the AWGN channel and use it for the fiber, we directly maximize the symbol-wise MI rate of the constellations for the fiber optic channel based on the EGN-model of fiber. In fact, we optimize various APSK constellations for both the AWGN and non-linear fiber channels and compare the MI performance of the optimized APSKs with other GS methods to demonstrate the superiority of our optimization scheme.

For the fiber channel, we shape APSK constellations at the maximum modified SNR of the system. By doing so, we maximize the MI rate while the impacts of shaping on NLIN are considered. Our results indicate that when we optimize APSK constellations specifically for the fiber channel, more improvement in the mutual information is achieved compared to optimizing APSK points for the AWGN channel and then using them on the fiber channel. Also, by comparing the mutual information performance of the optimal constellations with the widely used QAM constellations, we observe that the optimized APSK constellations can achieve much higher rates on both the AWGN and the fiber channel. These results are achieved by first formulating the mutual information maximization problem in the form of an optimization problem with a small number of optimization variables. And then efficiently solving this problem using the particle swarm optimization (PSO) [47].

In addition, by using the mutual information formula, we obtain an equation that results in the optimal radius of the APSK rings. Using this equation,

we analyze how the optimal radius of APSK rings are changing for both the AWGN and the non-linear fiber channels. We show that for short fiber links, optimized APSKs are close to those optimized for AWGN; however, when the fiber length increases, the radius of rings grows slower than those of the AWGN channel.

In addition to our GS method, we propose an HPGS method for the non-linear fiber channel based on the PFS architecture. We optimize the position and PMF of the points to maximize the GMI rate while the impacts of shaping on NLIN are considered. Using our method, we shape 32 and 64-point constellations for the fiber channel. Our reported GMIs outperform that of both QAMs and probabilistic-shaped QAMs in all scenarios. Moreover, our results show that constellations shaped for the fiber channel have higher GMI than constellations shaped for the AWGN channel.

Chapter 3

Geometric Shaped Constellations for the Fiber Channel

3.1 Introduction

GS is an efficient method to increase the AIR of the fiber channel. As mentioned in Chapter 2, implementation of GS is less complicated compared to PS. Moreover, geometrically shaped constellations are much more capable of mitigating fiber non-linear effects than the probabilistically shaped ones [23], [33], [34]. Despite the advantages of GS over PS, solving the optimization problem of (2.59) is challenging since it is non-convex [9]. However, putting some constraints on the position of the constellation points may reduce the optimization complexity significantly. In order to decrease the complexity burden, in this chapter, we assume that the constellation points are located on some rings. Constellations that consist of a few rings are known as amplitude-phase shift keying (ASPK). In the following, we show that APSK optimization is very efficient and that when we optimize APSK constellations, we outperform existing GS algorithms.

3.2 APSK Constellations

APSK constellations consists of R concentric rings, and on each ring, uniformly spaced phase shift keying (PSK) points are located. We can formulate an M-

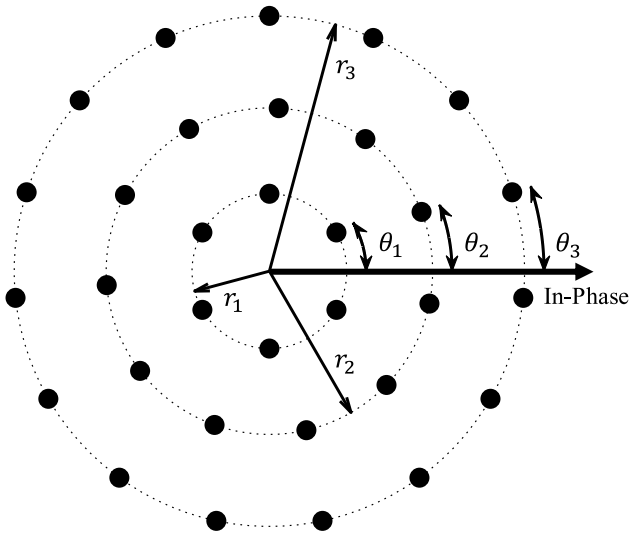

Figure 3.1: An example of a 32-APSK constellation with $n_1 = 6$, $\theta_1 = \frac{\pi}{6}$, $n_2 = 11$, $\theta_2 = \frac{13\pi}{110}$, $n_3 = 15$, and $\theta_3 = \frac{\pi}{10}$.

APSK constellation set as follows [48]:

$$\chi_{\text{APSK}} = \begin{cases} r_1 \exp\left(j\left(\frac{2\pi}{n_1} i + \theta_1\right)\right) & i = 0, \cdots, n_1 - 1 \\ r_2 \exp\left(j\left(\frac{2\pi}{n_2} i + \theta_2\right)\right) & i = 0, \cdots, n_2 - 1 \\ \vdots & \\ r_R \exp\left(j\left(\frac{2\pi}{n_R} i + \theta_R\right)\right) & i = 0, \cdots, n_R - 1 \end{cases}, \tag{3.1}$$

where n_k, r_k, and θ_k are the number of constellation points, radius, and phase offset of the k-th ring, respectively. Also, $\sum_{k=1}^{R} n_k = M$. Fig. 3.1 provides a demonstration of a 32-APSK based on (3.1), which consists of 3 rings.

Optimizing an M-APSK constellation according to (2.59) is equivalent to optimizing the parameters n_k, r_k, and θ_k for $k = 1, \cdots, R$. Hence, to simplify the mathematical representation of (3.1), we revise it as follows:

$$\vec{\chi}_{\text{APSK}} = \big[\underbrace{n_1, r_1, \theta_1}_{\text{ring 1}}, \underbrace{n_2, r_2, \theta_2}_{\text{ring 2}}, \cdots, \underbrace{n_R, r_R, \theta_R}_{\text{ring } R}\big]. \tag{3.2}$$

Considering (3.2), we substitute $\vec{\chi}_{\text{APSK}}$ with χ in (2.59). By doing so, for an M-point constellation, the number of optimization variables of (2.59) changes

from M complex numbers to $3R$ real numbers, which is a considerable reduction. Moreover, the complexity can be further decreased by using equation $n_1 = M - \sum_{k=2}^{R} n_k$ and by setting $\theta_1 = 0$. By doing so, we remove the two unknowns n_1 and θ_1 from (3.2).

Since the optimization problem of (2.59) is non-convex [9], using stochastic optimization algorithms is justified. Among many stochastic optimization algorithms, particle swarm optimization (PSO) [47] is a well-known approach that is computationally efficient and robust against its hyperparameters. Moreover, it has a good potential to quickly converge and escape from local maximums [49], [50]. Due to these advantages, in this thesis, we use PSO to find the optimized $\vec{\chi}^{*}_{\text{APSK}}$.

3.3 APSK Optimization Using PSO Algorithm

In PSO, we place several agents, or the particles, in the search space of the objective function. In each iteration, each particle, $\vec{a}_n$, computes the value of the objective function at its current location and moves in the search space according to a speed vector. The speed vector, $\vec{v}_n$, of each particle depends on the history of its best-found location, $\vec{p}_n^{best}$, and the best-found location among all particles, $\vec{g}^{best}$. After each particle moved based on the corresponding speed vector, the value of the objective function at the new location is computed, and the current location is replaced with the new location. Here, we update the speed vector $\vec{v}_n$ as follows:

$$\vec{v}_n(i+1) = w(i) \otimes \vec{v}_n(i) + \vec{U}_1 \otimes (\vec{p}_n^{best}(i) - \vec{a}_n(i)) + \vec{U}_2 \otimes (\vec{g}^{best}(i) - \vec{a}_n(i)), \tag{3.3}$$

where $w(i)$ is the adaptive weight of $\vec{v}_n(i)$ at the i-th iteration. Also, $\vec{U}_1$ and $\vec{U}_2$ are D-dimensional uniformly distributed random vectors in the intervals $(0, u_1)$ and $(0, u_2)$, respectively, and they determine the amount of movement in the direction of $\vec{p}_n^{best}$ and $\vec{g}^{best}$. Note that $\otimes$ represents the element wise multiplication of vectors.

Based on our problem definition in Chapter 2, we aim to find the optimal vector $\vec{\chi}_{\text{APSK}}$ in terms of R_{SMD}. As a result, we use (2.41) as the objective

function in the PSO algorithm and try to maximize it. In addition, since we removed n_1 and θ_1 from the unknowns of (3.2), the search space becomes a $(3R-2)$-dimensional space. According to the defined search space and objective function, we present our method to optimize $\vec{\chi}_{\text{APSK}}$ in Algorithm 1, where N_i and N_p are the number of iterations and particles, respectively. Moreover, the adaptive weight $w(i)$ is linearly decreased from its maximum to its minimum. Note that in the computation of $R_{\text{SMD}}(\vec{a}_n(i), \gamma_{(\vec{a}_n(i))})$, we round the value of n_k to the nearest integer number, and we scale the radii if energy of the constellation is not equal to one. Also, note that since the shaping SNR, γ, depends on the constellation points, γ must be computed for each particle separately.

The procedure of MI maximization for APSK constellations is provided in Algorithm 1; however, the value of hyperparameters are not determined. These hyperparameters include N_i, N_p, w_{max}, w_{min}, u_1, and u_2. In addition, the upper and lower bound of $\vec{\chi}_{\text{APSK}}$ entries and the maximum number of rings must be determined. In this thesis, we assume that the maximum number of rings is 7, i.e., $R = 7$, which makes the search space a 19-dimensional space. We set the minimum value of all entries of $\vec{\chi}_{\text{APSK}}$ zero. Also, the maximum values of n_k, r_k, and θ_k are set M, $\frac{3(\sqrt{M}-1)}{\sqrt{\frac{2}{3}(M-1)}}$ (three times the maximum amplitude of the M-QAM constellation), and $1.99 \times \pi$, respectively. The bounds of the search space entries are provided in Table 3.1. Moreover, for hyperpa-

Table 3.1: Bounds of $\vec{\chi}_{\text{APSK}}$ Entries

Parameter	**Minimum Value**	**Maximum Value**
n_k	0	M
r_k	0	$\frac{3(\sqrt{M}-1)}{\sqrt{\frac{2}{3}(M-1)}}$
θ_k	0	$1.99 \times \pi$

rameters of Algorithm 1, we use 150, 500, 0.9, 0.2, and 2 for N_p, N_i, w_{max}, w_{min}, u_1, and u_2, respectively. We obtained these values by performing several experiments on AWGN and fiber channel. In these experiments, we analyzed the convergence speed of PSO as well as its capability to escape from local optimums, and based on our analysis, we chose the value of hyperparameters

Algorithm 1 PSO Algorithm for MI Maximization

Output: optimized $\vec{\chi}_{\text{APSK}}$

Initialize the position vector of all particles randomly between their lower and upper bound.
Initialize the speed vector and $\vec{p}^{best}$ of all particles with zero vectors, and set the corresponding MI value to $-\infty$.
Initialize the position vector of $\vec{g}^{best}$ with a zero vector, and set the corresponding MI value to $-\infty$.
for $i = 1 : N_i$ **do**
  for $n = 1 : N_p$ **do**
    compute $R_{\text{SMD}}(\vec{a}_n(i), \gamma_{(\vec{a}_n(i))})$ using Equation (2.41), then set $R^n_{\text{SMD}} = R_{\text{SMD}}(\vec{a}_n(i), \gamma_{(\vec{a}_n(i))})$.
    if $R^n_{\text{SMD}} > R_{\text{SMD}}(\vec{p}^{best}_n, \gamma_{(\vec{p}^{best}_n)})$ **then**
      set $\vec{p}^{best}_n = \vec{a}_n(i)$.
    end if
    if $R^n_{\text{SMD}} > R_{\text{SMD}}(\vec{g}^{best}, \gamma_{(\vec{g}^{best})})$ **then**
      set $\vec{g}^{best} = \vec{a}_n(i)$.
    end if
  end for
  Set $w(i) = \frac{i\times(w_{max}-w_{min})+N_i\cdot w_{min}-w_{max}}{N_i-1}$.
  for $n = 1 : N_p$ **do**
    Compute $\vec{v}_n(i+1)$ using Equation (3.3).
    Set $\vec{a}_n(i+1) = \vec{a}_n(i) + \vec{v}_n(i+1)$.
    if Any entry in $\vec{a}_n(i+1)$ is larger/smaller than the upper bound/lower bound **then**
      set that entry of $\vec{a}_n(i+1)$ to its upper bound/lower bound.
    end if
  end for
end for
return $\vec{g}^{best}$

such that by using a reasonable number of particles, an acceptable convergence speed is achieved. Hyperparameters of Algorithm 1 are provided in Table 3.2.

Table 3.2: Hyperparameters of Algorithm 1

Hyperparameter	**Value**
N_p, number of particles	150
N_i, number of iterations	500
w_{max}	0.9
w_{min}	0.2
u_1	2
u_2	2

Although Algorithm 1 finds the near-optimal APSK constellation, it does not provide any mathematical intuition onto how the entries of $\vec{\chi}_{\text{APSK}}$ change in different situations. Since the radius of rings is one of the most important parameters that impact the minimum distance between constellation points, among entries of $\vec{\chi}_{\text{APSK}}$, we now focus on forming an equation for the radius of rings, and by using this equation, we analyze how the optimal radius of rings changes for both AWGN and fiber channel in Section 3.5 and Section 3.6.

3.4 Analysis of the Radius of APSK Rings

In the following, we find an equation that, subject to some approximations, results in obtaining the near-optimal r_ks for the given sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$. To achieve this equation, we need to find a closed-form equation that approximates MI. As mentioned before, by using the mismatched decoding technique and AWGN auxiliary channel, a lower bound for MI can be found (see (2.37)). Without loss of generality, we can formulate an AWGN channel as follows:

$$Y = \sqrt{\text{SNR}}X + N, \tag{3.4}$$

where $\mathbb{E}\{|X|^2\} = 1$, and N is a circularly symmetric Gaussian noise with zero mean and unit variance, i.e., $N \sim \mathcal{CN}(0, 1)$ so that SNR represents the average signal-to-noise ratio at the receiver. Considering (3.4), we can rewrite (2.37)

as follows:

$$R_{\mathrm{SMD}} = \mathbb{E}_{X,Y}\Big\{ \log_2 \Big(\frac{\frac{1}{\pi}\mathrm{e}^{-|Y-\sqrt{\mathrm{SNR}}X|^2}}{\sum_{i=1}^{M} \frac{1}{\pi}\mathrm{e}^{-|Y-\sqrt{\mathrm{SNR}}x_i|^2} P_X(x_i)} \Big) \Big\}. \tag{3.5}$$

By substituting (3.4) in (3.5) and setting $P_X(x_i) = \frac{1}{M}$ and averaging over X, R_{SMD} can be written as:

$$\begin{aligned} R_{\mathrm{SMD}} = \log_2 M - \\ \frac{1}{M}\mathbb{E}_N \Big\{ \underbrace{\sum_{j=1}^{M} \log_2 \Big(\sum_{i=1}^{M} \mathrm{e}^{-|\sqrt{\mathrm{SNR}}(x_j-x_i)+N|^2+|N|^2} \Big)}_{\text{loss}} \Big\}. \end{aligned} \tag{3.6}$$

The loss term shown in (3.6) can be used as a reference objective function to maximize R_{SMD}, i.e., minimizing the loss term is equivalent to maximizing R_{SMD}; however, this minimization is not feasible since the loss term depends on noise N. Here, we solve this problem through an approximation in the exponent of the exponential term. By expanding the exponent in (3.6), we have:

$$\begin{aligned} -|\sqrt{\mathrm{SNR}}(x_j - x_i) + N|^2 + |N|^2 = -\mathrm{SNR}|x_j - x_i|^2 - \\ 2\sqrt{\mathrm{SNR}}\,\mathrm{Re}\{(x_j - x_i)N\}. \end{aligned} \tag{3.7}$$

The first term in (3.7) will be the dominant term according to the dominated convergence theorem [51]. Moreover, in GS problems, the shaping SNR is usually larger than 10 dB; hence, we can ignore the second term of (3.7). We further discuss the accuracy of this approximation in the next two sections. By doing so, we define the approximated loss function as follows:

$$\text{loss function } (LF) = \sum_{j=1}^{M} \log_2 \Big(\sum_{i=1}^{M} \mathrm{e}^{-\mathrm{SNR}|x_j-x_i|^2} \Big). \tag{3.8}$$

As a result, the problem of finding the optimal set $\{r_k\}_{k=1}^{R}$ for the given sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ is equivalent to find the set $\{r_k\}_{k=1}^{R}$ such that (3.8) becomes minimum. The search space of this problem is an R-dimensional space, which cannot be solved analytically. However, proposing and using a model for the radius of rings can both decrease the complexity of the problem and give intuition about the optimization process.

To approximate a Gaussian source, we consider two regimes: low-SNR regimes and high-SNR regimes. In low-SNR regimes, the variance of the optimal Gaussian source is low, which results in a narrow Gaussian distribution. Hence, more rings should be placed at low-power levels than high-power ones. To do so, in low-power levels, the ring radius must grow slowly with the ring number, but for high-power levels, the radius must grow much faster. In other words, the ring radius increases super-linearly with the ring number. In high-SNR regimes, on the other hand, a broad Gaussian distribution source is the optimal source. To approximate a broad Gaussian distribution using a finite-size APSK, rings should be distributed such that the APSK resembles an almost-uniform distribution. To do so, depending on the number of points located on each ring ($\{n_k\}_{k=1}^{R}$), the ring radius can increase sub-linearly, linearly, or super-linearly with the ring number. To preserve these three behaviors, we assume that the growth of the ring radius with the ring number follows the following pattern:

$$r_k \propto k^{\tau} \qquad k = 1, 2, \cdots, R, \tag{3.9}$$

where parameter τ captures the speed of growth of the ring radius with the ring number. However, this pattern does not include the case that one constellation point is placed at the origin. This can be solved by starting the rings index from zero, i.e., $k = 0, 1, 2, \cdots, R-1$. In other words, if one constellation point is placed at the origin ($n_1 = 1$), the rings index start from 0 to $R-1$, otherwise, they start from 1 to R. Hence, we update (3.9) to get:

$$r_k = r_0(k - \Gamma)^{\tau} \qquad k = 1, 2, \cdots, R. \tag{3.10}$$

In (3.10), parameter Γ is equal to one if one constellation point is located at the origin, i.e., $\Gamma = 1$ if $n_1 = 1$, otherwise $\Gamma = 0$. Also, r_0 is a coefficient to adjust the average symbol energy of the constellation points. In the next two sections, we demonstrate the success of model (3.10).

We can analyze the behavior of r_ks based on (3.10) in four different cases:

- $\tau = 0$: radius of all rings is equal to r_0, which results in an M-PSK constellation.

- $0 < \tau < 1$: sub-linear behavior, the inner rings get loosen, and the outer rings get compacted.
- $\tau = 1$: linear behavior, the increase in the radii is completely linear.
- $\tau > 1$: super-linear behavior, the inner rings get compacted, and the outer rings get loosen.

By finding the optimal value of τ such that (3.8) becomes minimum, we show that the proposed model (3.10) is able to estimate the optimal radii with high accuracy.

To find the optimal value of τ, first, we rewrite (3.8) according to the model (3.10):

$$LF = \sum_{\substack{1\le k\le R\\ 0\le j\le n_k-1}} \log_2\Big(\sum_{\substack{1\le k'\le R\\ 0\le i\le n_{k'}-1}} \mathrm{e}^{-\gamma.r_0^2.|d_{k_j,k'_i}|^2}\Big), \tag{3.11}$$

where γ is the shaping SNR, and d_{k_j,k'_i} is the normalized distance to r_0, i.e.:

$$|d_{k_j,k'_i}|^2 = (k-\Gamma)^{2\tau} + (k'-\Gamma)^{2\tau} - 2[(k-\Gamma)(k'-\Gamma)]^{\tau}\cos\left(\varphi_{k_j} - \varphi_{k'_i}\right), \tag{3.12}$$

where

$$\varphi_{k_j} = \frac{2\pi}{n_k} j + \theta_k \quad , \quad \varphi_{k'_i} = \frac{2\pi}{n_{k'}} i + \theta_{k'}. \tag{3.13}$$

To compute the optimal τ, we solve:

$$\frac{\partial LF}{\partial \tau} = 0, \tag{3.14}$$

where $\frac{\partial LF}{\partial \tau}$ is provided in (3.15). The complete derivation of (3.15) is provided in Appendix A.

$$\frac{\partial LF}{\partial \tau} = \sum_{\substack{1\le k\le R\\ 0\le j\le n_k-1}} \Bigg(\sum_{\substack{1\le k'\le R\\ 0\le i\le n_{k'}-1}} \Bigg(\sum_{k''=1}^{R} \frac{n_{k''}}{M}(k''-\Gamma)^{2\tau}\cdot\big[(\frac{\partial\gamma}{\partial\tau} - 2\gamma\log_e(k''-\Gamma))|d_{k_j,k'_i}|^2$$
$$+ \gamma\cdot\frac{\partial(|d_{k_j,k'_i}|^2)}{\partial\tau}\big]\Bigg) \times \mathrm{e}^{-\frac{\gamma\cdot|d_{k_j,k'_i}|^2}{\sum_{k''=1}^{R}\frac{n_{k''}}{M}(k''-\Gamma)^{2\tau}}} \times \Bigg(-\log_{\mathrm{e}} 2 \times \sum_{\substack{1\le k'\le R\\ 0\le i\le n_{k'}-1}} \mathrm{e}^{-\frac{\gamma\cdot|d_{k_j,k'_i}|^2}{\sum_{k''=1}^{R}\frac{n_{k''}}{M}(k''-\Gamma)^{2\tau}}}\Bigg)^{-1}\Bigg). \tag{3.15}$$

Also, notice that the term $\frac{\partial \gamma}{\partial \tau}$ in (3.15) considers the impact of the constellation points on SNR, and it depends on the type of communication channel. For instance, SNR of the AWGN channels does not depend on the constellation points, which yields $\frac{\partial \gamma}{\partial \tau} = 0$. In the case of non-linear fiber channels, on the other hand, the modified SNR does depend on the constellation points, which results in $\frac{\partial \gamma}{\partial \tau} \neq 0$ (we discuss this fact in more depth in Section 3.6).

In the next two sections, we apply Algorithm 1 on two well-known communication channels, the AWGN and non-linear fiber channels, and show that the PSO-based algorithm outperforms other GS methods in terms of both MI performance and convergence speed. Also, we use (3.15) to analyze the behavior of the changes in the optimal radius of rings for different experiment setups.

3.5 APSK Optimization for the AWGN channel

In this section, first, we optimize 8, 16, 32, and 64-APSK constellations for the AWGN channels using Algorithm 1, and then, we compare them with QAM constellations and other GS methods in the literature.

3.5.1 MI Performance of the Optimized APSK Constellations

Since the optimizations are performed for AWGN channels, we assume that the shaping SNR γ is constant during the optimization process. Here, the values of shaping SNR for 32 and 64-APSKs are 14 dB and 15 dB, respectively. The MI performances of the optimized constellations as well as their shape are provided in Fig. 3.2. In all cases, the optimized APSKs outperform the QAM constellations for all values of SNR. In the case of 32-point APSK, Algorithm 1 converged to a structure with 4 rings. The numbers of points located at each ring are 1, 6, 11, and 14. For MI rates between 3.7 and 4.1 bits/symbol, this constellation achieves about 0.25 dB SNR gain over 32-QAM and maximum shaping gain of 0.069 bits/symbol (0.26 dB SNR gain) at an SNR of 11.62 dB.

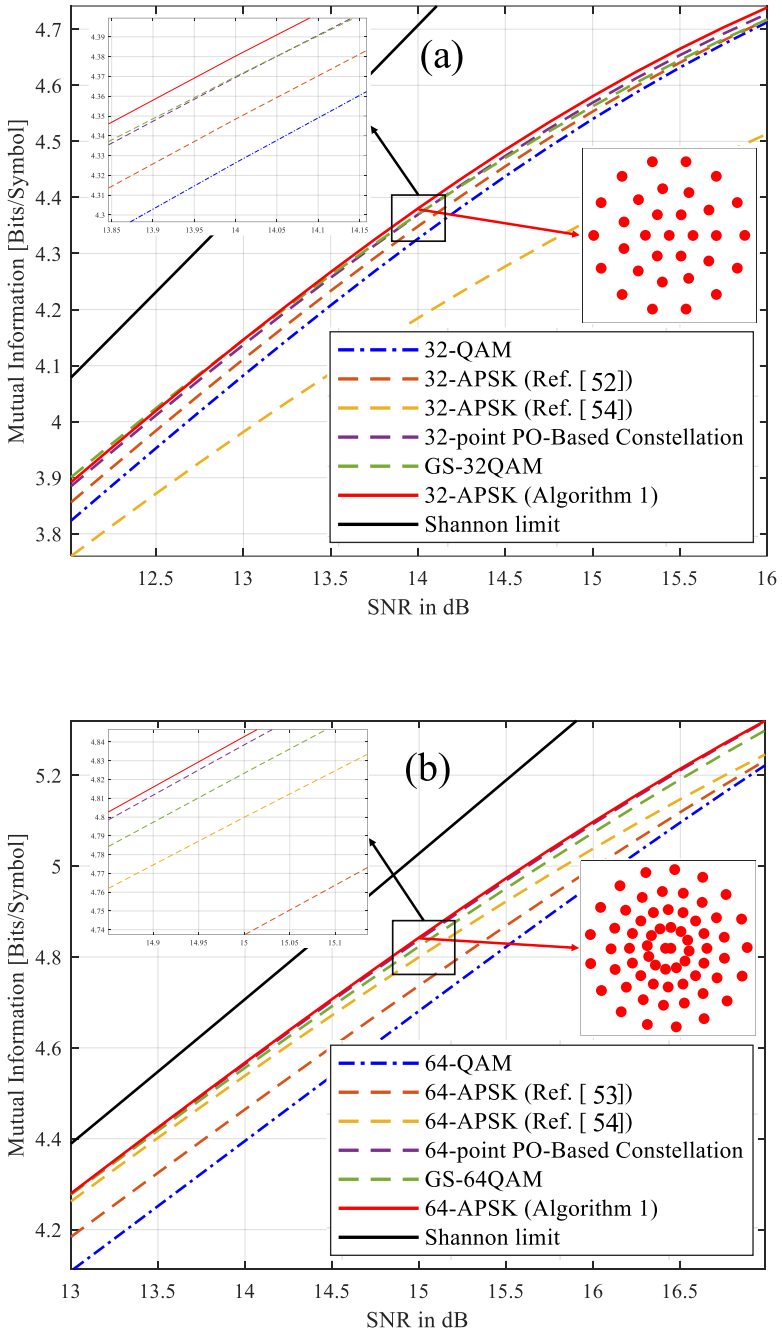

Figure 3.2: The MI performance of the optimized APSKs (using Algorithm 1) versus SNR for (a) 32-point APSK with shaping SNR of 14 dB and (b) 64-point APSK with shaping SNR of 15 dB.

For the 64-APSK, the optimized structure consists of 5 rings with $n_1 = 2$, $n_2 = 12$, $n_3 = 16$, $n_4 = 17$, and $n_5 = 17$. The maximum shaping gain of this constellation is 0.173 bits/symbol (0.6 dB SNR gain), which happens at an SNR of 13.64 dB, and it provides approximately 0.59 dB SNR gain over the 64-QAM for MI rates between 4.2 and 4.8 bits/symbol. Also, we optimize 8 and 16-point APSKs at shaping SNRs of 8 and 11.5 dB. For the 8-point APSK, among 7 possible rings, Algorithm 1 selected a 2-ring APSK with $n_1 = 1$ and $n_2 = 7$, with one point located at the origin. This optimized 8-APSK can provide the maximum shaping gain of 0.0592 bits/symbol at an SNR of 7.88 dB. Note that in this thesis, star-8-QAM is selected as the 8-QAM constellation. For the 16-point APSK, a 3-ring APSK with values of $n_1 = 1$, $n_2 = 6$, and $n_3 = 9$ has been selected. This constellation can achieve about 0.3 dB SNR gain over 16-QAM for MI rates between 2.8 and 3.2 bits/symbol and provides the maximum shaping gain of 0.072 bits/symbol at an SNR of 9.11 dB.

In addition, the MI performances of the other APSKs and GS methods published in the literature are provided in Fig. 3.2. We compare our optimized APSKs with proposed 32 and 64-point APSKs of [52], [53] and APSKs of [54]. As results suggest, our optimized APSKs outperform other constellations in all scenarios. Our 32 and 64-point APSKs outperform the APSKs of [52] and [53] by 0.032 and 0.106 bits/symbol at SNR values of 14 and 15 dB, respectively. A more detailed comparison of constellations is provided in the next subsection.

3.5.2 Comparison of Algorithm 1 with Other GS Methods

In this section, we compare Algorithm 1 with the recent GS methods in terms of MI performance and convergence speed. Here, we analyze shaping gains of the APSK constellations obtained from Algorithm 1, the Gaussian-like GS-QAM constellations [23], constellations optimized using pairwise optimization (PO) algorithm [26], [27], APSKs of [54], and APSKs with radius model of [45]. Note that [45] provides only the near-optimal radius of the APSK rings; hence, to compare our method with [45], we use $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ obtained

Table 3.3: Comparison of Shaping Gain in the AWGN Channel

M	SNR	Algorithm 1	Gaussian-like GS-QAM[23]	PO[26], [27]	APSK[45]	APSK[54]
8	7 dB	0.0568	0.0660	0.0589	0.0506	0.0568
	8 dB	0.0592	0.0620	0.0550	0.0221	0.0592
	9 dB	0.0551	0.0509	0.0450	0.0175	0.0551
16	10.5 dB	0.0655	0.0654	0.0421	0.0554	0.0536
	11.5 dB	0.0546	0.0538	0.0358	0.0377	0.0416
	12.5 dB	0.0406	0.0391	0.0273	0.0176	0.0259
32	13 dB	0.0638	0.0633	0.0546	0.0524	-0.1008
	14 dB	0.0539	0.0436	0.0432	0.0296	-0.1416
	15 dB	0.0407	0.0227	0.0292	0.0053	-0.1772
64	14 dB	0.1721	0.1609	0.1692	0.1652	0.1435
	15 dB	0.1614	0.1419	0.1570	0.1430	0.1184
	16 dB	0.1361	0.1124	0.1319	0.1067	0.0771

from Algorithm 1. In addition, the shaping gains are computed based on $\Delta_{\mathrm{MI}} = R^{*}_{\mathrm{SMD}} - R^{\mathrm{QAM}}_{\mathrm{SMD}}$, where R^{*}_{SMD} and $R^{\mathrm{QAM}}_{\mathrm{SMD}}$ are the MI rate of the shaped constellation and QAM constellation, respectively. Table 3.3 provides the shaping gains of Algorithm 1 and other mentioned methods. We consider constellations of 8, 16, 32, and 64 points. For each constellation size, we compare these methods at three values of SNR. These results show the success of our optimized APSKs.

In addition to the MI performance, the convergence speed of Algorithm 1 must also be considered since it is an iterative method. Fig. 3.3 demonstrates the convergence of Algorithm 1 for the optimized APSK constellations. For the constellation size of 8, 16, 32, and 64, to make sure that a good constellation is achieved, approximately 20, 50, 100, and 200 iterations are needed, respectively. This indicates that the required number of iterations almost linearly changes with the constellation size, which is much faster than PO algorithm that needs roughly M^2 iterations to converge to a good constellation [38].

3.5.3 Analysis of the Radius of APSK Rings

In this section, first, we explore the accuracy of Equation (3.15) and radius model of (3.10) in computing the near-optimal radius of APSK rings. After that, according to (3.15), we analyze how the optimal radius of rings are

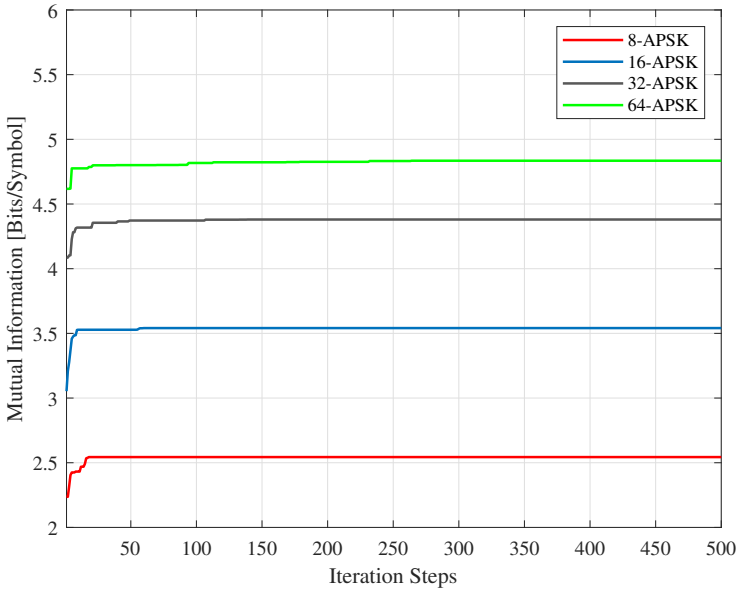

Figure 3.3: Convergence speed of Algorithm 1 for APSK constellations with size of 8, 16, 32, and 64. The optimization is performed for the AWGN channel.

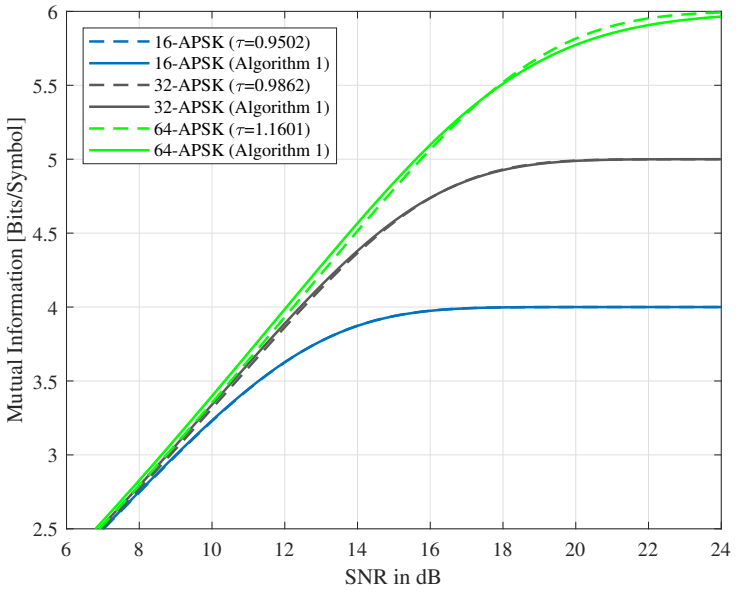

Figure 3.4: MI comparison between the APSK constellations obtained form Algorithm (1) and Equation (3.15) in the AWGN channel.

changing for different values of SNR.

To validate (3.15) and (3.10), we compare the MI performance of the optimized APSK constellations obtained in Section 3.5.1 with constellations of model (3.10), i.e., we use the optimized sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ that obtained from Algorithm (1) to compute the near-optimal radius of rings at the corresponding shaping SNRs according to Equation (3.15) and model (3.10), then, we compare the MI performance of these two constellation to explore the accuracy of (3.15) and (3.10). Notice that the values of shaping SNR are the same as Section 3.5.1 and are equal to 11.5, 14, and 15 dB for 16, 32, and 64-APSKs, respectively. For the 8-point APSK, since the structure obtained in Section 3.5.1 consists of two rings with one point at the origin, the optimal radius of the second ring does not depend on τ. Hence, in this section, we do not consider the optimized 8-APSK. The MI comparison of these two methods are provided in Fig. 3.4. For 16, 32, and 64-point APSKs, the optimal value of τ, according to (3.15), are 0.9502, 0.9862, and 1.1601, respectively. Based on Fig. 3.4, model (3.10) is quite capable of computing the near-optimal radius of APSK rings. In the case of 16-point APSK, model (3.10) strongly concurs with Algorithm (1). For 32 and 64-point APSKs, there is a tiny gap between the MI performance of Algorithm (1) and model (3.10), and the error between these two methods is less than 0.4% and 1%, respectively.

Since the proposed model for the radius of APSK rings is strongly in agreement with Algorithm (1), we can use this model to analyze how the optimal radius of APSK rings changes for different values of SNR. To do so, we consider the same sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ used in Fig. 3.4. Then, given these sets, we compute the optimal values of τ for different channel SNRs. Fig. 3.5 provides the optimal values of τ for SNR values between 5 and 20 dB. According to this figure, for 32 and 64-point APSKs, τ is monotonically decreasing in SNR. For 16-point APSK, τ is monotonically decreasing in SNR < 10 dB, and for SNR values of higher than 10 dB, τ increases with a very slow rate. As a result, generally speaking, τ increases as the channel SNR decreases. In other words, as the shaping SNR decreases, the optimized radius of APSK rings grows in a more super-linear manner. This is completely in agreement with the fact that

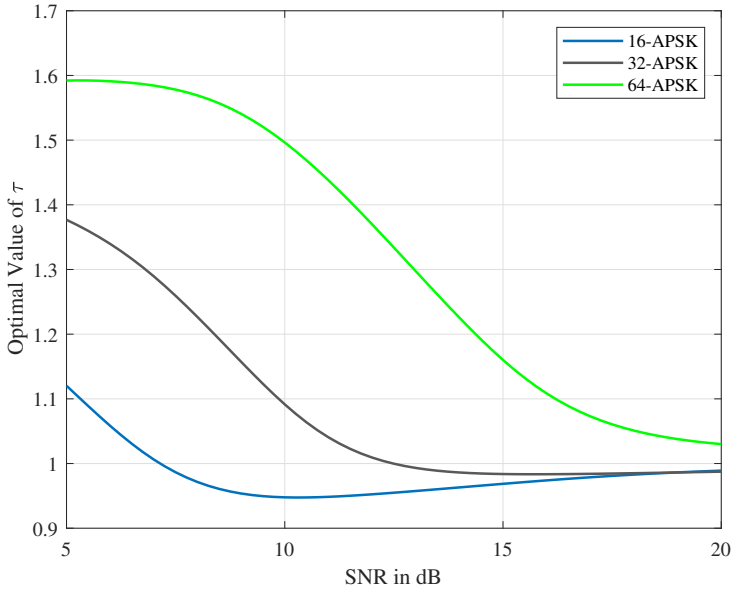

Figure 3.5: Optimal values of τ for 16, 32, 64-APSK constellations for the AWGN channel.

in the low-SNR regimes, APSK points try to mimic a low-variance Gaussian source.

3.6 APSK Optimization for the Non-Linear Fiber Channel

In this section, we optimize APSK constellations for the enhanced Gaussian noise (EGN) model [10] of the fiber channel. We also investigate the effects of GS in reach increase and MI gain. Finally, we analyze how the optimal radius of APSK rings changes in a fiber link and compare it with the AWGN channel.

As mentioned in Chapter 2, assuming all WDM channels are spaced equally and all with the same modulation format, equal symbol rate, and launch power, it can be shown that there is an optimal launch power $P_{\text{ch}}^{\text{opt}}$ at which the modified SNR of (2.28) becomes maximum [15]. By substituting the optimal power $P_{\text{ch}}^{\text{opt}}$ in (2.28), the maximum SNR can be obtained as follows [15]:

$$\text{SNR}_{\text{NL}}^{\text{max}} = \frac{2}{3} \times \sqrt[3]{\frac{1}{2 \cdot P_{\text{ASE}}^2 \cdot \eta}}. \tag{3.16}$$

3.6.1 Optimization Procedure for the EGN-Model

To optimize APSK constellations for the fiber channel based on the EGN-model, we try to maximize MI at the SNR of (3.16), i.e., we set the shaping SNR as follows:

$$\gamma = \mathrm{SNR}_{\mathrm{NL}}^{\max}(\chi). \tag{3.17}$$

Considering (3.17), we observe that the shaping SNR depends on the constellation χ in the fiber channel. Using the definition of (3.17), APSK constellations are optimized for the fiber channel in a manner that increase the MI rate while the impacts of shaping on the NLIN power are considered. Also, according to (3.17), we can compute $\frac{\partial \gamma}{\partial \tau}$, whose formula is provided in (3.18). Note that the complete derivation of (3.18) is provided in Appendix A. Interestingly, based on (3.18), the maximum value of (3.16) happens at $\tau = 0$, which means that PSK constellations introduce the least amount of NLIN into the system.

$$\frac{\partial \gamma}{\partial \tau} = -\frac{\gamma}{3\eta} \times \Bigg(\sum_{k=1}^{R} \sum_{k'=1}^{R} n_k n_{k'} \cdot \log_{\mathrm{e}} \big(\frac{k-\Gamma}{k'-\Gamma}\big) \cdot ((k-\Gamma)^2 (k'-\Gamma))^{2\tau}$$
$$\big[\frac{6\eta_\Psi}{M^2} \cdot (k-\Gamma)^{2\tau} + \frac{4(\eta_\Phi - 9\eta_\Psi)}{M^3} \cdot \sum_{k''=1}^{R} n_{k''} (k''-\Gamma)^{2\tau}\big] \Bigg) \times \big[\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau}\big]^{-4} \tag{3.18}$$

According to (3.16), we optimize APSK constellations for the fiber channel at the shaping SNR of (3.17). We discuss the results of the optimizations in Sections 3.6.3, 3.6.4. Also, in Section 3.6.5, we analytically compute and analyze the near-optimal radius of APSK for the fiber channel. Before that, we define the system parameters in the next section.

3.6.2 System Parameters and Simulation Procedure

We simulate a multi-span fiber system with parameters of Table 3.4. APSK / QAM symbols transmitted over all WDM channels, and x and y-polarization are i.i.d with uniform distribution. The central channel is the channel-under-test (CUT). Propagation of the signal over the fiber link is simulated using the split-step Fourier method (SSFM). The fiber link is made up of identical spans.

Table 3.4: System and Simulation Parameters

Parameter	**Value**
Modulation format	M-APSK/QAM
Polarization	Dual-polarization
Symbol rate	32 GBaud
WDM channels	9
WDM spacing	40 GHz
Pulse shape	Root-raised cosine (RRC)
RRC roll-off	0.01
Attenuation	0.22 dB/km
Dispersion	16.7 ps/nm/km
Non-linear coefficient	1.3 1/W/km
Span length	100 km
Amplification	EDFA
EDFA noise figure	5 dB
SSFM step size	0.1 km
Oversampling factor	16
APSK/QAM symbols per WDM ch.	2^{19}

At the end of each span, the loss is exactly compensated for by an erbium-doped fiber amplifier (EDFA), then, the ASE noise is added to the amplified signal. At the receiver, the central channel is filtered, and the chromatic dispersion is compensated for. After that, the signal passes through the matched filter, then, the filtered signal is downsampled to compute the received symbols. Finally, for each polarization, the MI rate between the transmitted and received symbols is computed according to the Monte Carlo simulations using (2.42). To compute the total MI rate, we sum over MI of x and y-polarization.

3.6.3 Reach Increase from GS

In this section, we optimize polarization-multiplexed (PM) APSK constellations at a specific distance (at the maximum modified SNR of that distance according to (3.16)) and compare their MI performance with PM-QAM constellations. We optimize 8, 16, 32, and 64-point PM-APSKs for spans 35 to 85 with a 10-span step, 15 to 55 with a 10-span step, 10 to 34 with an 8-span step, and 10 to 28 with a 6-span step, respectively. The MI performances of the optimized PM-APSK constellations and PM-QAM according to the EGN-model of fiber and SSFM simulations are provided in Fig. 3.6 and

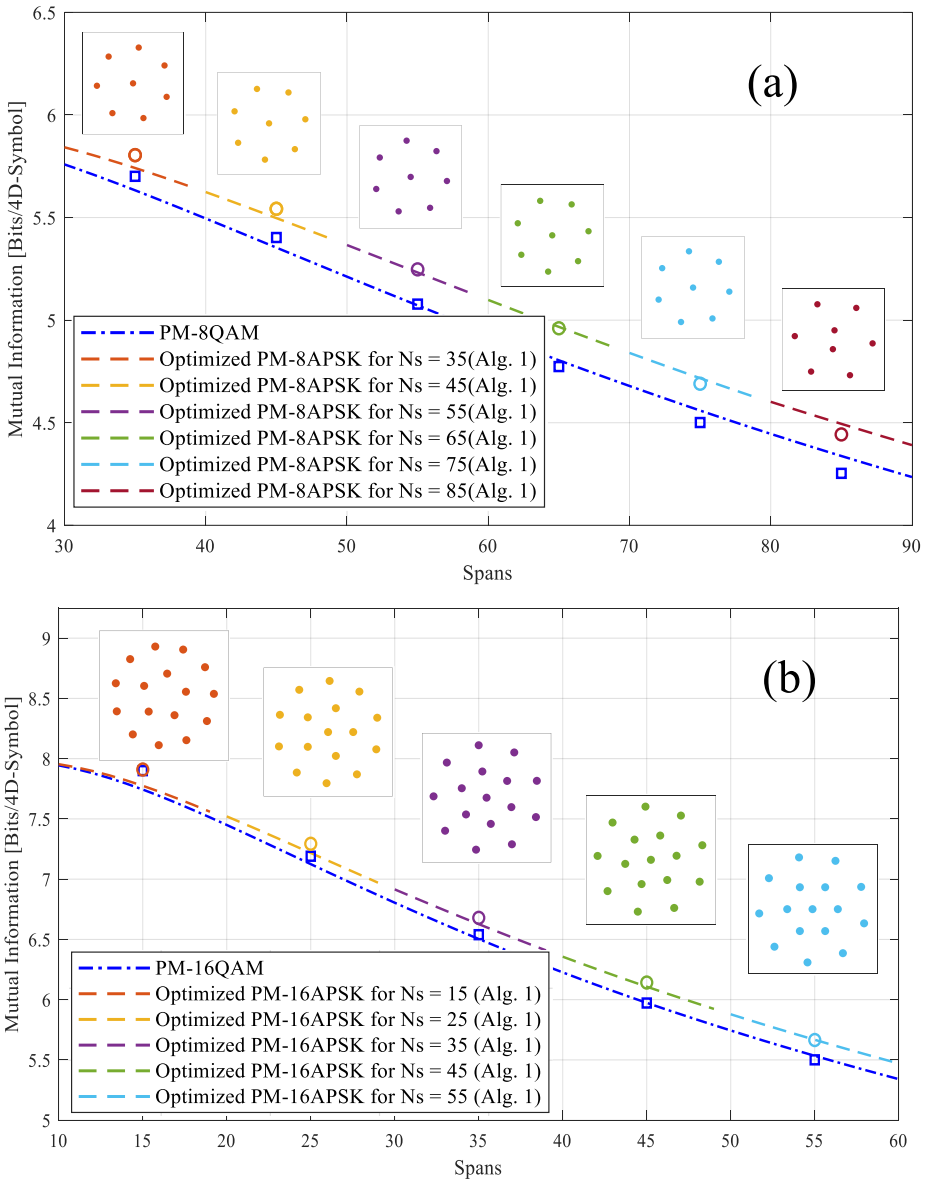

Figure 3.6: MI performance of the optimized PM-APSKs (using Algorithm 1) versus span number for (a) 8-point PM-APSKs and (b) 16-point PM-APSKs. Notice that the MI performance of the optimized PM-APSKs, and PM-QAMs according to the SSFM simulations are shown with circle, square markers, respectively.

Fig. 3.7. Based on Fig. 3.6 and Fig. 3.7, EGN-model is in good agreement with the SSFM simulations. The maximum error between these two methods is less than 2%, which shows that the EGN-model is a reliable tool for computing the MI performance of the fiber channel. In Fig. 3.6 (a), the MI performances of the optimized PM-8APSKs compared to the PM-star-8QAM constellation are

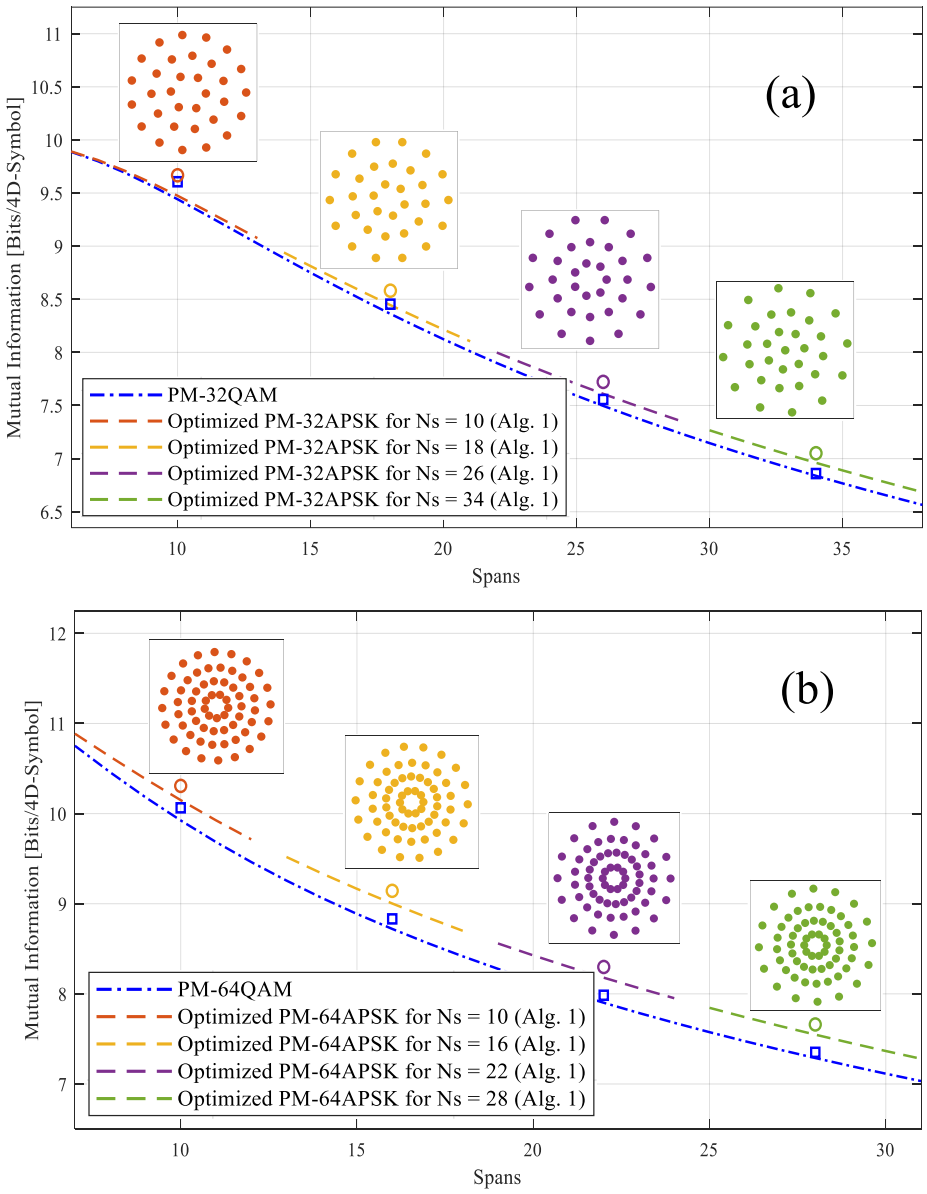

Figure 3.7: MI performance of the optimized PM-APSKs (using Algorithm 1) versus span number for (a) 32-point PM-APSKs and (b) 64-point PM-APSKs. Notice that the MI performance of the optimized PM-APSKs, and PM-QAMs according to the SSFM simulations are shown with circle, square markers, respectively.

provided. The MI rate of each optimized PM-APSK is shown within 10 spans. For instance, if a PM-8APSK is optimized for span 55, we show its MI rate for spans 50 to 60. For MI rates between 4.4 and 5.2 bits/4D-symbol, the optimized PM-8APSKs give approximately a 600 km (12.0%) reach improvement compared to PM-star-8QAM. For spans between 50 and 80, the shaping gain

of the optimized PM-8APSKs is almost 0.18 bits/4D-symbol. Also, according to the insets of Fig. 3.6 (a), Algorithm 1 choose 2-ring APSKs for spans 30 to 90; however, for span 85, the APSK structure changes from $(n_1 = 1, n_2 = 7)$ to $(n_1 = 2, n_2 = 6)$. This happens because the NLIN power increases with the fiber length, and Algorithm 1 tries to reduce NLIN by placing more constellation points in the low-power levels. In Fig. 3.6 (b), the MI performances of the optimized PM-16APSKs are provided. For MI rates between 5.5 and 6.5 bits/4D-symbol, approximately a 300 km (8.0%) reach improvement over PM-16QAM is achieved, and shaping gains of 0.17 bits/4D-symbol are obtained for spans 35 to 55. According to Fig. 3.7 (a), optimized PM-32APSKs give almost 200 km reach improvements over PM-32QAM for MI rates between 6.5 and 7.5 bits/4D-symbol. Also, our optimized PM-32APSKs provide shaping gains of 0.17 bits/4D-symbol for spans 25 to 35. According to Fig. 3.7 (b), using optimized PM-64APSKs, approximately 200 km reach improvements over PM-64QAM can be achieved for MI rates between 7.5 to 8.5 bits/4D-symbol as well as shaping gains of 0.31 bits/4D-symbol for spans 15 to 30. Considering insets of Fig. 3.6 (b), Fig. 3.7 (a), and (b), similar to 8-point PM-APSKs, as the fiber length increases, Algorithm 1 puts more points in the low-power levels and less point in the high-power levels. Moreover, the radius of APSK rings increases with ring number in a super-linear manner. We discuss these changes more in-depth in Section 3.6.5.

3.6.4 MI Gain from GS

In this section, we discuss shaping gains of the optimized 8, 16, 32, and 64-point PM-APSKs for spans 65, 35, 18, and 16, respectively. As well, we compare the performance of the optimized APSKs with Gaussian-like GS-QAMs, PO-based constellations, as well as the APSKs of [52]–[54]. Note that the shaping SNR for these GS methods is set to the maximum modified SNR of the corresponding QAM constellation. In addition, we discuss the effects of GS on NLIN power. The MI performances of the optimized PM-APSKs are provided in Fig. 3.8 and Fig. 3.9 for launch powers between -6 dBm to 4 dBm. Results of the SSFM simulations and EGN-model are shown with markers and

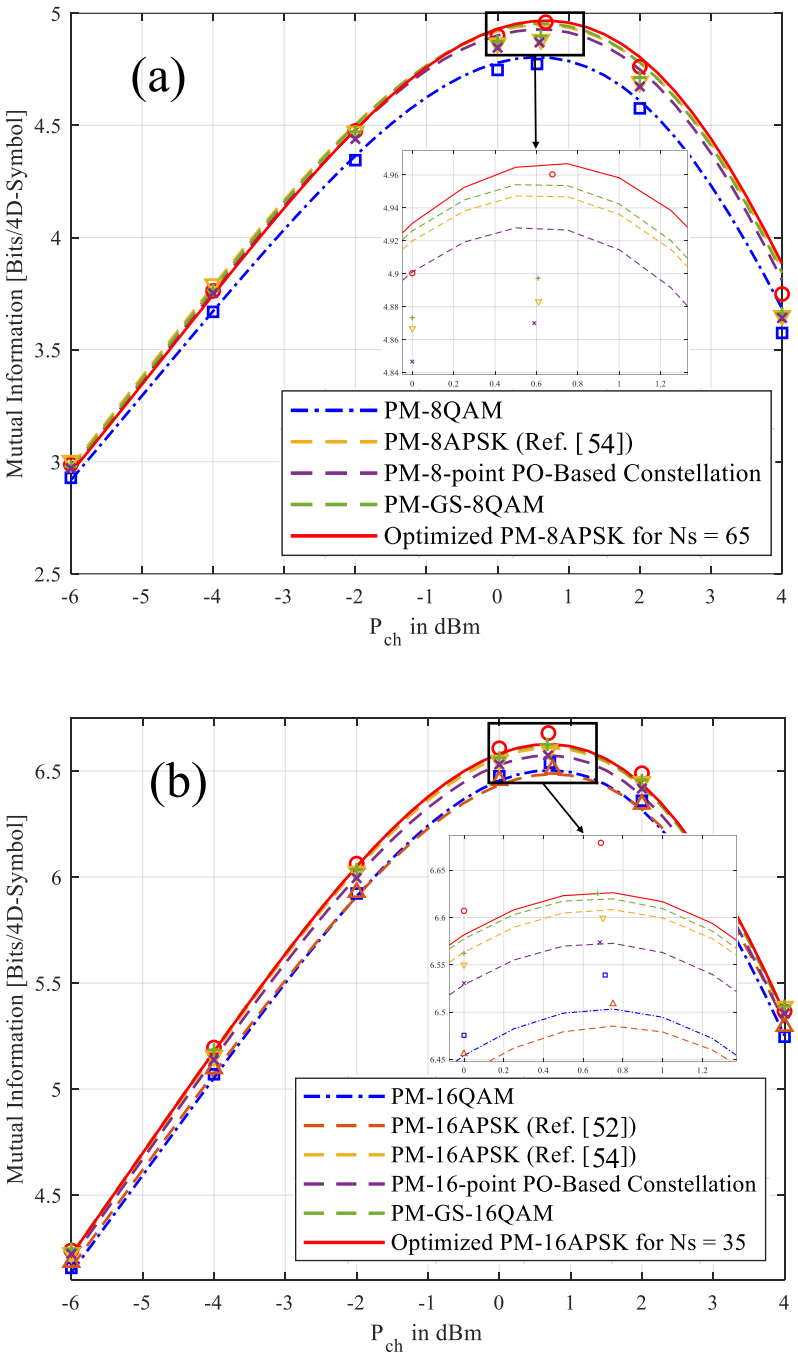

Figure 3.8: MI performance of the optimized PM-APSKs (using Algorithm 1) and other shaped constellations versus the launch power for (a) 8-point constellations at span 65 and (b) 16-point constellations at span 35. Note that SSFM simulations are shown with markers.

dashed lines, respectively, indicating that EGN-model concurs with SSFM simulations. According to Fig. 3.8 and Fig. 3.9, at the corresponding optimal launch power, the optimized 8, 16, 32, and 64-point PM-APSKs provide

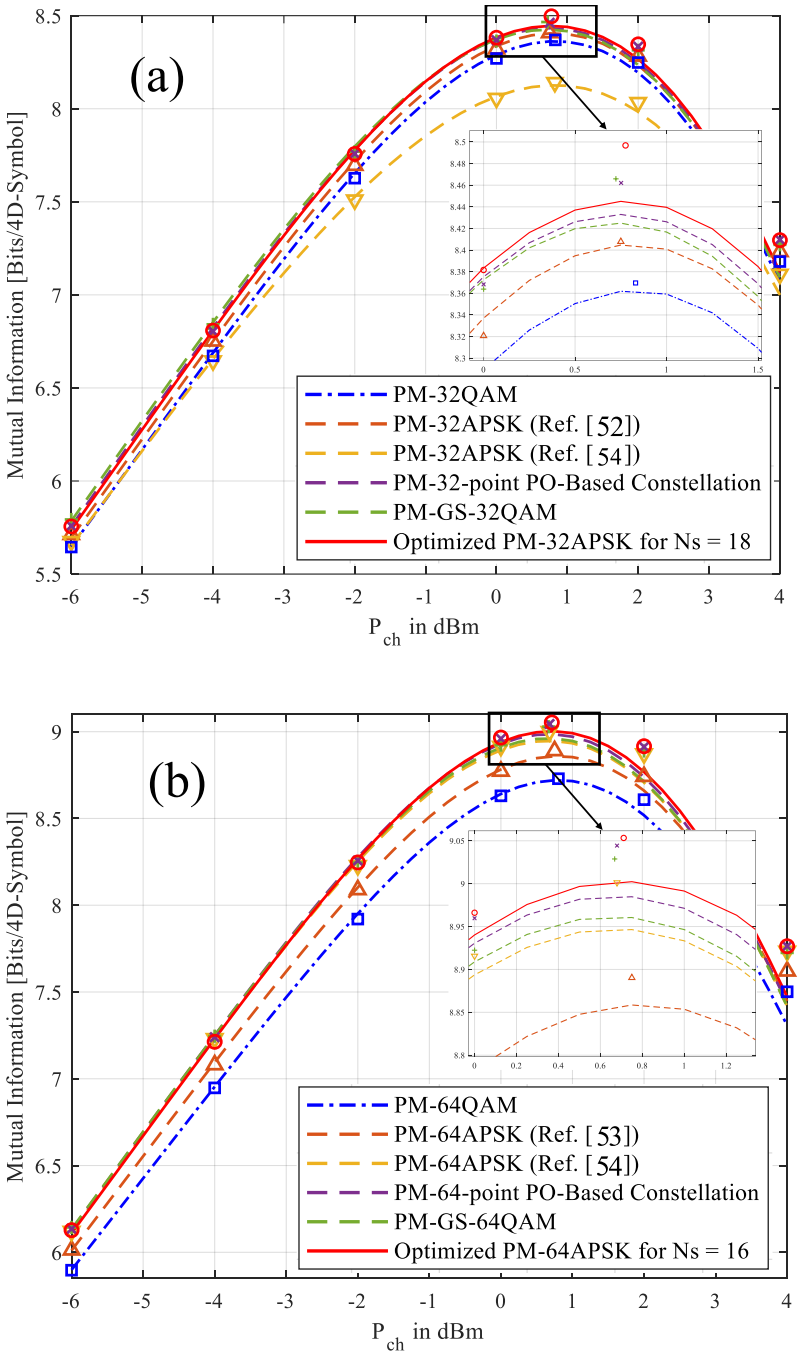

Figure 3.9: MI performance of the optimized PM-APSKs (using Algorithm 1) and other shaped constellations versus the launch power for (a) 32-point constellations at span 18 and (b) 64-point constellations at span 16. Note that SSFM simulations are shown with markers.

shaping gains of 0.1866, 0.1398, 0.1656, 0.3137 bits/4D-symbol over the PM-QAM constellations. In addition, our optimized PM-APSKs outperform other GS methods in all scenarios. Among PM 8, 16, and 32-point constellations,

Gaussian-like GS-QAMs have the best performance among benchmark algorithms, and our optimized PM-APSKs outperform them by 0.063, 0.054, and 0.031 bits/4D-symbol, respectively. For PM 64-point constellations, the PM-64-point-PO-based constellation has the best performance among benchmark algorithms, and our optimized APSK outperforms it by 0.009 bits/4D-symbol. Also, one can observe the impact of GS on the NLIN power. Based on (3.16), we can conclude that the maximum modified SNR is monotonically decreasing in η. As a result, modulations with higher NLIN power have lower $\mathrm{SNR}_{\mathrm{NL}}^{\max}$, or equivalently, have lower $P_{\mathrm{ch}}^{\mathrm{opt}}$. According to Fig. 3.8 and Fig. 3.9, the optimal launch power of PM-8APSK is higher than PM-star-8QAM, which shows that the NLIN power has decreased due to GS. In the case of 16, 32, and 64-point PM-APSKs, on the other hand, the optimal launch power is lower than the corresponding PM-QAM optimal launch power, which shows that in these cases, GS increases the NLIN power. To compare the shaped constellations in terms of NLIN, we plot $\mathrm{SNR}_{\mathrm{NL}}$ for average launch powers between -1.5 dBm to 3 dBm. The results are provided in Fig. 3.10. As Fig. 3.10 indicates, the optimized PM-64APSK introduces lower NLIN to the system than PM-64-point PO-based constellation, PM-64GS-QAM, and PM-64APSK of [54]. Although PM-64APSK of [53] introduces the least amount of NLIN among the shaped constellations, it does not provide MI gains as high as other shaped constellations. In other words, PM-64APSK of [53] does not shape the geometry of the constellation that much to be considered as a shaped constellation. Algorithm 1, however, optimizes APSKs in a manner that the optimized APSK both introduces low additional NLIN and provides very high MI gains. In Fig. 3.10, the optimized PM-64APSK has almost 0.05 dB higher SNR than PM-64-PO and almost 0.11 dB higher SNR than PM-64GS-QAM, and PM-64APSK of [54].

3.6.5 Analysis of the Radius of APSK Rings

In the following, first, we validate the radius model of (3.10) for the fiber channel according to Equations (3.15) and (3.18). Then, we discuss the differences between the AWGN and fiber channel in terms of the optimal radius of the

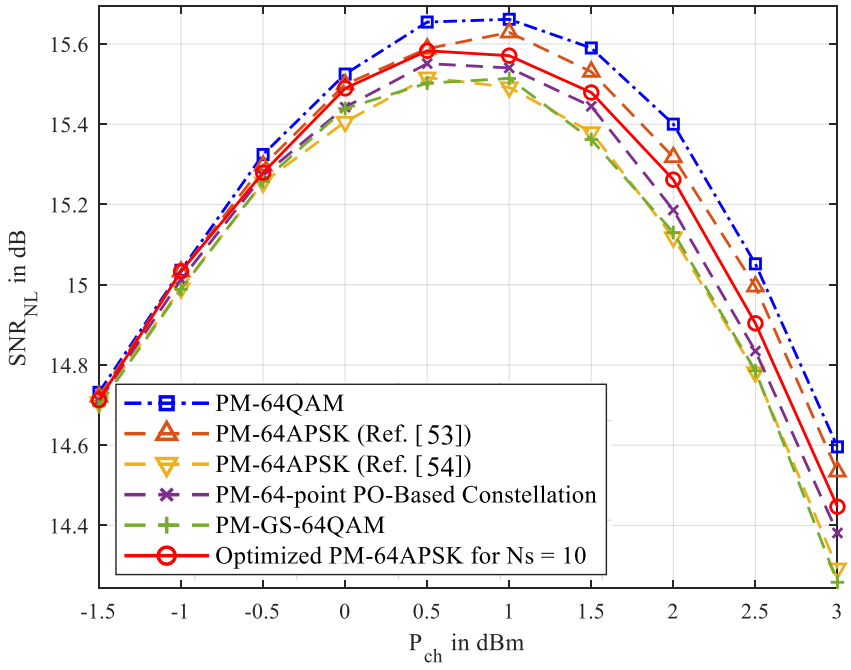

Figure 3.10: Values of $\mathrm{SNR}_{\mathrm{NL}}$ of 64-point constellations at span 10.

APSK rings.

To validate (3.15), we compare MI performance of the constellations obtained from Algorithm 1 with the same corresponding constellation but with the radii obtained from (3.10). We use the sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ of the optimized 8, 16, 32, and 64-point PM-APSK constellations for spans 85, 55, 34, and 28, respectively. The MI comparisons of Algorithm 1 and radius model of (3.10) are provided in Fig. 3.11. For 8, 16, 32, and 64-point PM-APSKs, the optimal value of τ, based on radius model of (3.10) and Equations (3.15) and (3.18), is 1.6224, 0.9495, 0.9564, and 1.1065, respectively. The maximum error between these two methods is less than 0.5%, which indicates that radius model of (3.10) strongly concurs with Algorithm 1.

To compare the optimal τ of the fiber and AWGN channel, for each M-APSK, we need to analyze identical sets of $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ in the fiber and AWGN channel. To do so, we use sets $\{n_k\}_{k=1}^{R}$ and $\{\theta_k\}_{k=1}^{R}$ of optimized 8, 16, 32, 64-point PM-APSKs for spans 85, 15, 10, and 16 obtained in Section 3.6.3, respectively. For 8, 16, 32, and 64-point APSKs, we compute the $\mathrm{SNR}_{\mathrm{NL}}^{\mathrm{max}}$ at spans 5 to 85. Then, for each constellation, we calculate the optimal

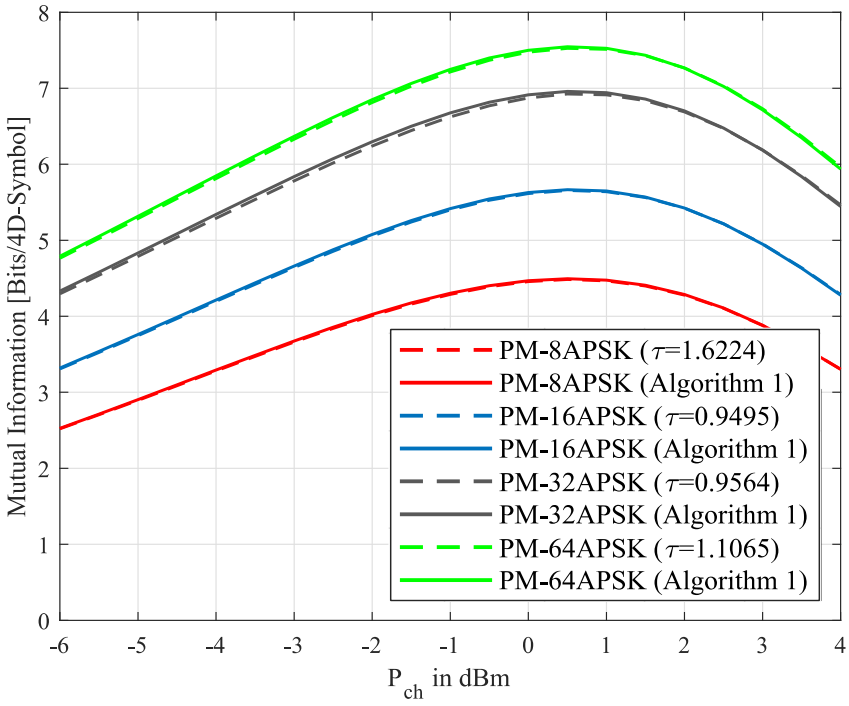

Figure 3.11: MI comparison between the PM-APSK constellations obtained form Algorithm 1 and Equation (3.15) in the fiber channel.

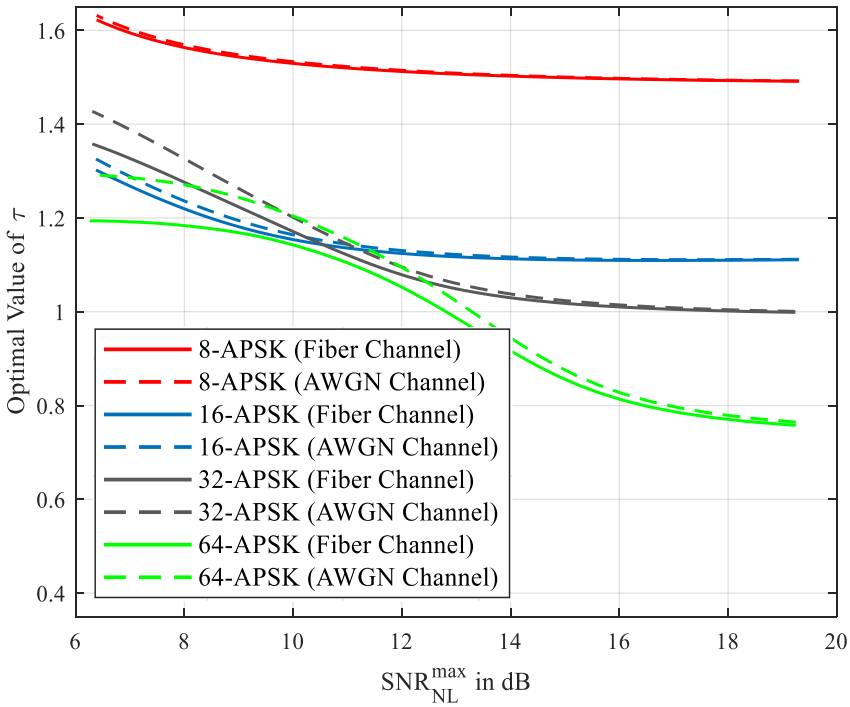

Figure 3.12: Comparison between the fiber channel and AWGN channel in terms of τ.

values of τ for the AWGN channel at the shaping SNRs equal to the obtained $\mathrm{SNR}_{\mathrm{NL}}^{\max}$s. By plotting the optimal values of τ for the fiber and AWGN channel in one figure, Fig. 3.12 is obtained. In the fiber channel, for all constellations, τ is monotonically decreasing in $\mathrm{SNR}_{\mathrm{NL}}^{\max}$ (or equivalently, monotonically increasing in span number), which shows that as the fiber length increases, the optimized radius of APSK rings grows faster. Also, in the AWGN channel, τ is monotonically decreasing in $\mathrm{SNR}_{\mathrm{NL}}^{\max}$; however, there are some differences between the fiber and AWGN channel. Based on Fig. 3.12, for high values of $\mathrm{SNR}_{\mathrm{NL}}^{\max}$ (for short fiber links), τ of the fiber and AWGN channel is almost equal, which means that for short fibers, the AWGN and fiber channels are similar. This was expected since, in the short fibers, ASE is dominant to NLIN. For low values of $\mathrm{SNR}_{\mathrm{NL}}^{\max}$, on the other hand, the difference between the fiber and AWGN channel becomes more evident. In 8 and 16-point APSKs, optimal values of τ of the AWGN channel are slightly greater than the optimal τs of the fiber channel. For 32 and 64-point APSKs, the optimal values of τ of the AWGN channel are much greater than the fiber channel. Hence, in general, the optimal values of τ for the AWGN channels are greater than the fiber channel. This is because for the fiber channel, Algorithm 1 is more cautious about increasing the radius of APSK rings in a super-linear manner. In the fiber channel, if we set the radius of APSK rings exactly equal to the radii optimized for AWGN channels, some constellation points are placed at very high-power levels that results in a significant increase in the NLIN power. Moreover, Fig. 3.12 clearly indicates that the constellations that are shaped based on the AWGN channel are not appropriate to be used in the highly nonlinear regimes of the fiber channel. To substantiate this claim, we compare the Kurtosis of the APSKs optimized for the AWGN and the fiber channel. The Kurtosis of a constellation (let say X) is defined as follows:

$$\mu_4 = \frac{\mathbb{E}\big\{|X - \mathbb{E}\{X\}|^4\big\}}{\Big(\mathbb{E}\big\{|X - \mathbb{E}\{X\}|^2\big\}\Big)^2}. \tag{3.19}$$

A larger μ_4 means increased NLIN in the fiber [24], [55], [56]. Hence, we expect that the APSKs optimized for the fiber channel have lower μ_4 than the

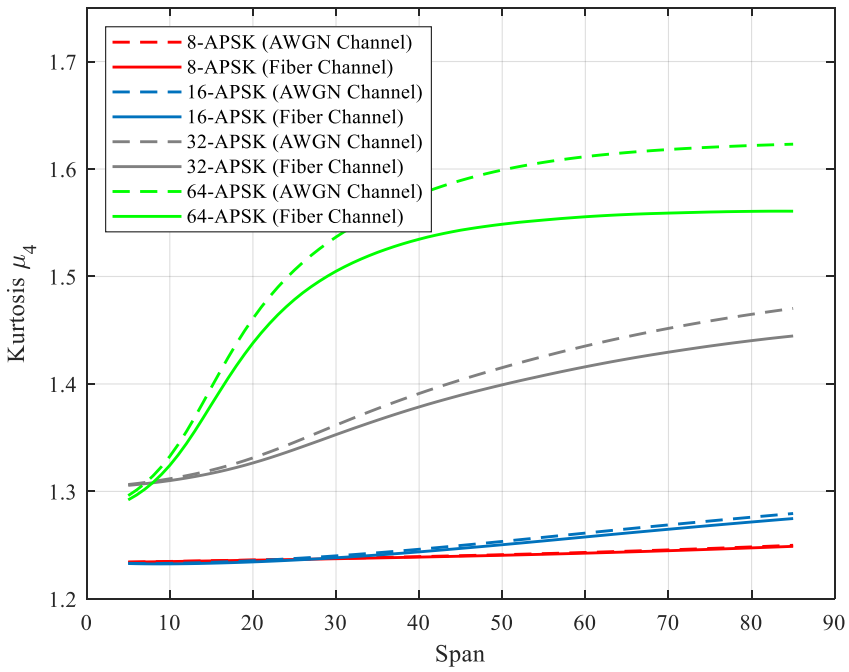

Figure 3.13: Kurtosis μ_4 of optimized APSKs for the fiber channel (solid lines) and the AWGN channel (dashed lines).

ones optimized for the AWGN channel. To show that, we compute μ_4 of 8, 16, 32, and 64-point APSKs optimized for fiber and AWGN channels for spans between 5 and 85. The results are provided in Fig. 3.13. Based on Fig. 3.13, as the span number increases (NLIN increases), the difference between μ_4 of the optimized APSKs for the AWGN and fiber channel increases as well. Meaning that at highly non-linear regimes of fiber, AWGN-based GS will result in higher NLIN than EGN-based GS.

Chapter 4

Hybrid Probabilistic-Geometric Shaped Constellations for the Fiber Channel

4.1 Introduction

In the previous chapter, we introduced a GS method that maximizes the MI rate of the fiber channel. However, the MI rate can be achieved if the optimal detection rule is employed. In most scenarios, the optimal detection rule of a geometrically shaped constellation is hard to find, and the detection rule of the BICM systems is employed instead. Consequently, in GS, there is a gap between the mutual information and generalized mutual information rate (the AIR of the BICM systems) due to employing sub-optimal detection. In addition, there is a rate loss in PS transmission systems due to using distribution matchers with finite blocklengths. As a result, in both GS and PS, the ultimate AIR cannot be achieved due to the imperfect implementation. To compensate for these losses, one can employ both PS and GS, which is known as hybrid probabilistic-geometric shaping (HPGS). However, to implement HPGS, a transmitter that is compatible with both PS and GS must be developed.

As mentioned in Chapter 2, designing a transmitter that both shapes and codes uniform data is a challenge. PAS [30] designs a transmitter that includes both PS and FEC coding. However, PAS is only suitable for square QAM constellations. For non-square QAMs, probabilistic fold shaping PFS [42] is

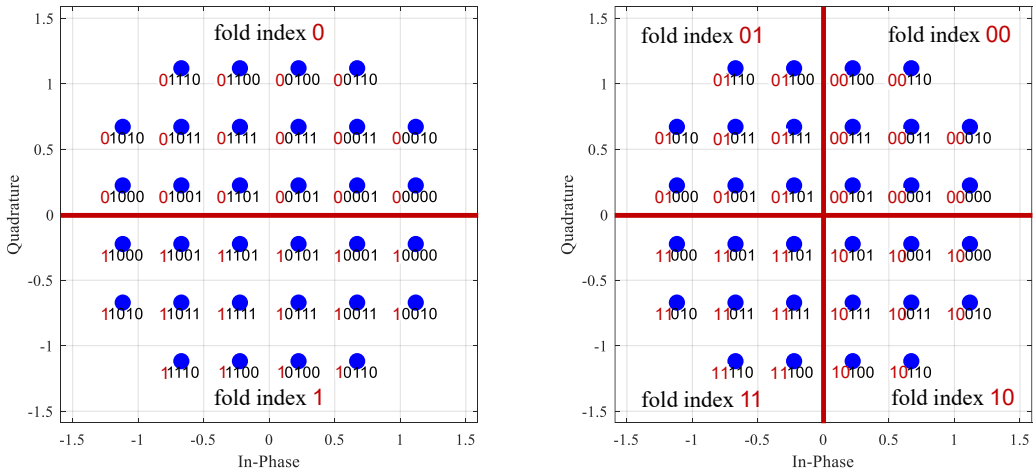

Figure 4.1: PFS of 32-QAM based on (a) 2-fold and (b) 4-fold structure.

employed.

PFS is similar to PAS; however, in PFS, two-dimensional DMs are used instead of one-dimensional ones. PFS uses the N-fold rotationally symmetrical property of the non-square QAMs to assign the FEC parity-check bits. In other words, PFS uses the $\log_2(N)$ bits out of $\log_2(M)$ of a constellation point for carrying the parity-check bits. Fig. 4.1 demonstrates 2-fold and 4-fold PFS of 32-QAM. In Fig. 4.1 (a), we consider that 32-QAM is symmetrical with respect to the in-phase axis (2-fold), and as shown in the figure, one bit out of five bits is used as the parity-check bits. Also, Fig. 4.1 (b) shows the 4-fold example of the shaping of 32-QAM, where 2 bits are reserved for parity-check bits. Although PFS is proposed for non-square QAM constellations, it can be used for geometrically-shaped constellations as well.

Since PFS considers only the N-fold rotationally symmetrical property of the non-square QAMs, it can be also used for geometrically-shaped constellations if points are N-fold rotationally symmetrical. Here, we consider 4-fold constellations, i.e., we shape the position of the points only for the first quadrant, and then, we fold them to obtain points of other quadrants. Also, we use PMFs from the family of MB distributions to shape the probability of

the points. Hence, we only need to find the optimal value of λ and use the power constraint of (2.61) to compute Δ. Moreover, during optimization, we assume that binary labeling is fixed, i.e., for a given binary labeling, we find the optimized position of the points of the first quadrant and optimal λ. We discuss our optimization scheme for HPGS in the next section.

4.2 Optimization Scheme

Our optimization scheme to find the near-optimal position of the whole points and λ is as follows: initialize the position of the points, binary labels, and λ with the corresponding parameters of probabilistic-shaped M-QAM. Then, for a point located in quadrant (I) (say x_n), we try to find two displacement vectors (Δx_n and $\Delta\lambda$) such that

$$\begin{aligned} x_n^{\text{new}} &= x_n^{\text{old}} + \Delta x_n, \\ \lambda^{\text{new}} &= \lambda^{\text{old}} + \Delta\lambda \end{aligned} \tag{4.1}$$

lead to the maximum increase in R_{BMD}. Then, using the optimal Δx_n^* and $\Delta\lambda^*$, we update the position of x_n and the MB distribution, respectively. After that, we follow the same procedure for point x_{n+1}. We repeat this procedure until the end of the last iteration. This procedure guarantees that the final constellation outperforms the corresponding probabilistic-shaped constellation. Our optimization scheme for HPGS is provided in Algorithm (2), where N_i is the number of iterations, and superscript (I) shows the corresponding values/sets of quadrant (I), and superscript $\mathcal{F}$ shows the folded values/sets of the corresponding values/sets of quadrant (I). Here, we use particle swarm optimization (PSO) [47] to find optimal Δx_n^* and $\Delta\lambda^*$. Based on our simulations, PSO outperforms other optimization methods such as gradient descent in terms of both computational efficiency and convergence speed.

The search space of particles for a point (say x_n) is as follows:

$$\vec{a}_{x_n} = [\Delta x_{x_n}^{\text{Re}}, \Delta x_{x_n}^{\text{Im}}, \Delta\lambda_{x_n}], \tag{4.2}$$

where $\Delta x_{x_n}^{\text{Re}}$ and $\Delta x_{x_n}^{\text{Im}}$ are the real and imaginary part of the displacement vector of point x_n. Also, the hyperparameters of the PSO algorithm and bounds of the entries of (4.2) are provided in Table 4.2 and Table 4.1, respectively.

Algorithm 2 Optimization Scheme for HPGS

Output: optimal χ^* and λ^*

Initialize $\chi^{(\mathrm{I})}$ with the position of the M-QAM first quadrant points.

Set $\mathcal{L}^{(\mathrm{I})}$ to the last $(m-2)$ bits of the optimized binary labeling of M-QAM obtained from binary switching algorithm (BSA) [40].

Initialize λ with the optimal λ of the probabilistically shaped M-QAM.

for $i = 1 : N_i$ **do**

  for $n = 1 : \frac{M}{4}$ **do**

    find optimal value of Δx_n and $\Delta\lambda$ such that $R_{\mathrm{BMD}}(\chi^{\mathcal{F}}, \mathcal{L}^{\mathcal{F}}, P_{X_{\lambda_n}}, \gamma^{\mathcal{F}})$ becomes maximum, where
    $\chi^{\mathcal{F}} = \big\{ \pm \mathrm{Re}(\chi_n^{(\mathrm{I})}) \pm j\mathrm{Im}(\chi_n^{(\mathrm{I})}) \big\}$,
    $\chi_n^{(\mathrm{I})} = \big\{ x_1, x_2, \cdots, x_n + \Delta x_n, \cdots, x_{\frac{M}{4}} \big\}$,
    $\mathcal{L}^{\mathcal{F}} = \big\{ [b_I, b_Q, \mathcal{L}^{(\mathrm{I})}] \big| b_I, b_Q = 0, 1 \big\}$,
    $P_{X_{\lambda_n}}$ is PMF of MB with $\lambda + \Delta\lambda$,
    $\gamma^{\mathcal{F}} = \mathrm{SNR}_{\mathrm{NL}}^{\max}(\chi^{\mathcal{F}}, P_{X_{\lambda_n}})$.
    Then, update the value of x_n and λ using optimal Δx_n^* and $\Delta\lambda^*$:
    $x_n = x_n + \Delta x_n^* \quad , \quad \lambda = \lambda + \Delta\lambda^*$.

    Scale the amplitudes of the points if energy of the constellation is not equal to one.

  end for

end for

$\chi^* = \big\{ \pm \mathrm{Re}(\chi^{(\mathrm{I})}) \pm j\mathrm{Im}(\chi^{(\mathrm{I})}) \big\}$,
$\lambda^* = \lambda$.

return χ^*, λ^*

Table 4.1: Bounds of $\vec{a}_{x_n}$ Entries

Parameter	**Minimum Value**	**Maximum Value**
$\Delta x_{x_n}^{\mathrm{Re}}$	$-\mathrm{Re}(x_n) + 0.001$	$\frac{3(\sqrt{M}-1)}{\sqrt{\frac{2}{3}(M-1)}}$
$\Delta x_{x_n}^{\mathrm{Im}}$	$-\mathrm{Im}(x_n) + 0.001$	$\frac{3(\sqrt{M}-1)}{\sqrt{\frac{2}{3}(M-1)}}$
$\Delta\lambda_{x_n}$	$-\lambda$	2.5

Table 4.2: Hyperparameters of Algorithm 2

Hyperparameter	**Value**
N_i, number of iterations	$10 \times M$
N_p, number of particles of PSO	10
N_i^{PSO}, number of iterations of PSO	60
w_{max}	0.9
w_{min}	0.2
u_1	2
u_2	2

4.3 HPGS Constellations for the Non-Linear Fiber Channel

In the previous section, we introduced our method to shape PMF and geometry of the constellation point at a given shaping SNR. As discussed in Chapter 2, the fiber channel is modulation-dependent, and during optimization of the constellation, impacts of shaping on NLIN must be considered. To do so, we follow the same procedure of Chapter 3, i.e., we set

$$\gamma = \mathrm{SNR}_{\mathrm{NL}}^{\max} = \frac{2}{3} \times \sqrt[3]{\frac{1}{2 \cdot P_{\mathrm{ASE}}^2 \cdot \eta}}. \tag{4.3}$$

In the following, we shape the geometry and PMF of the PM 32 and 64-point constellations and compare their GMI performance with both the corresponding PM-MQAMs and probabilistic shaped PM-MQAMs (PM-PS-MQAM). Also, system and simulation parameters are provided in Table 4.3. Propagation of the signal over the fiber link is simulated using the split-step Fourier method (SSFM). The fiber link is made up of identical spans. At the end of each span, the loss is exactly compensated for by an erbium-doped fiber amplifier (EDFA).

Table 4.3: System and Simulation Parameters

Parameter	**Value**
Modulation format	M-HPGS/PS-QAM/QAM
Polarization	Dual-polarization
Symbol rate	32 GBaud
WDM channels	9
WDM spacing	35 GHz
Pulse shape	Root-raised cosine (RRC)
RRC roll-off	0.01
Attenuation	0.22 dB/km
Dispersion	16.7 ps/nm/km
Non-linear coefficient	1.3 1/W/km
Span length	100 km
Amplification	EDFA
EDFA noise figure	5 dB
SSFM step size	0.1 km
Oversampling factor	16
APSK/QAM symbols per WDM ch.	2^{19}

4.3.1 Reach Increase from HPGS

The GMI performance of the HPGS constellations, PM-32-HPGS and PM-64-HPGS, versus the span number is provided in Fig. 4.2 (a) and (b), respectively. We optimize PM-32-HPGSs and PM-64-HPGSs for spans 12 to 32 and 5 to 20 with a 5-span step, respectively. GMI of each optimized PM-M-HPGS is shown within 5 spans. Based on Fig. 4.2 (a) and (b), HPGS constellations outperform both the corresponding PM-QAMs and PM-PS-QAMs in all cases. The reach improvements compared to PM-QAMs and PM-PS-QAMs are as high as 400 and 100 km, respectively. In Fig. 4.2 (a), for GMI rates between 6.5 and 8 bits/4D-symbol, PM-32-HPGSs give approximately a 400 and 100 km reach improvement compared to PM-32QAM and PM-PS-32QAM, respectively. For spans between 20 and 34, the optimized PM-32-HPGSs provide almost 0.462 and 0.085 bits/4D-symbol higher GMI rate than PM-32QAM and PM-PS-32QAM, respectively. Also, in Fig. 4.2 (b), PM-64-HPGSs provide approximately a 100 and 17 km reach improvement compared to PM-64QAM and PM-PS-64QAM for GMI rates between 9.5 and 11.5 bits/4D-symbol, respectively.

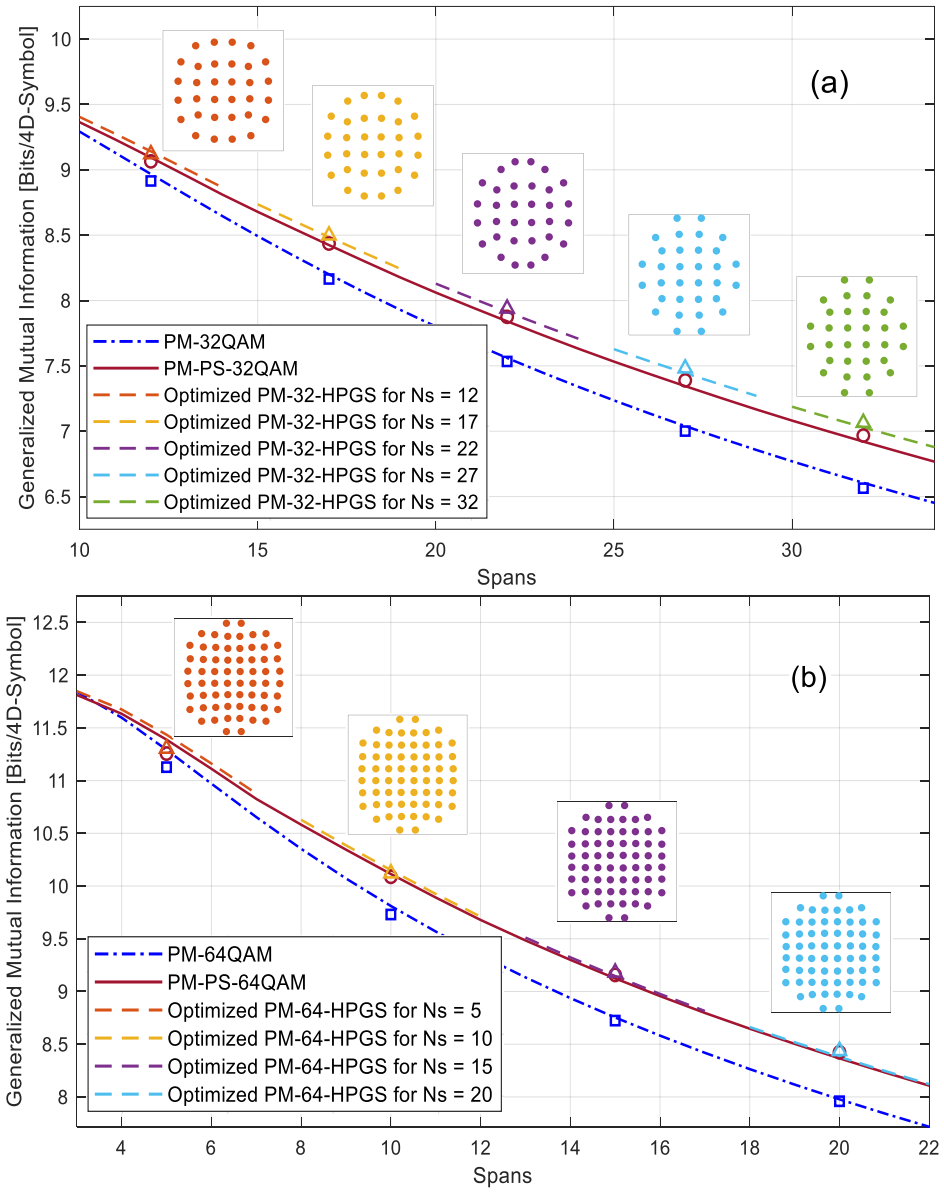

Figure 4.2: GMI performance of the (a) optimized PM-32-HPGSs and (b) PM-64-HPGSs versus span number. The GMI rate of the constellations according to the SSFM simulations are shown with markers. The optimized position of the points is shown in the insets of Fig. (a) and (b).

4.3.2 GMI Gain from HPGS

In Fig. 4.3 (a) and (b), the GMI performance of the optimized PM-32-HPGS and PM-64-HPGS versus P_{ch} is shown, respectively. Also, GMI of the HPGS constellations optimized for the AWGN channel(during optimization, γ is assumed to be constant) is provided. Based on Fig. 4.3 (a) and (b), HPGS

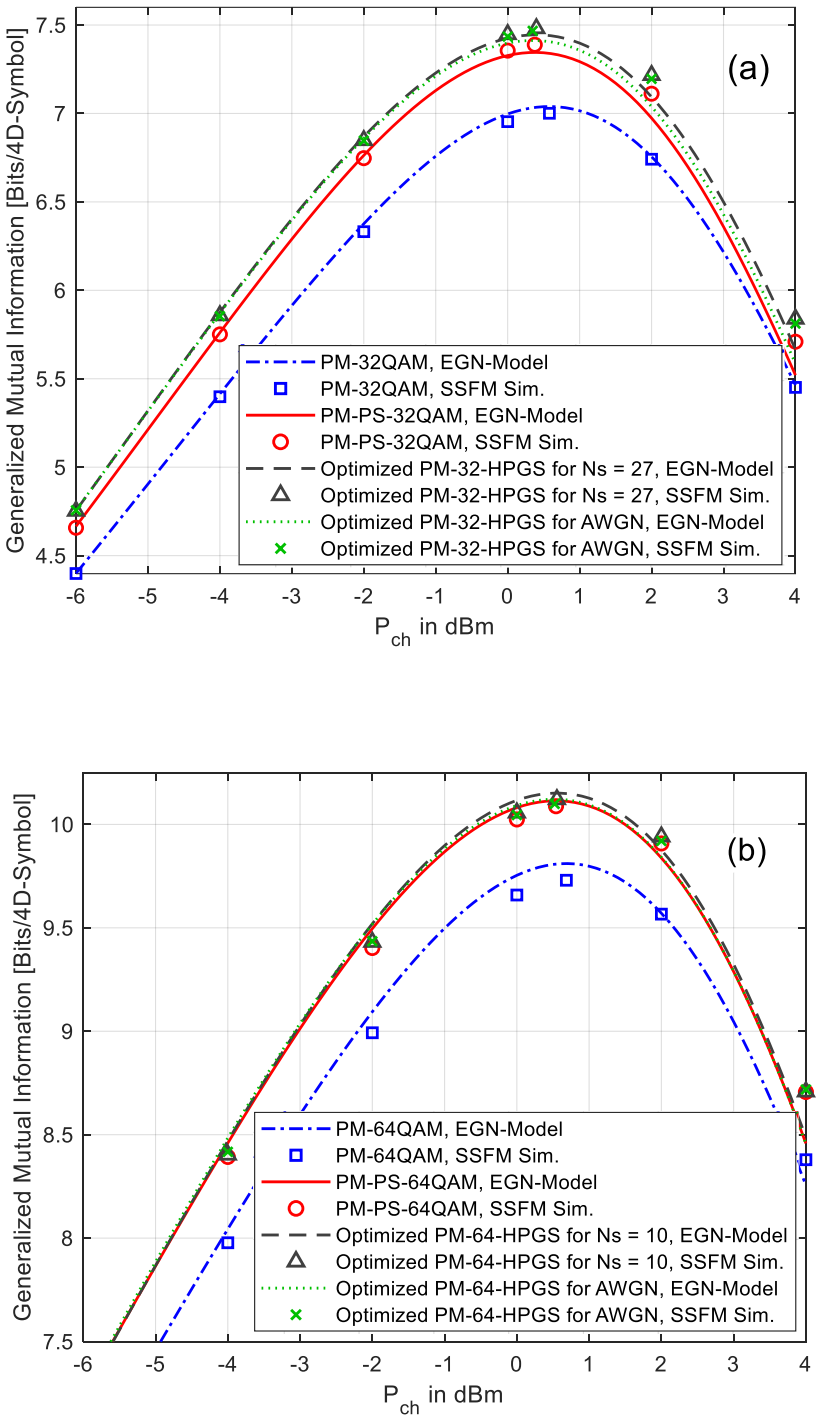

Figure 4.3: GMI performance of the (a) optimized PM-32-HPGSs and (b) PM-64-HPGSs versus the launch power for (a) PM-32-HPGS at span 27 and (b) PM-64-HPGS at span 10. The GMI rate of the constellations according to the SSFM simulations are shown with markers.

constellations outperform PM-QAMs and PM-PS-QAMs in all scenarios and provide gains up to 0.48 and 0.092 bits/4D-symbols, respectively. Moreover, in the non-linear regimes (high launch powers), constellations optimized for the fiber channel have much higher GMI than ones optimized for AWGN, but in the linear regimes (low launch powers), they exhibit similar performance. PM-32-HPGS provides gains of 0.48 and 0.092 bits/4D-symbols over PM-32QAM and PM-PS-32QAM, respectively. Also, PM-64-HPGS outperforms PM-64QAM and PM-PS-64QAM by 0.391 and 0.033 bits/4D-symbols, respectively. Note that the reported gains are computed at the corresponding optimal P_{ch}. The reason that PM-64-HPGSs provide less gain compared to PM-32-HPGSs is that for 64-point constellations, the gap to the channel capacity is very low, and hence, there is no room for significant gains. However, PM-64-HPGSs are still capable of outperforming PM-PS-64QAMs in all scenarios.

Chapter 5

Conclusion and Future Research Directions

5.1 Conclusions

In this thesis, we considered the optimization of the constellations for the non-linear fiber channel in order to increase AIR. We discussed that the fiber channel is a modulation-dependent channel, meaning that the type of modulation used to transmit information affects the degree to which the optical signal is distorted. This makes optimizing constellations for the fiber channel complicated since there is no closed-form solution to NLSE (the partial differential equation that governs the propagation of signals inside the optical fiber). In this work, we proposed constellation shaping methods to address this issue.

In Chapter 3, we introduced a GS method based on the PSO algorithm to maximize the MI rate. To reduce the complexity of the optimization problem, we assumed the APSK constraint on the position of the points. We showed that this constraint not only makes the convergence of the optimization algorithm faster but also results in higher MI rates than other GS methods. To optimize APSKs for the non-linear fiber channel, the shaping is done at the maximum modified SNR of the optical system. By doing so, our optimization algorithm maximizes the MI rate while the impacts of shaping on the NLIN power are considered. We indicated that this scheme provides significant MI gains and reach improvements. In addition, using the mutual information formula, we proposed a model for the radius of APSK rings, which in turn resulted in

an equation for the optimal radii. Using this equation, we analyzed how the optimal radius of the APSK rings changes for different shaping SNRs (for the AWGN channel) and different spans (for the non-linear fiber channel). We compared the optimal radius of the AWGN and fiber channel and concluded that, in the low-SNR regimes, in the AWGN channel, the optimal radius of the APSK rings grows faster than the optimal radius of the fiber channel. This indicates that in the highly non-linear regimes of the fiber channel, geometric shaping based on the AWGN is not valid, and shaping must be done for the fiber channel.

In Chapter 4, we proposed a hybrid shaping method by combining PS and GS. In our HPGS method, we start with the probabilistic-shaped QAM as the initial constellation, and at each step of the optimization, the position of one constellation point and parameter of the MB distribution are changed such that the GMI rate increases. This approach guarantees that the ultimate constellation provides higher GMI rates than the probabilistic-shaped QAM. To consider the impacts of shaping on the NLIN of the fiber, optimization is done at the maximum modified SNR. We compared AWGN-based and EGN-based HPGS in terms of the GMI rate and demonstrated that the AWGN-based shaping cannot provide GMI gains as high as the EGN-based shaping.

In both studies, the optimized constellations provide considerable gains over the baseline modulations. Our GS method provides reach improvements as high as 600 km over QAMs, and our HPGS method is capable of outperforming regular QAMs and probabilistic-shaped QAMs by 400 and 100 km, respectively.

5.2 Future Research Directions

As mentioned, modulations used in an optical fiber communication system have a considerable impact on the amount of nonlinearity introduced to the system. Hence, designing modulations in a manner that mitigates the interfering effects of the fiber channel is a great approach to improve the performance significantly. In this thesis, we employed the EGN-model of fiber to

design optimized constellations for the fiber channel. According to the EGN-model, fiber impairments manifest themselves as circularly symmetric additive Gaussian noise, and it considers the impacts of the modulations on fiber non-linearity in a random manner. However, based on NLSE, most of the fiber impairments such as non-linearity and GVD are in fact deterministic phenomena rather than random phenomena. Hence, optimizing constellation based on SSFM simulations may result in much better performance than the EGN-based optimization. The possible future works for SSFM-based constellation optimization are as follows:

- SSFM-based GS enables us to optimize four-dimensional constellation for the fiber channel. According to the Manakov equation (2.21), the total power of the x and y-polarization affects the non-linearity introduced to the corresponding WDM channel. In other words, the instantaneous power of y-polarization directly affects the non-linearity of x-polarization and vice versa. Hence, optimizing four-dimensional constellations will result in much more non-linearity mitigated constellations than two-dimensional ones. The EGN-model cannot be used for four-dimensional shaping since it assumes that the modulations in the x and y-polarization are identical. Consequently, SSFM-based optimizations must be used to design four-dimensional constellations.

- As mentioned in Chapter 2, to estimate the AIR of the fiber channel, mismatched decoding is used by assuming an AWGN auxiliary channel. Using SSFM simulations, one can find a more accurate auxiliary channel than the AWGN one. Using the accurate auxiliary channel, we have a more accurate estimation of the channel transition probability $f_{\mathbf{Y}|\mathbf{X}}$, which enables us to modify our shaping methods. For instance, using the accurate auxiliary channel, we can define a new minimum distance, which is more efficient than the traditional minimum Euclidean distance that the AWGN auxiliary channel suggests. Using this new distance, we will have much better insights into where to place the constellation points.

- SSFM-based PS may suggest that a different distribution is optimal for the fiber channel instead of the MB distribution. MB is optimal for PS assuming that the AWGN auxiliary channel is employed. If other auxiliary channels are employed, they may result in distributions that provide much lower gaps between the channel capacity and AIR.

The major issue of SSFM-based shaping is that SSFM simulations are extremely computationally complex, which makes SSFM-based optimization hard to implement. By investigating methods that can approximate SSFM simulations, one can implement the SSFM-based constellation optimization in a more efficient manner.

References

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

[2] G. Agrawal, *Nonlinear Fiber Optics*, ser. Optics and Photonics. Elsevier Science, 2013, ISBN: 9780123970237. [Online]. Available: https://books.google.ca/books?id=xNvw-GDVn84C.

[3] P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey, "Maximizing the optical network capacity," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 374, no. 2062, Mar. 2016, ISSN: 1364-503X. DOI: 10.1098/rsta.2014.0440. [Online]. Available: https://europepmc.org/articles/PMC4733919.

[4] G. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, 1998. DOI: 10.1109/18.720542.

[5] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949. DOI: 10.1109/JRPROC.1949.232969.

[6] ——, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[7] G. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, 1998. DOI: 10.1109/18.720542.

[8] "Signal shaping," in *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, Ltd, 2002, pp. 219–340, ISBN: 9780471439004.

[9] F. Steiner and G. Boecherer, "Comparison of geometric and probabilistic shaping with application to atsc 3.0," in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, Hamburg, Germany, 2017, pp. 1–6.

[10] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Express*, vol. 22, no. 13, pp. 16 335–16 362, Jun. 2014. DOI: 10.1364/OE.22.016335. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-22-13-16335.

[11] V. Curri, A. Carena, P. Poggiolini, G. Bosco, and F. Forghieri, "Extension and validation of the gn model for non-linear interference to uncompensated links using raman amplification," *Opt. Express*, vol. 21, no. 3, pp. 3308–3317, Feb. 2013. DOI: 10.1364/OE.21.003308. [Online]. Available: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-21-3-3308.

[12] D. Marcuse, C. Manyuk, and P. Wai, "Application of the manakov-pmd equation to studies of signal propagation in optical fibers with randomly varying birefringence," *Journal of Lightwave Technology*, vol. 15, no. 9, pp. 1735–1746, 1997. DOI: 10.1109/50.622902.

[13] O. Sinkin, R. Holzlohner, J. Zweck, and C. Menyuk, "Optimization of the split-step fourier method in modeling optical-fiber communications systems," *Journal of Lightwave Technology*, vol. 21, no. 1, pp. 61–68, 2003. DOI: 10.1109/JLT.2003.808628.

[14] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1524–1539, 2012. DOI: 10.1109/JLT.2012.2189198.

[15] P. 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*, vol. 32, no. 4, pp. 694–721, 2014. DOI: 10.1109/JLT.2013.2295208.

[16] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Express*, vol. 21, no. 22, pp. 25 685–25 699, Nov. 2013. DOI: 10.1364/OE.21.025685. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-21-22-25685.

[17] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No.01CH37270)*, vol. 5, San Antonio, 2001, 2992–2996 vol.5.

[18] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *Journal of Lightwave Technology*, vol. 23, no. 11, pp. 3755–3763, 2005.

[19] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3839–3852, 2013.

[20] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2315–2328, 2000.

[21] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 230–244, 2019. DOI: 10.1109/JLT.2019.2895770.

[22] D. M. Arnold, H. .-. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

[23] Z. Qu and I. B. Djordjevic, "On the probabilistic shaping and geometric shaping in optical communication systems," *IEEE Access*, vol. 7, pp. 21 454–21 464, 2019.

[24] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *Journal of Lightwave Technology*, vol. 34, no. 21, pp. 5063–5073, 2016. DOI: 10.1109/JLT.2016.2594271.

[25] A. Martinez, A. Guillen i Fabregas, G. Caire, and F. M. J. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2756–2765, 2009. DOI: 10.1109/TIT.2009.2018177.

[26] B. Chen, C. Okonkwo, D. Lavery, and A. Alvarado, "Geometrically-shaped 64-point constellations via achievable information rates," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, Bucharest, 2018, pp. 1–4.

[27] S. Zhang and F. Yaman, "Design and comparison of advanced modulation formats based on generalized mutual information," *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 416–423, 2018. DOI: 10.1109/JLT.2017.2779753.

[28] "Maximum entropy," in *Elements of Information Theory*. John Wiley & Sons, Ltd, 2005, ch. 12, pp. 409–425, ISBN: 9780471748823. DOI: https://doi.org/10.1002/047174882X.ch12. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/047174882X.ch12. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch12.

[29] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019. DOI: 10.1109/JLT.2019.2898855.

[30] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, 2015. DOI: 10.1109/TCOMM.2015.2494016.

[31] G. Böcherer, F. Steiner, and P. Schulte, "Fast probabilistic shaping implementation for long-haul fiber-optic communication systems," in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1–3. DOI: 10.1109/ECOC.2017.8346147.

[32] A. Amari, S. Goossens, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, C. M. Okonkwo, F. M. J. Willems, and A. Alvarado, "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightwave Technol.*, vol. 37, no. 23, pp. 5926–5936, Dec. 2019.

[33] K. Kojima, T. Yoshida, T. Koike-Akino, D. S. Millar, K. Parsons, M. Pajovic, and V. Arlunno, "Nonlinearity-tolerant four-dimensional 2A8PSK family for 5–7 bits/symbol spectral efficiency," *Journal of Lightwave Technology*, vol. 35, no. 8, pp. 1383–1391, 2017. DOI: 10.1109/JLT.2017.2662942.

[34] B. Chen, A. Alvarado, S. van der Heide, M. van den Hout, H. Hafermann, and C. Okonkwo, "Analysis and experimental demonstration of orthant-symmetric four-dimensional 7 bit/4d-sym modulation for optical fiber communication," *J. Lightwave Technol.*, vol. 39, no. 9, pp. 2737–2753, May 2021. [Online]. Available: http://www.osapublishing.org/jlt/abstract.cfm?URI=jlt-39-9-2737.

[35] S. Zhang and F. Yaman, "Constellation design with geometric and probabilistic shaping," *Optics Communications*, vol. 409, pp. 7–12, 2018, Advances in modulation and DSP for optical transmission systems, ISSN: 0030-4018. DOI: https://doi.org/10.1016/j.optcom.2017.08.063. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401817307538.

[36] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," in *2018 European Conference on Optical Communication (ECOC)*, Rome, 2018, pp. 1–3. DOI: 10.1109/ECOC.2018.8535453.

[37] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," in *45th European Conference on Optical Communication (ECOC 2019)*, Dublin, Ireland, 2019, pp. 1–4. DOI: 10.1049/cp.2019.0886.

[38] B. Moore, G. Takahara, and F. Alajaji, "Pairwise optimization of modulation constellations for non-uniform sources," *Canadian Journal of Electrical and Computer Engineering*, vol. 34, no. 4, pp. 167–177, 2009.

[39] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Increasing achievable information rates via geometric shaping," in *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3. DOI: 10.1109/ECOC.2018.8535358.

[40] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, vol. 7, no. 12, pp. 593–595, 2003. DOI: 10.1109/LCOMM.2003.821325.

[41] S. Zhang, Z. Qu, F. Yaman, E. Mateo, T. Inoue, K. Nakamura, Y. Inada, and I. B. Djordjevic, "Flex-rate transmission using hybrid probabilistic and geometric shaped 32qam," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, San Diego, CA, 2018, pp. 1–3.

[42] Z. Qu, S. Zhang, and I. B. Djordjevic, "Universal hybrid probabilistic-geometric shaping based on two-dimensional distribution matchers," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, San Diego, CA, 2018, pp. 1–3.

[43] E. Sillekens, D. Semrau, D. Lavery, P. Bayvel, and R. I. Killey, "Experimental demonstration of geometrically-shaped constellations tailored to the nonlinear fibre channel," in *2018 European Conference on Optical Communication (ECOC)*, Rome, 2018, pp. 1–3. DOI: 10.1109/ECOC.2018.8535199.

[44] Q. Xie, Z. Wang, and Z. Yang, "Simplified soft demapper for APSK with product constellation labeling," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2649–2657, 2012.

[45] Z. YANG, Q. XIE, K. PENG, and Z. WANG, "A novel BICM-ID system approaching shannon-limit at high spectrum efficiency," *IEICE Transactions on Communications*, vol. E94.B, no. 3, pp. 793–795, 2011. DOI: 10.1587/transcom.E94.B.793.

[46] Feng-Wen Sun and H. C. A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1714–1716, 1993.

[47] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, Perth, WA, Australia, 1995, 1942–1948 vol.4.

[48] Z. Liu, Q. Xie, K. Peng, and Z. Yang, "APSK constellation with gray mapping," *IEEE Communications Letters*, vol. 15, no. 12, pp. 1271–1273, 2011.

[49] Y. Duan, R. G. Harley, and T. G. Habetler, "Comparison of particle swarm optimization and genetic algorithm in the design of permanent magnet motors," in *2009 IEEE 6th International Power Electronics and Motion Control Conference*, Wuhan, 2009, pp. 822–825. DOI: 10.1109/IPEMC.2009.5157497.

[50] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, vol. 3, 1999, 1945–1950 Vol. 3. DOI: 10.1109/CEC.1999.785511.

[51] R. De Gaudenzi, A. Guillén i Fàbregas, and A. Martinez, "Turbo-coded APSK modulations design for satellite broadband communications," *International Journal of Satellite Communications and Networking*, vol. 24, no. 4, pp. 261–281, 2006. DOI: https://doi.org/10.1002/sat.841. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sat.841. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sat.841.

[52] R. De Gaudenzi, A. Guillen i Fabregas, and A. Martinez, "Performance analysis of turbo-coded apsk modulations over nonlinear satellite channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2396–2407, 2006. DOI: 10.1109/TWC.2006.1687763.

[53] K. P. Liolis and N. S. Alagha, "On 64-apsk constellation design optimization," in *2008 10th International Workshop on Signal Processing for Space Communications*, 2008, pp. 1–7. DOI: 10.1109/SPSC.2008.4686709.

[54] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010. DOI: 10.1109/JLT.2009.2039464.

[55] Y. C. Gultekin, A. Alvarado, O. Vassilieva, I. Kim, P. Palacharla, C. M. Okonkwo, and F. Willems, "Kurtosis-limited sphere shaping for nonlinear interference noise reduction in optical channels," *Journal of Lightwave Technology*, pp. 1–11, 2021, ISSN: 1558-2213. DOI: 10.1109/jlt.2021.3120915. [Online]. Available: http://dx.doi.org/10.1109/JLT.2021.3120915.

[56] M. N. Tehrani, M. Torbatian, H. Sun, P. Mertz, and K.-T. Wu, "A novel nonlinearity tolerant super-gaussian distribution for probabilistically shaped modulation," in *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3. DOI: 10.1109/ECOC.2018.8535379.

Appendix A

Derivation of Partial Derivatives of Equation (3.11)

In this appendix, we provide complete derivations of Equations (3.15) and (3.18).

A.1 Derivation of $\frac{\partial LF}{\partial \tau}$

By taking the partial derivative of (3.11) with respect to τ, we have:

$$\frac{\partial LF}{\partial \tau} = -\frac{1}{\log_{\mathrm{e}} 2} \sum_{\substack{1\le k\le R \\ 0\le j\le n_k-1}} \frac{\sum_{\substack{1\le k'\le R \\ 0\le i\le n_{k'}-1}} N_{k_j,k'_i}}{\sum_{\substack{1\le k'\le R \\ 0\le i\le n_{k'}-1}} D_{k_j,k'_i}}, \tag{A.1}$$

where

$$N_{k_j,k'_i} = \Big(\frac{\partial \gamma}{\partial \tau}.r_0^2.|d_{k_j,k'_i}|^2 + \gamma.\frac{\partial (r_0^2)}{\partial \tau}.|d_{k_j,k'_i}|^2 + \gamma.r_0^2.\frac{\partial (|d_{k_j,k'_i}|^2)}{\partial \tau}\Big) \times \mathrm{e}^{-\gamma.r_0^2.|d_{k_j,k'_i}|^2}, \tag{A.2}$$

and

$$D_{k_j,k'_i} = \mathrm{e}^{-\gamma.r_0^2.|d_{k_j,k'_i}|^2}. \tag{A.3}$$

In (A.2), three derivative terms are needed to be computed. The first term, $\frac{\partial \gamma}{\partial \tau}$, depends on the type of communication channel. For the non-linear fiber channel, we compute $\frac{\partial \gamma}{\partial \tau}$ in the next section. Term $\frac{\partial (r_0^2)}{\partial \tau}$ is computed based on the specified average symbol energy. Since we are using model (3.4), we have $\mathbb{E}\{|X|^2\} = 1$. In the case of APSK constellations, one can simply show that:

$$\mathbb{E}\{|X|^l\} = \sum_{k=1}^{R} \frac{n_k}{M} r_k^l \qquad \text{for even values of } l. \tag{A.4}$$

Using the proposed model of (3.10) and equation $\mathbb{E}\{|X|^2\} = 1$, we can compute r_0^2 as follows:

$$r_0^2 = \frac{1}{\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau}}, \tag{A.5}$$

which yields:

$$\frac{\partial(r_0^2)}{\partial\tau} = -\frac{2\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau}\log_{\mathrm{e}}(k-\Gamma)}{(\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau})^2}. \tag{A.6}$$

The third and last term, $\frac{\partial(|d_{k_j,k_i'}|^2)}{\partial\tau}$, considers the effects of τ on the normalized distance between two constellation points. According to (3.12), we have:

$$\begin{aligned}\frac{\partial(|d_{k_j,k_i'}|^2)}{\partial\tau} &= 2(k-\Gamma)^{2\tau}\log_{\mathrm{e}}(k-\Gamma) + 2(k'-\Gamma)^{2\tau}\log_{\mathrm{e}}(k'-\Gamma) \\ &\quad - 2[(k-\Gamma)(k'-\Gamma)]^{\tau} \times \log_{\mathrm{e}}((k-\Gamma)(k'-\Gamma))\cos(\varphi_{k_j} - \varphi_{k_i'}).\end{aligned} \tag{A.7}$$

By substituting Equations (A.5), (A.6), and (A.7) in Equations (A.2) and (A.3), and by substituting the obtained equations in (A.1), we reach to Equation (3.15).

A.2 Derivation of $\frac{\partial\gamma}{\partial\tau}$ Based on the EGN-Model of Fiber

Based on Equations (3.16) and (3.17), we can compute $\frac{\partial\gamma}{\partial\tau}$ using the chain rule:

$$\frac{\partial\gamma}{\partial\tau} = -\frac{\gamma}{3\eta} \times \frac{\partial\eta}{\partial\tau}, \tag{A.8}$$

where

$$\frac{\partial\eta}{\partial\tau} = \eta_{\Phi} \cdot \frac{\partial\Phi}{\partial\tau} + \eta_{\Psi} \cdot \frac{\partial\Psi}{\partial\tau}. \tag{A.9}$$

To compute $\frac{\partial\Phi}{\partial\tau}$ and $\frac{\partial\Psi}{\partial\tau}$, we need to rewrite Φ and Ψ based on the radius model of (3.10). Using (A.4), we have:

$$\Phi = \frac{\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{4\tau}}{(\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau})^2} - 2, \tag{A.10}$$

and

$$\Psi = \frac{\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{6\tau}}{(\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau})^3} - \frac{9\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{4\tau}}{(\sum_{k=1}^{R} \frac{n_k}{M}(k-\Gamma)^{2\tau})^2} + 12. \tag{A.11}$$

By taking the partial derivative of Φ and Ψ with respect to τ, we have:

$$\frac{\partial \Phi}{\partial \tau} = \Bigg[(\sum_{k=1}^{R} \frac{n_k}{M} 4(k-\Gamma)^{4\tau} \log_{\mathrm{e}}(k-\Gamma)) \cdot (\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau}) \\ - 4(\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{4\tau}) \cdot (\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau} \log_{\mathrm{e}}(k-\Gamma)) \Bigg] \times \Bigg[\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau} \Bigg]^{-3}, \tag{A.12}$$

and

$$\frac{\partial \Psi}{\partial \tau} = \Bigg[(\sum_{k=1}^{R} \frac{n_k}{M} 6(k-\Gamma)^{6\tau} \log_{\mathrm{e}}(k-\Gamma)) \cdot (\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau}) \\ - 6(\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{6\tau}) \cdot (\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau} \log_{\mathrm{e}}(k-\Gamma)) \Bigg] \times \Bigg[\sum_{k=1}^{R} \frac{n_k}{M} (k-\Gamma)^{2\tau} \Bigg]^{-4} \\ - 9 \cdot \frac{\partial \Phi}{\partial \tau}, \tag{A.13}$$

where $\frac{\partial \Phi}{\partial \tau}$ is defined in (A.12). By substituting Equations (A.12) and (A.13) in (A.9), and by substituting the obtained equation in (A.8), we reach to Equation (3.18).