Information theory and signal processing for the optical transmission system

Eric Sillekens

A dissertation submitted to the

University College London (UCL)

for the degree of

Doctor of Philosophy

Department of Electronic and Electrical Engineering

University College London

May 2020

I, Eric Sillekens, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In long distance optical transmission systems, the coherent optical transceiver with wavelength division multiplexing has become a mature technology. This work investigates signal processing to further improve the data throughput of such systems.

Using information theory, we have designed coded modulation schemes for the optical transmission system. We presented a modulation format for optical transmission systems which reduces the nonlinear distortion during transmission, increasing the achievable information rate. We designed machine learning techniques to increase the throughput in nonlinear fibre transmission, and carried out the first experimental demonstration of learned time-domain digital back-propagation.

Impact Statement

The world is relying on fibre optic transmission systems to form the backbone of the internet. The optical fibre can support large datarates and can carry signals over large distances, these properties making this the medium of choice for the vast majority of data transfer around the world. The insatiable demand for more bandwidth is rapidly filling the existing cables.

In this thesis we explore machine learning and information theory as concepts which will allow us to make better use of the existing infrastructure. Due to the increase in available compute power, techniques from information theory and machine learning that were previously deemed infeasible are now in the realm of possibilities.

We developed nonlinearity tolerant signal formats, in which the constellations onto which bits are encoded are optimised to increase achievable data-rates for the nonlinear fibre channel. The constellations were extensively tested, both in simulations and experimentally, using an optical testbed.

We investigated the adaptation of the digital filters used in the algorithms for fibre nonlinearity mitigation, using powerful machine learning techniques. As one major outcome of this work, we carried out the first experimental demonstration of learned time-domain digital back-propagation.

Acknowledgements

From the moment Prof. Robert Killey became my supervisor, he always supported and encouraged me, for which I am very grateful. I want to thank him for his support and guidance. I would like to acknowledge my second supervisor Dr Domaniç Lavery for his support and discussions and equally my tertiary supervisor Prof. Polina Bayvel for welcoming kindness and support, a thanks I would like to extend for the kindness and support I received from the whole optical networks group.

I would also like to acknowledge the supervisors I started with; Dr Benn Thomsen and Prof. Cyril Renaud, who have invited me for this wonderful journey and the industrial supervision I received from Dr Steve Debruslais at Xtera. I would like to acknowledge the financial support from EPSRC, Xtera and Huawei.

I am hugely indebted to Dr Alex Alvarado who shared his expertise on information theory with me to help me reach this level. Equally am I indebted to Dr Kai Shi, Dr Lídia Galdino and Dr Robert Maher who inducted me to the lab to conduct proper experiments. To Dr Zhe Li and Boris Karanov for improving my writing and Paris Andreades for always looking out for me. I would like to thank Thomas Gerard and Daniel Semrau for being good friends.

Additionally, I would like to thank Natalie Leijnen, for her unconditional support and love, she has got me through everything. I would like to thank my parents for teaching me to get what I want out of myself.

Contents

List of Figures

Chapter 1

Introduction

The world has increasingly become reliant on the internet. 51 % of CEOs expect that their companies will use more technology to meet their changing customer and stakeholder expectations [1]. It is no wonder that Cisco expects the total IP traffic to triple between 2017 and 2022, as shown in Fig. 1.1 from [2].

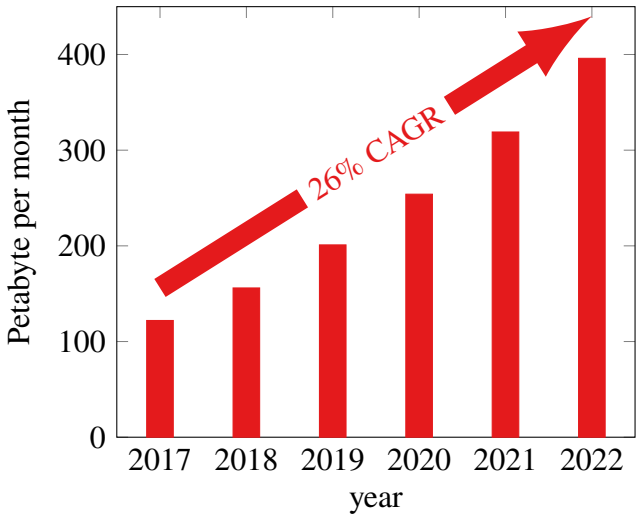


Figure 1.1: Cisco forecast of total IP traffic from Cisco VNI, 2018

The increase in people and devices connected to internet has carried through to increased demands on the backbone of the internet. However, major datacentre providers do not use the internet for their inter-datacentre communication, but pass the increased dataload through their own global network. This increase has lead to Microsoft, Facebook and Telxius jointly laying a new subsea cable across the Atlantic Ocean named MAREA [3]. This is an addition to their already global

distribution network shown in Fig. 1.2 from [4].

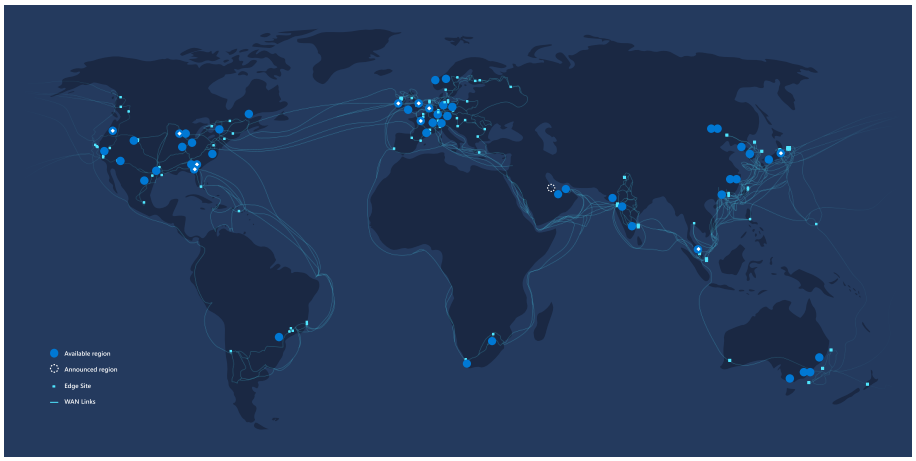


Figure 1.2: Microsofts global backbone network.

The infrastructure is based on optical fibres, since TAT-8 the de-facto standard of building long-haul communication infrastructure [5]. Where the TAT-8 was able to transmit 280 Mbit/s per fibre pair, the MAREA cable has achieved 26.2 Tbit/s per fibre pair, an almost $100,000\times$ increase. The first of the improvements came from stepping away from regenerating every span and using optical amplification to transport signals over larger distances. The TAT12/13 was the first transatlantic system that made use of Erbium-doped fibre amplifiers (EDFAs) [6], which, in turn, then easily allowed the use of more than one wavelength in the transmission system [7].

Another advance comes from moving from on-off keying [8] to phase and amplitude modulation [9], which, combined with polarisation multiplexing and pulse shaping to allow Nyquist-spaced WDM channels [10] has led to a year on year increase in capacity records [11, 12, 13, 14, 15, 16, 17, 18, 19, 20] as shown in Fig. 1.3.

However, we can see a saturation in the records obtained over the years. One solution is to use space division multiplexing [21], a technique that allows scaling of link capacities by many multiples. Although the technique is promising [22, 23], there are yet to be systems deployed using these technologies.

This work is investigating novel signal processing techniques that make more

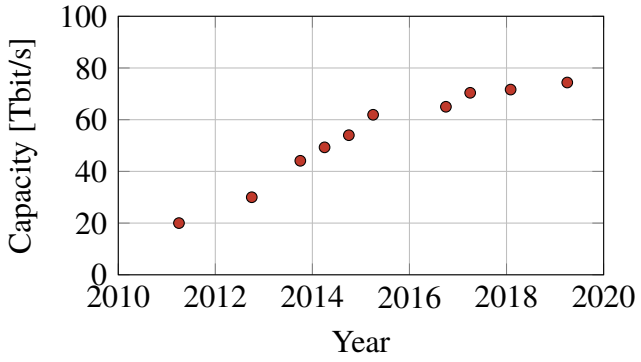


Figure 1.3: Transatlantic transmission records set in the last 10 years.

efficient use of the current deployable and deployed transmission systems. This is a promising approach. For example, it can be seen from [17] and [19] that more advance modulation formats and coded modulation are starting to be used in deployed systems.

Lately, machine learning has made its way into improving communication systems [24], and more specifically optical communication systems [25, 26]. These methods will allow us to go beyond the limits of linear systems [27], improve on the current limitations [28] and even approach the limit for the current techniques [29].

1.1 Optical transmission system

At the heart of our optical transmission system is the optical coherent transceiver, with which we can transmit and receive, for both polarisations, the phase and amplitude of our signal, using a dual polarisation (DP) IQ modulator (IQM) and a DP coherent receiver [30, 31].

1.1.1 Transmitter

The modulator converts two electrical signals onto the inphase (I) and quadrature (Q) components of the envelope of a continuous wave (CW) carrier. The two electrical signals drive two separate Mach-Zehnder modulators [32, 33] which are nested in another Mach-Zehnder structure, as shown in Fig. 1.4.

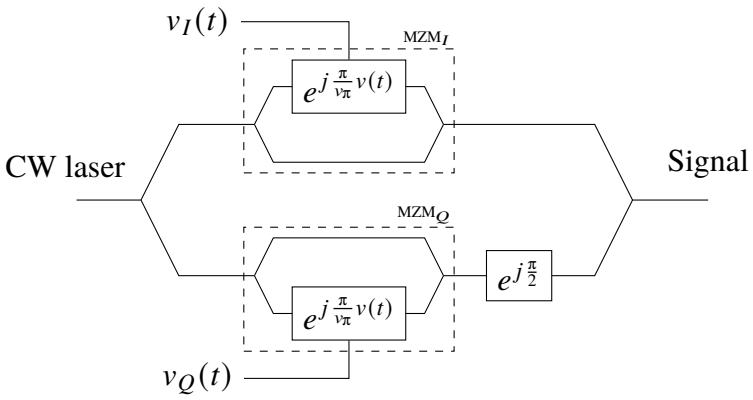


Figure 1.4: An IQ modulator as a nested structure of Mach-Zehnder modulators.

This has the effect of creating a signal

$$E_s(t) = v_I(t)\cos(\omega_c t + \phi_c(t)) + v_Q(t)\sin(\omega_c t + \phi_c(t)) \tag{1.1}$$

with the carrier phase ϕ_c and the carrier frequency ω_c many times larger than the modulated bandwidth of the envelope $A_s(t) = v_I(t) + jv_Q(t)$. Two of these structures can be nested again with a polarisation beam combiner to create a DP signal [34].

1.1.2 Receiver

The DP coherent receiver recovers both phase and amplitude. A single photodiode outputs a current I_{DD} proportional to the square of the electric field [35]

$$I_{\mathrm{DD}} \propto E_s E_s^* = \frac{1}{2}|A_s(t)|^2\left[1 + \cos(2\omega_c t + 2\phi_c(t))\right] \tag{1.2}$$

where the carrier at $2\omega_s$ is undetectable due to the limited bandwidth of the photodiode. This will only detect the amplitude of our envelope, but when we add a second laser we call the local oscillator (LO), we get

$$E_{\mathrm{LO}} = A_{\mathrm{LO}}\cos(\omega_{\mathrm{LO}} t + \phi_{\mathrm{LO}}(t)) \tag{1.3}$$

$$I_{\mathrm{coh}} \propto \frac{1}{2}|A_s|^2 + |A_{\mathrm{LO}}|^2 + \Re\left\{A_s A_{\mathrm{LO}}^* \exp\left(j(\omega_c - \omega_{\mathrm{LO}})t + j(\phi_c(t) - \phi_{\mathrm{LO}}(t))\right)\right\} \tag{1.4}$$

Mixing the LO and signal in an optical hybrid [36] allows us to recover both the

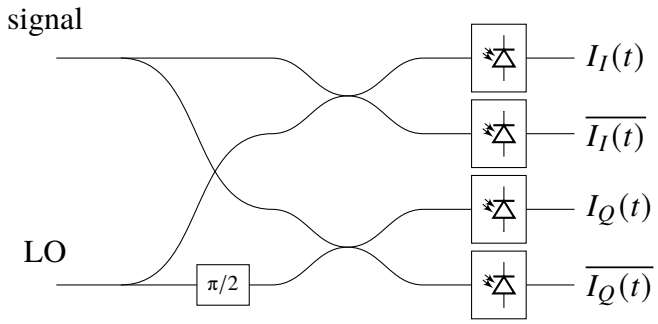


Figure 1.5: A coherent receiver will recover phase and amplitude information with the use of an LO.

inphase and quadrature information of the signal. In Fig 1.5, a coherent receiver is shown. Adding a polarisation beam splitter (PBS) cuts the signal into two orthogonal slices in the polarisation space and allowing them to be received separately. A digital linear transformation can then be used to retrieve the original two states of polarisation. This is achieved using digital signal processing (DSP) where a 2×2 multiple-input, multiple output (MIMO) adaptive filter is used to recover the sample phase and the polarisation rotation [37]. Then the frequency offset between the carrier and the LO $\omega_\Delta = \omega_c - \omega_{\mathrm{LO}}$ is removed and subsequently the laser phase noise $\phi_\Delta(t) = \phi_c(t) - \phi_{\mathrm{LO}}(t)$ is removed [38]. Optionally, pilot symbols can be inserted to increase the transmission quality [39, 40].

1.1.3 Optical transmission medium

The optical signals are carried over a fibre with very low losses [41] [42]. The electric field in the waveguide is described by solving the Maxwell equations

$$\nabla^2 E = \frac{-1}{c^2}\frac{\partial^2 E}{\partial t^2} + \mu_0 \frac{\partial^2 P}{\partial t^2} \tag{1.5}$$

with P the induced electric polarisation.

$$P(r,t) = \epsilon_0 \Bigg(\int_{-\infty}^{\infty} \chi_{\mathrm{L}}(t-t')E(r,t')dt' + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi_{\mathrm{NL}}(t-t_1, t-t_2, t-t_3)E(r,t_1)E(r,t_2)E(r,t_3)dt_1dt_2dt_3 \Bigg) \tag{1.6}$$

with χ_{L} and χ_{NL} the components of the first and third order suseptibility respectively [43]. When the third order part is neglected a solution can be found assuming our signal is

$$E(r,t) = \frac{1}{2}\left[A(r,t)\exp(j\omega_0 t) + A^*(r,t)\exp(-j\omega_0 t)\right] \tag{1.7}$$

which satisfies the Helmholtz equation in the frequency domain

$$\nabla^2 \tilde{E} + \epsilon(\omega)k_0^2 \tilde{E} = 0 \tag{1.8}$$

with $\tilde{}$ the Fourier transform, $k_0 = \omega/c$ and

$$\epsilon(\omega) = 1 + \tilde{\chi}_{\mathrm{L}}(\omega) + \frac{3}{4}\chi_{\mathrm{NL}}(\omega)|E(t)|^2 \tag{1.9}$$

After expanding the wavelength-dependent propagation constant in the direction of the fibre $\beta(\omega)$ in Taylor series, we arrive at the nonlinear Schrödinger equations (NLSE)

$$\beta(\omega) = \beta_0 + (\omega-\omega_0)\beta_1 + \frac{1}{2}(\omega-\omega_0)^2\beta_2 + \frac{1}{6}(\omega-\omega_0)^3\beta_3 \tag{1.10}$$

$$\frac{\partial A}{\partial z} + \beta_1\frac{\partial A}{\partial t} + \frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + \frac{\alpha}{2}A = j\gamma(\omega_0)|A|^2 A \tag{1.11}$$

$$\gamma(\omega_0) = \frac{n_2(\omega_0)\omega_0}{cA_{\mathrm{eff}}} \tag{1.12}$$

with α the fibre attenuation, n_2 the second-order nonlinear refractive index and A_{eff} the effective area. The equation can be rewritten with $T = t - z/v_g = t - \beta_1 z$ as

$$j\frac{\partial A}{\partial z} + \frac{j\alpha}{2}A - \frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} + \gamma|A|^2 A = 0 \tag{1.13}$$

If it is rewritten to separate the dispersion $\hat{D}$ and nonlinearity $\hat{N}$

$$\frac{\partial A}{\partial z} = (\hat{D} + \hat{N})A \tag{1.14}$$

$$\hat{D} = \frac{\beta_2}{2}\frac{\partial^2}{\partial T^2} - \frac{j\alpha}{2} \tag{1.15}$$

we can introduce the split-step Fourier method (SSFM) for the segment z to $z + h$ as follows [43]

$$A(z+h,T) = \exp\left(\frac{h}{2}\hat{D}\right)\exp\left(\int_z^{z+h}\hat{N}(z')dz'\right)\exp\left(\frac{h}{2}\hat{D}\right)A(z,T) \tag{1.16}$$

For a DP signal, we can use the Manakov equations [44] to obtain the nonlinear phase shift $\hat{N}$.

1.1.4 Optical amplification

For optical communication systems, EDFAs can be used to amplify the optical signal. These amplifiers have a broad gain and preserve the phase of the signals. The amplification stems from the erbium ions doped into the fibre which have their electrons raised to a higher energy level. Erbium is chosen because its gain spectrum coincides with the ultra-low loss region of the optical fibre. The gain can be calculated using a two-level model [45] with number of Erbium ions N_k in the k-th level

$$\frac{\partial N_2}{\partial t} = \left(\sigma_p^a N_1 - \sigma_p^e N_2\right)\frac{P_p}{a_p h\nu_p} + \left(\sigma_s^a N_1 - \sigma_s^e N_2\right)\frac{P_s}{a_s h\nu_s} - \frac{N_2}{\tau} \tag{1.17}$$

$$\frac{\partial N_1}{\partial t} = -\left(\sigma_p^a N_1 - \sigma_p^e N_2\right)\frac{P_p}{a_p h\nu_p} - \left(\sigma_s^a N_1 - \sigma_s^e N_2\right)\frac{P_s}{a_s h\nu_s} + \frac{N_2}{\tau} \tag{1.18}$$

with the σ_j^a and σ_j^e the absorption and emission spectra respectively, P_j the optical power, a_j the cross section and ν_j the frequency and τ the metastable lifetime. Now our signal and pump powers can be calculated using the steady state population inversion

$$N_2(z) = -\frac{T_2}{\Gamma_s a_s h\nu_s}\frac{\partial P_s}{\partial z} - \frac{sT_2}{\Gamma_p a_p h\nu_p}\frac{\partial P_p}{\partial z} \tag{1.19}$$

with $\Gamma_s a_s = \Gamma_p a_p$ the confinement factor and $s = 1$ or $s = -1$ for forward and backward pumping respectively. We can define the gain for amplifier length L with loss α and $g(z) = \left(\sigma_p^a N_1 - \sigma_p^e N_2\right)$

$$G = \exp\left(\int_0^L (g(z) - \alpha)dz\right) \tag{1.20}$$

In an EDFA the population inversion is achieved by pumping with another laser. The alternative is creating the population inversion in a semiconductor material using an electrical pump. The semiconductor optical amplifier (SOA) has shown to work over greater bandwidth [46], but suffers from nonlinearity. A lossless time-domain model of an SOA is described in [47], expressed as

$$\tau_S \frac{dh(t)}{dt} \approx G_0 - h(t) - \left[e^{h(t)} - 1\right]\frac{|E_{\text{in}}(t)|^2}{P_{\text{sat}}} \tag{1.21}$$

$$E_{\text{out}}(t) = E_{\text{in}}(t)\exp\left\{\left(\frac{1 - j\alpha_{\text{H}}}{2}\right)h(t)\right\} \tag{1.22}$$

Both amplifiers emit amplified spontaneous emissions (ASE) noise, where the noise spectral density per polarisation of the EDFA is

$$S_{\text{ASE}} = n_{sp} h\nu(G - 1) \tag{1.23}$$

the noise power of the SOA is

$$R_{\text{ASE}}(\Delta t) = n_{sp} h\nu \int_{-\infty}^{\infty} f(t) f^*(t - \Delta t)dt \tag{1.24}$$

with n_{sp} the spontaneous emission factor and $f(\cdot)$ the inverse Fourier transform of the SOA gain spectrum.

The signal distortion, as described by the NLSE, and the noise added by the amplifiers in the transmission system result in a limited achievable data rate. This thesis will investigate how to quantify the limitations and to use the metrics to increase the data rates through our transmission system.

1.2 Thesis outline

The continuation of this thesis is outlined as follows

- In chapter 2, we will introduce the fundamental information theoretical measures and use them to assess optical transmission systems.

- The same measures are used in chapter 3 to design constellations with a better performance when being transmitted over the nonlinear fibre channel. We will use the metrics explained in chapter 2.

- In chapter 4, machine learning is investigated. The relation between the previous metrics and machine learning costs are explained, and our knowledge on the nonlinear fibre channel is used to apply learning to improve a nonlinear mitigation technique.

- In chapter 5 this work draws conclusions on the information theoretical metrics for the optical transmission system and when to use them.

- The appendix A derives the MI for a continuous uniform distribution and ends at the 1.53 dB gap to capacity that constellation shaping closes.

1.3 Key contributions

i) In section 2.2.3 the mutual information and generalised mutual information are investigated as predictors for nonbinary and bit-interleaved coded modulation respectively.

ii) In section 2.4 the so-called Gaussian noise model is used to predict performance and then the generalised mutual information for an additive white Gaussian noise channel with the same signal to noise ratio is calculated. In section 3, this result is improved upon by including the modulation format dependent change in channel response into the achievable rate calculation.

iii) In section 3.2, probabilistically shaped constellations for the optical fibre channel are introduced. An heuristic expression describing the results of the global optimisation process is proposed .

iv) In section 3.3 1D geometrically shaped constellations are designed and experimentally evaluated. In agreement with the model, the fibre transmission system tailored constellations outperformed the constellations designed for the additive white Gaussian noise channel.

v) We have carried out the first experimental demonstration of learned digital back-propagation for fibre nonlinearity and dispersion compensation. In section 4.1 the process of training such a network is detailed.

1.4 List of publications

- E. Sillekens, R. van Uden, J. A. van Weerdenburg, M. Kuschnerov, H. de Waardt, A. Koonen, and C. Okonkwo, "Experimental demonstration of 8 state turbo trellis coded modulation employing 8 phase shift keying," in *European Conference on Optical Communication (ECOC)*. IEEE, 2015, pp. 1–3.

- D. Yu, J. A. van Weerdenburg, E. Silekens, R. van Uden, M. Tang, D. Liu, A. Velazquez-Benitez, P. Sillard, D. Molin, M. Bigot-Astruc *et al.*, "Trellis coded modulation transmission over 40km 6-LP mode fiber," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2016, pp. 1–3.

- E. Sillekens, A. Alvarado, C. Okonkwo, and B. C. Thomsen, "An experimental comparison of coded modulation strategies for 100 Gbit/s transceivers," *Journal of Lightwave Technology*, 2016.

- J. van Weerdenburg, R. van Uden, E. Sillekens, H. de Waardt, T. Koonen, and C. Okonkwo, "Cluster analysis of received constellations for optical performance monitoring," in *European Conference on Optical Communication (ECOC)*. VDE, 2016, pp. 1–3.

- D. S. Millar, L. Galdino, R. Maher, M. Pajovic, T. Koike-Akino, G. Saavedra, D. J. Elson, D. Lavery, K. Shi, M. S. Erkilinc *et al.*, "A simplified dual-carrier DP-64QAM 1 Tb/s transceiver," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2017, pp. 1–3.

- Z. Li, M. Erkilinc, K. Shi, E. Sillekens, L. Galdino, B. Thomsen, P. Bayvel, and R. Killey, "Performance improvement of electronic dispersion post-compensation in direct detection systems using DSP-based receiver linearization," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2017, pp. 1–3.

- K. Shi, E. Sillekens, and B. C. Thomsen, "246 GHz digitally stitched coherent receiver," in *Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2017, pp. M3D–3.

- Z. Li, M. Erkilinc, K. Shi, E. Sillekens, L. Galdino, B. Thomsen, P. Bayvel, and R. Killey, "112 Gb/s/λ WDM direct-detection Nyquist-SCM transmission at 3.15 (b/s)/Hz over 240 km SSMF enabled by novel beating interference compensation," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2017, pp. 1–3.

- Z. Li, M. S. Erkılınç, K. Shi, E. Sillekens, L. Galdino, B. C. Thomsen, P. Bayvel, and R. I. Killey, "SSBI mitigation and the Kramers–Kronig scheme in single-sideband direct-detection transmission with receiver-based electronic dispersion compensation," *Journal of Lightwave Technology*, vol. 35, no. 10, pp. 1887–1893, 2017.

- Z. Li, M. S. Erkilinc, K. Shi, E. Sillekens, L. Galdino, B. Thomsen, P. Bayvel, and R. Killey, "168 Gb/s/λ direct-detection 64-QAM SSB Nyquist-SCM trans-

mission over 80 km uncompensated SSMF at 4.54 b/s/Hz net ISD using a Kramers-Kronig receiver," in *European Conference on Optical Communication (ECOC)*, 2017.

- E. Sillekens, D. Semrau, G. Liga, N. Shevchenko, Z. Li, A. Alvarado, P. Bayvel, R. Killey, and D. Lavery, "A simple nonlinearity-tailored probabilistic shaping distribution for square QAM," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2018, pp. M3C–4.

- Z. Li, M. S. Erkilinc, K. Shi, E. Sillekens, L. Galdino, B. Thomsen, P. Bayvel, and R. Killey, "Joint optimisation of resampling rate and carrier-to-signal power ratio in direct-detection Kramers-Kronig receivers," in *European Conference on Optical Communication (ECOC)*. IEEE, 2017, pp. 1–3.

- Z. Li, M. S. Erkılınç, K. Shi, E. Sillekens, L. Galdino, T. Xu, B. C. Thomsen, P. Bayvel, and R. I. Killey, "Digital linearization of direct-detection transceivers for spectrally efficient 100 Gb/s/λ WDM metro networking," *Journal of Lightwave Technology*, vol. 36, no. 1, pp. 27–36, 2017.

- Z. Li, L. Galdino, T. Xu, M. S. Erkilinç, K. Shi, E. Sillekens, B. C. Thomsen, P. Bayvel, and R. Killey, "Performance of digital back-propagation in Kramers-Kronig direct-detection receivers," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2018, pp. 1–3.

- Z. Li, M. S. Erkılınç, K. Shi, E. Sillekens, L. Galdino, T. Xu, B. C. Thomsen, P. Bayvel, and R. I. Killey, "Spectrally efficient 168 Gb/s/λ WDM 64-QAM single-sideband nyquist-subcarrier modulation with Kramers–Kronig direct-detection receivers," *Journal of Lightwave Technology*, vol. 36, no. 6, pp. 1340–1346, 2018.

- L. Galdino, D. Lavery, Z. Liu, K. Balakier, E. Sillekens, D. Elson, G. Saavedra, R. Killey, and P. Bayvel, "The trade-off between transceiver capacity and symbol rate," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2018, pp. W1B–4.

- L. Galdino, A. Edwards, M. Ionescu, J. James, W. Pelouch, E. Sillekens, D. Semrau, D. Lavery, R. I. Killey, S. Barnes *et al.*, "120 Tbit/s transmission over single mode fibre using continuous 91 nm hybrid Raman-EDFA amplification," *arXiv preprint arXiv:1804.01845*, 2018.

- 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 *European Conference on Optical Communication (ECOC)*. IEEE, 2018, pp. 1–3.

- L. Galdino, D. Semrau, E. Sillekens, D. Lavery, R. Killey, and P. Bayvel, "Impact of transceiver subsystems on high-capacity optical transmission," in *Signal Processing in Photonic Communications*. Optical Society of America, 2018, pp. SpTh2G–1.

- Z. Li, M. Erkılınç, K. Shi, E. Sillekens, L. Galdino, T. Xu, B. Thomsen, P. Bayvel, and R. Killey, "DSP for single-sideband direct-detection systems," in *Signal Processing in Photonic Communications*. Optical Society of America, 2018, pp. SpM2G–5.

- D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel, "The ISRS GN model, an efficient tool in modeling ultra-wideband transmission in point-to-point and network scenarios," in *European Conference on Optical Communication (ECOC)*. IEEE, 2018, pp. 1–3.

- D. J. Elsonf, B. J. Puttnam, G. Rademacher, R. S. Luis, E. Sillekens, L. Galdino, D. Lavery, Y. Awaji, N. Wada, and P. Bayvel, "Impact of intercore crosstalk on achievable information rates," in *IEEE Photonics Society Summer Topical Meeting Series (SUM)*. IEEE, 2018, pp. 17–18.

- D. J. Elson, B. J. Puttnam, R. S. Luis, G. Rademacher, E. Sillekens, L. Galdino, D. Lavery, Y. Awaji, N. Wada, and P. Bayvel, "Nonlinearity mitigation in the presence of intercore-crosstalk," in *European Conference on Optical Communication (ECOC)*. IEEE, 2018, pp. 1–3.

- M. Ionescu, L. Galdino, A. Edwards, J. James, W. Pelouch, E. Sillekens, D. Semrau, D. Lavery, R. Killey, S. Barnes *et al.*, "91 nm C+L hybrid distributed Raman–Erbium-doped fibre amplifier for high capacity subsea transmission," in *European Conference on Optical Communication (ECOC)*. IEEE, 2018, pp. 1–3.

- H. Yuan, A. Saljoghei, T. Hayashi, T. Nakanishi, E. Sillekens, L. Galdino, P. Bayvel, Z. Liu, and G. Zervas, "Experimental investigation of static and dynamic crosstalk in trench-assisted multi-core fiber," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2019, pp. 1–3.

- D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel, "A modulation format correction formula for the Gaussian noise model in the presence of inter-channel stimulated raman scattering," *Journal of Lightwave Technology*, vol. 37, no. 19, pp. 5122–5131, 2019.

- M. Ionescu, D. Lavery, A. Edwards, E. Sillekens, L. Galdino, D. Semrau, R. Killey, W. Pelouch, S. Barnes, and P. Bayvel, "74.38 Tb/s transmission over 6300 km single mode fiber with hybrid EDFA/Raman amplifiers," in *Optical Fiber Communications Conference (OFC)*. IEEE, 2019, pp. 1–3.

- W. Yi, Z. Li, M. S. Erkılınç, D. Lavery, E. Sillekens, D. Semrau, Z. Liu, P. Bayvel, and R. I. Killey, "Performance of Kramers–Kronig receivers in the presence of local oscillator relative intensity noise," *Journal of Lightwave Technology*, vol. 37, no. 13, pp. 3035–3043, 2019.

- T. Gerard, H. Dzieciol, E. Sillekens, Y. Wakayama, A. Alvarado, R. I. Killey, P. Bayvel, and D. Lavery, "Coded modulation for 100G coherent EPON," *Journal of Lightwave Technology*, 2019.

- L. Galdino, D. Semrau, M. Ionescu, A. Edwards, W. Pelouch, S. Desbruslais, J. James, E. Sillekens, D. Lavery, S. Barnes *et al.*, "Study on the impact of nonlinearity and noise on the performance of high-capacity broadband hybrid

Raman-EDFA amplified system," *Journal of Lightwave Technology*, vol. 37, no. 21, pp. 5507–5515, 2019.

- E. Sillekens, W. Yi, D. Semrau, A. Ottino, B. Karanov, S. Zhou, K. Law, J. Chen, D. Lavery, L. Galdino *et al.*, "Experimental demonstration of learned time-domain digital back-propagation," *arXiv preprint arXiv:1912.12197*, 2019.

Chapter 2

Coded modulation for coherent optical transmission systems

In this chapter, we investigate coded modulation with its implications and limits. With coded modulation we refer to the mapping from bits to channel uses. For this analysis, we need to abstract the transmission system. The transmission systems for which we are designing the coded modulation scheme employ coherent reception. In such systems, information can be transmitted using optical phase, as both the in-phase and the quadrature components of the light source are modulated. Then, using matched filters at the transmitter and receiver, the channel is set to transmit time-discrete symbols. We wish to transmit the maximum number of bits, using a sequence of symbols.

To put a quantitative measure on the amount of information we can reliably transmit, we have to make use of information theory [48]. For this purpose, we use bits as a measure of information, not to be confused with the information source that can be zero or one. For the noiseless case, if we take a source X with different events, in this case symbols $\{x_1, x_2, \ldots, x_M\}$, and their probabilities $p_1, p_2, \ldots, p_M$, the entropy

$$\mathrm{H}(X) = -\sum_{i=1}^{M} p_i \log_2(p_i) \tag{2.1}$$

measures the amount of information this source can describe per symbol.

For the binary case, where we only have two outputs with probabilities p and

$1 - p$, the binary entropy is given by

$$\mathrm{H}_b(p) = -(p \log_2(p) + (1-p) \log_2(1-p)) \tag{2.2}$$

And if both outputs are equiprobable ($p = {}^1/_2$), the source has an entropy of 1 bit per symbol, as shown in Fig. 2.1.

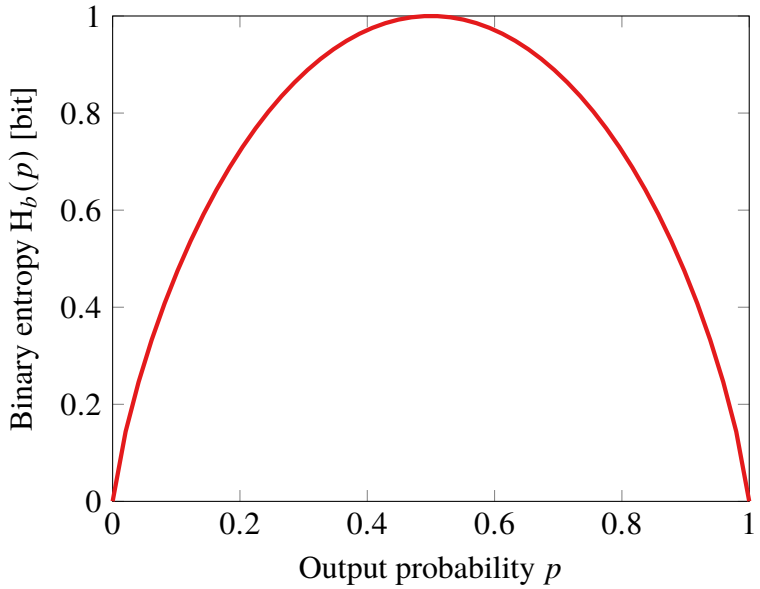


Figure 2.1: Binary entropy as function of output probability. Where no information is needed to describe a variable that is always the same, i.e. $p = 1$ or $p = 0$. Maximum information is in a variable that has both outputs equiprobable.

If the source does not produce ones and zeros with equal probability, less information is needed to describe the outputs. For example, if the source only produces ones, then, just guessing ones perfectly describes the outputs and no information is transmitted. And in Fig. 2.1, we can see that $\mathrm{H}_b(1) = 0$.

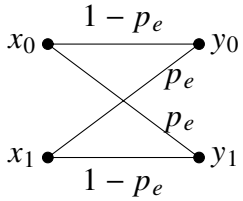


Figure 2.2: Binary symmetric channel

The same measure can be used to describe the channel. If we use the binary

symmetric channel, shown in Fig. 2.2, as an example, the events in this case are; making an error with p_e and not making an error with $1 - p_e$. We can use Eq. (2.1) to calculate how much information is necessary to describe the errors, i.e., the conditional entropy of x given y, $\mathrm{H}(X|Y) = \mathrm{H}_b(p_e)$. Therefore, the amount of information R we can transmit per symbol is 1 [49]

$$\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y|X) \tag{2.3}$$

$$\begin{aligned} R &= \mathrm{H}(X) - \mathrm{H}(X|Y) \\ &= \mathrm{H}(Y) - \mathrm{H}(Y|X) \end{aligned} \tag{2.4}$$

the information that is transmitted information $\mathrm{H}(X)$ minus the information loss from the channel $\mathrm{H}(X|Y)$. The average of this rate is the maximum rate a binary code can achieve.

In our case, we have a channel with continuous probability functions. Therefore, we will use the continuous version of Eq. (2.1);

$$\mathrm{H}(X) = -\int_{-\infty}^{\infty} p_X(x) \log_2(p_X(x)) dx \tag{2.5}$$

$$\mathrm{H}(X,Y) = -\int\int_{-\infty}^{\infty} p_{X,Y}(x,y) \log_2(p_{X,Y}(x,y)) dy dx \tag{2.6}$$

and with a one-dimensional Gaussian distribution with variance σ^2, we can calculate the entropy as;

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x^2/2\sigma^2)} \tag{2.7}$$

$$\mathrm{H}(X) = \log_2 \sqrt{2\pi\, \mathrm{e}\, \sigma^2} \tag{2.8}$$

if the channel is additive $Y = X + N$ and both X and N are independent Gaussian

distributions, the capacity can be calculated using

$$C = \mathrm{H}(Y) - \mathrm{H}(Y|X) = \mathrm{H}(Y) - \mathrm{H}(N) \tag{2.9}$$

$$= \frac{1}{2}\log_2\left(2\pi\,\mathrm{e}(\sigma_X^2 + \sigma_N^2)\right) - \frac{1}{2}\log_2\left(2\pi\,\mathrm{e}\,\sigma_N^2\right) \tag{2.10}$$

$$= \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \tag{2.11}$$

commonly referred to as the Shannon capacity. Defining signal-to-noise ratio (SNR) as $\mathrm{SNR} = \frac{E[|X|^2]}{E[|N|^2]} = \frac{\sigma_X^2}{\sigma_N^2}$, we obtain the $C = \frac{1}{2}\log_2(1 + \mathrm{SNR})$ per real dimension. When we are using both the inphase and the quadrature to transmit data, the capacity is doubled to $\log_2(1 + \mathrm{SNR})$. This is the maximum information rate we can achieve for the additive white Gaussian noise (AWGN) channel. However, in practice we are transmitting a non-Gaussian signal.

2.1 Coded modulation

From Eq. (2.4), we can expect that there is a dependency of transmitted distribution X on the achievable information rate (AIR), where, for the actual data we want to transmit, we cannot put any restriction on it and therefore have to assume the worst, i.e, maximum entropy. The approach used to map the data to channel uses, which we refer to as coded modulation (CM) [50], has a great impact on the AIR.

CM generally consists of two parts, a forward error correction (FEC) and a modulation format, where both aspects together determine the throughput and the performance of the channel. There are two main strategies for designing the CM for a channel. In the first one, the FEC and modulation format are designed as one block. Here, the FEC has full knowledge of the channel and uses this to correct the errors and decode the message. The other strategy is to split up the FEC and the modulation format. This can be achieved by designing a FEC that maps from bits to bits and then maps those bits to symbols separately, and subsequently, at the receiver, converts the received symbols back to bit likelihoods for the FEC to decode. The second strategy is also known as bit interleaved coded modulation (BICM) [51, 52].

For both strategies, we can measure an upper bound for performance in the form

of an AIR, where for the nonbinary case, the AIR is measured between symbols in and symbols out, the BICM strategy has to take the bit mapping into account.

Both AIRs measure an achievable rate in bits, where the mutual information (MI) measures the rate for nonbinary FEC, the generalised mutual information (GMI) measures the rate for BICM. These rates can be calculated as follows [53]. For this, we are using a discrete set transmitted symbols, i.e., X produces symbols from the set $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$ with probabilities $p_1, p_2, \ldots, p_M$ and received symbols $Y \in \mathbb{R}$. Similarly to [53] we will combine Eq. (2.4) and (2.6) to calculate the MI

$$\begin{aligned} I(X;Y) &\triangleq \mathbb{E}_{\mathcal{X},\mathcal{Y}}\left[\log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right] \\ &= \sum_{\mathcal{X}} \int_{-\infty}^{\infty} p_{X,Y}(x,y) \log_2 \left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) dy \end{aligned} \tag{2.12}$$

and the GMI from [51, 52] as follows;

$$G = \sum_{k=1}^{m} I(B_k;Y) \tag{2.13}$$

where B_k denotes the k^{th} bit.

A Monte-Carlo integration of the metric is detailed in [53, Eq. (34)]. This integral can be calculated with samples generated with a normal distribution or samples that have been experimentally obtained. The simplified equation is

$$\lambda_k[t] = \log\left(\frac{\sum_{\mathcal{X}_{b_k=1}} p_{X|Y}(x|y[t])}{\sum_{\mathcal{X}_{b_k=0}} p_{X|Y}(x|y[t])}\right) \tag{2.14}$$

$$G \approx \frac{1}{N_s} \sum_{t=1}^{N_s} \sum_{k=1}^{m} \log_2\left(1 + \mathrm{e}^{(1-2*b_k[t])\lambda_k[t]}\right) \tag{2.15}$$

where $b_k[t]$ and $\lambda_k[t]$ are the bit and the log-likelihood ratio (LLR) for the t-th sample of Y respectively and $\mathcal{X}_{b_k=\{1,0\}} = \{x \in \mathcal{X} | b_k = \{1,0\}\}$.

2.2 Comparison

We have carried out an investigation comparing binary FEC and nonbinary FEC in [54] for the optical channel, where we found that the implementation of the scheme has a bigger impact than the difference in performance bound.

Nonbinary FEC with symbol-wise decoder structure

Binary FEC with bit-wise decoder structure

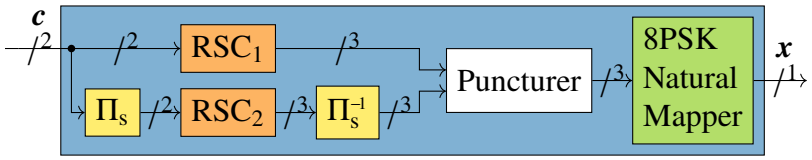
Figure 2.3: The two CM strategies considered in this work. One is the nonbinary FEC where the error correction and symbol mapping are applied as a single block. The other is BICM where a separate bitwise error correcting code is combined with a bit-to-symbol mapper.

The two strategies are shown in Fig. 2.3, where for the nonbinary FEC, turbo trellis-coded modulation (TTCM) [55] was implemented, while the binary FEC was implemented as low-density parity check (LDPC) [56] codes. Both strategies were using 8-phase-shift keying (PSK) for the modulation format.

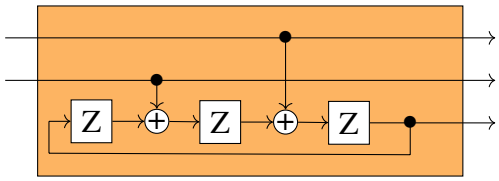
2.2.1 TTCM

The TTCM was implemented according to [55, 54], as shown in Fig. 2.4. Two identical recursive systematic convolutional (RSC) encoders were used, where one RSC encoder works on the bitstreams directly and the other encoder works on symbol by symbol interleaved bitstreams and directly deinterleaved after encoding, since the RSC code is systematic, meaning it will pass the data bits through unaltered , as can be seen in Fig 2.4(b). The puncturer will select alternating symbols from both RSC encoders, discarding the other. Now we have for every symbol the unaltered two input bits and a third parity bit. The RSC encoder is designed for 8-PSK with natural mapping, meaning the first two bits will add a π and $\pi/2$ radian phase shift respectively and the parity bit an additional $\pi/4$.

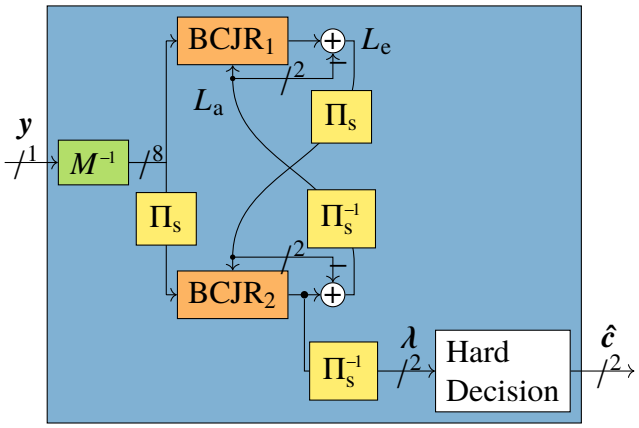
At the receiver, Fig. 2.4(c), the received noisy symbol will be converted into



a) NB-FEC Encoder



b) RSC Encoder



c) NB-FEC Decoder

Figure 2.4: The encoder a) consists of two R=2/3 RSC encoders b), where one is interleaved. Alternating symbols from both encoders are transmitted, the others discarded in the puncturer. The decoder c) has two BCJR decoders that operate using symbol observations, one for each RSC encoder at the transmitter.

8 symbol log likelihoods (LLs) and symbols split according to what encoder has generated them and padding with zeros for the other symbols. Two Balh, Cocke, Jelinek and Raviv (BCJR) [57] decoders were used to obtain symbol-wise a-priori information on the message bits, which is then shared with the other decoder. For consecutive passes, the a-priori information the decoder has used is subtracted from the information passed to the other receiver. After 10 iteration between the two decoders, the two message-bits per symbol are extracted and after hard decision compared to the transmitted bits.

2.2.2 LDPC

For the bit-wise binary coded modulation, an LDPC [56] was selected. We used the $R = 2/3$ LDPC code from the DVB-S2 standard [58, 59]. The encoder is shown in Fig. 2.5 and the decoder has the same structure reversed. The incoming bit stream is divided into three equal streams, each being encoded with the LDPC code, and consecutively combined together. A large bit-wise interleaver will spread the bits from all encoders through the frame, after which the bits are passed to the bit-to-symbol mapper. The same modulation format was used as in the TTCM implementation, but binary reflected Gray code (BRGC) [60] was used to map bits to symbols as it is the optimal for BICM [61].

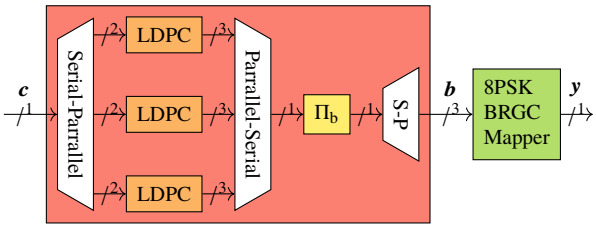


Figure 2.5: LDPC encoder implementation where three off-the-shelf LDPC encoders are interleaved and mapped to symbols using a BRGC. The receiver has the same structure in reverse.

The choice of interleaving three LDPC codewords was made to make both strategies more comparable. In Fig. 2.6, we compared the bit-error rate (BER) of a single LDPC codeword with an equal length TTCM codeword and three interleaved LDPC codewords with a three times longer TTCM codeword. We can see that increasing the TTCM codeword length has little impact on its performance, meaning it is sufficiently long for a good performance comparison. Interleaving three LDPC codewords has a different effect, whilst it does not change the SNR at which it falls off by much, the error floor is decreased to be more in line with the performance of the TTCM. Now for both schemes a hard-decision low-overhead outer FEC can be assumed to clear the remaining errors.

2.2.3 Theoretical comparison

For the comparison, we started by looking at the achievable information rates for the two strategies. In Fig. 2.7, the MI and GMI for 8-PSK are shown versus SNR. The

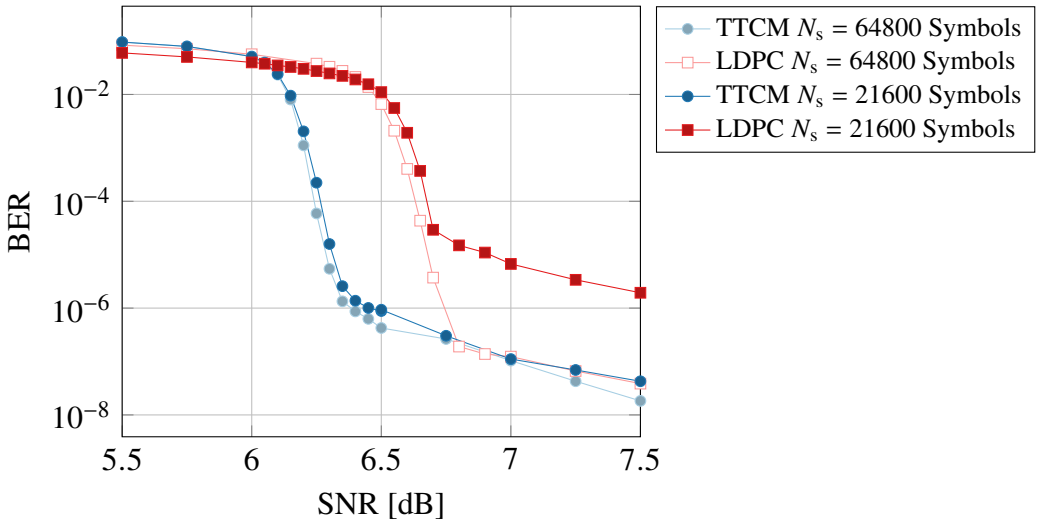


Figure 2.6: Impact of codeword length on the performance, where for the TTCM a three times longer interleaver was implemented. The LDPC length was increased by creating three codewords and interleaving them.

MI is obtain by numerically evaluating Eq. (2.12) for 8-PSK and the GMI is obtained by evaluating Eq. (2.13) for 8-PSK with BRGC. We see a very small difference in performance for both the strategies. This difference can be better seen in terms of BER. After decoding, the 8-PSK with $R = 2/3$ rate will output 2 message bits. The amount of information loss for these bits with a given error rate can be measured with Eq. (2.2). This gives a minimum BER for a given AIR, where

$$\text{AIR} \geq R * \log 2(M) * (1 - \text{H}_b(p)) \tag{2.16}$$

which can be solved numerically. This bound is called the distortion bound [62].

In Fig. 2.8, BER of the two strategies are compared versus SNR. First the minimum BER and the BER for the AWGN are shown. While, for the minimum BER, the two strategies only have a performance difference of 0.1 dB, the two implementations differ by approximately 0.4 dB SNR. Therefore, the implementations have a theoretically different implementation penalty which makes them hard to compare.

The experimentally obtained waveforms from a 1000 km recirculating loop are also shown in Fig. 2.8. Additional noise loading was needed to increase the BER to where the FEC breaks down. We explored two different methods of noise

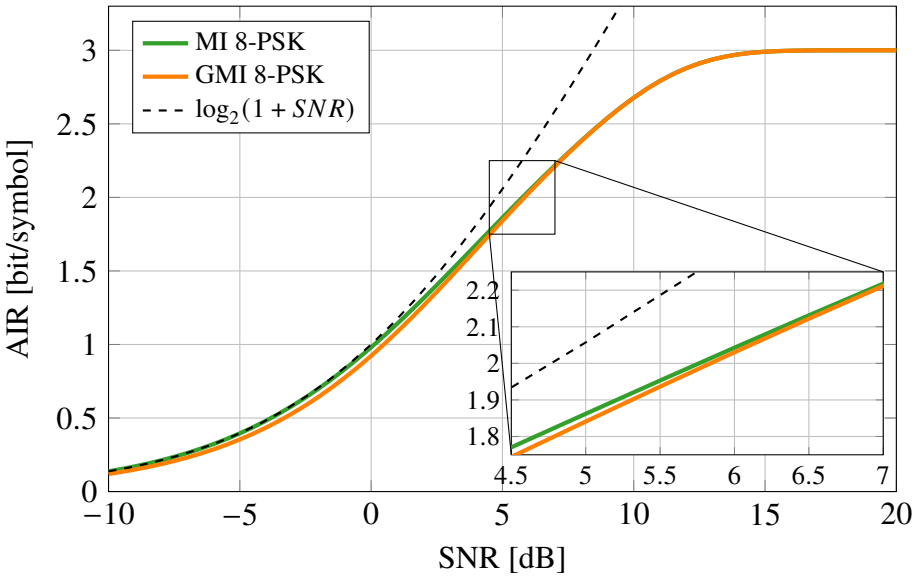


Figure 2.7: The achievable information rate for 8-PSK for nonbinary and binary FEC, i.e., MI and GMI respectively.

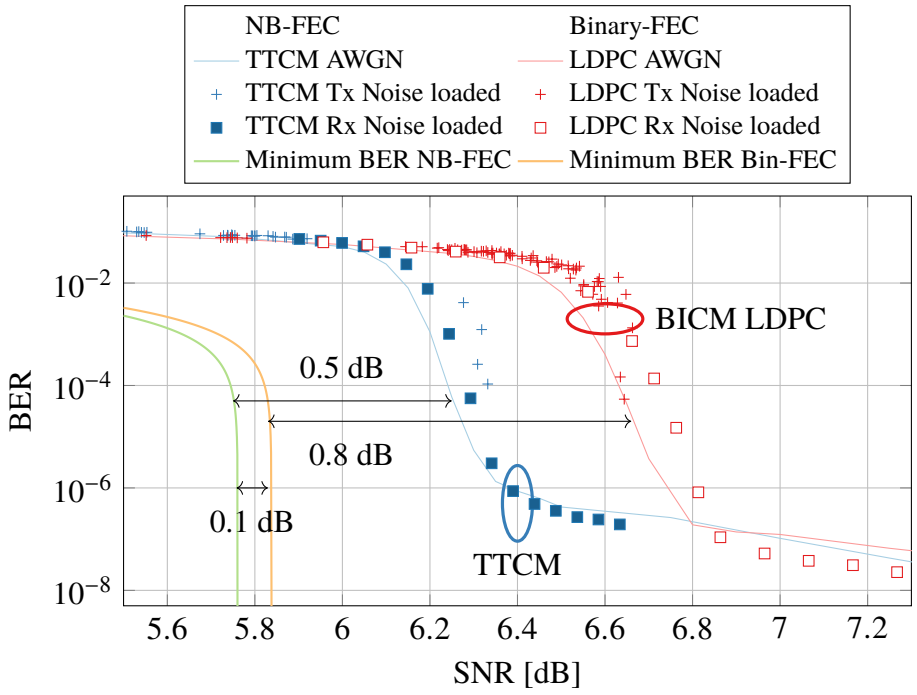


Figure 2.8: The BER of the TTCM and BICM LDPC compared to the theoretical performance.

loading the experimental data. In the first, we added the noise optically after the transmitter, resulting in the noise co-propagating through the optical fibre for the whole of the transmission distance. The second is where we add the noise digitally

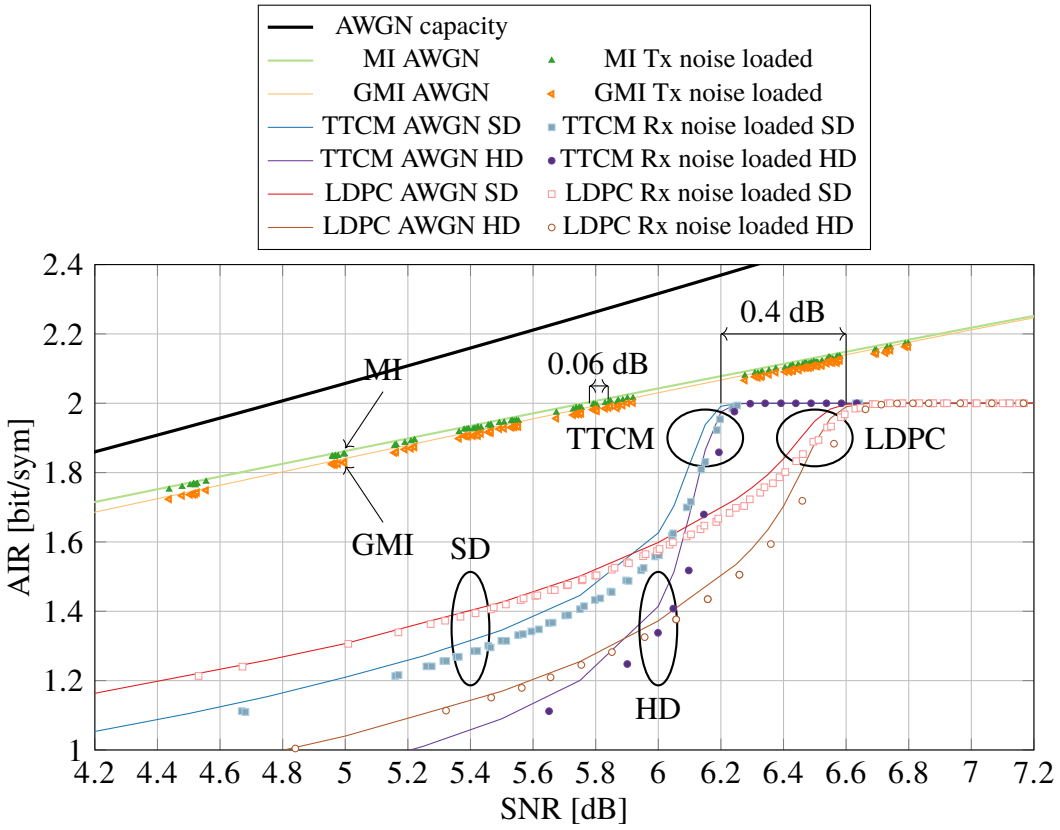


Figure 2.9: Performance of the coded modulation schemes, the theoretical curves for AWGN and the markers showing the realisations using experimentally obtained data.

after equalisation. Here, adding noise can be done repeatedly with great accuracy in the resulting noise variance. Both methods perform approximately equally, but the receiver noise loading gives us a finer control, allowing to show the fall of the BER curve with more detail. It also allows to show the error floor of the experiment.

The same information is also shown in the form of AIR versus SNR in Fig. 2.9. The reverse of what was at the distortion bound from Eq. (2.16), the binary entropy can also by used to show the post-FEC BER as an AIR. The result for AIR after hard decision (HD) are indicated with HD in the legend.

Since both the algorithms that decode the codewords also provide the soft information on the message bits in the form of LLRs, we can use Eq. (2.15) to calculate the post-FEC AIR as well. These results are indicated with soft decision (SD) in the legend.

In Fig. 2.9, these results are shown together with the MI and GMI for 8-PSK and

the AWGN capacity, $\log_2(1+\text{SNR})$. We observe the same 0.4 dB SNR performance difference between the two strategies as we have seen earlier. In this figure, we can also see the gap to the AWGN channel capacity, which has a greater SNR difference than the two strategies compared, a conclusion we could only make with the metrics introduced in this chapter.

2.3 Modulation format comparison

The same metrics can also be used to predict the performance across multiple constellation formats and different rates. Where Fig. 2.9 shows only a single constellation format for a single code rate, Fig. 2.10 shows the same plot for multiple code rates and another constellation format.

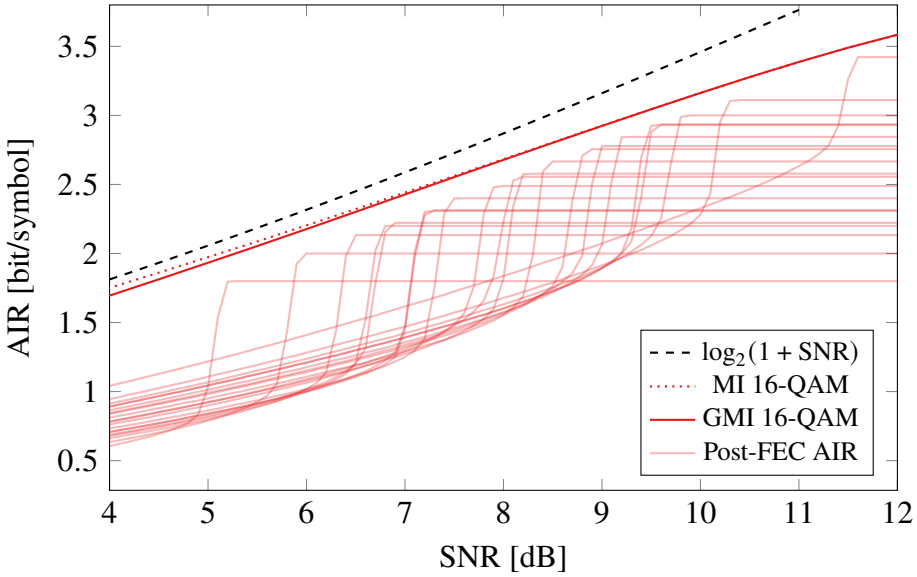


Figure 2.10: The DVB-S2X LDPC codes on 16-QAM, the AIR calculated from the post-FEC BER is shown.

In Fig. 2.10, we show the post-FEC AIR, from Eq. (2.16) for all DVB-S2X [59] LDPC codes with a 16-quadrature amplitude modulation (QAM) BRGC modulation. Alongside the AWGN capacity, the MI and the GMI, the lines show the upper bound for the assumed outer HD-FEC. The different LDPC codes with varying rates converge for a high SNR to their rate, e.g., $R = 2/3$, multiplied by the number of coded bits, i.e., $\log_2(16) = 4$. However, if we are not close to the AWGN capacity, there is most likely a code rate that will achieve a higher net rate. Therefore, these

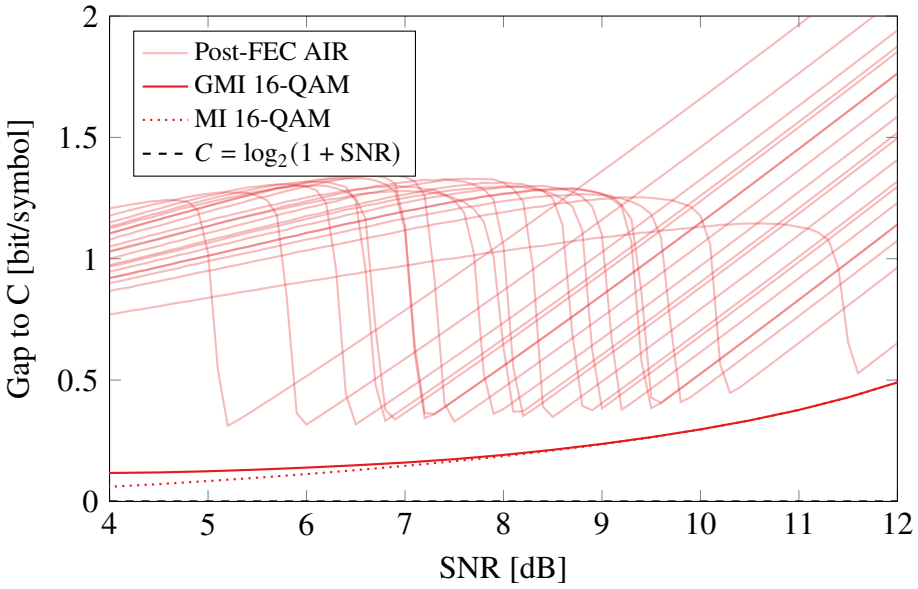


Figure 2.11: Gap to AWGN capacity for the DVB-S2X rates on 16-QAM

codes can be more accurately compared if they are shown as gap to the AWGN capacity. This is shown in Fig. 2.11, here all codes from the DVB-S2X standard [59], are evaluated on BRGC 16-QAM.

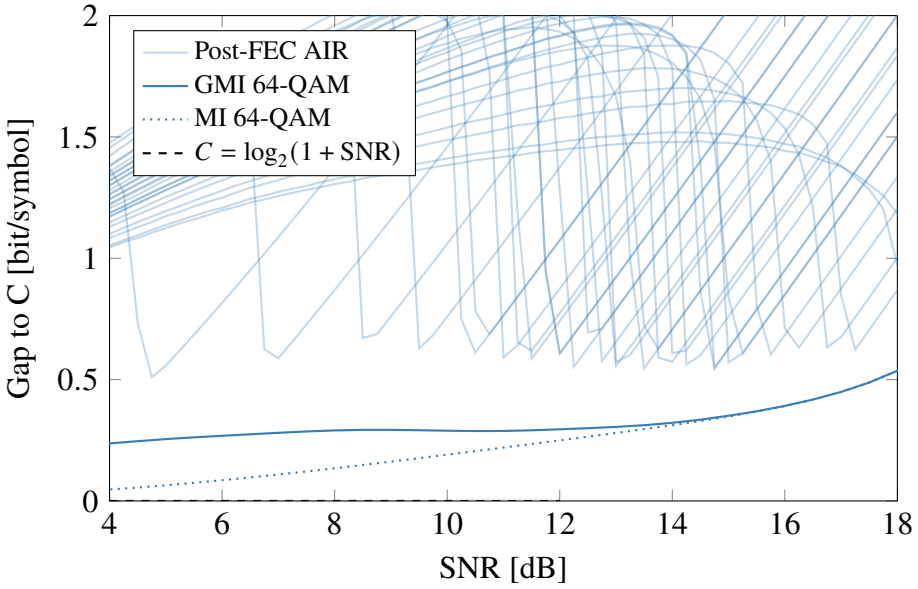


Figure 2.12: Gap to AWGN capacity for the DVB-S2 and DVB-S2X rates on 64-QAM

The same analysis is also carried out for 64 and 256-QAM and shown in Fig 2.12 and Fig 2.13 respectively. Now all LDPC codes from both the DVB-S2 [58] and its extension DVB-S2X [59] are evaluated. Both the MI and GMI are shown here

as well. From theory [52], the LDPC should be bounded by the GMI and not the MI. This can be seen for the lower SNRs, where the gap to AWGN capacity is increased and follows the GMI curve. Even in this comparison, the difference in implementation has a significant effect as well, because every LDPC is a different implementation from the same family.

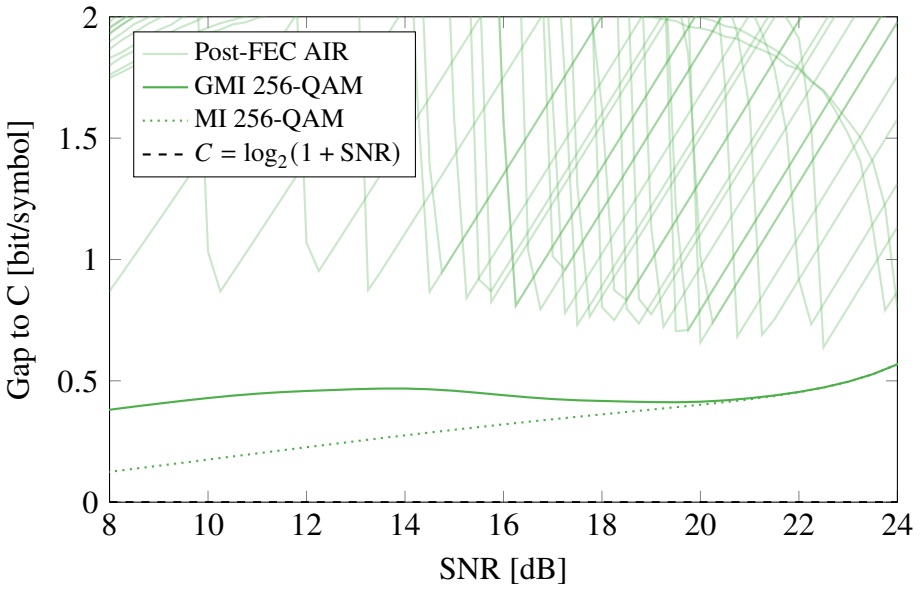


Figure 2.13: Gap to AWGN capacity for the DVB-S2 and DVB-S2X rates on 256-QAM

The gap to AWGN capacity can be used to compare constellation formats and select the format that has the best performance. In Fig. 2.14, the gap to the AWGN capacity is shown for the square QAM varying from 16 to 1,048,875 points. Both the MI and GMI are calculated and the gap to $\log_2(1 + \text{SNR})$ is shown. Because all integrals are evaluated for the AWGN channel, the Shannon capacity is the upper bound in performance. Therefore none of the results are negative. Note that the MI and GMI shown in Fig. 2.11, 2.12 and 2.13 are the same results as shown in this figure.

Furthermore, in Fig. 2.14, the MI is shown for continuous uniform distribution. Because we did not assign bits to the continuous distribution, we can only calculate the MI. The calculation is shown in Appendix. A. We can see that the MI for all the modulation formats follow the MI for a continuous uniform distribution until the discrete constellation format is limited by the number of points in the constellation.

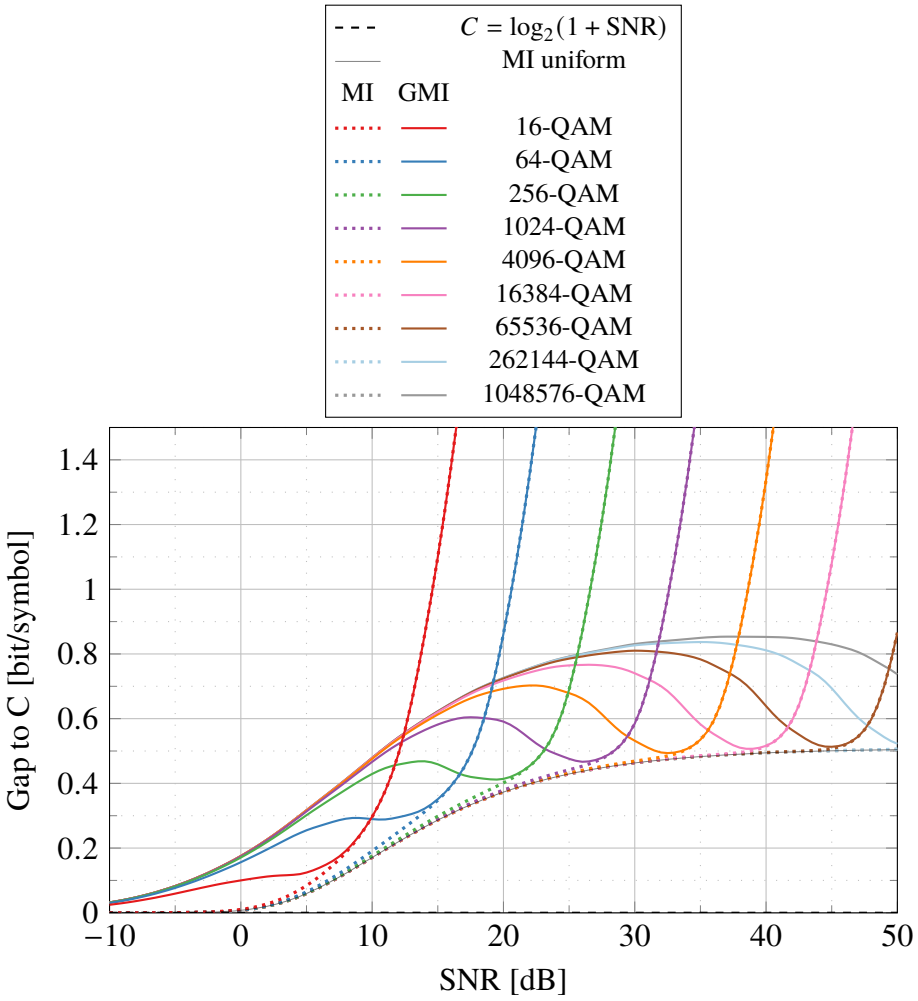


Figure 2.14: Gap to capacity for multiple constellation formats, where the curves for MI follow the MI curve for a continuous uniform distribution, and the GMI gets closer to capacity for a range of SNRs.

2.4 Nonlinear channel with the GN model

Until now, we have analysed the AIR for the AWGN channel. If we approximate the optical fibre transmission system as an AWGN channel, the performance can be predicted with the expression [63, 64]:

$$\mathrm{SNR}_i \approx \frac{P_i}{P_{\mathrm{ASE}} + \eta(f_i) P_i^3} \tag{2.17}$$

where P_i is the launched power, P_{ASE} the ASE power and $\eta(f_i)$ the nonlinear coefficient and f_i the centre frequency for channel i.

This expression can be extended to include effects such as Raman scattering

and modulation format dependency. If we follow the solution proposed in [65], we start a heuristic model of the fibre attenuation from [66, Ch. 3],

$$\alpha(\lambda) \approx \alpha_{\mathrm{IR}}(\lambda) + \alpha_{\mathrm{RS}}(\lambda) + \alpha_{\mathrm{im}} \tag{2.18}$$

$$\alpha_{\mathrm{IR}}(\lambda) = 7.81 \cdot 10^{11} \exp\left(\frac{-48.48 \cdot 10^{-6}}{\lambda}\right) \text{ dB} \tag{2.19}$$

$$\alpha_{\mathrm{RS}}(\lambda) = 0.148 \left(\frac{1550\text{nm}}{\lambda}\right) \text{ dB} \tag{2.20}$$

$$\alpha_{\mathrm{im}} = 0.024 \text{ dB} \tag{2.21}$$

where $\alpha_{\mathrm{IR}}(\lambda)$ is the infrared absorption, $\alpha_{\mathrm{RS}}(\lambda)$ is the Rayleigh scattering and α_{im} the imperfection losses. The result is shown in Fig. 2.15.

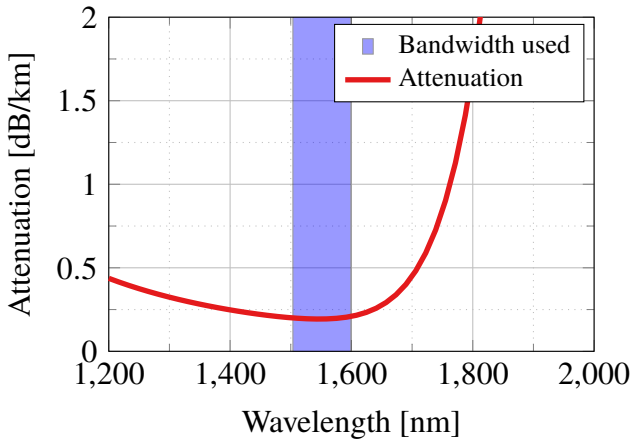


Figure 2.15: Fibre loss from the model, the highlighted area is the bandwidth used in this evaluation.

We present an example where we investigate the AIR for a 40-Gbd 300-channel transmission system, transmitted over ten 100-km spans of single mode fibre with a launch power of 0 dBm per channel. The system occupies 12 THz of bandwidth, which will have inter-channel stimulated Raman scattering (ISRS). This launched power is above the optimum launched power and shows a significant amount of nonlinear interference (NLI) and ISRS [67]. Using the fibre loss from (2.18), we can calculate the apparent loss, i.e., the loss of optical power from attenuation and Raman scattering combined, with the Raman gain by solving the differential

equation;

$$\frac{\partial P_i}{\partial z} = -\underbrace{\sum_{k=i+1}^{M} \frac{f_k}{f_i} g_{\mathrm{r}}(\Delta f) P_k P_i}_{\text{ISRS loss}} + \underbrace{\sum_{k=1}^{i-1} g_{\mathrm{r}}(\Delta f) P_k P_i}_{\text{ISRS gain}} - \alpha\left(f_i\right) P_i, \tag{2.22}$$

where P_i and P_k are the powers of channels i and k, the latter being at frequency f_k, M is the total number of channels $\Delta f$is the frequency spacing between the channel i and k, $g_{\mathrm{r}}(\Delta f)$ is the ISRS gain coefficient which is dependent on the frequency spacing between the channels and $\alpha(f_i)$ is the fibre attenuation in the absence of ISRS. The result is plotted in Fig. 2.16.

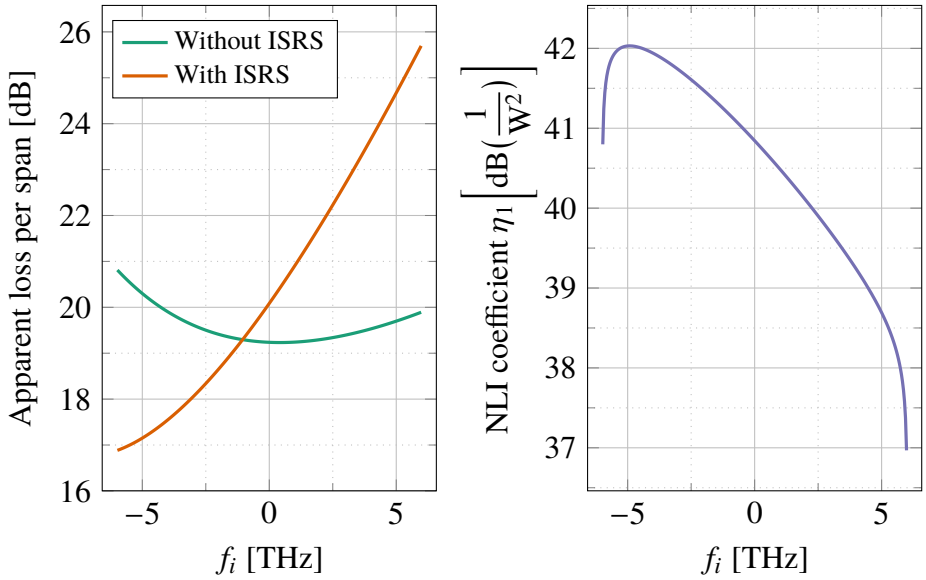


Figure 2.16: Apparent loss and nonlinear coefficient for a 300 channel transmission versus the centre frequency of every channel f_i. The apparent loss without ISRS is the fibre attenuation, while with ISRS, a differential equation is used to calculated the apparent loss for every channel, which in turn is used to calculate the nonlinear coefficient.

In Fig. 2.16, the nonlinear coefficient $\eta(f_i)$ used in (2.17) is shown as calculated from [65]:

$$\eta_n\left(f_i\right) \approx \sum_{j=1}^{n} \left[\frac{P_{i,j}}{P_i}\right]^2 \cdot \left[\eta_{\mathrm{SPM,j}}\left(f_i\right) n^{\epsilon} + \eta_{\mathrm{XPM,j}}\left(f_i\right)\right], \tag{2.23}$$

where η_{SPM} is the contribution to the nonlinear interference coefficient from self-phase modulation (SPM) and η_{XPM} is the contribution from cross-phase modulation

from all the other wavelength division multiplexing (WDM) channels. P_i is the power of channel i launched into the first span, and $P_{i,j}$ is the power of channel i launched into the jth span. The contributions from the SPM are assumed to contribute coherently with coherence factor ϵ. Now we can use the closed form expressions from [65], $\eta_{\mathrm{SPM}}(f_i)$ and $\eta_{\mathrm{XPM}}(f_i)$ for a link with fibre attenuation α, group velocity dispersion β_2, group velocity dispersion slope β_3, nonlinear coefficient γ, and SRS gain slope C_{r}:

$$\begin{aligned}\eta_{\mathrm{SPM}}\left(f_i\right) \approx & \frac{4}{9}\frac{\gamma^2}{B_i^2}\frac{\pi}{\phi_i\bar{\alpha}\left(2\alpha+\bar{\alpha}\right)} \\ & \cdot\left[\frac{T_i-\alpha^2}{\alpha}\mathrm{asinh}\left(\frac{\phi_i B_i^2}{\pi a}\right)+\frac{A^2-T_i}{A}\mathrm{asinh}\left(\frac{\phi_i B_i^2}{\pi A}\right)\right],\end{aligned} \tag{2.24}$$

with $\phi_i = \frac{3}{2}\pi^2\left(\beta_2 + 2\pi\beta_3 f_i\right)$, $A = \alpha + \bar{\alpha}$ and $T_i = \left(\alpha + \bar{\alpha} - P_{\mathrm{tot}}C_{\mathrm{r}}f_i\right)^2$, in which

$$\begin{aligned}\eta_{\mathrm{XPM}}\left(f_i\right) \approx & \frac{32}{27}\sum_{k=1,k\neq i}^{N_{\mathrm{ch}}}\left(\frac{P_k}{P_i}\right)^2\frac{\gamma^2}{B_k\phi_{i,k}\bar{\alpha}\left(2\alpha+\bar{\alpha}\right)} \\ & \cdot\left[\frac{T_k-\alpha^2}{\alpha}\mathrm{atan}\left(\frac{\phi_{i,k}B_i}{\alpha}\right)+\frac{A^2-T_k}{A}\,\mathrm{atan}\left(\frac{\phi_{i,k}B_i}{A}\right)\right],\end{aligned} \tag{2.25}$$

with $\phi_{i,k} = 2\pi^2\left(f_k - f_i\right)\left[\beta_2 + \pi\beta_3\left(f_i + f_k\right)\right]$ and tuning parameter $\bar{\alpha} = \alpha$ set to the attenuation.

Using the results shown in Fig. 2.16, we are one step further to calculating (2.17). We calculate P_{ASE} by assuming a gain-flattened EDFA with a noise figure, of 4.5 dB. We can calculated the ASE power per span as;

$$P_{\mathrm{ASE},i} = 2n_{\mathrm{sp}}(L_{\mathrm{app}} - 1)h\nu B_i \tag{2.26}$$

with $n_{\mathrm{sp}} = \frac{1}{2}10^{\frac{4.5}{10}}$ and $L_{\mathrm{app}} = \frac{P_i(z=0)}{P_i(z=l)}$ from (2.22) with l the fibre length, $h\nu$ the photon energy and B_i the channel bandwidth.

The SNR we can now calculate is shown in Fig. 2.17. This includes the ISRS for the apparent loss and the nonlinear phase modulation. The result is an SNR

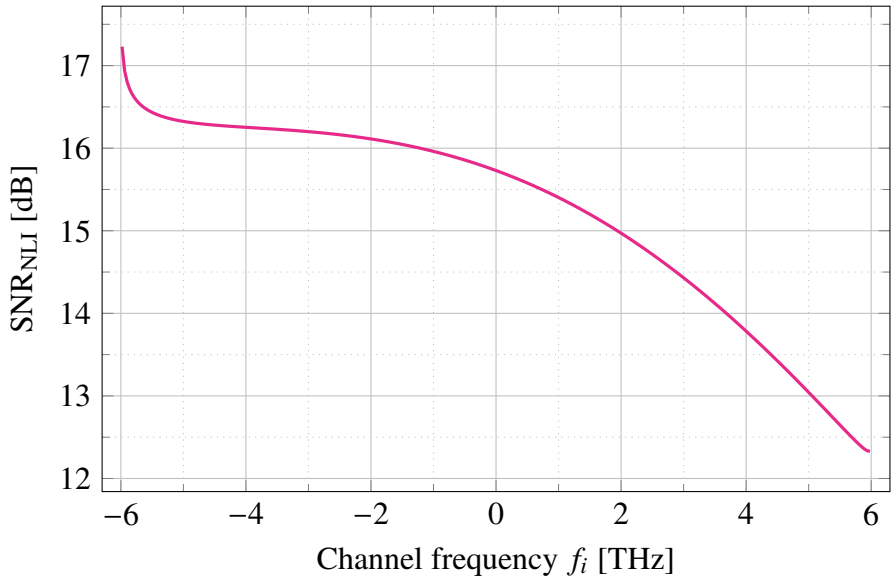


Figure 2.17: The calculated SNR per channel using the ISRS GN model. The higher frequencies have a higher apparent loss due to Raman pumping and therefore a lower SNR.

where we treat the resulting interference as noise.

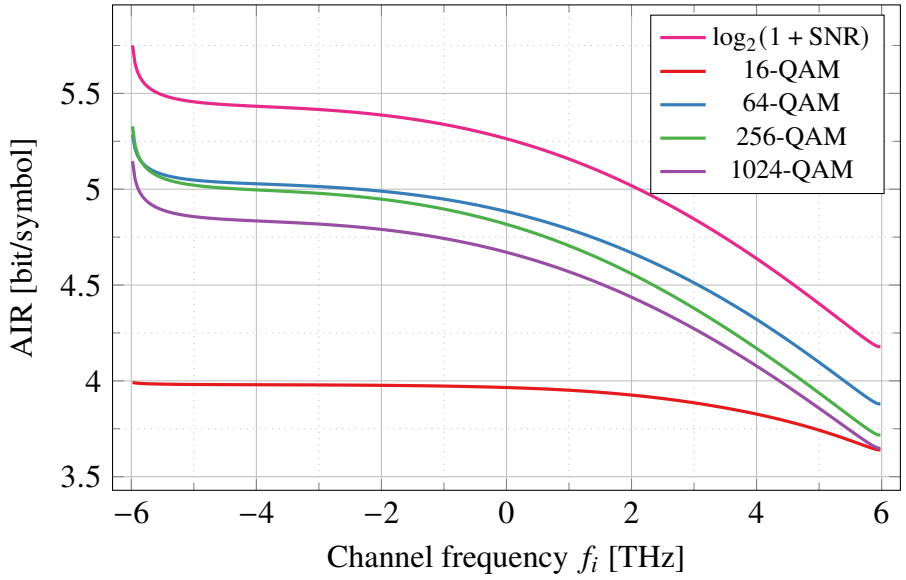


Figure 2.18: The calculated AIR per channel for the AWGN assumption.

The SNR is shown as AIR in Fig. 2.18. To convert the predicted SNR into an AIR, there a few options. One of them is to look at the AWGN capacity $\log_2(1 + \text{SNR})$. And since we treat the channel as if it was a AWGN channel, the resulting SNR is the upper bound of the AIR [68]. Another method is to use (2.13)

to calculate the performance of a modulation format for a given SNR. The results for this calculation are also shown in Fig. 2.18 for 16, 64, 256 and 1024-QAM and we can see that for this system 64-QAM performs the best for most of the channels.

This analysis is valid for the $\log_2(1 + \text{SNR})$ case, because it assumes a transmitted Gaussian distribution and the Gaussian noise (GN) model has the same assumption. However, when we change the modulation format, we also change the nonlinear response of the channel [69, 70]. This is something we will exploit in the next chapter.

2.5 Summary

This chapter has explored methods of measuring the achievable information rate of an AWGN channel. We implemented a variety of CM methods and demonstrated accurate prediction of their performance. The results of this work will be exploited in the next chapter to design signal constellation which are optimised for the nonlinear channel, increasing achievable information rates.

Chapter 3

Constellation shaping for the optical transmission system

From the last chapter, we carry a few concepts over to the next chapter, SNR, MI and GMI. We ended the last chapter by concluding that the channel response is affected by the transmitted signal itself. The excess kurtosis of the modulation formats has an impact on the amount of nonlinear interference introduced by the channel. The excess kurtosis is defined as the 4-th order moment normalised for a Gaussian distribution being zero;

$$\Phi(X) = \frac{\mathbb{E}[|X|^4]}{(\mathbb{E}[|X|^2])^2} - 2 \tag{3.1}$$

which we can use to expand Eq. (2.23) to take the effect of the constellation into account. Then, following [71], the correction will be

$$\begin{aligned}\eta_{\text{corr.},n}(f_i) \approx \frac{80}{81}\Phi \sum_{k=1,k\neq i}^{N_{\text{ch}}} \left(\frac{P_k}{P_i}\right)^2 \frac{\gamma^2}{B_k} \Bigg\{ & \frac{1}{\phi_{i,k}\bar{\alpha}\,(2\alpha+\bar{\alpha})} \\ & \cdot \left[\frac{T_k-\alpha^2}{\alpha}\text{atan}\left(\frac{\phi_{i,k}B_i}{\alpha}\right) + \frac{A^2-T_k}{A}\,\text{atan}\left(\frac{\phi_{i,k}B_i}{A}\right)\right] \\ & + \frac{2\pi\tilde{n}T_k}{|\phi|\,B_k^2\alpha^2A^2}\left[(2\,|\Delta f|-B_k)\log\left(\frac{2\,|\Delta f|-B_k}{2\,|\Delta f|+B_k}\right)+2B_k\right]\Bigg\},\end{aligned} \tag{3.2}$$

with Φ the excess kurtosis of the constellation format used in all channels. Typical values for this metric range between -1 and 0. The excess kurtosis for

uniform QAM modulation is shown in table. 3.1. P_i and P_k are the launched power of the i-th and k-th channel with frequencies f_i and f_k respectively, $\Delta f = f_k - f_i$, fibre attenuation α, group velocity dispersion β_2, group velocity slope β_3, nonlinear coefficient γ, channel bandwidth B_i and $\phi_i = \frac{3}{2}\pi^2 \left(\beta_2 + 2\pi\beta_3 f_i\right)$, $\phi_{i,k} = 2\pi^2 \left(f_k - f_i\right) \left[\beta_2 + \pi\beta_3 \left(f_i + f_k\right)\right]$, $A = \alpha + \bar{\alpha}$, $T_i = \left(\alpha + \bar{\alpha} - P_{\text{tot}} C_{\text{r}} f_i\right)^2$.

Table 3.1: Excess kurtosis of selected modulation formats.

Modulation format	Excess kurtosis Φ
QPSK	-1
16-QAM	-0.6800
64-QAM	-0.6190
256-QAM	-0.6050
1024-QAM	-0.6012
uniform distribution	-0.6000
Gaussian distribution	0

Use of the model with the modulation format correction allows us to predict the performance for difference modulation formats. In Fig. 3.1, we have used the same parameters as used in Fig. 2.17 and applied the correction for a number of constellation formats. We can see that switching to quadrature phase shift keying (QPSK) results in the greatest performance increase and the other formats have a lower but still significant impact compared to Gaussian modulation, for which the correction is 0 and leading therefore to the same result as without the correction.

With this knowledge we show the difference in AIR for the different modulation formats in Fig. 3.2. The results from Fig. 2.18 are shown as dotted results. It is clear that the model without the modulation format correction underestimates the performance of the nonlinear fibre channel. Because QPSK is limited by the cardinality for every channel, it always achieves about 2 bit/symbol. The AIR for 16-QAM is also severely limited by the cardinality, therefore despite having a SNR gain, shows little improvement with the better model.

The 1024-QAM channel is too far from capacity for this SNR range to be the best performing constellation format, both with and without considering the effect of the constellation format on the nonlinear fibre channel. Both 64-QAM and 256-

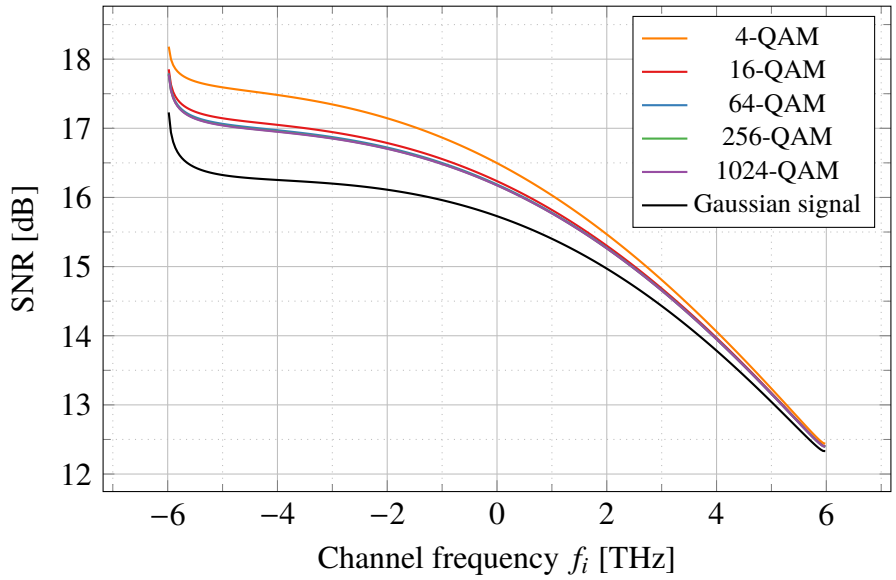


Figure 3.1: SNR for with the modulation format correction incorporated into the ISRS GN model.

QAM perform similarly. The greatest performance increase is seen at the lower frequencies, where the nonlinear interference is the dominating noise source.

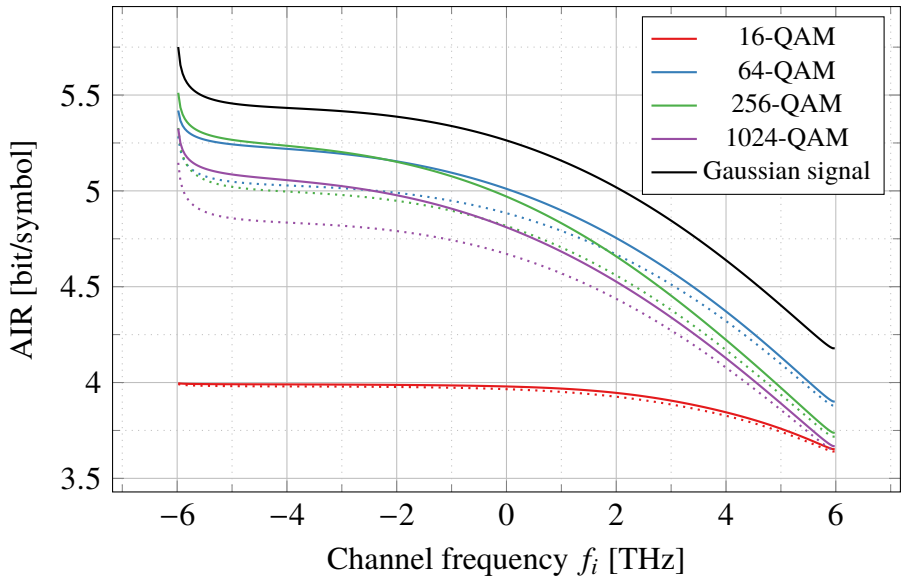


Figure 3.2: AIR with the modulation format correction, using the SNR from the modulation format corrected ISRSGN model.

The system performance is still worse than that achieved with the Gaussian signal. To address this, first, we will make our signal more Gaussian-like to close the gap for the linear channel. Then we are going to achieve a better result for the

nonlinear channel by taking the difference in performance from fibre nonlinearity into account.

3.1 Constellation shaping for the linear channel

Previously we have observed that there is a fixed gap between the performance of square modulation formats and the performance in MI of a Gaussian signal for the AWGN channel. We have seen that the practical square uniform modulation formats approach the performance of a uniform distribution. So, intuitively, a more Gaussian shaped constellation should approach the AWGN capacity.

In this work, we will use two methods to achieve a more Gaussian shaped constellation whilst still being able to transmit any bit message. The first is probabilistic shaping and the other geometric shaping. The probabilistic shaping still uses a square QAM, but changes the probabilities of the individual points [72], whereas the geometric shaping keeps the probabilities uniform, but changes the geometric positions of the constellation points [73].

3.1.1 1D constellation shaping

Probabilistic shaping

For the probabilistic constellation shaping, it is easiest to stick to the one-dimensional constellation diagrams. If we follow [72, Eq. (12)], we find the optimal constellation for the AWGN channel to be

$$p_X(x_i) = \frac{\exp(-\lambda||x_i||^2)}{\sum_{j=1}^{M} \exp(-\lambda||x_j||^2)} \tag{3.3}$$

this distribution can be applied to the uniform square QAM constellation. This is also known as the Maxwell-Boltzmann (MB) distribution.

One method to achieve an arbitrary distribution is using constant composition distribution matching (CCDM) from [74], such that, by using arithmetic codes [75, 76], with decoder at the transmitter and the encoder at the receiver, the equiprobable bits are transformed into symbols with the desired distribution. When the symbols only describe the amplitude information of the 1 dimensional constellation, the sign

information can be used to transmit arbitrary bits. This will also ensure a zero mean realised constellation. To ensure the distribution of the symbols does not change at the receiver, a FEC code is implemented between the channel and the CCDM. For the FEC to not change the distribution, a systematic code has to be chosen. The symbols are passed through the systematic part of the code and therefore remain unchanged. The parity bits are then used for the sign bits [77]. With an M-QAM constellation, $m = \log_2(M)$ bits can be assigned to the 2-dimensional constellation. With $\frac{m}{2} - 1$ bit per dimension used by the output of the CCDM, the last bit is for the output of the FEC. When a higher rate FEC is desired, additional bits can be generated, passed through the systematic part of the FEC and transmitted as sign bits.

Geometric shaping

Another method of closing this gap to capacity is to move equiprobable constellation points, such that their density appears more Gaussian shaped. This is known as geometric constellation shaping and has been shown to approach capacity [73].

In this work, we have numerically optimised the locations of the constellation points for a given channel. In the case of the AWGN channel, this channel is described by its SNR. Using a fast calculation of GMI, employing the Gauss-Hermite quadrature integration of the integral [53, Eq. (42)-(43)], the calculation is fast enough to approximate the gradient with exploratory steps. Then a gradient descent or a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is used to obtain an optimised constellation [78, 79, 80, 81].

Comparison between probabilistic and geometric constellation shaping

In Fig. 3.3 and Fig. 3.4, a comparison is made between uniform square QAM, a 1D geometrically shaped constellation and a probabilistically shaped constellation. For all three cases the same rate was achieved, the square QAM and the geometrically shaped constellation were both corrected using the same LDPC code and because the two constellations had the same number of points, with each point having a unique

bit label, the same rate was achieved. For the probabilistically shaped constellation, a higher rate FEC was needed, as the CCDM uses part of the overhead to achieve the shaped distribution.

For the 64-QAM comparison in Fig. 3.3, a net rate of $2/3$ was chosen. For the square and the geometrically shaped (GS) constellations, the $R = 2/3$ LDPC code from [58] was used, resulting in 4 bit per symbol. For the probabilistically shaped (PS) constellation, the CCDM, using 1.75 bit per 1D symbol and a bypass rate of 0.25 bit per 1D symbol, resulted in the same 4 bit per symbol using the $R = 3/4$ LDPC from [58]. The minimum SNR required is given by $4 = \log_2(1 + SNR)$, which is $2^4 - 1 \approx 11.76$ dB.

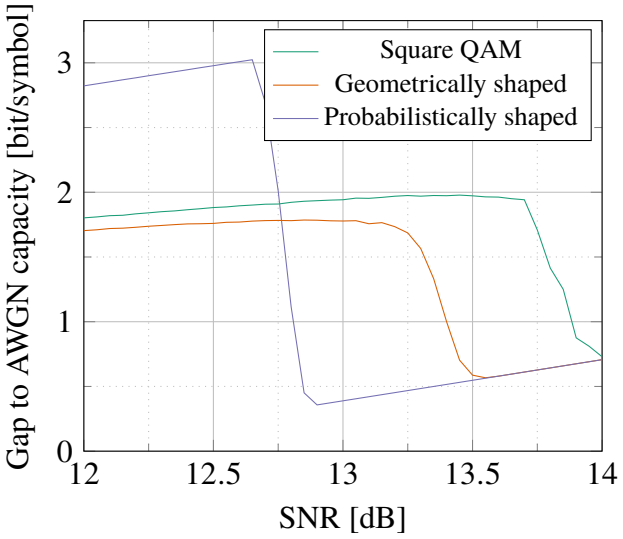


Figure 3.3: Gap to capacity comparing probabilistic and geometric shaping with uniform square QAM. For all formats, 8 points per dimension and a net rate $2/3$ was chosen. This is equivalent to 64-QAM.

For the 1024-QAM comparison in Fig. 3.4, a net rate $4/5$ was chosen. The square and GS constellations were coded with the $R = 4/5$ LDPC code from [58]. The PS constellation has a CCDM with $3\frac{5}{6}$ bit per 1D constellation and the bypass rate $1/6$ for the $R = 5/6$ LDPC from [58]. Both schemes now achieve a rate of 8 bit per symbol, with the required SNR of $2^8 - 1 \approx 24.1$ dB.

In Fig. 3.6, Fig. 3.7 and Fig. 3.8, the 1D constellations are shown with the noise for the given SNR. For every pixel, the probability density function (PDF)

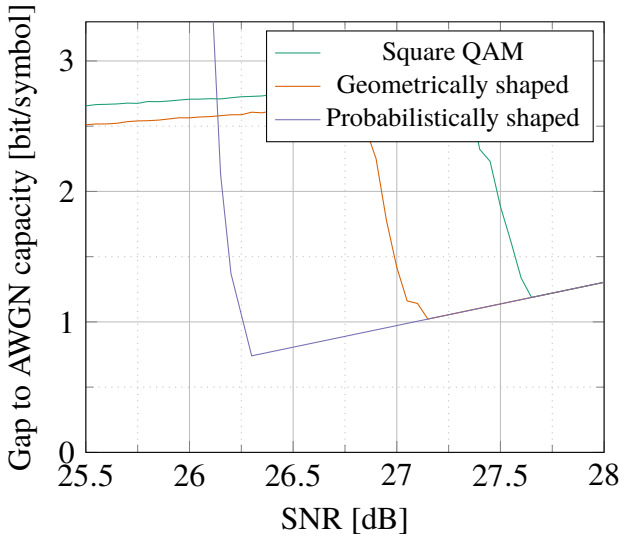


Figure 3.4: Gap to capacity comparing probabilistic and geometric shaping with uniform square QAM. For all formats, 32 points per dimension and a net rate $4/5$ was chosen. This is equivalent to 1024-QAM.

of the received signal was integrated and normalised to the maximum, $1/M$ at the maximum SNR. This has the consequence that the lower cardinality constellations have a lower intensity.

For the uniform square QAM in Fig. 3.6, we can see that if the SNR is low enough, the PDF of the received signal resembles a uniform distribution with a roll-off at the extrema.

For the geometrically-shaped QAM in Fig. 3.7, the constellations were optimised for every SNR. The resulting constellations sometimes have the points collapsed onto a single coordinate, leaving the decoder unable to predict the differing bit between the bit labels, trading it for more certainty on the other bits.

For the probabilistically shaped QAM in Fig. 3.8, the λ parameters in equation (3.3) have been numerically optimised with respect to MI for every SNR. The resulting constellation was then normalised taking the new distribution into account.

We can see that the uniform QAM has more uniform-like received distribution, whereas both the shaped QAM have a more Gaussian-like distribution. Where the MI optimised PS has a smoother appearance, the GMI optimised GS shows structures to increase the distance between bits. This allows for efficient bit demapping. Note

that the GS constellation can be demapped with a single symbol, the PS CCDM spreads the information over as many symbols as the size of the CCDM.

In Fig. 3.5, the results of both PS and GS are shown in the same format as Fig. 2.14. The MI of the PS is always lower than the GMI of the GS, this gap is partially from the bit to symbol mapping that is included in the GS as it is designed to find a trade-off that includes this mapping.

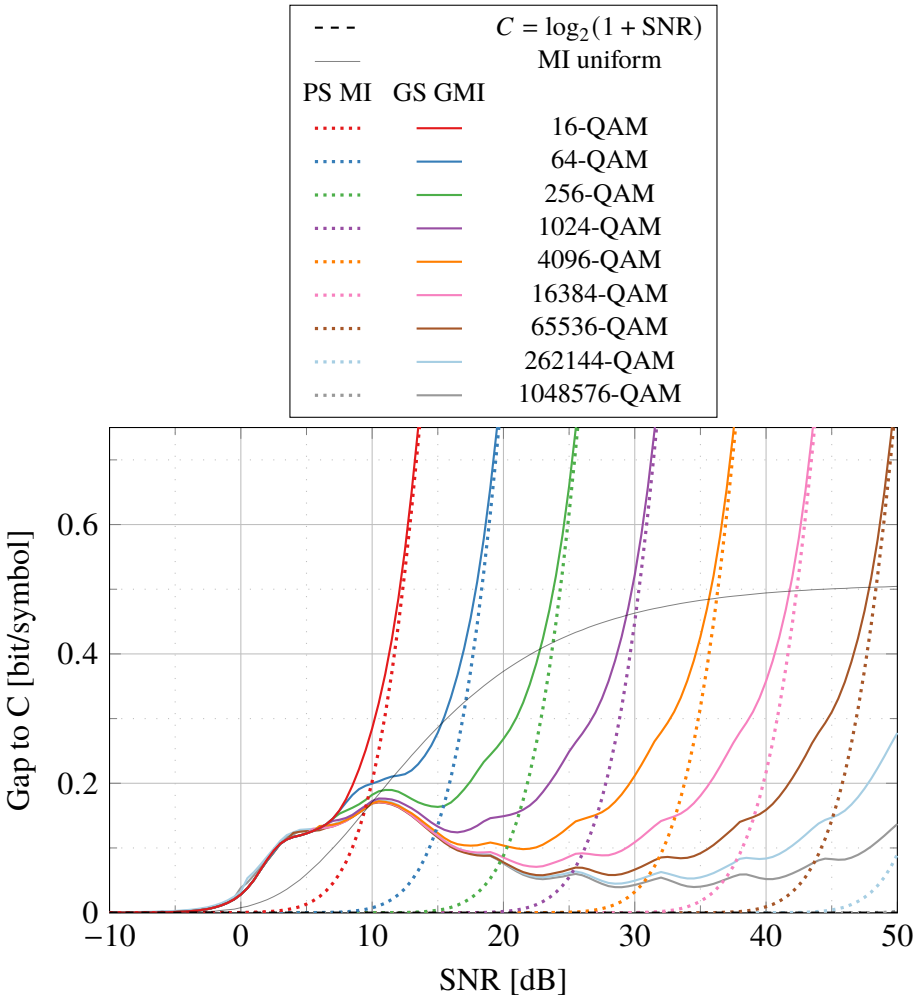


Figure 3.5: Gap to capacity for 1D optimised constellations. For the PS constellations the MI gap is shown, for GS constellations the GMI gap.

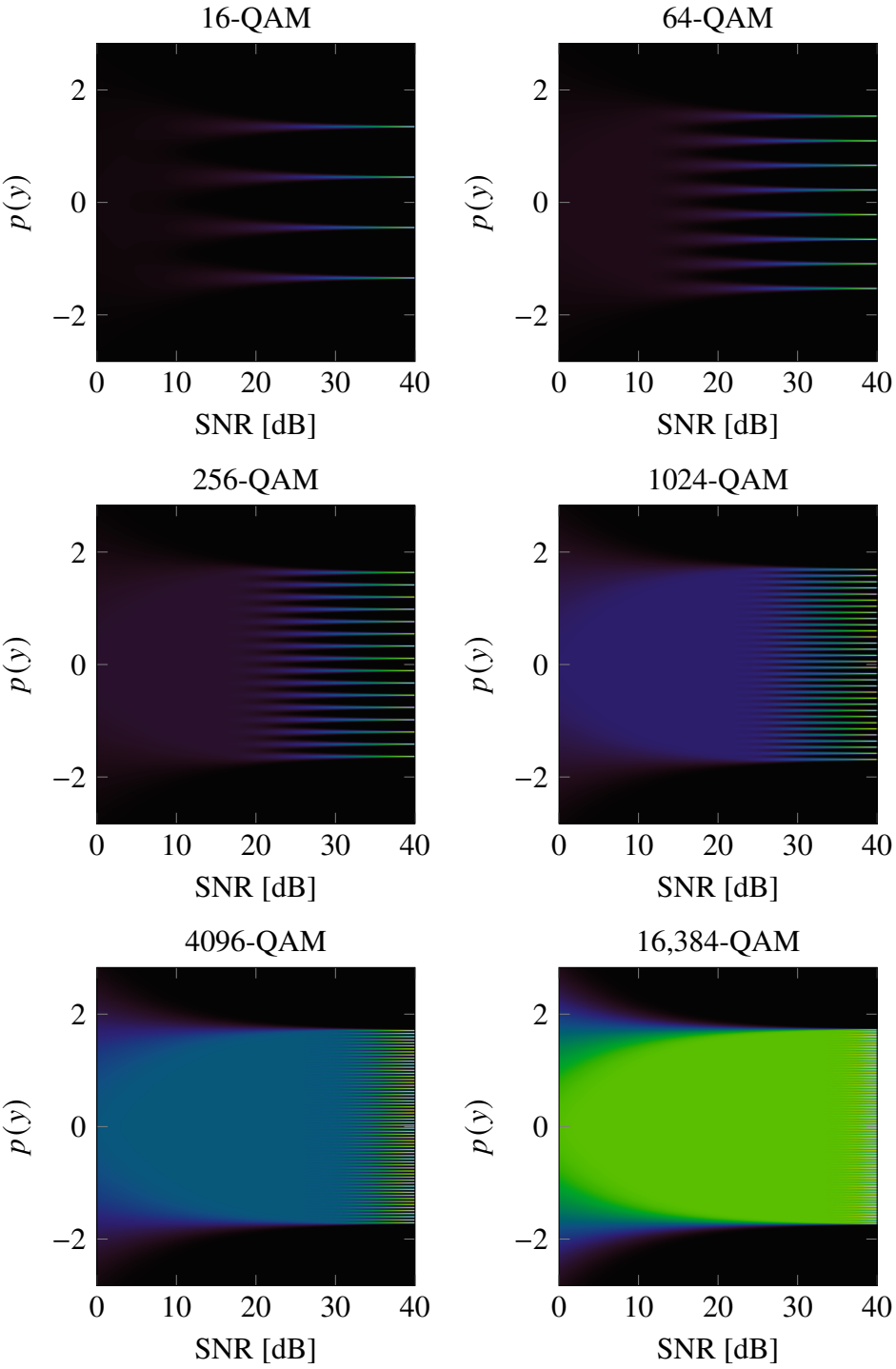


Figure 3.6: Received probability distribution for uniform QAM.

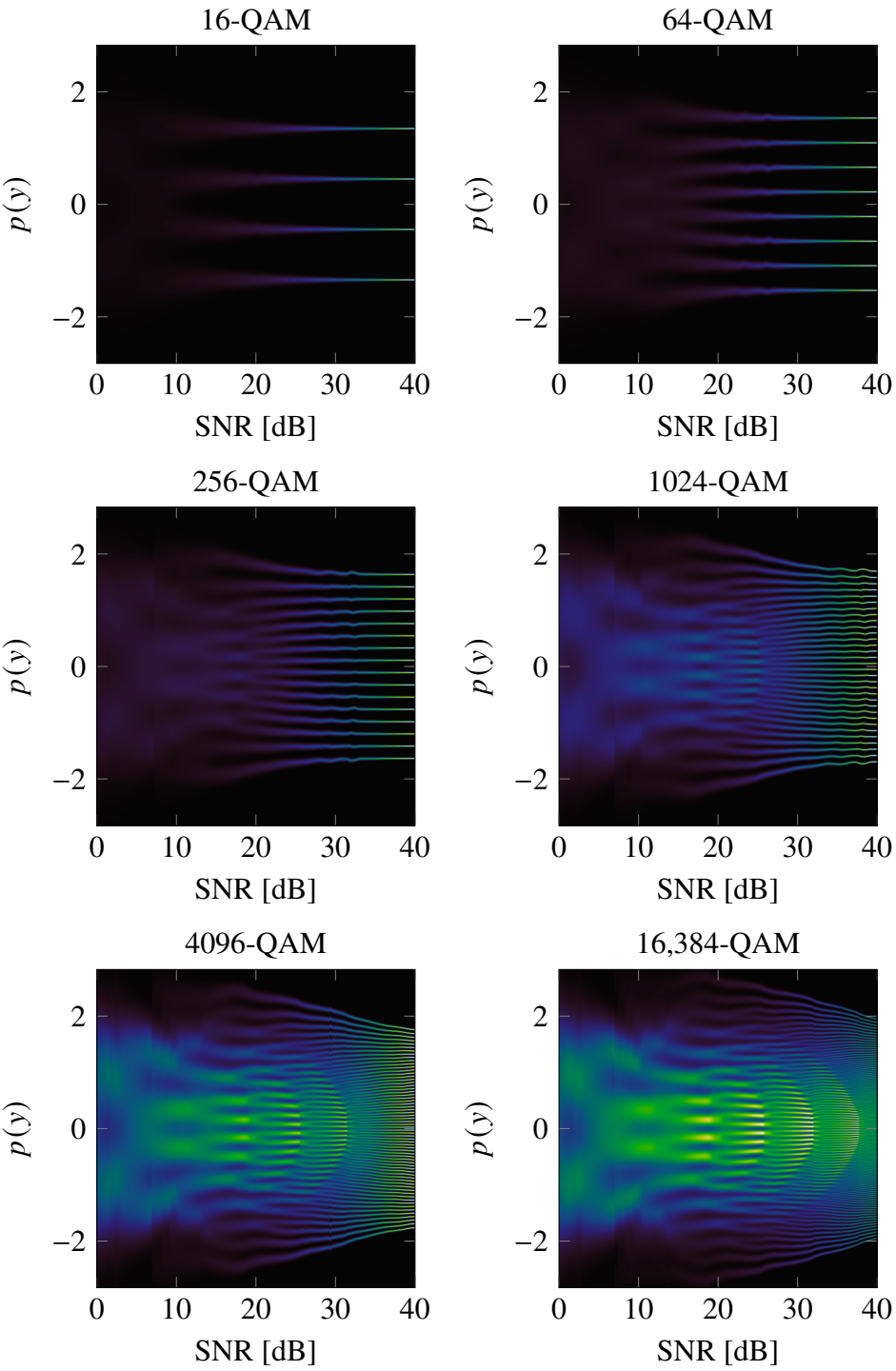


Figure 3.7: Received probability distribution for GS QAM.

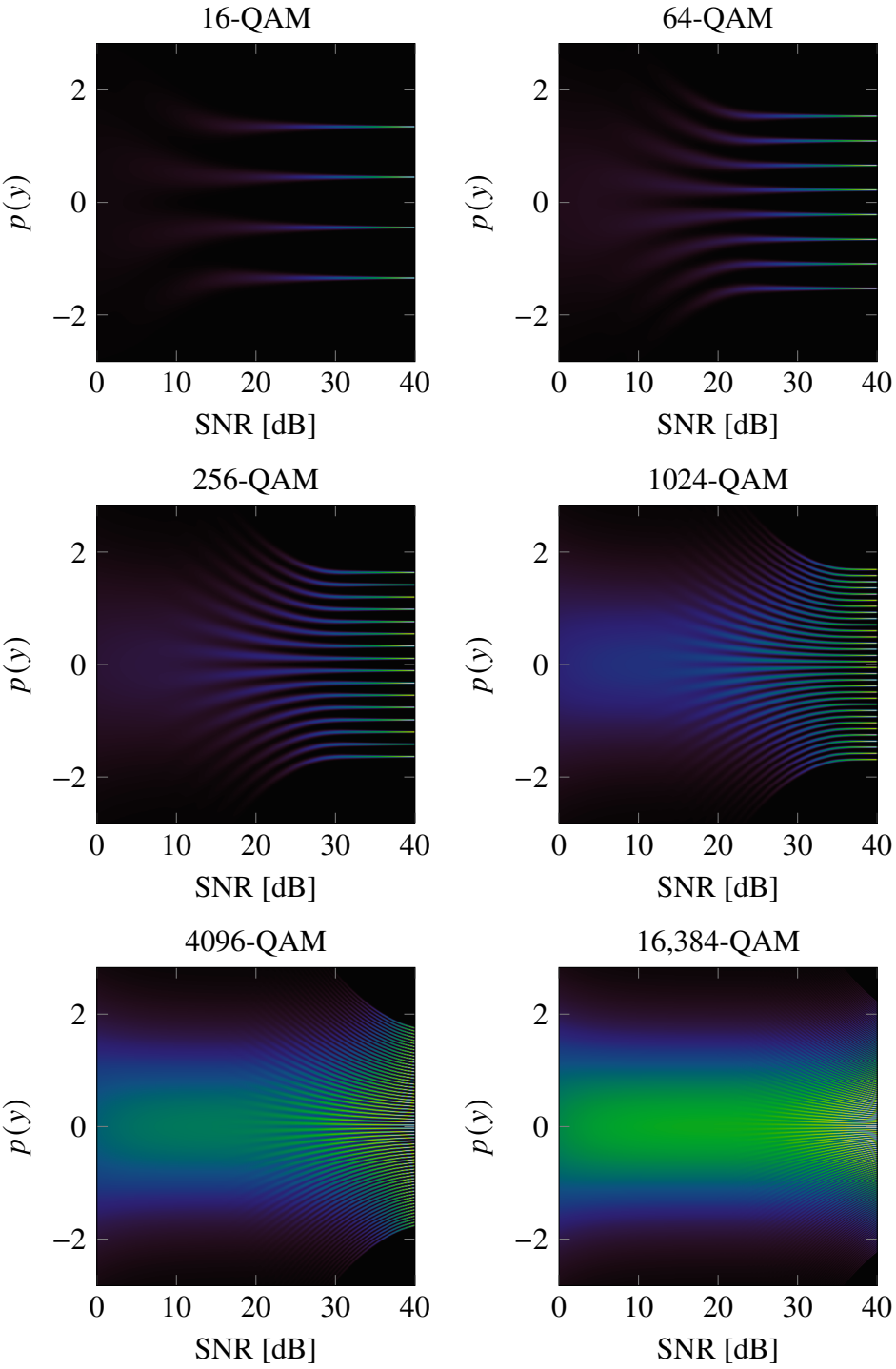


Figure 3.8: Received probability distribution for PS QAM.

3.2 Probabilistic shaping for the nonlinear optical channel

The shaping strategies investigated in this chapter were, until now, tailored to the AWGN channel. In this section we will investigate how to apply these methods for the nonlinear optical transmission system. We have proposed a simple distribution that outperforms the optimum for the AWGN channel in [82].

We used an analytical model similar to equation (2.17), but with a simpler modulation format correction than equation (3.2). Similar to [83], we assume that the nonlinear interference noise can be written as

$$\eta_{\text{tot}} P^3 \approx \left(\eta_1 + \eta_2 \Phi(X)\right) P^3 \tag{3.4}$$

with real numbers η_1 and η_2, the launch power P and the excess kurtosis $\Phi(X) \triangleq \frac{\mathbb{E}\left[|X|^4\right]}{\mathbb{E}\left[|X|^2\right]^2} - 2$ of the complex constellation.

If we consider the SNR at optimum launch power;

$$\text{SNR} = \frac{P}{P_{\text{ASE}} + \eta_{\text{tot}} P^3} \tag{3.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}P}\text{SNR} = \frac{P_{\text{ASE}} - 2\eta_{\text{tot}} P^3}{\left(P_{\text{ASE}} + \eta_{\text{tot}} P^3\right)^2} = 0 \tag{3.6}$$

$$P_{\text{ASE}} = 2\eta_{\text{tot}} P_{\text{opt}}^3 \tag{3.7}$$

$$P_{\text{opt}} = \left(\frac{P_{\text{ASE}}}{2\eta_{\text{tot}}}\right)^{\frac{1}{3}} = \left(\frac{P_{\text{ASE}}}{2\eta_1(1 + \frac{\eta_2}{\eta_1}\Phi(X))}\right)^{\frac{1}{3}} \tag{3.8}$$

$$\text{SNR}_{\text{opt}} = \frac{\left(\frac{\frac{1}{2\eta_1} P_{\text{ASE}}}{(1+\frac{\eta_2}{\eta_1}\Phi(X))}\right)^{\frac{1}{3}}}{P_{\text{ASE}} + \frac{P_{\text{ASE}}}{2}} \tag{3.9}$$

we can obtain the following relationship between the SNR at optimum launch power between an input distribution A and an input distribution B:

$$\frac{\text{SNR}_{\text{opt,A}}}{\text{SNR}_{\text{opt,B}}} = \left(\frac{1 + c\Phi(X_{\text{B}})}{1 + c\Phi(X_{\text{A}})}\right)^{\frac{1}{3}} \tag{3.10}$$

where $c \triangleq \frac{\eta_2}{\eta_1}$ is a measure for the relative impact of the modulation format on the nonlinear interference. Note that the P_{ASE} is assumed independent of modulation format and both distributions have the same parameters η_1 and η_2. This investigation uses a Gaussian constellation as a reference distribution as its excess kurtosis is given by $\Phi(X_{\mathrm{Gaussian}}) = 0$.

Using fast Gauss-Hermite quadratures to calculate the MI, similar to [84] described in appendix B, numerical optimisation was used to generate probabilities for constellation points based on their power, i.e., $|x_i|^2$, finding a probability for every ring in the constellation. The trade off between the SNR gain from excess kurtosis and the MI loss from suboptimal shaping for AWGN will yield a net MI gain. Heuristically, we found that the independent ring optimizations did not find any improvements over

$$p_X(x_i) = \frac{\exp(-\nu_1|x_i|^2 - \nu_2|x_i|^4)}{\sum_{j=1}^{M} \exp(-\nu_1|x_j|^2 - \nu_2|x_j|^4)} \tag{3.11}$$

where the optimization parameters ν_1 and ν_2 significantly reduce the computational complexity compared to the unconstrained problem.

We first evaluated the MI for different square QAM modulation formats and optimized the shaped formats for $c = 0.69$, a value obtained using equation (3.10) from simulation (see Section 3.2.1). We assume an AWGN channel with the noise variance corrected by the constellation format excess kurtosis. The results of the optimization process are plotted as MI versus the optimum SNR for a Gaussian-modulated signal as shown in Fig 3.9. The plotted MI values are calculated for an AWGN with the SNR deviation according to equation (3.10). The results for DP 64QAM, 256QAM and 1024QAM are shown for uniform distribution, the MB distribution and the proposed optimized distribution. For both the MB and the proposed optimized distributions, the ring powers are numerically optimized for each value of Gaussian-modulated SNR. The MI difference is lower than zero for the cases where the gain from the nonlinear channel response, compared to Gaussian signal transmission, is more than the gap to the AWGN capacity for the auxiliary channel.

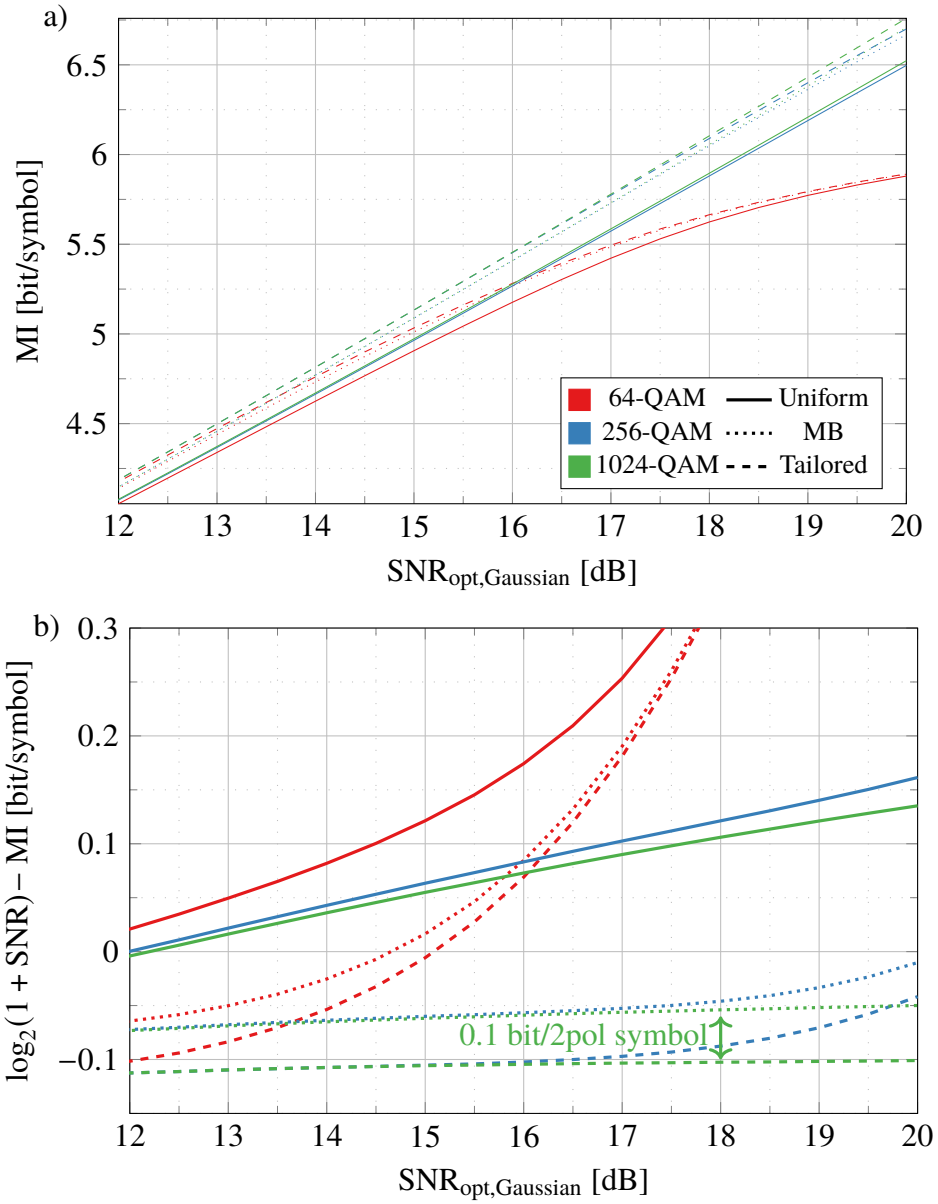


Figure 3.9: a) MI for optimised PS constellation. b) MI difference with respect to Gaussian transmitted signal.

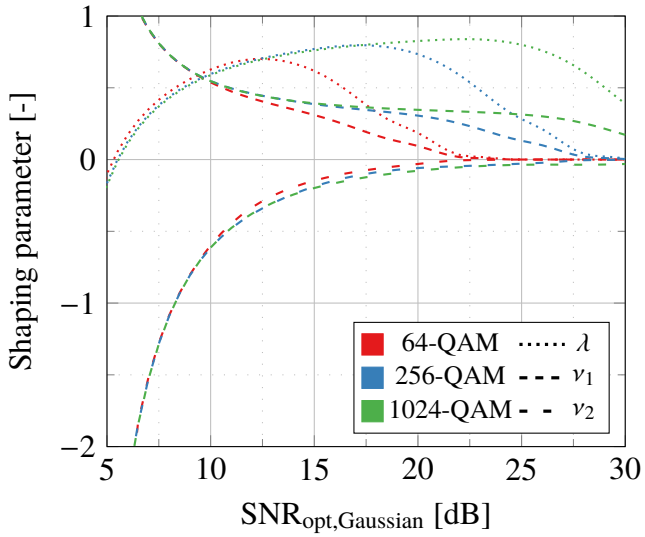


Figure 3.10: The shaping parameters obtained by numerical optimisation. The probability mass function of the optimized and MB distribution tailored to $\text{SNR}_{\text{opt,Gaussian}}$=18 dB.

In Fig. 3.10, the shaping parameters used to obtain the results in Fig. 3.9 are shown. In this figure, the parameters λ from equation (3.3) and ν_1, ν_2 from equation (3.11) are shown. These parameters are obtained using numerical optimisation. The values for 64, 256 and 1024 QAM are similar for the low SNR regime and they converge all to zero for the high SNR regime, where they are limited by their cardinality and equiprobable has the highest entropy. The lower cardinality constellations are affected at a lower SNR by this limitation.

In Fig. 3.11, the distribution comparing the MB and nonlinear fibre channel optimised shaped constellations are shown. For both 64 and 256 QAM we can see that high probabilities around the origin, $|x_i| = 0$ are reduced and spread out, as well as the tail of the distribution being moved towards the origin. This results in a flatter distribution and contributes to a lower excess kurtosis $\Phi(X)$, which will yield a more favourable response of the channel. However the overall shape is still applied, also benefiting from the effects of shaping the constellation.

3.2.1 Numerical simulation demonstration

We simulated a single span transmission link based on 200 km ultra-low-loss single-mode fibre with an attenuation of 0.165 dB/km, a dispersion coefficient

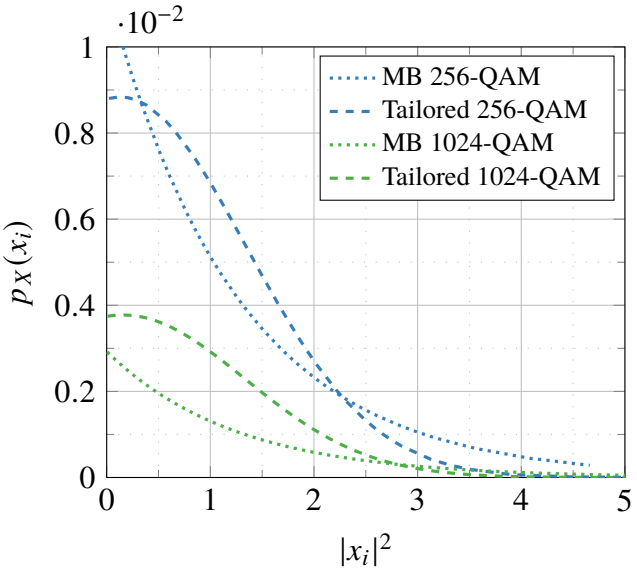


Figure 3.11: comparing the distributions for MB and nonlinear fibre channel optimised shaped constellations.

of 16.3 ps/nm/km and a nonlinear parameter of 1.2 /W/km, followed by an Erbium-doped fibre amplifier with a noise figure of 5 dB. The transmitter generates a dual-polarization 33 GHz-spaced 5×33 GBd WDM signal, yielding 5×400 Gbit/s. The optimum transmission performance is achieved by sweeping the optical launch power per channel. For Gaussian modulation the simulated transmission system achieves an SNR of 18 dB at optimal launch power.

The simulation results are shown in Fig. 3.12. The shift in optimal launched power can be observed. Both shaping strategies have shaping gain over uniform QAM, however the trade-off between shaping gain and nonlinear channel response has the additional MI gain as observed in Fig. 3.9. The observed SNR is shown in Fig. 3.13. The uniform QAM shows the best SNR performance and Gaussian signal the worst. Both shaping strategies show a response between the two cases, showing there is trade-off with potential observable benefits if an experiment is conducted.

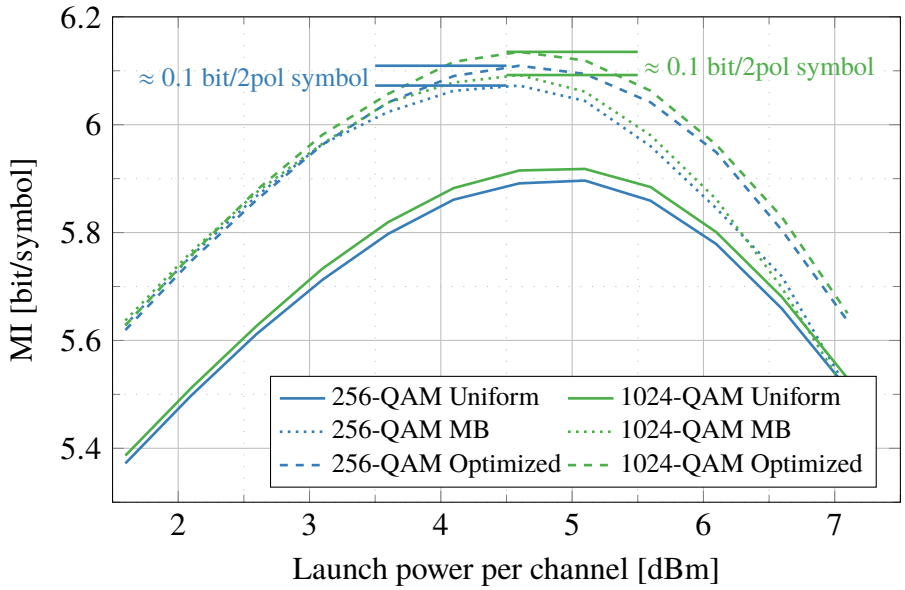


Figure 3.12: Simulation results based 200 km single span transmission over ultra low loss fibre.

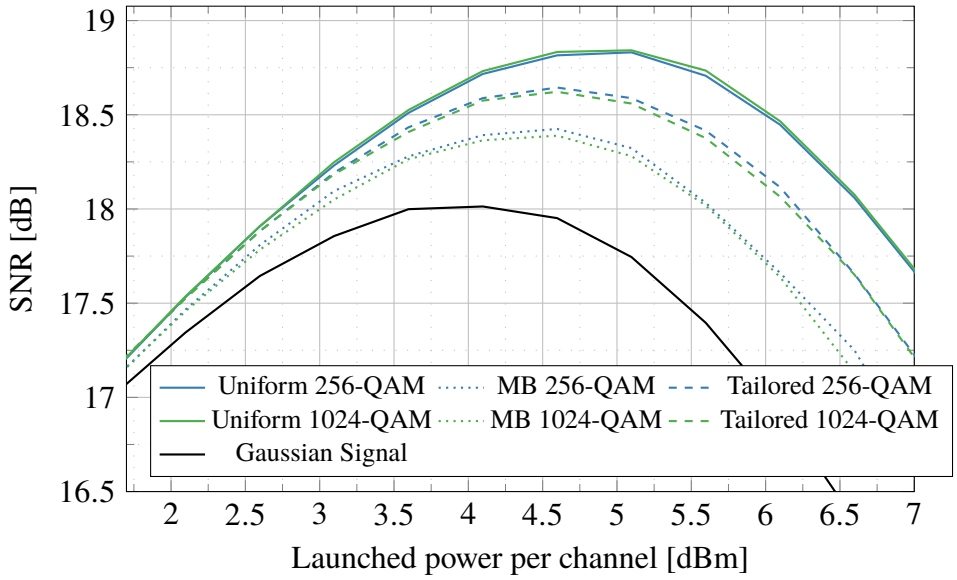


Figure 3.13: The achieved SNR from the simulation results.

3.3 Experimental demonstration of 1D geometric shaping

In the previous section, we have seen that shaping for nonlinear fibre channel has a benefit that is measurable in simulation. In his section we demonstrate the effects of constellation shaping experimentally.

In [85], we used 1D GS for the experimental demonstration, because this shaping method preserves the corners of the 2D constellation. These features allow the conventional blind DSP to recover the signal after optical fibre transmission. When using PS, the insertion of pilots is needed [86].

Using a quasi-Newton algorithm, the constellations are optimised for GMI, this being the measure that predicts an ideal BICM [53] and optimising the constellation for GMI will therefore increase the throughput of our system. At every step, the SNR was changed according to Eq. (3.10) for the excess kurtosis of the constellation at that optimisation step.

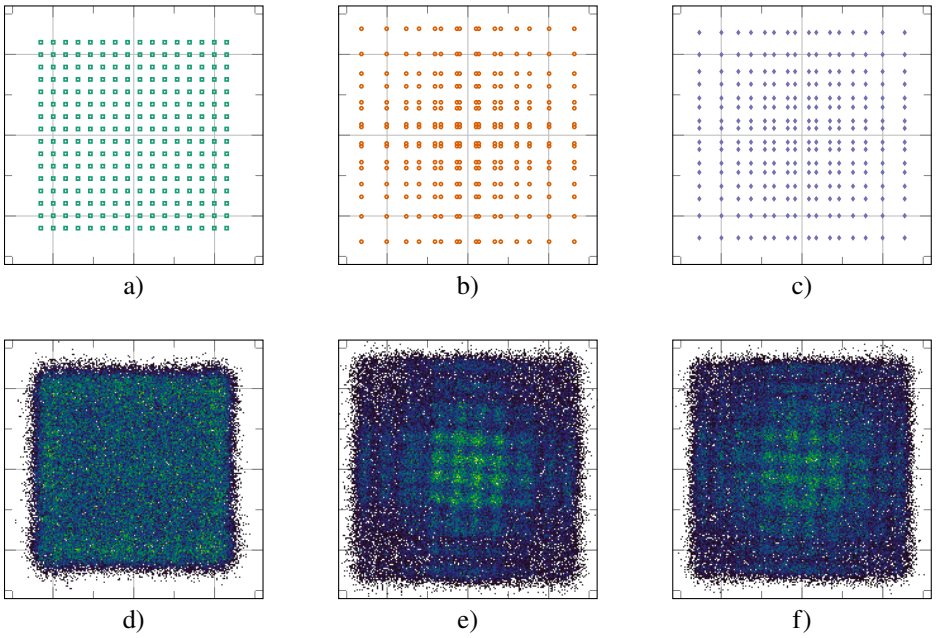


Figure 3.14: Geometrically-shaped constellation diagrams for 256-QAM, normalised to unit power: a) uniform, b) tailored to the AWGN channel, and c) tailored to the nonlinear fibre channel. Their respective received constellations after transmission over 160 km are shown in d), e) and f).

The constellations in Fig. 3.14 were designed with Fig. 3.14.a) as a reference

format. Uniform square QAM was used to obtain the system parameters of the experimental single-span 160 km long link for the central of three 35.2 GHz-spaced 35 GBd channels. The noise figure of the amplifiers was extracted from a low launch power measurement. The nonlinear coefficient of the fibre was 1.2 $W^{-1}km^{-1}$. The optimum SNR for Gaussian modulation was approximately 18 dB and for the nonlinearity-tailored constellation the ratio $c = \frac{\eta_1}{\eta_2} = 0.55$ was used.

In Fig. 3.15, the constellations are optimised for different SNRs at optimum launched power. Similar to section 3.1.1, we can see that the biggest gain is achieved by shaping, however a greater gain can be achieved when the nonlinear fibre response is taken into account. At 18 dB SNR, a 0.2 bit/symbol throughput increase can be expected for the AWGN-tailored constellation, and an additional 0.02 bit/symbol increase for the nonlinearity-tailored constellation.

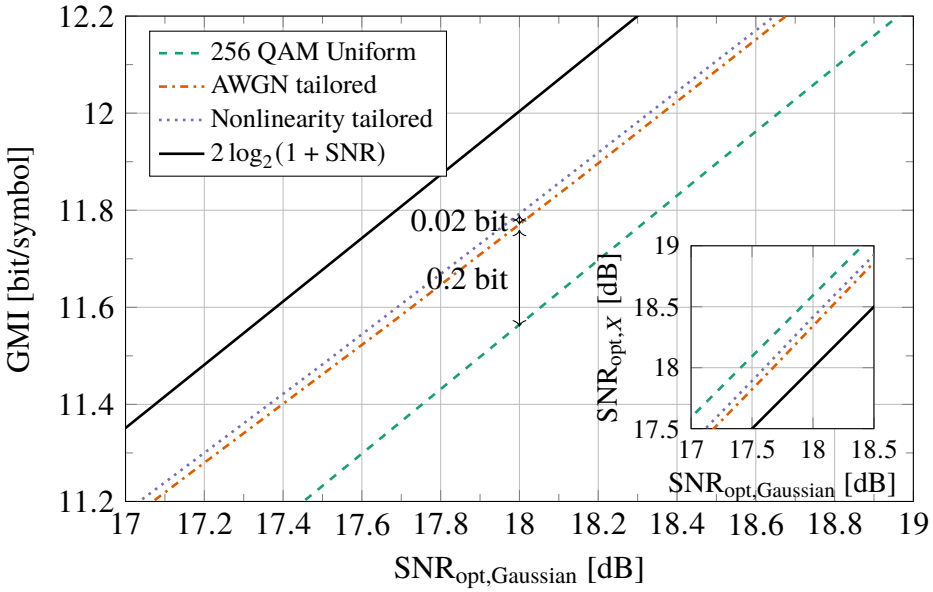


Figure 3.15: GMI performance of the GS constellation. In the inset, the relative SNR for the different constellations is shown. For the Gaussian distribution, the theoretical performance is shown.

Experimental setup

The experimental setup is shown in Fig. 3.16. The 3×35 GBd superchannel was transmitted over a single span of 160 km of standard single mode fibre (SSMF) to focus on the impact of NLI. The launch power into the fibre span was swept to

investigate the nonlinear tolerance of the constellations designed.

The transmitter consisted of the channel under test in which a single 100 kHz-linewidth external cavity laser (ECL) was modulated using a DP IQ modulator. Two additional ECLs were modulated by a separate modulator to form the aggressor channels. Eight 8-bit digital-to-analogue converter (DAC) channels operating at 87.5 GS/s, with effective number of bits (ENOB) of ∼4 bits at 17.5 GHz were used to generate independent channels. Both signal paths are amplified using an EDFA and combined by a 50/50 coupler.

Before the span, an EDFA and variable optical attenuator (VOA) followed by a 5% power tap with an optical power meter (OPM) were used to control the launch power into the span. The 160 km span of SSMF was followed by an EDFA, the output of which was passed to the receiver.

The receiver had a band-pass filter and EDFA followed by a DP coherent receiver with a separate ECL as local oscillator. After 160 GS/s analogue-to-digital converters (ADCs), offline DSP was used. After electronic dispersion compensation, a radially directed equaliser and a decision directed carrier phase estimator were used to recover the symbols. The SNR and GMI were extracted from the received symbols.

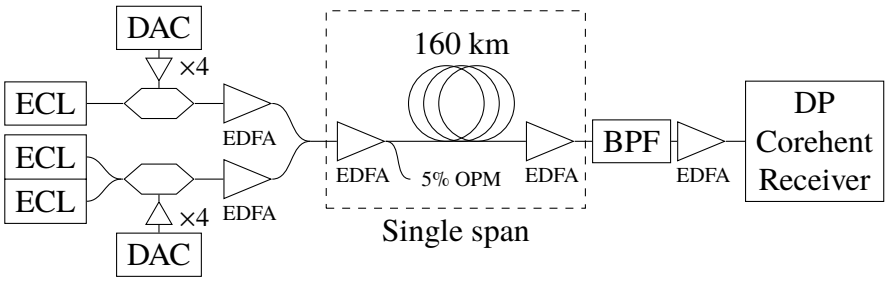


Figure 3.16: Experimental setup used to show shaping for the nonlinear fibre channel. A three channel superchannel was created, by surrounding a channel under test by two channels generated by a second DAC and modulator.

Experimental Results

The received constellations at optimum launch powers are shown in Fig. 3.14 d), e) and f). In contrast to the uniform QAM scatter plot, the constellations tailored for the AWGN channel, shown in Fig. 3.14 e), and for the nonlinear fibre channel,

shown in Fig. 3.14 f), set the lower energy points closer together, making these points more pronounced. The denser points should not be mistaken for probabilistically shaped constellations; all constellation points are equiprobable, the overlapping noise distributions around these points resulting in increased density of samples within the central area. Furthermore, it can be clearly seen that, as a result of fibre nonlinearity, the relative phase rotation between the central and outer points is higher in the AWGN tailored constellation than in the nonlinearity tailored constellation.

The experimentally-measured GMIs and SNR are shown in Fig. 3.17 and Fig. 3.18, respectively. The markers are experimentally obtained and the lines are from the model. The back-to-back (BtB) SNR and η_{tot} derived from the model are shown in Tab. 3.2.

Table 3.2: Back-to-back SNR and η_{tot} for the constellations.

	BtB SNR [dB]	η_{tot} [dB]
Uniform	22.78	27.61
AWGN tailored	21.63	28.23
Nonlinearity tailored	22.01	28.08

The uniform QAM exhibited the highest BtB SNR and the lowest η_{tot}, resulting in the highest SNR at optimal launch power. However, this modulation format has no shaping gain and consequently has the lowest GMI of the three constellation formats evaluated, with an optimum value of 11.6 bit/symbol.

The AWGN tailored format has the lowest SNR across all launch powers, but outperforms uniform QAM (i.e. higher GMI) at launch powers below the optimal. At higher launch powers the model does not predict any gains, while the experimentally observed performance was marginally higher.

Due to lower excess kurtosis, the nonlinearity tailored constellation has a η_{tot} lower than the AWGN tailored constellation as predicted by Eq. (3.4). Additionally, the lower excess kurtosis results in a higher BtB SNR because of the lighter tailed distributions of the constellation and resulting reduced quantisation noise. The nonlinearity tailored constellation offers a trade-off between shaping gain and the impact of the nonlinear interference. It achieved a GMI of 11.7 bit/symbol after transmission over the 160 km link, a > 0.1 bit/symbol increase over the other two

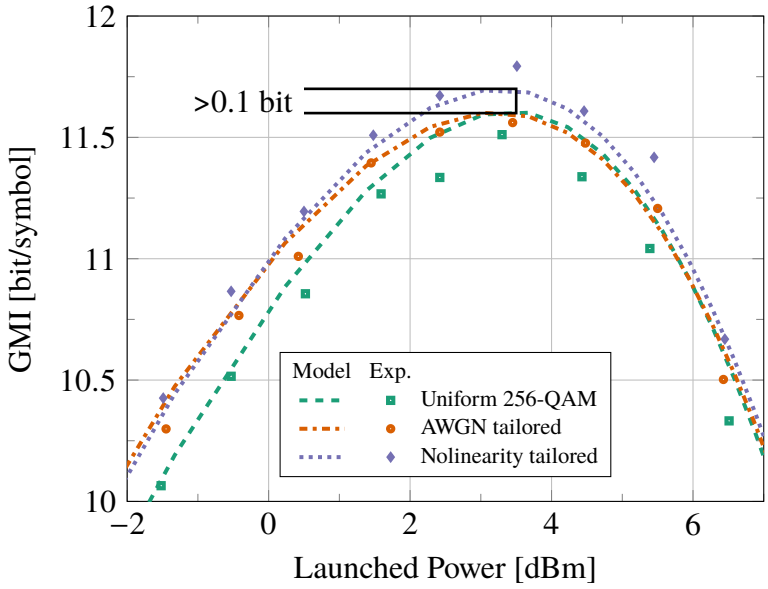


Figure 3.17: The GMI versus the launch power for all constellations. The model is shown with the lines and markers are experimental results.

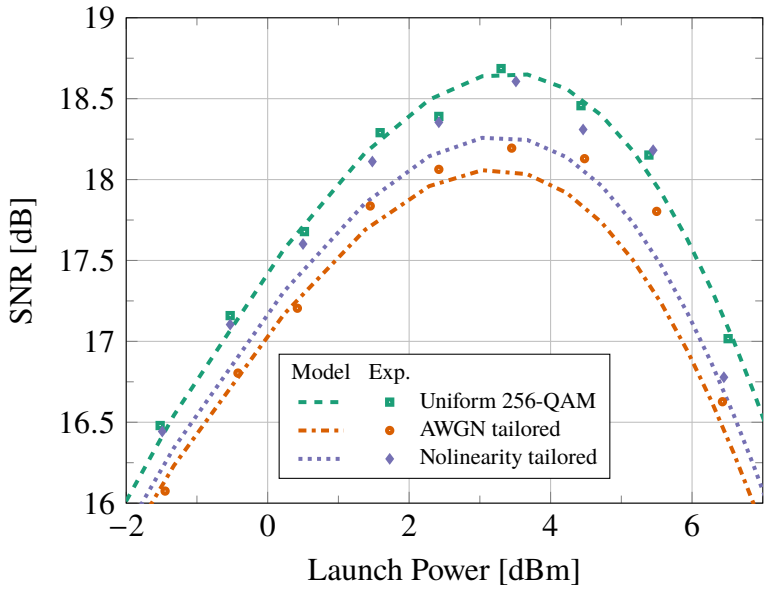


Figure 3.18: The SNR versus the launch power for all constellations. The model is shown with the lines and markers are experimental results.

constellations.

3.4 Summary

Combining our knowledge of information theory with the response of the nonlinear fibre channel, we investigated the impact of constellation formats on nonlinear

interference and proposed a method to design constellation diagrams for the non-linear fibre channel. We have proposed an heuristic expression that describes the best-performing probabilistically shaped constellation we have found for the non-linear fibre channel that reduces the optimisation space to two variables. We have experimentally demonstrated 1D geometric shaping and predicted the achievable information rate using the metrics introduced earlier. In the next chapter, we use the same metrics for machine learning.

Chapter 4

Machine learning for the optical transmission system

In this chapter we use the same performance metrics, SNR, MI and GMI, to train supervised learning algorithms. These metrics are related to the three main cost functions in machine learning, so optimising for one is equivalent to optimising for the other.

For explaining the concepts, we can introduce a universal function approximator $\mathcal{U}(y, w) = \hat{x}$, which can replicate any function mapping input y to output x if the weights w are tuned correctly [87]. The tuning can be done by minimising a cost or a loss function, which we can defined for approximated outputs $\hat{x}$ and desired outputs x.

The cost which is related to SNR is the mean squared error (MSE). When our universal function $\mathcal{U}(y) = \hat{x}$ has to reproduce the transmitted signal x from the received signal y, an intuitive choice for the error function would be the MSE [88].

$$J_{\text{MSE}} = |y - x|^2 \tag{4.1}$$

$$\text{SNR} = \frac{E\left[|x|^2\right]}{E\left[|y - x|^2\right]} \tag{4.2}$$

$$\text{SNR} = \frac{E\left[|x|^2\right]}{E\left[J_{\text{MSE}}\right]} \tag{4.3}$$

It can be seen that the SNR is inversely proportional to the expected value of the MSE.

Our aim is to maximise the SNR using the weights w in our function $\mathcal{U}(y, w) = \hat{x}$.

$$\arg\max_{w} \text{SNR} = \arg\min_{w} E\left[J_{\text{MSE}}\right] \tag{4.4}$$

This results in a simplified gradient for the gradient descent

$$\frac{\partial}{\partial w} J_{\text{MSE}} = \frac{\partial}{\partial w} \left|\mathcal{U}(y, w) - x\right|^2 \tag{4.5}$$

$$= 2\left(\mathcal{U}(y, w) - x\right) \mathcal{U}'(y, w) \tag{4.6}$$

With this, we can train the algorithms to approximate our desired waveform.

Sometimes, the exact waveform is not important. When we wish to communicate, we are sending and receiving messages. These messages are made up of bits, which we can model as Bernoulli distributed random variables with probability $\phi \in [0, 1]$. Usually, the probability parameter that describes the Bernoulli distribution for the message is $\phi = 0.50$, i.e., both 0 and 1 are equally likely.

$$p_B(b = 0) = \phi \tag{4.7}$$

$$p_B(b = 1) = 1 - \phi \tag{4.8}$$

Now after observing the output of our algorithm, $z = \mathcal{U}(y, w)$, we can update the probability we know of our received bits. We define an unnormalised probability $\tilde{p}_{X|Y}(x|y)$ analogous to [89, Ch. 6].

$$\log\left(\tilde{p}_{B|Y}(b|y)\right) = bz \tag{4.9}$$

$$\tilde{p}_{B|Y}(b|y) = \exp(bz) = \begin{cases} \exp(0) & \text{if } b = 0 \\ \exp(z) & \text{if } b = 1 \end{cases} \tag{4.10}$$

$$p_{B|Y}(b|y) = \frac{\tilde{p}_{B|Y}(b|y)}{\sum_{b \in \{0,1\}} \tilde{p}_{B|Y}(b|y)} \tag{4.11}$$

Because we are using a Bernoulli distribution, the common choice is to then use a

logit as the link function. This function converts the probability into an LLR

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \tag{4.12}$$

with its inverse also known as the sigmoid function $\sigma(\cdot)$;

$$\text{logit}\big(p_{B|Y}(b|y)\big) = \sigma^{-1}(z) = \log\left(\frac{p_{B|Y}(b|y)}{1-p_{B|Y}(b|y)}\right) \tag{4.13}$$

$$\sigma(z) = \frac{\exp(z)}{\exp(z)+\exp(0)} = \frac{1}{1+\exp(-z)} \tag{4.14}$$

We can see that equation (4.13) calculates the same metric as equation (2.14) and we could use the output of the network directly in our FEC. These functions are typically used to separate the linear layers, which then allows neural networks to be universal function approximators [87]. These functions to separate layers are also referred to as activation functions.

Now, if we combine equation (4.11) and (4.14) we get

$$p_{B|Y}(b|y) = \begin{cases} \sigma(-z) & \text{if } b = 0 \\ \sigma(z) = 1-\sigma(-z) & \text{if } b = 1 \end{cases} \tag{4.15}$$

$$= \sigma((2b-1)z) \tag{4.16}$$

Then we can define the binary cross-entropy cost function J_{BCE} as

$$J_{\text{BCE}} = -\log(p_{B|Y}(b|y)) \tag{4.17}$$

$$= -\log\left(\sigma((2b-1)z)\right) \tag{4.18}$$

$$= \log\left(\frac{1}{\sigma((2b-1)z)}\right) = \log(1+\exp(-(2b-1)z)) \tag{4.19}$$

with the gradient

$$\frac{\partial}{\partial w} J_{\mathrm{BCE}} = \frac{-1}{\sigma((2b-1)z)} \cdot \sigma((2b-1)z)\sigma(-(2b-1)z) \tag{4.20}$$

$$\frac{\partial}{\partial w} J_{\mathrm{BCE}} = \begin{cases} \dfrac{\sigma(-z)\sigma(z)}{\sigma(-z)} & \text{if } b = 0 \\ \dfrac{\sigma(-z)\sigma(z)}{\sigma(z)} & \text{if } b = 1 \end{cases} \tag{4.21}$$

Also note that equation (4.19) is similar to equation (2.15) to the point that minimising the binary cross entropy is equivalent to maximising the GMI if the outputs z are treated as log-likelihood ratios.

If we want to make symbol decision to optimise the MI, we can use a similar set of equations, but for a multinoulli distribution [90]. The softmax is the activation function and the categorical cross entropy is the loss function for the set of outputs $\boldsymbol{z} = [z_1, z_2, \ldots, z_M]$, where $z_i = \mathcal{U}_i(y, w_i)$.

$$\begin{aligned} p_{X|Y}(x_i|y) &= \mathrm{softmax}_i\,(\boldsymbol{z}) \\ &= \frac{\exp(z_i)}{\sum_{j=1}^{M} \exp(z_j)} \end{aligned} \tag{4.22}$$

$$J_{\mathrm{CCE}} = -\sum_{j=1}^{M} t_j \log\big(\mathrm{softmax}_j\,(\boldsymbol{z})\big) \tag{4.23}$$

$$t_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{4.24}$$

where $\boldsymbol{t} = [t_1, t_2, \ldots, t_M]$ is the one-hot vector of the transmitted symbols. The softmax function normalises the symbol log likelihoods such that their likelihoods sum to one.

This means that, provided our whole signal processing chain is differentiable, we can use a gradient descent method to adapt the digital signal processing to our simulation or experimental data. The classical machine learning approach is to add an artificial neural network (ANN) [91, 92, 93] or a deep nerual network (DNN) [94, 95]. In this work, we focus on a single processing block at the receiver, though end-to-end optimisation is also a promising technique [96]. We also focus

on differentiable DSP algorithms, though other techniques fall under the machine learning (ML) umbrella as well [25].

4.1 Learned digital back-propagation

One of the problems to which we can apply ML techniques is digital back-propagation (DBP) [97, 98, 99]. This is a technique derived from the SSFM to compensate fibre nonlinearity [100] applied in the time domain [101], known as time-domain digital back-propagation (TD-DBP). The motivation for compensating both dispersion and nonlinearity in the time domain is to avoid the need for repeated Fourier transforms, thus reducing complexity. In this section, we investigate the learning of finite impulse response (FIR) filter responses to compensate chromatic dispersion in the time-domain as the linear step in a neural network and applying the phase shifts to compensate the Kerr nonlinearity as activation functions between them to achieve DBP [98].

In Fig. 4.1, the building block of the TD-DBP is shown. It consists of a linear step, in which the chromatic dispersion for the step is applied, following which a power-dependent phase shift is applied to compensate for the Kerr effect. Both effects can be described by their respective transformations [35, Ch. 8];

$$E(\omega, z + \Delta z) = e^{\alpha \Delta z} e^{jK(\omega T)^2} E(\omega, z) \tag{4.25}$$

$$E(t, z + \Delta z) = e^{-j\gamma \Delta z |E_z(t)|^2} E(t, z) \tag{4.26}$$

where α is fibre loss, Δz is fibre step length, $K = \frac{\beta_2 \Delta z}{2T^2}$, ω angular frequency, T sampling period, and β_2 group velocity dispersion, γ the nonlinearity coefficient and $|E_z(t)|^2$ the normalised, step-averaged, instantaneous optical power. In the fibre, these effects manifest themselves simultaneously, however in the compensation they need to be applied separately. The solution is to split the link into many shorter steps, where the error from separating the two effects is small enough.

The convolutional step to apply the chromatic dispersion compensation is sep-

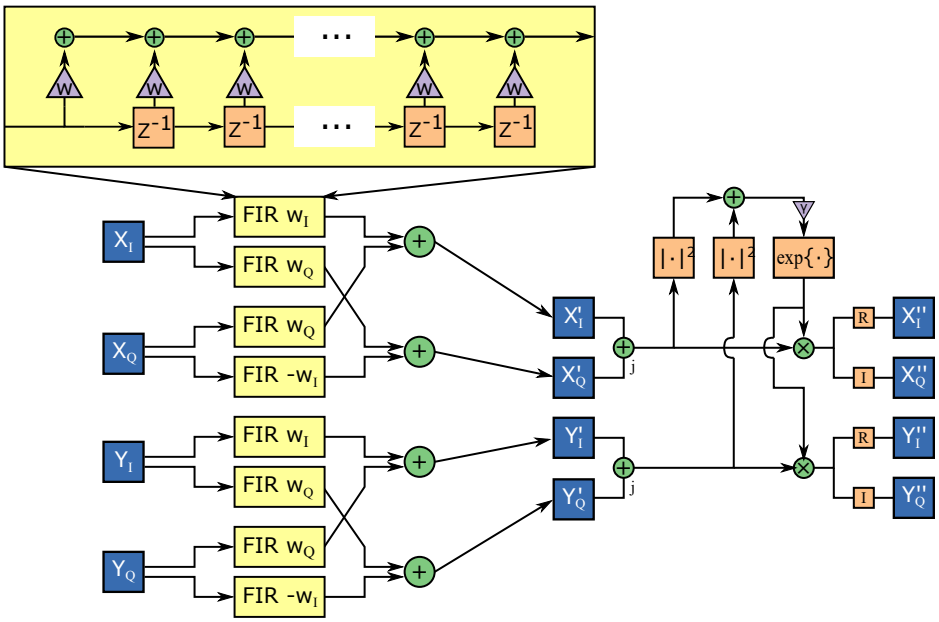


Figure 4.1: The building block of the TD-DBP algorithm. The chromatic dispersion is compensated with a complex FIR filter where the real and the imaginary components are applied separately. The second part is applying the phase shift to compensate the Kerr nonlinearity.

arated into its real and imaginary components

$$(X_\mathrm{I} + jX_\mathrm{Q}) * (W_\mathrm{I} + jW_\mathrm{Q}) = \begin{array}{c}(X_\mathrm{I} * W_\mathrm{I} - X_\mathrm{Q} * W_\mathrm{Q}) \\ +j(X_\mathrm{I} * W_\mathrm{I} + X_\mathrm{Q} * W_\mathrm{I})\end{array} \tag{4.27}$$

Because both polarisations incur the same chromatic dispersion, the same filter response can be used for both polarisations. To reduce training time and prevent the algorithms from getting stuck in local minima, the chromatic dispersion filters are initialised with analytical filters.

If we start with the chromatic dispersion as described in equation (4.25) the obvious solution is to sample N taps for the transformation in the time domain as electronic dispersion compensation (EDC) [37],

$$H_\mathrm{EDC}(e^{j\omega T}) = e^{jK(\omega T)^2} \tag{4.28}$$

$$h[n] = \sqrt{\frac{j}{4K\pi}} e^{-j\frac{n^2}{4K}}, \quad -\left\lfloor \frac{N}{2} \right\rfloor \leq n \leq \left\lfloor \frac{N}{2} \right\rfloor \tag{4.29}$$

$$N = 2\lfloor 2K\pi \rfloor + 1 \tag{4.30}$$

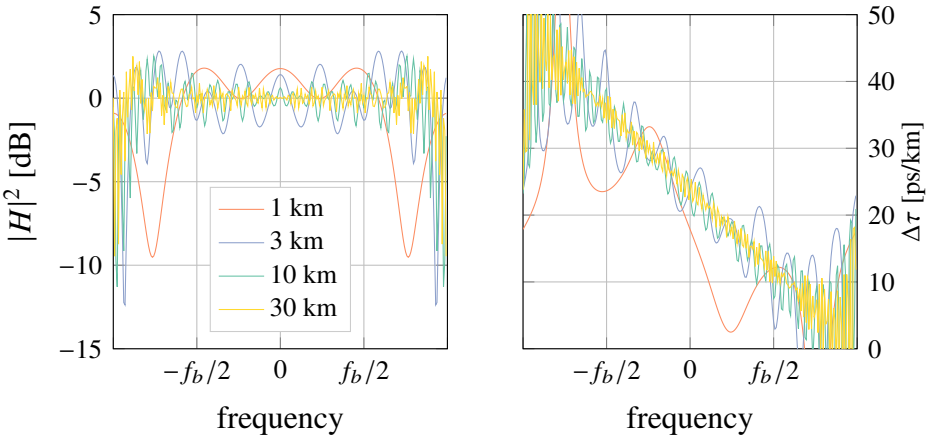


Figure 4.2: The amplitude response and group delay of directly sampled chromatic dispersion compensation filters, for symbol rate f_b 56 GBd, sample rate $f_s = 2f_b$, 17 ps/nm/km dispersion at 1550 nm.

However, because chromatic dispersion is inherently a frequency domain effect, the direct sampling approach is not very accurate, especially when dealing with very short filters as can be seen in Fig. 4.2. An alternative method is to integrate in the frequency domain at specific time delays [102]

$$\begin{aligned} D[n] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_{\text{EDC}}(e^{j\omega T})e^{jn\omega T}d(\omega T) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega T(K\omega T+n)}d(\omega T) \end{aligned} \tag{4.31}$$

$$D[n] = \frac{e^{-j\left(\frac{n^2}{4K}+\frac{3\pi}{4}\right)}}{4\sqrt{\pi K}}\left(\text{erf}\left(\frac{e^{j\frac{3\pi}{4}}(2K\pi - n)}{2\sqrt{K}}\right) + \text{erf}\left(\frac{e^{j\frac{3\pi}{4}}(2K\pi + n)}{2\sqrt{K}}\right)\right) \tag{4.32}$$

where the erf function is defined as

$$\text{erf}(\alpha) = \frac{2}{\sqrt{\pi}}\int_0^{\alpha} e^{-t^2}dt \tag{4.33}$$

This can be further improved upon by limiting the bandwidth over which the chromatic dispersion is compensated [103], where filters are designed that minimise the error within a bandwidth, whilst limiting the out-of-band gain to improve stability.

The step lengths were chosen such that every step had an equal power drop in the fibre; the first compensation steps were very long and the last steps, corresponding to the start of the span with the highest power, had very short steps. The shorter

the step, the fewer the number of taps are required, as given by equation (4.30). For every span, the same set of steps, with their filter weights, were used.

In this work, deep learning of the filter weights (equivalent to a neural network's layer parameters) is implemented in Tensorflow using the RAdam optimiser [104]. Identical complex FIR filter weights are applied to both polarisations in each step, reducing the overall number of weights to be optimized. Initialisation of the time-domain filter taps was carried out via numerical simulation of the fibre transmission link for a single channel. The forward propagation was modelled using a small NLSE fibre step size (100 m) at a launch power of 5 dBm. This is beyond optimum launch power for linear compensation and was chosen to demonstrate the nonlinear compensation abilities.

4.1.1 Experimental setup

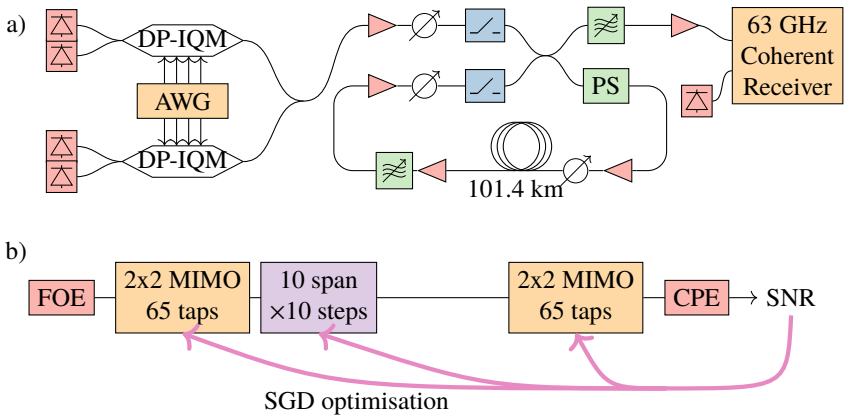


Figure 4.3: a) Experimental configuration with 4×64 GBd channels and 101.4 km recirculating loop. b)Function diagram of the receiver DSP for the L-TDDBP.

The experimental setup is shown in Fig. 4.3(a). A fibre transmission distance of 1014 km was emulated using a recirculating loop. The waveform of the 64-QAM 64-GBd channel under test (CUT) was generated offline and sent to two channels of a 33-GHz 92-GSa/s arbitrary waveform generator (AWG) and, using a dual-polarization IQM), modulated the outputs of two <100 kHz ECLs. Two additional 64-QAM 64-GBd aggressor channels were modulated using an additional AWG with a dual polarization IQM onto two ECLs and interleaved with the other channels to achieve uncorrelated sequences between neighbouring WDM channels. The

recirculating loop with a 101.4 km span, a polarisation scrambler (PS) and three EDFAs and an optical band-pass filter had the signal circulating 10 times, totalling a 1014 km transmission. At the receiver an optical band pass filter followed by an EDFA extracts the CUT for detection with a coherent receiver employing 63-GHz bandwidth 160-GS/s analogue-to-digital converters.

For this experimental demonstration, 10 steps per span were chosen for the learned TD-DBP. To take fibre loss into account, non-uniform FIR filter lengths were employed, implementing steps with equal power differences between their inputs and outputs. The resulting logarithmic step size distribution is better tailored to the exponentially decaying signal power and requires fewer steps per span [105]. Additionally, the nonuniform step size will give each filter a different starting position. For the fibre nonlinearity compensation of 10 spans of 101.4 km each, the parameters were α of 0.16 dB/km, β_2 of -20.18 ps^2/km and γ_{DBP} of 0.8 W^{-1}km^{-1}. The 10 FIR filters used in the TD-DBP employed a total of 270 complex-valued tap weights at a sampling rate of 128 GSa/s. For the FD-DBP, 50 equidistant steps/span were used. This requires $2 \times 10 \times 50$ FFT operations per polarisation, while in the TD-DBP scheme the use of FFT operations is circumvented, with the aim of lowering the computational complexity.

Next, for the processing of experimental data, the filter weights from simulation were used for initialisation. To prevent the dispersion filters from learning the response of the transceiver impairments, an additional 2x2 MIMO filter was added before applying digital back propagation, as shown in Fig. 4.3(b). Thus, the resulting structure has two linear MIMO equalisers, compensating for polarisation mode dispersion (PMD) and transmitter and receiver impairments. Note that in this way, using the automatic differentiation in Tensorflow, the filter that is applied prior to the link compensation is also optimized through gradient descent. A root-raised cosine (RRC) filter was applied before the MIMO blocks. The carrier phase estimation was achieved by inserting pilot symbols. One in 32 symbols was a known QPSK symbol. Interpolation of the phase between the pilot symbols was performed using a Wiener filter [106, Eq. (32)], following which a mean-squared-error cost is

calculated.

During the training procedure, first the linear filters at both sides of the link compensation were optimised. Subsequently, all filters were updated on each optimisation step. For the FD-DBP, the α, γ and launch power were swept for optimisation, after which pilot-aided DSP was applied. For the experimental waveform, a single randomly generated 2^{16}-symbol waveform was used. We split the bit sequence and corresponding received waveforms into two datasets. The first 52224 symbols (80%) were used as training data for updating the filter weights. The remaining 13312 symbols (20%) were used as testing data, to obtain results reported in the figures presented.

4.1.2 Results

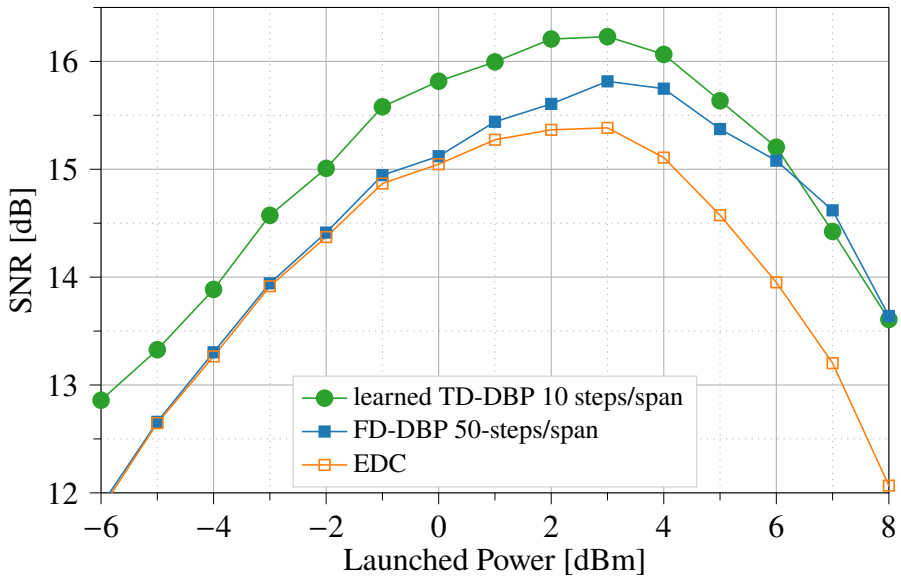


Figure 4.4: SNR vs. launched power for learned TD-DBP compared to conventional FD-DBP and linear EDC only methods.

The launched power was increased with 1 dB increments from -6 to +8 dBm per channel. The resulting SNR is here defined as $\frac{\mathrm{E}[|X|^2]}{\mathrm{E}[|X-Y|^2]}$, where X and Y are the transmitted and received signal respectively. Fig. 4.4 shows a comparison of achieved SNR for TD-DBP, FD-DBP and EDC. The TD-DBP and FD-DBP are implemented using 10 and 50 steps per span respectively. Both schemes provide similar performance improvements from non-linearity compensation, with slightly

higher accuracy in the high power regime for the conventional FD-DBP scheme, due to the larger number of steps used. However, the TD-DBP achieves a higher SNR in the low power regime, suggesting a better linear compensation.

Fig. 4.5 compares TD-DBP with two learned linear compensation strategies. The figure shows the learned DBP performance for two cases, the proposed non-linear mitigation scheme, and the same scheme with $\gamma_{\text{DBP}} = 0$, i.e., providing only linear compensation and a scheme where the whole chromatic dispersion is compensated in a single filter. For a low launch power into the fibre, the first two schemes show comparable performance, while a nonlinearity mitigation gain of 0.3 dB was achieved at optimal launch powers. Using a single filter achieves better linear gain, but converges to the $\gamma_{\text{DBP}} = 0$ method in the high launched power regime.

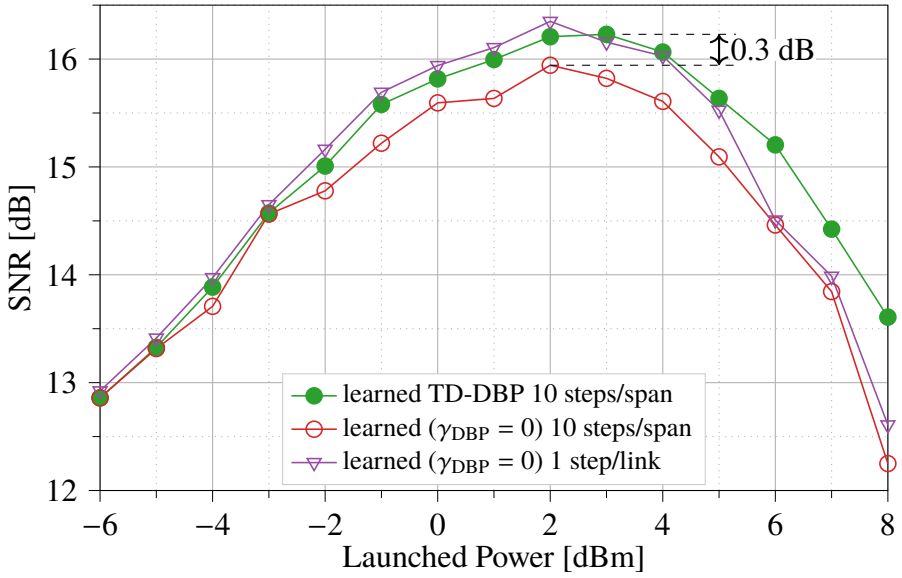


Figure 4.5: TD-DBP compared to the same structure with $\gamma_{\text{DBP}} = 0$ with the same number of steps and a single step for the whole link.

To confirm that the algorithm is performing digital back-propagation, i.e., approximating the SSFM model, the amplitude response and group delay of the 10 individual filter used each span are plotted in Fig. 4.6. The expected response is an all-pass filter (H) with a linear group delay ($\Delta\tau$), compensating for chromatic dispersion. It can be seen that, while the individual filters have significant ripples, the combined filter, in Fig. 4.7, has an almost perfect response within the signal

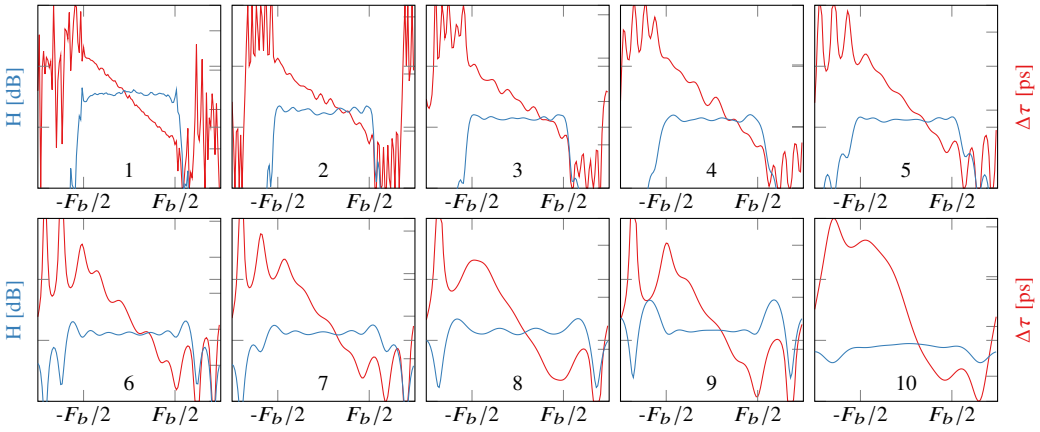


Figure 4.6: Amplitude response and group delay of the 10 individual filters used every span. The first filters cover more distance and have more tap weights than the latter filters.

bandwidth, with a flat amplitude response and a smooth group delay.

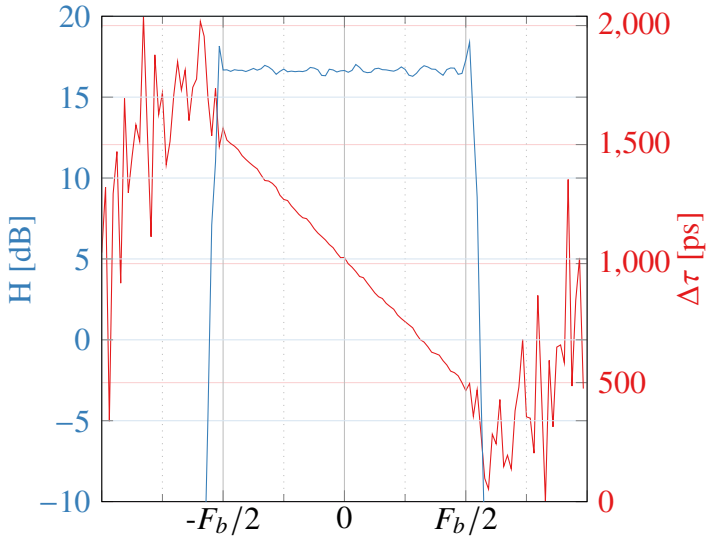


Figure 4.7: Combined response of all 10 cascaded filters, showing a flat amplitude response and smooth linear group delay. Showing a bandpass filter with accurate chromatic dispersion compensation.

We have also trained a single convolutional layer for the whole link to apply the chromatic dispersion compensation and shown the results in Fig. 4.5. This method outperforms all other methods over launch powers up to the optimal power. However for higher launch powers, its performance converges to the linear compensation results. When looking at the auto-correlation of the single learned filter in Fig. 4.8, we can see spikes where the filter has compensated a reflection. This suggests that

not all the linear effects are compensated for in our demonstration of the TD-DBP approach, and combining this result with nonlinear compensation will increase the performance even further.

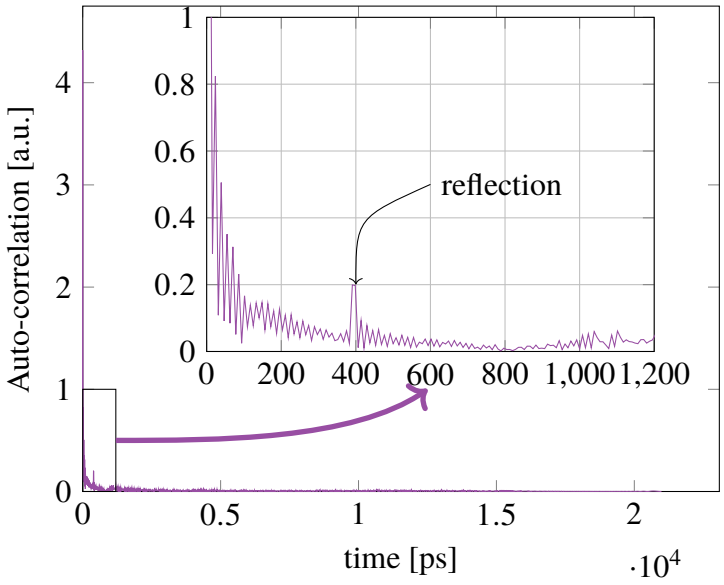


Figure 4.8: The auto-correlation of the learned single filter used to compensate chromatic dispersion.

We suspect the difference in performance between the single layer and the deep network to partially be attributable to gradient propagation through the many layers, in this work over 100. A method to combat this is proposed in [107], where residual links are bypassing the layer. The results are equivalent due to the universal function approximators the work is based on. Trying to learn $x_{k+1} = \mathcal{U}\{x_k\}$ will be equivalent to learning $x_{k+1} = x_k + \mathcal{V}\{x_k\}$ if $\mathcal{V}\{x\} \triangleq \mathcal{U}\{x\} - x$. These layers have the gradient

$$\frac{\partial x_{k+1}}{\partial x_k} = \mathcal{U}'\{x_k\} \tag{4.34}$$

$$\frac{\partial x_{k+1}}{\partial x_k} = 1 + \mathcal{V}'\{x_k\} \tag{4.35}$$

but if two layers are applied sequentially, the gradient becomes

$$x_{k+2} = \mathcal{U}\{\mathcal{U}\{x_k\}\} \tag{4.36}$$

$$\frac{\partial x_{k+2}}{\partial x_k} = \mathcal{U}'\{x_{k+1}\}\mathcal{U}'\{x_k\} \tag{4.37}$$

$$\frac{\partial x_{k+2}}{\partial x_k} = 1 + \mathcal{V}'\{x_{k+1}\}\,(1 + \mathcal{V}'\{x_k\}) \tag{4.38}$$

We can see the direct function will have a multiplicative term for every layer, but the residual link will propagate the constant through all the layers and therefore have a more direct link with the error function.

For the DBP, we can treat the effects of chromatic dispersion and nonlinear phase shift as perturbations to our signal. This is a different approach than the perturbation DBP [108, 109], although it shares some concepts.

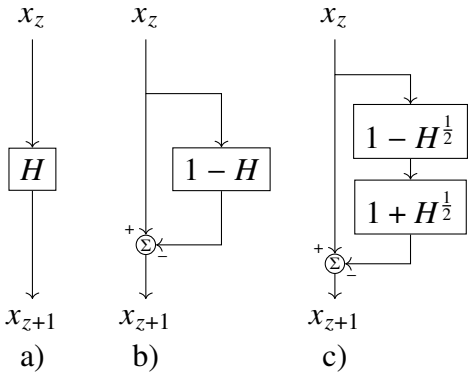


Figure 4.9: The proposed digital back-propagation block. a) a single dispersion block b) add the residual link c) split the block into to half steps.

In Fig. 4.9, the proposed block with the residual link is shown (d). The block is designed initially with a dispersion block in the frequency domain $x_{k+1} = H(x) = x\mathrm{e}^{-jK(\omega T)^2}$, which can either be applied directly (a) or with a residual link (b). Next,

in (c) we have split the dispersion step such that we have

$$\begin{aligned} x_{k+1} &= x_k - x_k \left(1 - \mathrm{e}^{\frac{-jK(\omega T)^2}{2}}\right)\left(1 + \mathrm{e}^{\frac{-jK(\omega T)^2}{2}}\right) \\ &= x_k - x_k\left(1 - \left(\mathrm{e}^{\frac{-jK(\omega T)^2}{2}}\right)^2\right) \\ &= x_k \mathrm{e}^{-jK(\omega T)^2} \end{aligned} \tag{4.39}$$

Now we have split the step, we can apply the nonlinear phase shift in the middle of the block, effectively recreating a split step method with a residual link. This block is shown in Fig. 4.10. The nonlinear phase shift as function $g(\cdot)$ is applied in the middle of the step and two half inverse phase shifts $g^{-\frac{1}{2}}(\cdot)$ are applied to cancel out the signal in the residual link as

$$\begin{aligned} x_{z+h} &= x_z - g^{-\frac{1}{2}}\left(g\left(g^{-\frac{1}{2}}(x_z) + H^{\frac{1}{2}}x_z\right)\right) \\ &\quad - H^{\frac{1}{2}} g\left(g^{-\frac{1}{2}}(x_z) + H^{\frac{1}{2}}x_z\right) \end{aligned} \tag{4.40}$$

$$\begin{aligned} &= x_z - g^{-\frac{1}{2}}\left(g\left(g^{-\frac{1}{2}}(x_z)\right)\right) \\ &\quad - g^{-\frac{1}{2}}\left(g\left(H^{\frac{1}{2}}x_z\right)\right) + H^{\frac{1}{2}} g\left(g^{-\frac{1}{2}}(x_z)\right) \\ &\quad + H^{\frac{1}{2}} g\left(H^{\frac{1}{2}}x_z\right) \end{aligned} \tag{4.41}$$

$$\approx x_z - g^{-\frac{1}{2}}\left(g\left(g^{-\frac{1}{2}}(x_z)\right)\right) + H^{\frac{1}{2}} g\left(H^{\frac{1}{2}}x_z\right) \tag{4.42}$$

$$= H^{\frac{1}{2}} g\left(H^{\frac{1}{2}}x_z\right) \tag{4.43}$$

This method has shown the potential for lowering DBP complexity and whilst there still is a challenge to scale the problem, we have conjectured a method to alleviate this limitation.

4.2 Summary

We have linked the metrics introduced earlier to the cost functions associated with machine learning. Using these, we experimentally demonstrated, for the first time, learned time-domain digital back-propagation and showed its capability to perform nonlinearity compensation.

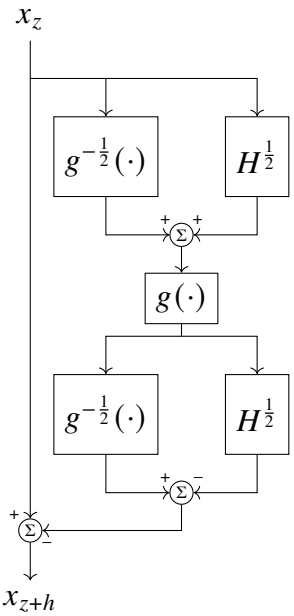


Figure 4.10: The proposed digital back-propagation block.

Chapter 5

Conclusions

This thesis has described the research on improving the achievable information rates of the optical transmission system using information theory and machine learning. Information theory allows us to put precise metrics on communication systems. The concepts of uncertainty and reliability in optical communication systems were described, which allow us to predict the performance of different coded modulation schemes and evaluate where the system can be improved to increase datarates.

Although the two coded modulation strategies considered in this study were predicted to have similar performance, one of them outperformed the other more than expected. This indicated that it had a lower implementation penalty, and this was precisely measured with the metrics proposed in chapter 2.

Efficient methods to calculate the achievable information rate for optical transmission systems have been used to optimise the transmitted constellations. In chapter 3, the constellations tailored for the nonlinear fibre channel outperformed the optimum constellations for the additive white Gaussian noise (AWGN) channel, demonstrating that we can increase achievable data throughput by tailoring our algorithms for the nonlinear fibre channel.

A simple heuristic expression for probabilistic shaping was introduced. The constellations adhering to this expression were straightforward to obtain and were the amongst the best-performing of all those found in the numerical optimisation. In the case of transmission over nonlinear fibre, these constellations were found to outperform the constellations optimised for the additive white Gaussian noise

(AWGN) channel.

For the experimental demonstration of the shaped constellations, geometrically shaped constellations were designed. 1D geometrical shaping was chosen for the demonstration because it preserves properties of uniform square quadrature amplitude modulation (QAM) that aid the conventional digital signal processing (DSP) to recover the signal. The shaped constellations performed as predicted by the metrics introduced in chapter 2.

The same information theoretical metrics were aligned with cost functions for machine learning, which we next investigated for the compensation of fibre nonlinearity and dispersion. These metrics give a clear insight when choosing a cost function for the machine learning problem at hand.

We experimentally demonstrated learned digital back-propagation for the first time. After the similarities between time-domain digital back-propagation and a convolutional neural network were noticed, machine learning was used to optimise the filters used in the back-propagation algorithms. An advantage of implementing the back-propagation in the time domain is that it avoids the need for multiple Fourier transforms, which are required in the case of the conventional split-step Fourier method. We have implemented the proposed algorithms to work in an experimental testbed and demonstrated that it was effective in performing nonlinearity compensation. The compensation technique not only compensated for fibre nonlinearity, it also compensated transceiver impairments, resulting in a bigger overall gain.

The results indicated that the deep learning structure can be improved upon when compared to shallow learning. We have conjectured a method that addresses this problem and should give us a further improvement on the results achieved.

Future work could also be carried out assessing the complexity of the learned time domain digital back-propagation, and comparing it with that of the conventional frequency-domain digital back-propagation (DBP).

The extension of the digital back-propagation using machine learning to wider bandwidth signals, including inter-channel nonlinearity compensation, will be a

future direction of the research.

Appendix A

MI for continuous uniform distribution

The channel law of AWGN channel is given by

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(-\frac{|y-x|^2}{2\sigma_z^2}\right) \tag{A.1}$$

The input distribution is assumed to be

$$p_X(x) = \frac{1}{b-a}, \quad x \in [a,b] \tag{A.2}$$

Thus, the output of distribution $p_Y(y)$ of channel in (A.1) is defined as

$$\begin{aligned} p_Y(y) &\triangleq \int_{\text{supp}\,[p_X(x)]} p_{X,Y}(x,y)\,dx && \text{(A.3)} \\ &= \int_a^b p_{Y|X}(y|x)\,p_X(x)\,dx && \text{(A.4)} \\ &= \frac{1}{(b-a)\sqrt{2\pi\sigma_z^2}} \int_a^b \exp\left(-\frac{|y-x|^2}{2\sigma_z^2}\right) dx && \text{(A.5)} \end{aligned}$$

Accounting that

$$\int_a^b \exp\left(-\frac{|y-x|^2}{2\sigma_z^2}\right) dx = \frac{\sqrt{2\pi\sigma_z^2}}{2}\left[\text{erf}\left(\frac{b-y}{\sqrt{2}\,\sigma_z}\right) - \text{erf}\left(\frac{a-y}{\sqrt{2}\,\sigma_z}\right)\right] \tag{A.6}$$

with

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \tag{A.7}$$

we then have

$$p_Y(y) = \frac{1}{2\,(b-a)} \left[\operatorname{erf}\left(\frac{b-y}{\sqrt{2}\,\sigma_z} \right) - \operatorname{erf}\left(\frac{a-y}{\sqrt{2}\,\sigma_z} \right) \right] \tag{A.8}$$

Applying $a = -\sigma_x\sqrt{3}$ and $b = \sigma_x\sqrt{3}$, it reads

$$p_Y(y) = \frac{1}{4\sqrt{3}\,\sigma_x} \left[\operatorname{erf}\left(\frac{\sigma_x\sqrt{3}+y}{\sqrt{2}\,\sigma_z} \right) + \operatorname{erf}\left(\frac{\sigma_x\sqrt{3}-y}{\sqrt{2}\,\sigma_z} \right) \right] \tag{A.9}$$

Now the mutual information can be found,

$$I(X;Y) = h(Y) - h(X|Y) \tag{A.10}$$

First,

$$h(X|Y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} p_{X,Y}(X,Y) \log\left(p_{Y|X}(y|x) \right) dy dx \tag{A.11}$$

$$= \int_{\mathcal{X}} p_X(x) \left(\underbrace{-\int_{\mathcal{Y}} p_{Y|X}(y|x) \log\left(p_{Y|X}(y|x) \right) dy}_{\frac{1}{2}\log\left(2\pi\,\mathrm{e}\,\sigma_z^2\right)} \right) dx \tag{A.12}$$

$$= \int_{\mathcal{X}} p_X(x) dx \left(\frac{1}{2} \log(2\pi\,\mathrm{e}\,\sigma_z^2) \right) = \frac{1}{2} \log(2\pi\,\mathrm{e}\,\sigma_z^2) \tag{A.13}$$

And

$$h(Y) = -\int_{\mathcal{Y}} p_Y(y) \log(p_Y(y)) dy \tag{A.14}$$

which we integrate using Monte-Carlo integration, i.e., via

$$\int_{\mathcal{Y}} p_Y(y) g(y) dy \approx \frac{1}{N_s} \sum_{n=1}^{N_s} g(y_{(n)}), \tag{A.15}$$

with samples of Y $y_{(n)}$ with $n = 1, \ldots, N_s$ and real values function

$$g(y) = -\log\left(\frac{1}{4\sqrt{3}\sigma_x}\left[\operatorname{erf}\left(\frac{\sigma_x\sqrt{3}+y}{\sqrt{2}\,\sigma_z}\right)+\operatorname{erf}\left(\frac{\sigma_x\sqrt{3}-y}{\sqrt{2}\,\sigma_z}\right)\right]\right). \tag{A.16}$$

Assuming $h(Y) \sim h(X) = \frac{1}{2}\log(12\sigma_x^2)$ as $\frac{\sigma_x^2}{\sigma_z^2} \to \infty$, therefore

$$I(X;Y) \sim \frac{1}{2}\log(12\sigma_x^2) - \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_z^2) = \frac{1}{2}\log\left(\frac{6}{\pi\,\mathrm{e}}\,\frac{\sigma_x^2}{\sigma_z^2}\right), \tag{A.17}$$

which differs from the Gaussian case $I(X;Y) \sim \frac{1}{2}\log\left(\frac{\sigma_x^2}{\sigma_z^2}\right)$ by

$$\frac{6}{\pi\,\mathrm{e}} \approx -1.5329\,\mathrm{dB} \tag{A.18}$$

Appendix B

Efficient MI and GMI estimations

For the MI and GMI estimations on the AWGN channel a Gauss quadrature integration is an efficient method. If we transmit using the M-ary constellation $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$ and a channel with noise variance σ_z^2. This is the AWGN channel with an SNR of $\frac{\mathbb{E}_X[|x|^2]}{\sigma^2}$.

For the mutual information, we assume all constellation points are equally likely and we introduce $d_{ij} = x_i - x_j$, $m = \log_2(M)$, $\mathcal{I}_{b_k}^{(i)}$ the constellation points where the k-th bit is the same as the k-th bit of the i-th constellation point and z is zero mean Gaussian distributed.

$$I = m - \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}_{\mathcal{Z}}\left[\log_2\left(\sum_{j=1}^{M}\exp\left(\frac{|d_{ij}|^2 + 2\Re\{z d_{ij}\}}{-\sigma_z^2}\right)\right)\right] \tag{B.1}$$

$$\begin{aligned} G = m - \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{m} E_{\mathcal{Z}}\Bigg[& \log_2\left(\sum_{j=1}^{M}\exp\left(\frac{|d_{ij}|^2 + 2\Re\{z d_{ij}\}}{-\sigma_z^2}\right)\right) \\ & - \log_2\left(\sum_{p\in\mathcal{I}_b^{(i)}}\exp\left(\frac{|d_{ip}|^2 + 2\Re\{z d_{ip}\}}{-\sigma_z^2}\right)\right)\Bigg] \end{aligned} \tag{B.2}$$

Now the expected value for $z = z_i + jz_q$ can be evaluated using the Gauss-Hermite quadrature

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(z_i^2+z_q^2)} f(z_i + jz_q)\, dz_i dz_q \approx \sum_{i=1}^{n}\sum_{q=1}^{n} \alpha_{l_i}\alpha_{l_q} f(\zeta_{l_i} + j\zeta_{l_q}) \tag{B.3}$$

with the weights α and sample points ζ. The MI can be estimated as

$$I = m - \frac{1}{M\pi} \sum_{i=1}^{M} \sum_{l_i=1}^{n} \sum_{l_q=1}^{n} \alpha_{l_i} \alpha_{l_q} \log_2 \left(\sum_{j=1}^{M} \exp \left(\frac{|d_{ij}|^2 + 2\Re\{(\zeta_{l_i} + j\zeta_{l_q}) d_{ij}\}}{-\sigma_z^2} \right) \right) \tag{B.4}$$

and the GMI as

$$\begin{aligned} G = m &- \frac{m}{M\pi} \sum_{i=1}^{M} \sum_{l_i=1}^{n} \sum_{l_q=1}^{n} \alpha_{l_i} \alpha_{l_q} \log_2 \left(\sum_{j=1}^{M} \exp \left(\frac{|d_{ij}|^2 + 2\Re\{(z_i + jz_q) d_{ij}\}}{-\sigma_z^2} \right) \right) \\ &- \frac{1}{M\pi} \sum_{i=1}^{M} \sum_{k=1}^{m} \sum_{l_i=1}^{n} \sum_{l_q=1}^{n} \alpha_{l_i} \alpha_{l_q} \log_2 \left(\sum_{p \in \mathcal{I}_{b_k}^{(i)}} \exp \left(\frac{|d_{ip}|^2 + 2\Re\{(z_i + jz_q) d_{ip}\}}{-\sigma_z^2} \right) \right) \end{aligned} \tag{B.5}$$

For the PS shaping, an efficient estimation for a single real dimension is

$$I = - \sum_{i=1}^{M} p_i \mathbb{E}_{\mathcal{Z}} \left[\log_2 \left(\sum_{j=1}^{M} p_i \exp \left(\frac{|d_{ij}|^2 + 2z d_{ij}}{-\sigma_z^2} \right) \right) \right] \tag{B.6}$$

$$I = - \frac{1}{\sqrt{\pi}} \sum_{i=1}^{M} p_i \sum_{l=1}^{n} \alpha_l \log_2 \left(\sum_{j=1}^{M} p_i \exp \left(\frac{|d_{ij}|^2 + 2\zeta_l d_{ij}}{-\sigma_z^2} \right) \right) \tag{B.7}$$

where $p_x(x_i) = p_i$ and the SNR is $\frac{2\mathbb{E}[|x|^2]}{\sigma^2}$. Note that the fist part is a scaled subsection of the MI calculation. And the weights with the sample points used in this work can be found in table B.1.

ζ	α
0.3429	6.1086×10^{-1}
1.0366	2.4014×10^{-1}
1.7567	3.3874×10^{-2}
2.5327	1.3436×10^{-3}
3.4362	7.6404×10^{-6}

Table B.1: Gauss-Hermite sampling points and weights of order 10. Half of the points are shown, the other half is identical but with $\zeta = -\zeta$.

Glossary

ADC analogue-to-digital converter 67

AIR achievable information rate 31

AWG arbitrary waveform generator 78

ANN artificial neural network 74

ASE amplified spontaneous emissions 20

AWGN additive white Gaussian noise 31

BCJR Balh, Cocke, Jelinek and Raviv 34

BFGS Broyden-Fletcher-Goldfarb-Shanno 52

BER bit-error rate 35

BICM bit interleaved coded modulation 31

BRGC binary reflected Gray code 35

BtB back-to-back 68

CCDM constant composition distribution matching 51

CM coded modulation 31

CUT channel under test 78

PSK phase-shift keying 33

QAM quadrature amplitude modulation 39

QPSK quadrature phase shift keying 49

RRC root-raised cosine 79

RSC recursive systematic convolutional 33

SD soft decision 38

SOA semiconductor optical amplifier 20

SPM self-phase modulation 44

SSFM split-step Fourier method 19

SSMF standard single mode fibre 66

SNR signal-to-noise ratio 31

TD-DBP time-domain digital back-propagation 75

TTCM turbo trellis-coded modulation 33

VOA variable optical attenuator 67

WDM wavelength division multiplexing 45

Bibliography

[1] D. M. Nally, "Redefining business success in a changing world," *19th Annual Global CEO Survey. Pricewaterhouse-Coopers. New York, Madrid*, 2016. [Online]. Available: http://www.pwc.com/gx/en/ceo-survey/2016/landing-page/pwc-19th-annual-global-ceo-survey.pdf

[2] "Cisco visual networking index: Forecast and trends, 2017–2022," Cisco Systems, Inc., White Paper, 2018.

[3] "Microsoft, facebook and telxius complete the highest-capacity subsea cable to cross the atlantic," Microsoft Corporate Website., Press Release, 2017.

[4] Y. Khalidi, "How microsoft builds its fast and reliable global network," Microsoft, Blog, 2017. [Online]. Available: https://azure.microsoft.com/en-gb/blog/how-microsoft-builds-its-fast-and-reliable-global-network/

[5] P. Runge and P. Trischitta, "The SL undersea lightwave system," *Journal of Lightwave Technology*, vol. 2, no. 6, pp. 744–753, dec 1984.

[6] P. Trischitta, M. Colas, M. Green, G. Wuzniak, and J. Arena, "The TAT-12/13 cable network," *IEEE Communications Magazine*, vol. 34, no. 2, pp. 24–28, 1996.

[7] N. S. Bergano, "Wavelength division multiplexing in long-haul transoceanic transmission systems," *Journal of Lightwave Technology*, vol. 23, no. 12, pp. 4125–4139, dec 2005.

[8] S. Bigo, W. Idler, J.-C. Antona, G. Charlet, C. Simonneau, M. Gorleir, M. Molina, S. Borne, C. de Barros, P. Sillard, P. Tran, R. Dischler,

W. Poehlmann, P. Nouchi, and Y. Frignac, "Transmission of 125 WDM channels at 42.7 gbit/s (5 tbit/s capacity) over 12×100 km of TeraLight/spl trade/ ultra fibre," in *European Conference on Optical Communication (ECOC)*, 2001, PD.M.1.1.

[9] A. H. Gnauck, G. Raybon, S. Chandrasekhar, J. Leuthold, C. Doerr, L. Stulz, A. Agarwal, S. Banerjee, D. Grosz, S. Hunsche, A. Kung, A. Marhelvuk, D. Maywar, M. Movassaghi, X. Liu, C. Xu, X. Wei, and D. M. Gill, "2.5 Tb/s (64×42.7 Gb/s) transmission over 40×100 km NZDSF using RZ-DPSK format and all-Raman-amplified spans," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2002, PD.FC2.

[10] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, "On the performance of nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," *Journal of Lightwave Technology*, vol. 29, no. 1, pp. 53–61, jan 2011.

[11] J.-X. Cai, Y. Cai, C. Davidson, A. Lucero, H. Zhang, D. Foursa, O. Sinkin, W. Patterson, A. Pilipetskii, G. Mohs, and N. Bergano, "20 Tbit/s capacity transmission over 6,860 km," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2011, PD.PB4.

[12] M. Mazurczyk, D. G. Foursa, H. G. Batshon, H. Zhang, C. R. Davidson, J.-X. Cai, A. Pilipetskii, G. Mohs, and N. S. Bergano, "30 Tb/s transmission over 6,630 km using 16QAM signals at 6.1 bits/s/Hz spectral efficiency," in *European Conference on Optical Communication (ECOC)*, 2012, PD.Th.3.C.2.

[13] M. Salsi, P. Tran, S. Bigo, P. Brindel, L. Schmalen, J. Renaudier, H. Mardoyan, R. Rios-Muller, G. Charlet, and A. Ghazisaeidi, "38.75 Tb/s transmission experiment over transoceanic distance," in *European Conference on Optical Communication (ECOC)*, 2013, PD.3.E.2.

[14] J.-X. Cai, H. Zhang, H. G. Batshon, M. Mazurczyk, O. V. Sinkin, Y. Sun, A. Pilipetskii, and D. G. Foursa, "Transmission over 9,100 km with a ca-

pacity of 49.3 Tb/s using variable spectral efficiency 16 QAM based coded modulation," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2014, PD.Th5B.4.

[15] J.-X. Cai, Y. Sun, H. G. Batshon, M. Mazurczyk, H. Zhang, D. G. Foursa, and A. N. Pilipetskii, "54 Tb/s transmission over 9,150 km with optimized hybrid Raman-EDFA amplification and coded modulation," in *European Conference on Optical Communication (ECOC)*, sep 2014, PD3.3.

[16] J. X. Cai, H. G. Batshon, M. Mazurczyk, H. Zhang, Y. Sun, O. V. Sinkin, D. G. Foursa, and A. Pilipetskii, "64QAM based coded modulation transmission over transoceanic distance with >60 Tb/s capacity," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2015, PD. Th5C.8.

[17] A. Ghazisaeidi, I. F. d. Jauregui Ruiz, R. Rios-Muller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet, and J. Renaudier, "65Tb/s transoceanic transmission using probabilistically-shaped DM-64QAM," in *European Conference on Optical Communication (ECOC)*, Sep. 2016, PD.Th.3.C.4.

[18] J.-X. Cai, H. G. Batshon, M. V. Mazurczyk, O. V. Sinkin, D. Wang, M. Paskov, W. Patterson, C. R. Davidson, P. Corbett, G. Wolter, T. Hammon, M. Bolshtyansky, D. Foursa, and A. Pilipetskii, "70.4 Tb/s capacity over 7,600 km in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2017, PD.Th5B.2.

[19] J.-X. Cai, H. G. Batshon, M. V. Mazurczyk, O. V. Sinkin, D. Wang, M. Paskov, W. W. Patterson, C. R. Davidson, P. C. Corbett, G. M. Wolter, T. E. Hammon, M. A. Bolshtyansky, D. G. Foursa, and A. N. Pilipetskii, "70.46 Tb/s over 7,600 km and 71.65 Tb/s over 6,970 km transmission in C+L band using coded modulation with hybrid constellation shaping and nonlinearity com-

pensation," *Journal of Lightwave Technology*, vol. 36, no. 1, pp. 114–121, jan 2018.

[20] M. Ionescu, D. Lavery, A. Edwards, E. Sillekens, L. Galdino, D. Semrau, R. Killey, W. Pelouch, S. Barnes, and P. Bayvel, "74.38 Tb/s transmission over 6300 km single mode fiber with hybrid EDFA/Raman amplifiers," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2019.

[21] P. J. Winzer, "Spatial multiplexing in fiber optics: The 10x scaling of metro/core capacities," *Bell Labs Technical Journal*, vol. 19, pp. 22–30, 2014.

[22] B. J. Puttnam, R. S. Luis, W. Klaus, J. Sakaguchi, J.-M. D. Mendinueta, Y. Awaji, N. Wada, Y. Tamura, T. Hayashi, M. Hirano, and J. Marciante, "2.15 Pb/s transmission using a 22 core homogeneous single-mode multi-core fiber and wideband optical comb," in *European Conference on Optical Communication (ECOC)*, sep 2015.

[23] D. Soma, Y. Wakayama, S. Beppu, S. Sumita, T. Tsuritani, T. Hayashi, T. Nagashima, M. Suzuki, H. Takahashi, K. Igarashi, I. Morita, and M. Suzuki, "10.16 Peta-bit/s denseSDM/WDM transmission over low-DMD 6-mode 19-core fibre across C+L band," in *European Conference on Optical Communication (ECOC)*, Sep. 2017.

[24] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, apr 2017.

[25] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, Jan. 2019.

[26] J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected

PDM-QAM signals," *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 868–875, feb 2017.

[27] 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, feb 2010.

[28] L. Galdino, D. Semrau, D. Lavery, G. Saavedra, C. B. Czegledi, E. Agrell, R. I. Killey, and P. Bayvel, "On the limits of digital back-propagation in the presence of transceiver noise," *Optics Express*, vol. 25, no. 4, p. 4564, feb 2017.

[29] R. Dar and P. J. Winzer, "On the limits of digital back-propagation in fully loaded WDM systems," *IEEE Photonics Technology Letters*, vol. 28, no. 11, pp. 1253–1256, jun 2016.

[30] S. Betti, F. Curti, G. D. Marchis, and E. Iannone, "A novel multilevel coherent optical system: 4-quadrature signaling," *Journal of Lightwave Technology*, vol. 9, no. 4, pp. 514–523, apr 1991.

[31] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation," *Journal of Lightwave Technology*, vol. 24, no. 1, pp. 12–21, jan 2006.

[32] L. Zehnder, "Ein neuer interferenzrefraktor," p. 275–285, 1891.

[33] L. Mach, "Über einen interferenzrefraktor," *Zeitschrift für Instrumentenkunde*, vol. 12, no. 3, p. 89–93, 1892.

[34] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A review of lithium niobate modulators for fiber-optic communications systems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 1, pp. 69–82, jan 2000.

[35] G. P. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. John Wiley and Sons, 1992.

[36] L. G. Kazovsky, L. Curtis, W. C. Young, and N. K. Cheung, "All-fiber 90° optical hybrid for coherent communications," *Applied Optics*, vol. 26, no. 3, p. 437, feb 1987.

[37] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164–1179, Sep. 2010.

[38] T. Pfau, S. Hoffmann, and R. Noe, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for *m* -qam constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989–999, Apr. 2009.

[39] M. Mazur, J. Schröder, A. Lorences-Riesgo, T. Yoshida, M. Karlsson, and P. A. Andrekson, "Overhead-optimization of pilot-based digital signal processing for flexible high spectral efficiency transmission," *Optics Express*, vol. 27, no. 17, pp. 24 654–24 669, Aug 2019.

[40] Y. Wakayama, E. Sillekens, L. Galdino, D. Lavery, R. Killey, and P. Bayvel, "Increasing achievable information rates with pilot-based dsp in standard intradyne detection," in *European Conference on Optical Communication (ECOC)*, 2019, W.1.B.

[41] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," *Proceedings of the Institution of Electrical Engineers*, vol. 113, no. 7, pp. 1151–1158, July 1966.

[42] F. P. Kapron, D. B. Keck, and R. D. Maurer, "Radiation losses in glass optical waveguides," *Applied Physics Letters*, vol. 17, no. 10, pp. 423–425, 1970.

[43] G. P. Agrawal, *Nonlinear Fiber Optics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[44] P. K. A. Wai and C. R. Menyak, "Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence," *Journal of Lightwave Technology*, vol. 14, no. 2, pp. 148–157, 1996.

[45] E. Desurvire, *Erbium-doped fiber amplifiers : principles and applications / Emmanuel Desurvire.* Wiley, 1994.

[46] J. Renaudier *et al.*, "First 100-nm continuous-band WDM transmission system with 115 tb/s transport over 100 km using novel ultra-wideband semiconductor optical amplifiers," in *European Conference on Optical Communication (ECOC)*, Sept 2017, Th.PDP.A.3.

[47] D. Cassioli, S. Scotti, and A. Mecozzi, "A time-domain computer simulator of the nonlinear response of semiconductor optical amplifiers," *IEEE J. Quantum Electron.*, vol. 36, no. 9, pp. 1072–1080, Sept 2000.

[48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, apr 2005.

[50] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[51] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[52] A. Guillén i Fàbregas, A. Martinez, 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, Jun. 2009.

[53] A. Alvarado, T. Fehenberger, B. Chen, and F. M. Willems, "Achievable information rates for fiber optics: Applications and computations," *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 424–439, Jan. 2018.

[54] E. Sillekens, A. Alvarado, C. M. Okonkwo, and B. C. Thomsen, "An experimental comparison of coded modulation strategies for 100 gb/s transceivers," *Journal of Lightwave Technology*, vol. 34, no. 24, pp. 5689–5697, Dec. 2016.

[55] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded mudulation using punctured component code," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 206–218, 1998.

[56] R. Gallagher, *Low-density parity-check codes*. Cambridge, Mass: MIT-Press, 2003.

[57] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[58] *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 1: DVB-S2*, European Standard ETSI EN 302 307-1, 2005.

[59] *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: DBV-S2 Extensions (DVB-S2X)*, European Standard ETSI EN 302 307-2, 2014.

[60] G. Frank, "Pulse code communication," U.S. Patent 2,632,058, Mar., 1953.

[61] A. Guillén i Fàbregas, A. Martinez, and G. Caire, *Bit-Interleaved Coded Modulation*, ser. Foundations and trends in communications and information theory. Now Publishers, 2008, vol. 5:1-2.

[62] A. Leven, F. Vacondio, L. Schmalen, S. Brink, and W. Idler, "Estimation of soft FEC performance in optical transmission experiments," *IEEE Photonics Technology Letters*, vol. 23, no. 20, pp. 1547–1549, Oct. 2011.

[63] A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *European Conference on Optical Communication (ECOC)*, Sep. 1993, MoC2.4.

[64] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *Journal of Lightwave Technology*, vol. 30, no. 24, pp. 3857–3879, dec 2012.

[65] D. Semrau, G. Saavedra, D. Lavery, R. I. Killey, and P. Bayvel, "A closed-form expression to evaluate nonlinear interference in raman-amplified links," *Journal of Lightwave Technology*, vol. 35, no. 19, pp. 4316–4328, Oct. 2017.

[66] G. E. Keiser, *Optical Fiber Communications*. Mcgraw Hill Higher Education, Nov. 2008.

[67] D. Semrau, E. Sillekens, P. Bayvel, and R. I. Killey, "Modeling and mitigation of fiber nonlinearity in wideband optical signal transmission [invited]," *Journal of Optical Communications and Networking*, vol. 12, no. 6, p. C68, apr 2020.

[68] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Optics Express*, vol. 23, no. 7, pp. 9183–9191, Apr. 2015.

[69] A. Mecozzi and R. Essiambre, "Nonlinear shannon limit in pseudolinear coherent systems," *Journal of Lightwave Technology*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.

[70] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Optics Express*, vol. 22, no. 13, pp. 16 335–16 362, Jun. 2014.

[71] D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel, "A modulation format correction formula for the gaussian noise model in the presence of inter-channel stimulated raman scattering," *Journal of Lightwave Technology*, vol. 37, no. 19, pp. 5122–5131, Oct. 2019.

[72] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for gaussian channels," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[73] F.-W. 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, Sep. 1993.

[74] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[75] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[76] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, Jun. 1987.

[77] 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, Dec. 2015.

[78] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[79] R. Fletcher, "A new approach to variable metric algorithms," *The Computer Journal*, vol. 13, no. 3, pp. 317–322, Mar. 1970.

[80] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Mathematics of Computation*, vol. 24, no. 109, pp. 23–23, Jan. 1970.

[81] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–647, Sep. 1970.

[82] E. Sillekens, D. Semrau, G. Liga, N. A. Shevchenko, Z. Li, A. Alvarado, P. Bayvel, R. I. Killey, and D. Lavery, "A simple nonlinearity-tailored probabilistic shaping distribution for square qam," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, Mar. 2018, M3C.4.

[83] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *IEEE International Symposium on Information Theory*, Jun. 2014, pp. 2794–2798.

[84] A. Alvarado, F. Brannstrom, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Information Theory Workshop*. IEEE, Oct. 2011.

[85] 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 *European Conference on Optical Communication (ECOC)*, 2018, Tu3G.3.

[86] J. Renner *et al.*, "Experimental comparison of probabilistic shaping methods for unrepeated fiber transmission," *Journal of Lightwave Technology*, vol. 35, no. 22, pp. 4871–4879, 2017.

[87] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, dec 1989.

[88] S. M. Stigler, "Gauss and the invention of least squares," *Ann. Statist.*, vol. 9, no. 3, pp. 465–474, May 1981.

[89] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[90] J. S. Bridle, "Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters," in *Advances in neural information processing systems*, 1990, pp. 211–217.

[91] M. Martinez-Ramon, A. Artes-Rodriguez, A. Navia-Vazquez, and A. R. Figueiras-Vidal, "Adaptively combined lms and logistic equalizers," *IEEE Signal Processing Letters*, vol. 11, no. 10, pp. 777–779, Oct. 2004.

[92] H. Zhao and J. Zhang, "Adaptively combined fir and functional link artificial neural network equalizer for nonlinear communication channel," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 665–674, Apr. 2009.

[93] S. Sygletos, A. Redyuk, and O. Sidelnikov, "Nonlinearity compensation techniques using machine learning," in *OSA Advanced Photonics Congress (AP)*. Optical Society of America, 2019, SpT2E.2.

[94] S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, "Reconstruction of binary signals using an adaptive radial-basis-function equalizer," *Signal Process.*, vol. 22, no. 1, pp. 77–93, Jan. 1991.

[95] T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization: Deep learning for fiber-optic nonlinearity compensation," 2019, arXiv:1911.10131.

[96] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.

[97] C. Häger and H. D. Pfister, "Deep learning of the nonlinear schrödinger equation in fiber-optic communications," in *IEEE ISIT*, Jun. 2018, pp. 1590–1594.

[98] ——, "Nonlinear interference mitigation via deep neural networks," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2018, W3A.4.

[99] ——, "Wideband time-domain digital backpropagation via subband processing and deep learning," in *European Conference on Optical Communication (ECOC)*, Sep. 2018.

[100] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.

[101] C. Fougstedt, M. Mazur, L. Svensson, H. Eliasson, M. Karlsson, and P. Larsson-Edefors, "Time-domain digital back propagation: Algorithm and finite-precision implementation aspects," in *Optical Fiber Communications Conference (OFC)*. Optical Society of America, 2017, W1G.4.

[102] A. Eghbali, H. Johansson, O. Gustafsson, and S. J. Savory, "Optimal least-squares fir digital filters for compensation of chromatic dispersion in digital coherent optical receivers," *Journal of Lightwave Technology*, vol. 32, no. 8, pp. 1449–1456, Apr. 2014.

[103] A. Sheikh, C. Fougstedt, A. G. i Amat, P. Johannisson, P. Larsson-Edefors, and M. Karlsson, "Dispersion compensation FIR filter with improved robustness to coefficient quantization errors," *Journal of Lightwave Technology*, vol. 34, no. 22, pp. 5110–5117, Nov. 2016.

[104] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2019, arXiv:1908.03265.

[105] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "Suppression of spurious tones induced by the split-step method in fiber systems simulation," *IEEE Photonics Technology Letters*, vol. 12, no. 5, pp. 489–491, may 2000.

[106] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *Journal of Lightwave Technology*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.

[107] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, arXiv:1512.03385.

[108] X. Liang and S. Kumar, "Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links," *Optics Express*, vol. 22, no. 24, pp. 29 733–29 745, Dec. 2014.

[109] W. Yan, Z. Tao, L. Dou, L. Li, S. Oda, T. Tanimura, T. Hoshida, and J. C. Rasmussen, "Low complexity digital perturbation back-propagation," in *European Conference on Optical Communication (ECOC)*, Sep. 2011, Tu.3.A.2.