Ph.D. Thesis

Advanced Forward Error Correction for Future Long-Haul Optical Communication Systems

Paolo Leoni

der Bundeswehr
Universität München

Faculty of Electrical Engineering and Information Technology
Institute of Information Technology
Chair of Communication Systems

Supervisor and first reviewer	Univ.-Prof. Dr.-Ing. Berthold Lankl
Second reviewer	Prof. Dr.-Ing. Norbert Hanik

April 2024

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN

Advanced Forward Error Correction for Future Long-Haul Optical Communication Systems

Paolo Leoni

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr-Ing.)

angenommenen Dissertation.

Promotionsausschuss

1. Prüfer:	Univ.-Prof. Dr.-Ing. Berthold Lankl
2. Prüfer:	Prof. Dr.-Ing. Norbert Hanik

Die Dissertation wurde am 29. September 2023 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 21. Februar 2024 angenommen. Die mündliche Prüfung fand am 23. April 2024 statt.

Success is not final, failure is not fatal:
it is the courage to continue that counts.
Falsely attributed to Winston Churchill [1].

Abstract

In 2010, when I started my Ph.D., this was the general consensus of the research community on forward error correction (FEC) for 100G long-haul coherent optical communications systems [2]:

> Currently the Optical Internetworking Forum (OIF) has locked in on dual-polarisation quaternary phase shift keying (DP-QPSK) following digital coherent receivers for 100 Gb/s transport systems.
>
> System designers have shifted their interest to intensively search for even more powerful FEC schemes having an NCG of over 10 dB. ... The candidates for this application are soft-decision and iterative decoding.
>
> The FEC redundancy in high-speed optical communications is limited by, among others, the availability of high-speed analog devices, the associated optical components, and the complexity of the digital circuitry. The industry consensus is that the maximum practical redundancy is currently not beyond 20 percent for 100 Gb/s digital coherent systems.

The purpose of this work is to challenge this approach, in particular the choice of quadrature phase-shift keying (QPSK), the use of iterative decoding alone and the limit of 20% redundancy, which corresponds to a symbol rate of $\approx$ 31 GBd. The other building blocks, namely polarisation division multiplexing (PDM), coherent digital signal processing (DSP) and soft-decision (SD), will be maintained.

I will present a solution for 100G long-haul non differentially encoded (NDE) coherent optical communications systems that achieves a net coding gain (NCG) over the uncoded QPSK of 11.8 dB. This solution is based on a *larger* constellation which allocates space for a *higher* redundancy of $\approx$ 60%, however having a *smaller* symbol rate of $\approx$ 28 GBd and hence occupying a *narrower* bandwidth. It uses available off-the-shelf building blocks with a complexity that, by now, should be more than feasible.

I will also present a related solution for 100G long-haul differentially encoded (DE) coherent optical communications systems that achieves a NCG over the uncoded QPSK of 10.2 dB.

These solutions are also perfectly scalable to 100G+ long-haul coherent optical communications systems, especially when the latters are realised increasing the symbol rate and hence the occupied bandwidth and not the size of the constellation.

Kurzfassung

Im Jahr 2010, als ich meine Promotion begann, war dies der allgemeine Konsens der Forschungsgemeinschaft zur Vorwärtsfehlerkorrektur für kohärente optische 100G Langstreckenkommunikationssysteme [2]:

> Derzeit hat sich das Optical Internetworking Forum (OIF) auf die Quadratur-Phasenumtastung mit dualer Polarisation nach digitalen kohärenten Empfängern für 100 Gb/s Transportsysteme festgelegt.
>
> Systementwickler haben ihr Interesse auf die intensive Suche nach noch leistungsfähigeren Vorwärtsfehlerkorrektur-Schemata mit einem Nettokodierungsgewinn von über 10 dB verlagert. ... Die Kandidaten für diese Anwendung sind weiche Entscheidung und iterative Decodierung.
>
> Die Vorwärtsfehlerkorrektur-Redundanz in der optischen Hochgeschwindigkeitskommunikation ist unter anderem durch die Verfügbarkeit von analogen Hochgeschwindigkeitsgeräten, die zugehörigen optischen Komponenten und die Komplexität der digitalen Schaltungen begrenzt. Die Industrie ist sich einig, dass die maximale praktische Redundanz bei digitalen kohärenten Systemen mit 100 Gb/s derzeit nicht über 20% liegt.

Mit dieser Arbeit soll dieser Ansatz in Frage gestellt werden, insbesondere die Wahl der Quadratur-Phasenumtastung, die Verwendung der iterativen Decodierung allein und die Begrenzung der Redundanz auf 20%, welche einer Symbolrate von ≈ 31 GBd entspricht. Die anderen Bausteine, nämlich Polarisationsmultiplexing, kohärente digitale Signalverarbeitung und weiche Entscheidung, werden beibehalten.

Ich werde eine Lösung für nicht differenziell kodierte kohärente optische 100G Langstreckenkommunikationssysteme vorstellen, welche einen Nettokodierungsgewinn von 11,8 dB gegenüber der uncodierten Quadratur-Phasenumtastung erreicht. Diese Lösung basiert auf einer *größeren* Konstellation, welche Platz für eine *höhere* Redundanz von $\approx 60\%$ bietet, jedoch eine *kleinere* Symbolrate von ≈ 28 GBd aufweist und daher eine *geringere* Bandbreite belegt. Sie verwendet verfügbare Standardbausteine mit einer Komplexität, welche mittlerweile mehr als machbar sein sollte.

Ich werde auch eine verwandte Lösung für differenziell kodierte kohärente optische 100G Langstreckenkommunikationssysteme vorstellen, welche einen Nettokodierungsgewinn von 10,2 dB gegenüber der uncodierten Quadratur-Phasenumtastung erreicht.

Diese Lösungen sind auch perfekt auf kohärente optische 100G+ Langstreckenkommunikationssysteme skalierbar, insbesondere wenn die letzteren durch Erhöhung der Symbolrate und damit der belegten Bandbreite und nicht der Größe der Konstellation realisiert werden.

Contents

1. Introduction

In the early days of optical communications (1970s), the challenges were mostly in the Physics and Electronics fields: first the invention, then the development and the production of light amplification by stimulated emission of radiations (LASERs), modulators, optical fibres, optical filters, optical amplifiers and photodiodes (PDs), among others. These early systems, although very complex from the Physics and Electronics perspective, were quite trivial from the Communication Theory point of view:

- at the transmitter, the optical power was modulated with a straightforward format called on-off keying (OOK): a "1" was encoded as "light on", a "0" as "light off";
- at the receiver, the optical power was detected and a straightforward hard-decision (HD) was taken: power above a certain threshold meant "1", power below the same threshold meant "0";
- impairments due to imperfections in the transmitter and/or the receiver as well as to propagation itself were left uncompensated for;
- longer reaches, lower bit error rates (BERs) or higher bit rates were achieved improving the devices: stabler LASERs, faster modulators, optical fibres with lower losses, optical filters with steeper frequency responses, optical amplifiers with lower noise figures and so on.

At some point, however, the constantly growing demand for capacity and reach made resorting to techniques from other kinds of communications (wired, such as communications over twisted pair or coaxial cable, as well as wireless, such as satellite or mobile communications) more and more attractive, if not unavoidable, in order to compensate for some of the impairments mentioned above. Two factors in particular proved essential [3]:

1. the advent of coherent detection, which made it possible to recover the whole electric field (amplitude and phase) in both polarisations, and
2. the advance in very large scale integration (VLSI), which made it possible to implement compensating algorithms at the rates needed in optical communications.

Just to give a rough idea of the time frame, in 1977 General Telephone and Electronics deployed the first *6 Mbps* fiber-optic system, transmitting the world's first live telephone traffic [4, 5]; in 2009 (32 years later), Verizon (operator) and Nortel (supplier, now Ciena) deployed the first commercial *100 Gbps* system (around 16.7 *thousand* times

faster), using quadrature phase-shift keying (QPSK) and polarisation division multiplexing (PDM) [6–8]. Since then, more and more techniques have been imported, adapted and applied from other kinds of communications to optical communications, especially in digital signal processing (DSP).

This work goes in the same direction, focusing on forward error correction (FEC) and advocating for a paradigm shift. In particular,

1. it shows that a certain technique, introduced by Ungerböck with its 1982 seminal paper [9] (see also [10, 11]), can be successfully imported, adapted and applied in 100G long-haul coherent optical communications systems at reasonable complexity. In his paper, Ungerböck proposed a very specific solution to a very specific problem, but his idea to use a "larger than strictly needed" constellation to allocate space for stronger codes is very general and can be used also in other applications;

2. it investigates the opportunity to use more advanced channel models which include also the (residual) phase noise beside the additive one. In fact, optical communications systems are impaired by the phase noise more than other kinds of communications mentioned above.

In both cases, the purpose is to propose new, alternative solutions for 100G long-haul non differentially encoded (NDE) and differentially encoded (DE) coherent optical communications systems that outperform current, traditional ones based on "as large as strictly needed" constellations (QPSK), codes with lower redundancies (20%), which are allocated only in wider occupied bandwidths corresponding to larger symbol rates (31.44 GBd) and channel models with only additive noise.

1.1. Motivation

Achieving the targeted BER for a lower signal-to-(*additive*-)noise ratio (SNR) is important in all communications systems, because the "freed" SNR can be used in a variety of ways to either increase the performance (e.g., in terms of reach or targeted BER) or lower the costs (e.g., by relaxing the components requirements).

In optical communications this is particularly important because the *optical* signal-to-noise ratio (OSNR) cannot be increased indefinitely by increasing the transmitted optical power, since beyond a certain threshold non-linearities will limit the performance of the system. Moreover, if the targeted BER can be achieved for a lower SNR, the transmitted optical power can be reduced; a reduced transmitted optical power, in turn, stimulates less non-linearities, which reduces the amount of noise and leads to an increased SNR.

Occupying a narrower bandwidth by placing most of the redundancy in the constellation and thus reducing the symbol rate is also beneficial: it reduces the speed at which electronic components must operate, it reduces the penalties accumulated when passing through optical filters and reconfigurable optical add-drop multiplexers (ROADMs) and it frees up bandwidth which can be used to pack neighbouring channels more densely, thus increasing the overall spectral efficiency (SE) of wavelength division multiplexing

(WDM) systems. In particular, the chosen FEC solution results in a standardised occupied bandwidth; in other words, the proposed solutions are backward compatible systems able to operate on existing networks.

1.2. Method

"Blindly" importing a technique from one field to another often leads to only suboptimal results: the general idea should remain more or less the same, but its details must be adapted to the conditions peculiar to the new application field for the import to be successful.

1.2.1. CE and IDD

The traditional approach when dealing with FEC is to consider it as a black-box to add at the beginning and at the end of an already existing system. The constellation has already been chosen, among those that carry the required number of bits per symbol m: for instance,

- an *uncoded* 100G long-haul coherent optical communications system employing PDM and with a symbol rate $f_s = 25$ GBd needs a constellation with $m = 2$ bits/symbol (QPSK is the natural choice).

When FEC is added to an existing system, the symbol rate and hence the occupied bandwidth is traditionally increased by the redundancy of the FEC solution:

- adding a FEC solution with 20% overhead (OH) to the previous example mantaining a constellation with $m = 2$ bits/symbol means that it now needs a symbol rate $f_s = 30$ GBd.

Its redundancy is, in other words, allocated in the bandwidth. This choice was reasonable for the first optical communications systems, for bandwidth was an abundant and readily available resource. In this work this technique will be referred to as "bandwidth expansion (BE)". Note that this technique offers a flexible choice of the amount of redundancy, for the occupied bandwidth can assume (in principle) any value. Such approach sees FEC solutions with lots of redundancy less favourably.

Ungerböck's alternative approach instead is to place the redundancy in the constellation (not in the bandwidth) and to choose the constellation and the FEC solution at the same time (not one after the other):

- adding a FEC solution with 50% OH to the previous example mantaining a symbol rate $f_s = 25$ GBd means that it now needs a constellation with $m = 3$ bits/symbol (which 8-ary constellation should then be used?).

It is true that larger constellations achieve higher pre-FEC BERs for the same SNR; however, stronger codes translate higher pre-FEC BERs to the same post-FEC BERs. Moreover, if the "constellation-FEC solution" combination is chosen wisely, the stronger code

more than compensates for the increased required SNR and the combination achieves the targeted post-FEC BERs for a lower SNR. Note that if the redundancy is allocated in the constellation, the occupied bandwidth (which is slowly becoming a scarser and scarser resource) must not be increased. In this work, this technique will be referred to as "constellation expansion (CE)", as opposed to BE. Note also that this technique offers a rigid choice of the amount of redundancy, for the size of the constellation can assume only values equal to powers of two. Such approach sees FEC solutions with lots of redundancy more favourably.

Obviously, a hybrid approach is also possible, which places the redundancy in both the bandwidth and the constellation:

- adding a FEC solution with 80% OH to the previous example could mean that it now needs a constellation with $m = 3$ bits/symbol *and* a symbol rate $f_s = 30$ GBd, although other combinations are also possible.

This technique takes the best of both worlds, offering a flexible choice of the amount of redundancy and seeing favourably FEC solutions with lots of redundancy.

The solutions proposed in this work will follow this hybrid approach. Beside PDM, they:

- present a FEC solution with 60% OH, need a constellation with $m = 3$ bits/symbol and a symbol rate $f_s = 27.95$ GBd.

However, using a "larger than strictly needed" constellation opens up a series of issues which must be addressed for the constellation-FEC solution combination to outperform the traditional approach:

- the choice of the constellation is not obvious anymore;
- special precautions must be taken when matching a constellation larger than the QPSK with a binary code, for the Euclidean distance between constellation points does not necessarily correspond to the Hamming distance between code word bits.

The solutions proposed in this work will make use of

- mutual information (MI) to choose the constellations and
- iterative demapping and decoding (IDD) to exploit the full capacity of the chosen constellations.

1.2.2. Channel models with phase noise

With respect to the other kinds of communications mentioned above, optical communications are characterised by:

- larger linewidths of the LASERs modulated at the transmitter and used as local oscillator (LO) at the receiver;

- the presence of non-linear effects in the optical fibre (hence, non localised) arising when the optical power is too high.

Both larger linewidths and (some of) the non-linear effects result in more phase noise, which hence is a bigger issue in optical communications than in other kinds of communications. However, channel models used for FEC usually consider only *additive* noise, so the question arises whether these models are still "good enough" to be used for FEC in optical communications or whether performance can be improved by using models which also take into account the *phase* noise, thus better representing the channel "seen by the soft-decision (SD) demapper and FEC decoder".

The importance of this step should not be underestimated: the choice of the channel model influences both the choice of the constellation and the calculation of the log-likelihood ratios (LLRs).

This work will investigate whether and, if so, how the phase noise must be taken into account in both steps.

1.2.3. Stronger codes, smaller symbol rates, narrower occupied bandwidths and higher complexity

With respect to the other kinds of communications mentioned above, optical communications are characterised also by

- lower targeted post-FEC BERs, nowadays 10^{-15};
- higher speeds at which electronic components have to work.

Very low targeted post-FEC BERs can only be achieved with capacity-approaching iterative codes such as turbo and low-density parity-check (LDPC) codes. In particular, since complexity is a delicate issue in optical communications, LDPC codes are very interesting for these systems because of their inherent suitability for parallelisation. Iterative codes, however, have one main problem: error floors.

The solutions proposed in this work will make use of

- a very hardware-aware standardised iterative SD LDPC code to quickly achieve the targeted BER at a reasonable complexity, followed by
- a very well-known classic algebraic HD Reed-Solomon (RS) code to remove the error floor;
- a smaller symbol rate and hence a narrower occupied bandwidth, which reduces the speed required of the electronic components.

Finally, the limit represented by complexity is a "soft" one: with time, speeds, sizes and power consumptions that seem unfeasible today will become feasible in the future. The limit represented by the MI, on the other hand, is a "hard" one: no matter how sofisticated the FEC solution, the Shannon limit of a given constellation cannot be beaten. To go beyond the latter, a larger constellation is the only solution.

1.3. Structure

Beside this first, introductory chapter and the final, conclusive one, this work has five central chapters:

- in chapter 2 five models will be presented for the channel "seen by the SD demapper and FEC decoder". Four of them model long-haul NDE and DE coherent optical communications systems considering either only the additive noise or both the additive and the phase noise; as such, they depend either only on the SNR or on both the SNR and the signal-to-*phase*-noise ratio (SPNR). The fifth model is an approximated version of the one modelling long-haul NDE coherent optical communications systems with both noises;
- in chapter 3 two estimators of the SNR and of the SPNR will be presented, based on the approximated model from chapter 2;
- in chapter 4 five laboratory experiments will be presented, representing 100G long-haul NDE and DE coherent optical communications systems operating in highly non-linear regime and thus impaired by (homogeneous) intra-channel non-linearities and/or homogeneous or heterogeneous inter-channel non-linearities. This serves two purposes:
  1. to validate the models from chapter 2;
  2. to determine in which ranges the SNR and more importantly the SPNR of practical long-haul NDE and DE coherent optical communications systems lie, applying to them the estimators of chapter 3. This step is particulary important because the choice of the constellations for the proposed solutions and the calculation of the LLRs depend on the SNR and on the SPNR ranges.

  It will be shown that the models are indeed correct and that the SPNR lies in three ranges (18 dB, 21 dB and 30 dB), depending on the experiment;
- in chapter 5 the MI of the QPSK (which is used in the traditional approach) and eight other constellations (which can be used in the alternative approach) will be presented, over the two NDE and DE models with both noises from chapter 2. This serves two purposes:
  1. to understand in which SPNR range the phase noise must be considered and in which it can be neglected (if at all) and what is the case for practical long-haul NDE and DE coherent optical communications systems (thanks to the results from chapter 4);
  2. to choose the best constellations for the proposed solutions.

  It will be shown that for each model there is indeed a "wPWGN-poor" region, in which the phase noise can be neglected, and that in both the NDE case and the DE case the SPNR lies in this region. Moreover, it will also be shown that, in this region, the 8-star and the 2 amplitude- 4 phase-shift keying (2A4P) are the best 8-ary constellations for NDE and DE systems, respectively;

- in chapter 6 the proposed solutions will be presented, based on the two NDE and DE models with only additive noise from chapter 2 (thanks to the results from chapter 5). The proposed solutions will also be compared in terms of minimum SNR required to achieve "quasi error-free" communication and complexity with a traditional one which reuses as many building blocks as possible but needs a *larger* symbol rate and hence occupies a *wider* bandwidth. It will be shown that
  - the proposed solution for 100G long-haul NDE coherent optical communications systems achieves a net coding gain (NCG) over the uncoded QPSK of 11.8 dB, 1.0 dB from the theoretical limit of the 8-star and 0.6 dB *beyond* the theoretical limit of the QPSK, however at the cost of an increased complexity: the traditional solution requires a 0.7 dB higher OSNR but also 0.2x the complexity;
  - the proposed solution for 100G long-haul DE coherent optical communications systems, instead, achieves a NCG over the uncoded QPSK = 10.2 dB, 1.0 dB from the theoretical limit of the 2A4P and 0.1 dB from the theoretical limit of the QPSK, even in this case at the cost of an increased complexity: the traditional solution requires a 0.1 dB higher OSNR but also 0.2x the complexity.

As it can be seen, the various arguments are interdependent and the reasoning sometimes flows in a circular fashion, which makes its linear exposition particularly challenging. Figure 1.1 shows a mind map of the chapter structure of this work highlighting these interdependencies, which will be helpful in managing the material.

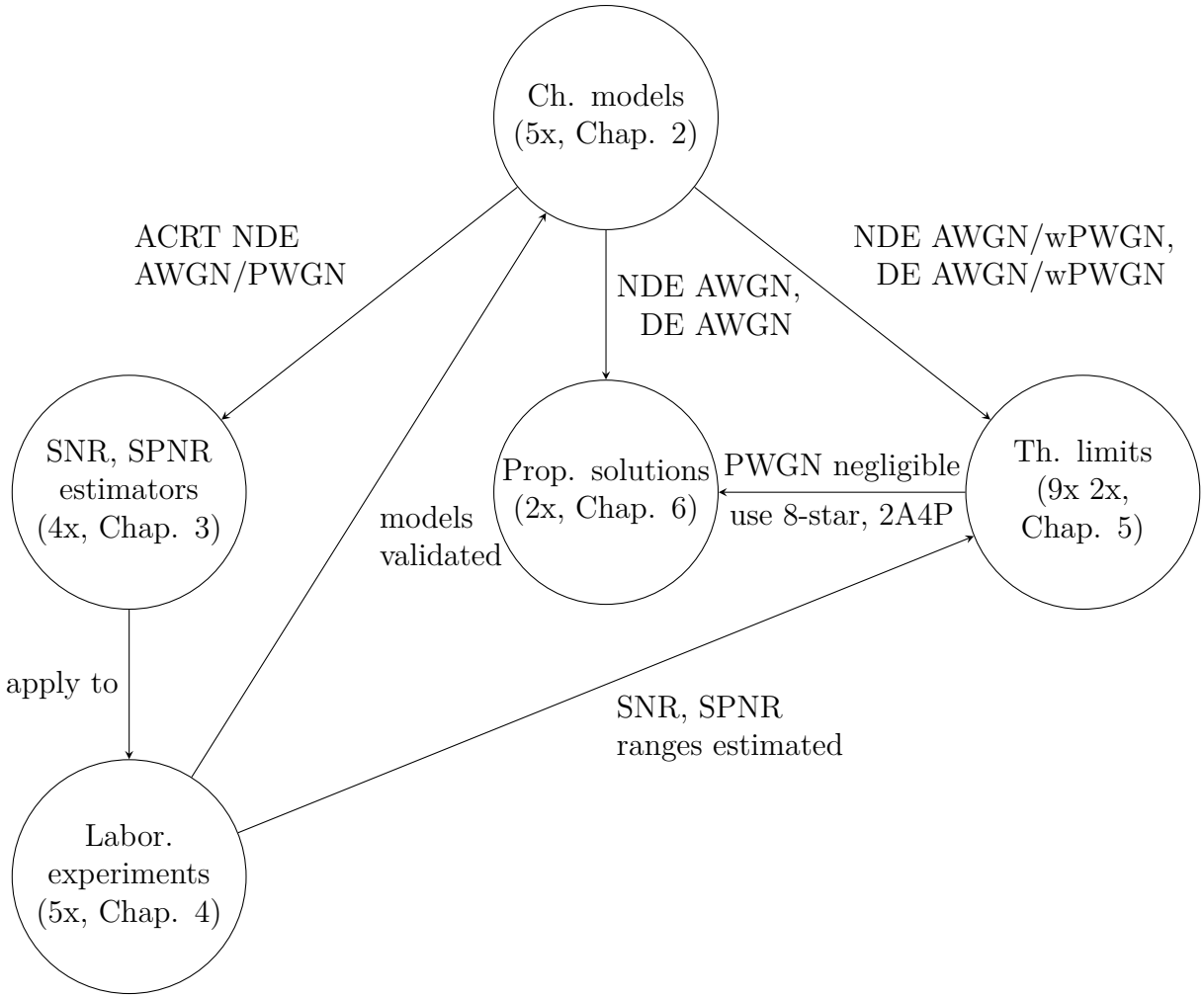

Figure 1.1.: Mind map of the chapter structure of this work, highlighting interdependencies and the circular nature of the exposition.

2. Statistical Models

In this chapter I present four models representing the channels *seen by the soft-decision (SD) demappers and forward error correction (FEC) decoders* of long-haul non differentially encoded (NDE) and differentially encoded (DE) coherent optical communications systems operating in both linear and non-linear regime as well as a fifth one helpful to derive two estimators of two parameters on which the first four models depend.

Figure 2.1 shows a mind map of the models. The first model is

1. the NDE additive white Gaussian noise (AWGN) model.

I will then extend this model in two directions, by adding the wrapped phase white Gaussian noise (wPWGN) and/or by introducing the DE function, thus obtaining three more models:

2. the NDE AWGN/ wPWGN model,
3. the DE AWGN model and
4. the DE AWGN/ wPWGN model.

Between the second and the third model I will also derive a fifth model, an approximated version of the second one based on stronger assumptions:

5. the approximated, counter-rotated and translated (ACRT) NDE AWGN/ phase white Gaussian noise (PWGN) model.

Figure 2.1 shows also where each model will be used:

- I will use the ACRT NDE AWGN/ PWGN model in chapter 3 to derive two efficient, minimum-variance unbiased (MVU) estimators of the signal-to-(*additive*-)noise ratio (SNR) and of the signal-to-*phase*-noise ratio (SPNR), which are the ratios between the energy per transmitted symbol E_s and the AWGN variance N_0 and the PWGN variance Θ_0, respectively (see section 3.3 for the proper definition). With respect to the NDE AWGN/ wPWGN model and to the DE AWGN/ wPWGN model, this model greatly simplified the derivation of the estimators;
- I will use the NDE AWGN/ wPWGN model and the DE AWGN/ wPWGN model in chapter 5 to estimate the theoretical limits in the form of the mutual information (MI) of the quadrature phase-shift keying (QPSK) as well as of various alternative constellations, for very broad ranges of the SNR and of the SPNR. Being exact, the reliability of these models and hence of the theoretical limits is not limited to certain ranges of the parameters;

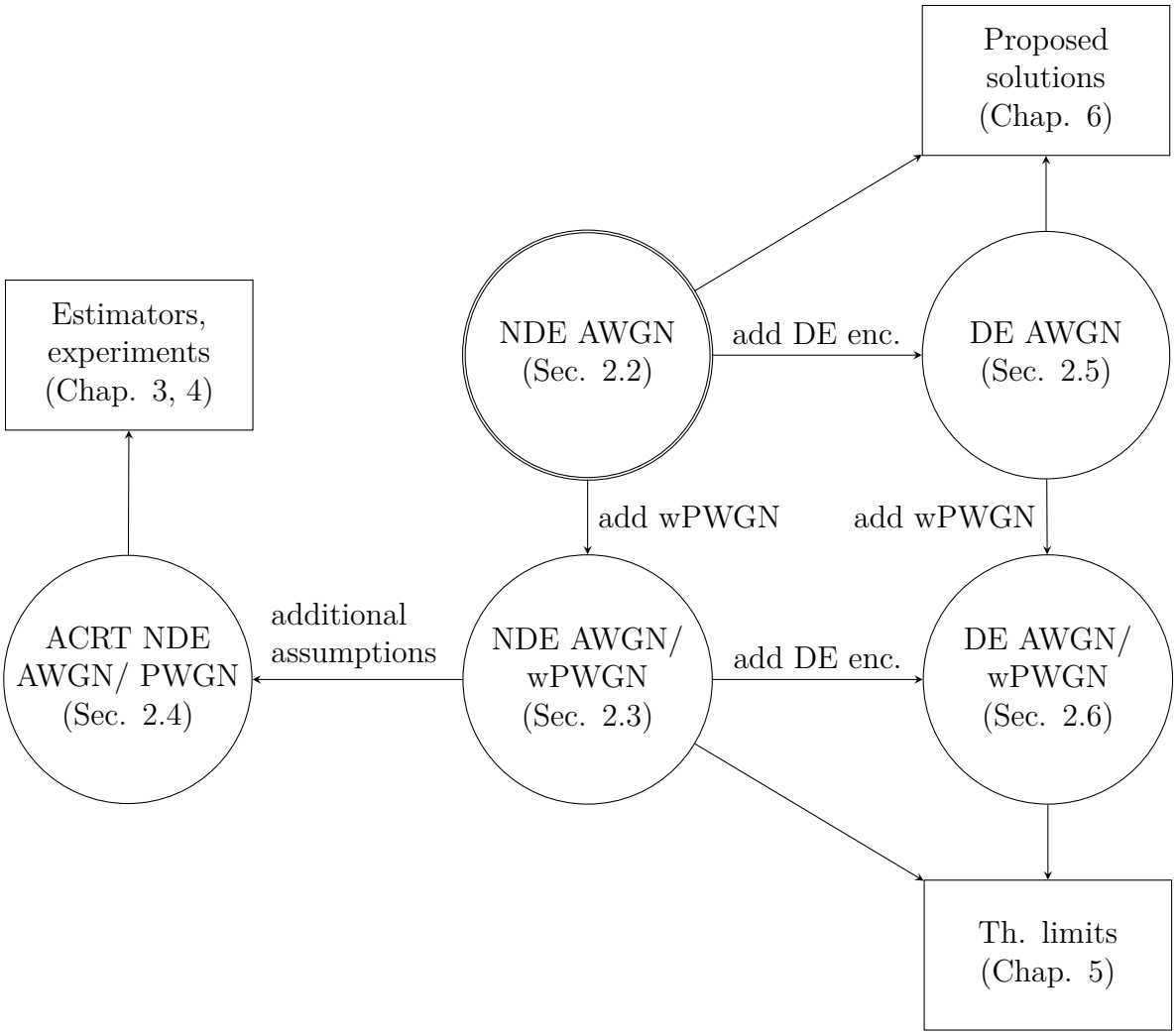

Figure 2.1.: Mind map of the models used in this work, indicating which one is derived, how, from which one and used where (see also table 2.1). The starting point is framed by a double line.

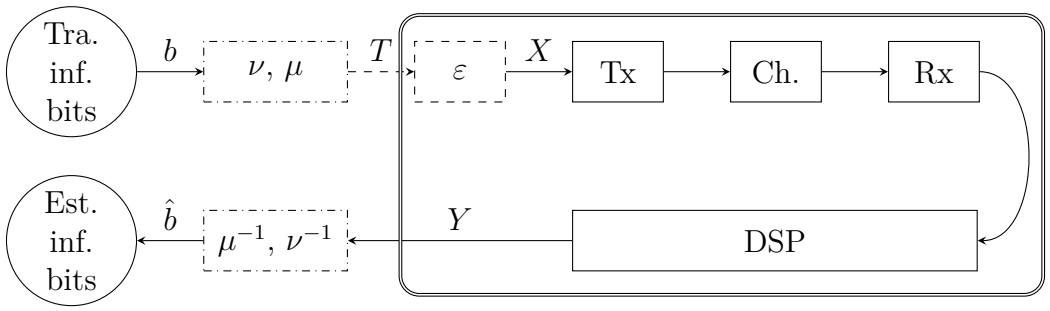

Figure 2.2.: Canonical block diagram of an optical communications system employing SD FEC. The DE encoder (dashed) is present only if the optical communications system is DE. The part framed by a double line is the channel seen by the SD demapper and FEC decoder and represented by the models presented in this chapter. The dash-dotted blocks will be replaced with the proposed FEC solution presented in chapter 6, see figure 6.1.

- I will use the NDE AWGN model and the DE AWGN model in chapter 6 to obtain the extrinsic log-likelihood ratios (LLRs) implemented by the SD demappers of the proposed solutions. These models offer formulas in closed form for the probability density functions (pdfs), which simplifies the derivation of the extrinsic LLRs.

Part of the material presented in this chapter has been published on the IEEE Photonics Technology Letters (PTL) [12, 13], at the Optical Fiber Conference (OFC) [14], at the European Conference on Optical Communications (ECOC) [15], at the Signal Processing in Photonic Communications (SPPCom) conference [16] as well as at the International Conference on Transparent Optical Networks (ICTON) [17].

2.1. Channel seen by the SD demapper and FEC decoder

In order to avoid misunderstandings I would like to emphasize at the very beginning what I mean by "seen by the SD demapper and FEC decoder". The models introduced in this chapter include, *but are not limited to*, the optical fibre. In fact, they include also what comes before and after it: the optoelectronic transmitter and receiver as well as the digital signal processing (DSP) algorithms.

Figure 2.2 shows a high level block diagram of such a system. In a nutshell, this is what happens in each "macroblock":

- the FEC encoder(s) ν encode(s) the transmitted *information* bits b into transmitted *encoded* bits b_e (not shown in figure 2.2). The mapper μ maps the transmitted encoded bits b_e to either transmitted *symbols* X or transmitted *transitions* T, depending on whether the system is NDE or DE.

  In the solutions I will propose in chapter 6 this dash-dotted block will contain two FEC encoders, for an inner SD code and an outer hard-decision (HD) code;

- if the system is DE, a DE encoder ε is present, which encodes the transmitted transitions T into transmitted symbols X.

  One of the solutions I will propose in chapter 6 will have a DE encoder;

- in the transmitter Tx a modulator modulates one (or two) polarisation(s) of a light amplification by stimulated emission of radiation (LASER) with the transmitted symbols X.

  In the solutions I will propose in chapter 6 the transmitter will modulate two polarisations, i.e. the system will use polarisation division multiplexing (PDM).

  The modulated LASER then travels through

- the optical channel Ch, made up of a certain number of optical fibre spools and optical amplifiers;

- in the receiver Rx the received signal is combined with a local oscillator (LO), photodiodes (PDs) generate electrical signals from the modulated LASER and analog-to-digital converters (ADCs) sample them;

- the DSP compensates for most of the effects introduced by the transmitter, the optical channel and the receiver and calculates the received symbols Y;

- the SD demapper μ^{-1} calculates the channel LLRs $\Lambda_{\mu^{-1}}$ from the received symbols Y.

  If the system is DE, the SD demapper μ^{-1} implements also the DE decoding.

  The FEC decoder(s) ν^{-1} decode(s) the channel LLRs $\Lambda_{\mu^{-1}}$ into estimated information bits $\hat{b}$.

  In the solutions I will propose in chapter 6 this dash-dotted block will contain two FEC decoders, for an inner SD code and an outer HD code. Moreover, the inner SD code and the SD demapper will perform iterative demapping and decoding (IDD) refining these LLRs before decoding them.

The blocks inside the double line constitute the channel seen by the SD demapper and FEC decoder and represented by the models presented in this chapter.

Since, as said, the DSP is aimed at resolving most of the problems introduced by the transmitter, the optical fibre and the receiver, most of the effects and of the countermeasures will cancel each other out and, as such, will not appear in the models. For instance, a carrier recovery (CR) in the DSP will take care of the phase and frequency noises of the LASERs in both the transmitter and the receiver, hence neither the former nor the latter will be modeled.

What will appear in the models are *residual* additive and/or phase noises, left over after (or in some cases even added by) the DSP. These models will depend on the variances of these residual noises.

2.2. NDE AWGN model

The first approach one encounters when studying FEC models the received symbol Y as the transmitted one X impaired by the AWGN N:

$$Y = X + N. \tag{2.1}$$

The transmitted symbol X is usually assumed to be *discrete*, complex and uniformly distributed:

$$p_X(x) = P[X = x] = \frac{1}{|\mathcal{X}|}, \quad x \in \mathcal{X} \subset \mathbb{C}, \tag{2.2}$$

where $\mathcal{X}$ is the constellation and $|\mathcal{X}|$ its cardinality (i.e., the number of possible transmitted symbols). I assume the transmitted symbol X to be complex in order to model long-haul coherent optical communications systems with in-phase and quadrature components. Clearly, each transmitted symbol X carries

$$m = \log_2(|\mathcal{X}|) \tag{2.3}$$

transmitted bits (could be *information* bits or *encoded* bits, depending on whether the system is uncoded or coded). The average energy per transmitted symbol is usually indicated as

$$E_s = E[|X|^2]. \tag{2.4}$$

However, there is the (very important) special case of the "Gaussian constellation", in which the transmitted symbol X is assumed to be *continuous*, complex and Gaussianly distributed with zero mean and the average energy per transmitted symbol E_s as complex variance:

$$p_X(x) = \frac{1}{\pi E_s} e^{-\frac{|x|^2}{E_s}}, \quad x \in \mathbb{C}. \tag{2.5}$$

In this case, the transmitted constellation is the whole complex plane.

Since in this work I will almost always consider discrete transmitted symbols, from now on I will only show formulas for this case, leaving to the reader's goodwill the task to substitute (2.2) with (2.5) and adapt the formulas consequently in the case of the Gaussian constellation (e.g., replacing summations with integrals when marginalizing a pdf).

The AWGN N is assumed to be *continuous*, complex and Gaussianly distributed with zero mean and complex variance N_0:

$$p_N(n) = \frac{1}{\pi N_0} e^{-\frac{|n|^2}{N_0}}, \quad n \in \mathbb{C}. \tag{2.6}$$

Again, I assume the AWGN N to be complex in order to model long-haul coherent optical communications systems with noise on both the in-phase and the quadrature components.

Being the sum of a discrete (or, in one case, continuous) and complex random variable and a continuous and complex one, the received symbol Y will also be a *continuous* and complex random variable.

2.2.1. NDE AWGN model: Statistical assumptions

The random variable N does not model one additive noise in particular, but rather the summation of a variety of independent *residual* additive noises. Potential sources include, but are not limited to:

- digital-to-analog converters (DACs), modulators, filters and amplifiers in the transmitter;
- optical amplifiers producing amplified spontaneous emission (ASE);
- non-linear effects in the optical fibre, which are traditionally left uncompensated – although recently solutions have been proposed to partly or fully compensate them, e.g. digital back-propagation (DBP) [18];
- filters, PDs and ADCs in the receiver;
- the DSP itself.

In other words, this model resorts to the central limit theorem, which indicates that the (properly normalised) summation of many independent random variables will tend to have a Gaussian distribution, even if the original random variables themselves are not normally distributed. Note that this independence of the various additive noises, which leads to the "G" in AWGN, must not be confused with the uncorrelation of the various random variables N_i impairing various transmitted symbols X_i, which is represented by the "W" in AWGN.

The latter is another classic assumption in the FEC community. Again, this does not mean assuming that the effects introduced by the transmitter, the optical fibre and the receiver do not have memory, often they do. It means assuming that this memory will be (almost) completely removed by the DSP, after which it is hence safe to assume that only (almost) uncorrelated noises are left over. To some extent, the presence of residual correlation after the DSP can be considered as an indicator of a sub-optimal choice of the DSP algorithms and/or of their parameters. Some unmodeled residual correlation after the DSP might even be consciously accepted (or at least tolerated), as the price to pay for a simple and hence usable model: a *mismatched* receiver will have sub-optimal performance, and the question then becomes whether this performance degradation can be accepted or not.

A great deal of work has been done in the last decades to model the non-linear effects in the optical fibre. Early studies, which did *not* include DSP and focussed mostly on links with dispersion compensating fibre (DCF), showed that non-linear noise is not always Gaussianly distributed [19–21]. More recent studies, however, which did include DSP and focussed mostly on uncompensated links, showed instead that many non-linear effects, with the exception of cross-phase modulation (XPM), can be modelled as AWGN *after* ideal DSP [22–24]. Note also that the non-linear response of silica occurs over a time scale of 60-70 fs and can hence be considered instantaneous for pulses > 1 ps (corresponding to 1 TBaud) [25, Chapter 2].

In chapter 4 I will prove these assumptions valid for intra- and/or homogeneous inter-channel non-linearities by means of laboratory experiments.

2.2.2. NDE AWGN model: Distribution of the received symbol

I can easily write the conditional pdf of the received symbol Y conditioned on the transmitted symbol X in closed form:

$$p_{Y|X}(y|x) = p_N(y-x) = \frac{1}{\pi N_0} e^{-\frac{|y-x|^2}{N_0}}, \quad y \in \mathbb{C}, x \in \mathcal{X}. \tag{2.7}$$

The fact that (2.7) is in closed form is very important; as soon as wPWGN is introduced, as I will show in section 2.3, this will not be the case anymore, which will remarkably increase the complexity of the treatise.

2.3. NDE AWGN/wPWGN model

The first direction in which I will extend the NDE AWGN model is adding the phase noise, obtaining one of the so-called *partially coherent* models [26]. This extension is motivated by the will to have a "less mismatched" receiver: in fact, long-haul coherent optical communications systems are more impaired by phase noise than other kinds of communications systems, since LASERs have larger linewidths than electrical oscillators and some non-linear effects in the optical fibre can manifest as phase noise.

In particular, the received symbol Y is modelled as the transmitted one X impaired by the AWGN N and by the wPWGN Θ:

$$Y = Xe^{+j\Theta} + N. \tag{2.8}$$

Note that the order in which the AWGN N and the wPWGN Θ are applied to the transmitted symbol is irrelevant, since

$$Y = (X+N)e^{+j\Theta} = Xe^{+j\Theta} + Ne^{+j\Theta} = Xe^{+j\Theta} + N', \tag{2.9}$$

where the random variable N' is distributed as the random variable N [26].

The AWGN N is still assumed to be continuous, complex and Gaussianly distributed according to (2.6).

The wPWGN Θ is assumed to be continuous, *real* (note that $e^{+j\Theta}$, not Θ, is complex), Gaussianly distributed with zero mean and variance Θ_0 and *wrapped* (since it is an angle):

$$p_\Theta(\theta) = \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta-2\pi k)^2}{\Theta_0}}, \quad \theta \in]-\pi, +\pi]. \tag{2.10}$$

Note that the random variable Θ is limited between $-\pi$ and $+\pi$ (any other interval given by a translation of this interval by multiple integers of $\pm 2\pi$ would also do) and that within this interval the tails of the replicas of the Gaussian bell centered around $2\pi k, k \neq 0$ will add to the one centered around 0. More on this nontrivial pdf can be read in [26]; for instance that

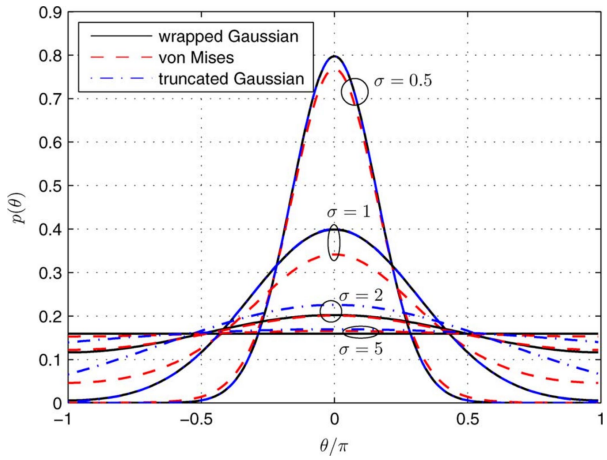

Figure 2.3.: Wrapped Gaussian pdf with zero mean and various variances (together with other pdfs not considered in this work) [26].

> the wrapped Gaussian approaches an uniform distribution for large Θ_0 and can be approximated by a Gaussian distribution for small Θ_0,

as shown in figure 2.3 [26].

Being the sum of a discrete (or, in one case, continuous) and complex rotated random variable and a continuous and complex one, the received symbol Y will also be a continuous and complex random variable.

2.3.1. NDE AWGN/wPWGN model: Statistical assumptions

The random variable Θ does not model one phase noise in particular, but rather the summation of a variety of independent *residual* phase noises. Potential sources include, but are not limited to:

- LASERs in both the transmitter and the receiver;
- non-linear effects in the optical fibre;
- the DSP itself.

Resorting again to the central limit theorem the random variable Θ is hence assumed to be wrapped Gaussianly distributed.

The LASER phase noise is per se obviously not white, but the model assumes that its version processed by the DSP and hence seen by the SD demapper and FEC decoder will be.

As said in section 2.2.1, most of the non-linearities *after* ideal DSP can be modelled as AWGN, with the notable exception of XPM. Modelling work showed that in systems *without* DSP XPM is indeed Gaussianly distributed [27], but is not white [28–30]. After a linear DSP that preserves the Gaussianity and removes the memory it is hence reasonable

to model XPM as wPWGN. Moreover, should the need arise (e.g. real systems using non-ideal DSP), the presence of an ideal interleaver-deinterleaver pair can be assumed [26]. Note that XPM being correlated *before* DSP leads to only moderate spectral broadening of the transmitted signal due to phase noise, which in turn means that effects of filtering and sampling a signal whose spectrum is broadened (such as signal distortions, energy loss and increased captured noise power) can be neglected [26].

This model assumes that the wPWGN Θ is independent of the transmitted symbol X,

$$p_{\Theta|X}(\theta|x) = p_{\Theta}(\theta). \tag{2.11}$$

At first sight, this could probably appear as a somewhat shaky assumption: higher launch powers do cause more non-linear effects, so a symbol with higher energy can cause more non-linearities than a symbol with lower energy. However, this doubt might arise only for the non-linear effects due to the propagation of the central channel: the phase noise due to the propagation of the neighbouring channels depends on the transmitted symbols in the neighbouring channels, which are however not represented by X. And even for the former, if for instance DBP is used, the effect (and hence the dependence) of the transmitted symbol on the phase noise is removed. This effect can also be mitigated by choosing constellations with points with constant amplitude, e.g. QPSK. Finally, even in this case an ideal interleaver-deinterleaver pair can be assumed if needed [26].

In chapter 4 I will prove these assumptions valid for intra- and/or homogeneous inter-channel non-linearities and to an extent also for heterogeneous inter-channel non-linearities by means of laboratory experiments.

2.3.2. NDE AWGN/wPWGN model: Distribution of the received symbol

In order to be able to write the conditional pdf of the received symbol Y conditioned on the transmitted symbol X in closed form, I can condition it also on the wPWGN Θ:

$$p_{Y|X,\Theta}(y|x,\theta) = p_N(y - xe^{+j\theta}) = \frac{1}{\pi N_0} e^{-\frac{\left|y - xe^{+j\theta}\right|^2}{N_0}}, \quad y \in \mathbb{C}, x \in \mathcal{X}, \theta \in]-\pi, +\pi]. \tag{2.12}$$

Thanks to (2.11) and Bayes I obtain the conditional pdf of the received symbol Y conditioned on the transmitted symbol X:

$$\begin{aligned} p_{Y|X}(y|x) &= \int_{-\pi}^{+\pi} p_{Y|X,\Theta}(y|x,\theta) \cdot p_{\Theta|X}(\theta|x) d\theta = \\ &= \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y - xe^{+j\theta}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta - 2\pi k)^2}{\Theta_0}} d\theta, \quad y \in \mathbb{C}, x \in \mathcal{X}, \end{aligned} \tag{2.13}$$

which is quite cumbersome and cannot be solved analytically.

The pdf of the received symbol Y $p_Y(y)$ is obtained again thanks to Bayes multiplying (2.13) by (2.2) and summing over all $x \in \mathcal{X}$.

2.4. ACRT NDE AWGN/PWGN model

As I showed in the previous section, extending the NDE AWGN model by adding the wPWGN leads to a conditional pdf of the received symbol Y conditioned on the transmitted symbol X (2.13) which is very hard to handle. As I will show in chapter 3, estimators of the SNR and of the SPNR based on this model could be derived, which however proved unstable when applied to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems; moreover, it is not possible to calculate their Cramér-Rao lower bound (CRLB) in closed form, which prevents me from knowing whether these estimators are efficient or not.

As such, the need arises for an approximated model; the insight that typically comes with being able to find an analytical solution to such estimation problems will prove worthy of the sacrifice in accuracy. This requires two additional steps:

1. assuming that the wPWGN Θ is *small*, the latter can be replaced by its Taylor expansion centred at zero (i.e., with its Maclaurin expansion), stopped at its second term (the linear one),

$$e^{+j\Theta} \approx 1 + j\Theta. \tag{2.14}$$

   By doing so I can model the received symbol Y as the transmitted one X impaired by the AWGN N and by the PWGN Θ:

$$Y \approx X(1 + j\Theta) + N = X + jX\Theta + N. \tag{2.15}$$

   The PWGN Θ is assumed to be continuous, real, Gaussianly distributed with zero mean and variance Θ_0 and *not wrapped* anymore:

$$p_\Theta(\theta) = \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta)^2}{\Theta_0}}, \quad \theta \in \mathbb{R}; \tag{2.16}$$

2. assuming that the transmitted symbol X has *unitary amplitude* (which, incidentally, also excludes the case $X = 0$) and that it is *known*,

$$|X| = 1, \quad X \text{ known}, \tag{2.17}$$

   the received symbol Y can be divided by the transmitted symbol X (i.e., the modulation can be "removed") and translated by 1.

   By doing so the counter-rotated and translated (CRT) received symbol Y' is given by the sum of the PWGN Θ multiplied by j and the counter-rotated AWGN N',

$$Y' = \frac{Y}{X} - 1 \approx j\Theta + \frac{N}{X} = j\Theta + N', \tag{2.18}$$

   where thanks to (2.17) the random variable N' is distributed as the random variable N (with a reasoning similar to the one in section 2.3).

The set of the CRT received symbols $Y'_i, i = 1, \ldots, L$, i.e. the CRT received constellation, lies on the origin and is symmetric with respect to the real axis $y = 0$ as well as to the imaginary axis $x = 0$: it is an ellipse, whose minor axis depends on the AWGN variance N_0 and whose major axis depends on the latter and on the PWGN variance Θ_0. The in-phase component of the CRT received symbol Y' is hence given only by the in-phase component of the counter-rotated AWGN N', whereas its quadrature component is given by the sum of the PWGN Θ and the quadrature component of the counter-rotated AWGN N':

$$\Re(Y') \approx \Re(N'), \tag{2.19}$$
$$\Im(Y') \approx \Theta + \Im(N'). \tag{2.20}$$

The condition (2.17) corresponds to using a phase-shift keying (PSK) modulation and transmitting only pilot tones. Obviously, a communications system in which the receiver already knows the transmitted symbols is useless, and in fact in the following I will not propose solutions based on this model, rather I will use it "only" to derive estimators of the SNR and of the SPNR, which I will then apply to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems, in which clearly I know the transmitted symbols.

2.4.1. ACRT NDE AWGN/PWGN model: Statistical assumptions

Recall that, if the counter-rotated AWGN N' is complex, its two in-phase and quadrature components $\Re(N')$ and $\Im(N')$ are uncorrelated (and hence independent), continuous, *real* and Gaussianly distributed with zero mean and variance $N_0/2$ each:

$$p_{\Re(N')}(\Re(n)) = p_{\Im(N')}(\Im(n)), \quad n \in \mathbb{C}. \tag{2.21}$$

Moreover, this model assumes that the PWGN Θ is independent of the quadrature component of the counter-rotated AWGN $\Im(N')$ (or, more in general, of the whole counter-rotated AWGN N'), so that their sum is continuous, real and still Gaussianly distributed with zero mean and variance equal to the sum of their variances Θ_0 and $N_0/2$:

$$p_{\Theta+\Im(N')}(\Im(n)) = p_{\Theta}(\Im(n)) * p_{\Im(N')}(\Im(n)), \quad n \in \mathbb{C}. \tag{2.22}$$

This corresponds to assuming that each one of the sources listed in sections 2.2.1 and 2.3.1 generates noises that, singularly taken, are either an additive or a phase noise, but not both at the same time.

2.4.2. ACRT NDE AWGN/PWGN model: Distribution of the CRT received symbols

Thanks to (2.19) I can easily write the pdf of the in-phase component of the CRT received symbol Y',

$$p_{\Re(Y')}(\Re(y)) = p_{\Re(N')}(\Re(y)) = \frac{1}{\sqrt{2\pi(\frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Re(y))^2}{(\frac{1}{2}N_0)}}, \quad y \in \mathbb{C}. \tag{2.23}$$

Similarly, thanks to (2.20) I can write the pdf of the quadrature component of the CRT received symbol Y',

$$p_{\Im(Y')}(\Im(y)) = p_{\Theta+\Im(N')}(\Im(y)) = \frac{1}{\sqrt{2\pi(\Theta_0 + \frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Im(y))^2}{(\Theta_0+\frac{1}{2}N_0)}}, \quad y \in \mathbb{C}. \tag{2.24}$$

Finally, I can write the pdf of the CRT received symbol Y' as the *joint* pdf of its in-phase and quadrature components,

$$\begin{aligned} p_{Y'}(y) &= p_{\Re(Y')}(\Re(y)) \cdot p_{\Im(Y')}(\Im(y)) = \\ &= \frac{1}{\sqrt{2\pi(\frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Re(y))^2}{(\frac{1}{2}N_0)}} \cdot \frac{1}{\sqrt{2\pi(\Theta_0 + \frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Im(y))^2}{(\Theta_0+\frac{1}{2}N_0)}}, \quad y \in \mathbb{C}. \end{aligned} \tag{2.25}$$

Note that it is not necessary anymore to go through the intermediate step of conditioning on the transmitted symbol X, because now the latter is known, it is not a random variable anymore.

If now I have a series of *independent* CRT received symbols $Y'_i, i = 1, \ldots, L$, I can write their *joint* pdf as:

$$\begin{aligned} p_{Y'_1,Y'_2,\ldots,Y'_L}(y_1, y_2, \ldots, y_L) &= \prod_{i=1}^{L} p_{Y'_i}(y_i) = \\ &= \prod_{i=1}^{L} \frac{1}{\sqrt{2\pi(\frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Re(y_i))^2}{(\frac{1}{2}N_0)}} \cdot \frac{1}{\sqrt{2\pi(\Theta_0 + \frac{1}{2}N_0)}} e^{-\frac{1}{2}\frac{(\Im(y_i))^2}{(\Theta_0+\frac{1}{2}N_0)}}, \\ & y_1, y_2, \ldots, y_L \in \mathbb{C}. \end{aligned} \tag{2.26}$$

The independency of the CRT received symbols $Y'_i, i = 1, \ldots, L$ is given under the assumption that the overall discrete-time equivalent impulse response of the system fulfills the Nyquist criterion for the absence of intersymbol interference (ISI). In chapter 4 I will prove these assumptions valid for intra- and/or homogeneous inter-channel non-linearities and to an extent also for heterogeneous inter-channel non-linearities by means of laboratory experiments.

2.5. DE AWGN model

The second direction in which I will extend the NDE AWGN model is adding the DE. This extension is motivated by the will to investigate long-haul DE coherent optical communications systems.

In this case the bits are mapped to transmitted *transitions* and the SD demapper μ^{-1} must consider *two* consecutive received symbols to calculate the extrinsic LLRs for the bits mapped to *one* transmitted transition (see again section 2.1 and figure 2.2). In other words, the SD demapper and the FEC decoder see a channel with *one* input, the transmitted transition, and *two* outputs, the two received symbols. Note that this does not mean that the symbol rate must be doubled: in a continuous transmission, every received symbol serves in two symbol periods, as first (previous) received symbol in one symbol period and as second (current) received symbol in the next one.

In principle, it is possible to DE encode the amplitude and/or the phase:

- the DE of the amplitude is helpful against fading and is for instance employed in radio communications;
- the DE of the phase is helpful against cycle slip and is for instance employed in satellite and optical communications. Clearly,
- the two operations can also be combined, if needed.

In this work I will only refer to the DE of the phase.

Each of the two received symbols $Y_i, i = 1, 2$ is modelled as the transmitted one $X_i, i = 1, 2$ impaired by the AWGN $N_i, i = 1, 2$ as in section 2.2:

$$Y_i = X_i + N_i, \quad i = 1, 2. \tag{2.27}$$

The first (previous) transmitted symbol X_1 is still assumed to be discrete, complex and uniformly distributed according to (2.2). The second (current) transmitted symbol X_2, instead, depends on the latter and on the transmitted transition T according to the DE function ε:

$$X_2 = \varepsilon(X_1, T) = \begin{cases} |X_2| & = |T| \\ \angle X_2 & = \angle X_1 + \angle T. \end{cases} \tag{2.28}$$

The second (current) transmitted symbol X_2 hence has amplitude given by the amplitude of the transmitted transition T and phase given by the *sum* of the phase of the first (previous) transmitted symbol X_1 and the phase of the transmitted transition T. Obviously, the phase of the first (previous) transmitted symbol X_1, considered given in the current symbol period, was actually calculated in the previous one; similarly, the phase of the second (current) transmitted symbol X_2, calculated in the current symbol period, will actually be considered given in the next one.

Similarly to (2.2), the transmitted transition T is assumed to be discrete, complex and uniformly distributed:

$$p_T(t) = P[T = t] = \frac{1}{|\mathcal{T}|}, \quad t \in \mathcal{T} \subset \mathbb{C}, \tag{2.29}$$

where $\mathcal{T}$ is the transition constellation and $|\mathcal{T}|$ its cardinality (i.e., the number of possible transmitted transitions). Note that in some applications the transition constellation $\mathcal{T}$ might be different from the constellation $\mathcal{X}$, since the former must include a 0° phase transition. Clearly, each transmitted transition T carries

$$m = \log_2(|\mathcal{T}|) \tag{2.30}$$

transmitted bits.

Each of the AWGN $N_i, i = 1, 2$ is still assumed to be continuous, complex and Gaussianly distributed according to (2.6). This model was already introduced in the past [31–34]; here I simply re-obtained it in a different way.

A final note on the choice of the letter t: in this work I will not present functions of the continuous time, hence I can use this letter for the transmitted transition without ambiguity.

2.5.1. DE AWGN model: Statistical assumptions

The first (previous) transmitted symbol X_1 is obvioulsy independent of the transmitted transition T,

$$p_{X_1|T}(x_1|t) = p_{X_1}(x_1). \tag{2.31}$$

The same holds true for the first (previous) received symbol Y_1,

$$p_{Y_1|X_1,T}(y_1|x_1, t) = p_{Y_1|X_1}(y_1|x_1). \tag{2.32}$$

2.5.2. DE AWGN model: Distribution of the received symbols

Thanks to (2.32) I can write the conditional pdf of the first (previous) received symbol Y_1 conditioned on the first (previous) transmitted symbol X_1 *and* on the transmitted transition T in closed form:

$$p_{Y_1|X_1,T}(y_1|x_1, t) = p_N(y_1 - x_1) = \frac{1}{\pi N_0} e^{-\frac{|y_1 - x_1|^2}{N_0}}, \quad y_1 \in \mathbb{C}, x_1 \in \mathcal{X}. \tag{2.33}$$

Using the DE function (2.28) I can also write the conditional pdf of the second (current) received symbol Y_2 conditioned on the first (previous) transmitted symbol X_1 *and* on the transmitted transition T in closed form:

$$\begin{aligned} p_{Y_2|X_1,T}(y_2|x_1, t) &= p_{Y_2|X_2}(y_2|\varepsilon(x_1, t)) = p_N(y_2 - \varepsilon(x_1, t)) = \\ &= \frac{1}{\pi N_0} e^{-\frac{|y_2 - \varepsilon(x_1,t)|^2}{N_0}}, \quad y_2 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}. \end{aligned} \tag{2.34}$$

I can then write the *joint* conditional pdf of the first (previous) received symbol Y_1 *and* the second (current) received symbol Y_2 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T as

$$\begin{aligned} p_{Y_1,Y_2|X_1,T}(y_1,y_2|x_1,t) &= p_{Y_1|X_1,T}(y_1|x_1,t)\cdot p_{Y_2|X_1,T}(y_2|x_1,t) = \\ &= \frac{1}{\pi N_0} e^{-\frac{|y_1-x_1|^2}{N_0}} \cdot \frac{1}{\pi N_0} e^{-\frac{|y_2-\varepsilon(x_1,t)|^2}{N_0}}, \quad y_1,y_2 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}. \end{aligned} \tag{2.35}$$

Finally, thanks to (2.31) and Bayes I obtain the joint conditional pdf of the first (previous) received symbol Y_1 and the second (current) received symbol Y_2 conditioned on the transmitted transition T:

$$\begin{aligned} p_{Y_1,Y_2|T}(y_1,y_2|t) &= \sum_{x_1\in\mathcal{X}} p_{Y_1,Y_2|X_1,T}(y_1,y_2|x_1,t)\cdot p_{X_1|T}(x_1|t) = \\ &= \sum_{x_1\in\mathcal{X}} \frac{1}{\pi N_0} e^{-\frac{|y_1-x_1|^2}{N_0}} \cdot \frac{1}{\pi N_0} e^{-\frac{|y_2-\varepsilon(x_1,t)|^2}{N_0}} \cdot \frac{1}{|\mathcal{X}|}, \quad y_1,y_2 \in \mathbb{C}, t \in \mathcal{T}. \end{aligned} \tag{2.36}$$

The joint pdf of the first (previous) received symbol Y_1 and the second (current) received symbol Y_2 is obtained again thanks to Bayes multiplying (2.36) by (2.29) and summing over all $t \in \mathcal{T}$.

2.6. DE AWGN/wPWGN model

For the last model I will further extend either the NDE AWGN/ wPWGN model of section 2.3 adding the DE or, equivalently, the DE AWGN model of section 2.5 adding the phase noise.

Each of the two received symbols $Y_i, i = 1,2$ is modelled as the transmitted one $X_i, i = 1,2$ impaired by the AWGN $N_i, i = 1,2$ and by the wPWGN $N_i, i = 1,2$:

$$Y_i = X_i e^{+j\Theta_i} + N_i, \quad i = 1,2. \tag{2.37}$$

The first (previous) transmitted symbol X_1 is still assumed to be discrete, complex and uniformly distributed according to (2.2). The transmitted transition T is still assumed to be discrete, complex and uniformly distributed according to (2.29). The second (current) transmitted symbol X_2 still depends on the former and on the latter according to the DE function (2.28). Each of the AWGN $N_i, i = 1,2$ is still assumed to be continuous, complex and Gaussianly distributed according to (2.6). Each of the wPWGN $\Theta_i, i = 1,2$ is still assumed to be continuous, real and Gaussianly distributed according to (2.10).

2.6.1. DE AWGN/wPWGN model: Statistical assumptions

The first wPWGN Θ_1 is obvioulsy independent of the transmitted transition T,

$$p_{\Theta_1|X_1,T}(\theta_1|x_1,t) = p_{\Theta_1|X_1}(\theta_1|x_1). \tag{2.38}$$

2.6.2. DE AWGN/wPWGN model: Distribution of the received symbols

In order to be able to write the conditional pdf of the first (previous) received symbol Y_1 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T in closed form, I can condition it also on the first wPWGN Θ_1 and use (2.32):

$$p_{Y_1|X_1,T,\Theta_1}(y_1|x_1,t,\theta_1) = p_{Y_1|X_1,\Theta_1}(y_1|x_1,\theta_1) = p_N(y_1 - x_1 e^{+j\theta_1}) = \\ = \frac{1}{\pi N_0} e^{-\frac{\left|y_1 - x_1 e^{+j\theta_1}\right|^2}{N_0}}, \quad y_1 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}, \theta_1 \in]-\pi, +\pi]. \tag{2.39}$$

Thanks to (2.38), (2.11) and Bayes I obtain the conditional pdf of the first (previous) received symbol Y_1 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T:

$$p_{Y_1|X_1,T}(y_1|x_1,t) = \int_{-\pi}^{+\pi} p_{Y_1|X_1,T,\Theta_1}(y_1|x_1,t,\theta_1) \cdot p_{\Theta_1|X_1,T}(\theta_1|x_1,t)d\theta_1 = \\ = \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_1 - x_1 e^{+j\theta_1}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_1 - 2\pi k)^2}{\Theta_0}} d\theta_1, \\ y_1 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}. \tag{2.40}$$

Similarly, in order to be able to write the conditional pdf of the second (current) received symbol Y_2 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T in closed form, I can condition it also on the second wPWGN Θ_2 and use the DE function (2.28):

$$p_{Y_2|X_1,T,\Theta_2}(y_2|x_1,t,\theta_2) = p_{Y_2|X_2,\Theta_2}(y_2|\varepsilon(x_1,t),\theta_2) = p_N(y_2 - \varepsilon(x_1,t)e^{+j\theta_2}) = \\ = \frac{1}{\pi N_0} e^{-\frac{\left|y_2 - \varepsilon(x_1,t)e^{+j\theta_2}\right|^2}{N_0}}, \quad y_2 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}, \theta_2 \in]-\pi, +\pi]. \tag{2.41}$$

Thanks to the DE function (2.28), (2.11) and Bayes I obtain the conditional pdf of the second (current) received symbol Y_2 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T:

$$p_{Y_2|X_1,T}(y_2|x_1,t) = \int_{-\pi}^{+\pi} p_{Y_2|X_1,T,\Theta_2}(y_2|x_1,t,\theta_2) \cdot p_{\Theta_2|X_1,T}(\theta_2|x_1,t)d\theta_2 = \\ = \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_2 - \varepsilon(x_1,t)e^{+j\theta_2}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_2 - 2\pi k)^2}{\Theta_0}} d\theta_2, \\ y_2 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}. \tag{2.42}$$

I can then write the *joint* conditional pdf of the first (previous) received symbol Y_1 *and* the second (current) received symbol Y_2 conditioned on the first (previous) transmitted symbol X_1 *and* the transmitted transition T as

$$\begin{aligned} p_{Y_1,Y_2|X_1,T}(y_1,y_2|x_1,t) &= p_{Y_1|X_1,T}(y_1|x_1,t)\cdot p_{Y_2|X_1,T}(y_2|x_1,t) = \\ &= \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_1-x_1 e^{+j\theta_1}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_1-2\pi k)^2}{\Theta_0}} d\theta_1 \cdot \\ &\quad \cdot \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_2-\varepsilon(x_1,t)e^{+j\theta_2}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_2-2\pi k)^2}{\Theta_0}} d\theta_2, \\ &\quad y_1,y_2 \in \mathbb{C}, x_1 \in \mathcal{X}, t \in \mathcal{T}. \end{aligned} \tag{2.43}$$

Finally, thanks to (2.31) and Bayes I obtain the joint conditional pdf of the first (previous) received symbol Y_1 and the second (current) received symbol Y_2 conditioned on the transmitted transition T:

$$\begin{aligned} p_{Y_1,Y_2|T}(y_1,y_2|t) &= \sum_{x_1\in\mathcal{X}} p_{Y_1,Y_2|X_1,T}(y_1,y_2|x_1,t)\cdot p_{X_1|T}(x_1|t) = \\ &= \sum_{x_1\in\mathcal{X}} \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_1-x_1 e^{+j\theta_1}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_1-2\pi k)^2}{\Theta_0}} d\theta_1 \cdot \\ &\quad \cdot \int_{-\pi}^{+\pi} \frac{1}{\pi N_0} e^{-\frac{\left|y_2-\varepsilon(x_1,t)e^{+j\theta_2}\right|^2}{N_0}} \cdot \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Theta_0}} e^{-\frac{1}{2}\frac{(\theta_2-2\pi k)^2}{\Theta_0}} d\theta_2 \cdot \frac{1}{|\mathcal{X}|}, \\ &\quad y_1,y_2 \in \mathbb{C}, t \in \mathcal{T}. \end{aligned} \tag{2.44}$$

The joint pdf of the first (previous) received symbol Y_1 and the second (current) received symbol Y_2 is obtained again thanks to Bayes multiplying (2.44) by (2.29) and summing over all $t \in \mathcal{T}$.

2.7. Summary

In this chapter I explained what I mean by the channels *seen by the SD demappers and FEC decoders* of long-haul NDE and DE coherent optical communications systems operating in both linear and non-linear regime and I presented a total of five models that represent them.

In these models the transmitted symbols are impaired by *residual* additive and/or phase white (wrapped) Gaussian noises; as such, the models (and everything that is based on them) depend on the variances of said noises. Some of these models include the DE to represent long-haul DE coherent optical communications systems: in these systems the DE decoding is performed by the SD demapper.

Table 2.1 reports schematically these models, their assumptions and their purposes.

Name (used for)	Formula	Assumption
NDE AWGN (proposed solutions, Chap. 6)	$Y = X + N$	X uniform, N AWGN
NDE AWGN/wPWGN (th. limits, Chap. 5)	$Y = Xe^{+j\Theta} + N$	Θ wPWGN, Θ indep. of X
ACRT NDE AWGN/PWGN (estimators, Chap. 3)	$Y' \approx j\Theta + N'$,	Θ small and PWGN, X known and $\lvert X\rvert = 1$, Θ indep. of $\Im(N')$
DE AWGN (proposed solutions, Chap. 6)	$Y_i = X_i + N_i, \quad i = 1, 2$ with $X_2 = \varepsilon(X_1, T)$	T uniform
DE AWGN/wPWGN (th. limits, Chap. 5)	$Y_i = X_i e^{+j\Theta_i} + N_i, \quad i = 1, 2$ with $X_2 = \varepsilon(X_1, T)$	

Table 2.1.: Models presented in this chapter and used in this work (see also figure 2.1).

3. Estimators of the SNR and of the SPNR

In this chapter I present two pairs of estimators of the additive white Gaussian noise (AWGN) variance N_0 and the phase white Gaussian noise (PWGN) variance Θ_0 based on two of the models presented in chapter 2, namely

- the non differentially encoded (NDE) AWGN/ wrapped phase white Gaussian noise (wPWGN) model presented in section 2.3 and
- the approximated, counter-rotated and translated (ACRT) NDE AWGN/ PWGN model presented in section 2.4.

The first pair of estimators proved problematic when applied to the received symbols resulting from the laboratory experiments representing 100G long-haul NDE and differentially encoded (DE) coherent optical communications systems of chapter 4, as I will describe. The need hence arose for the second pair of estimators, which do not suffer from the same problems.

3.1. Estimators based on the NDE AWGN/wPWGN model

One possible way to approach the estimation problem mentioned above is to notice that some statistical quantities of the NDE AWGN/ wPWGN model (2.8), which I report here for simplicity,

$$Y = Xe^{+j\Theta} + N,$$

are related to each other. Hence I can first estimate via Monte Carlo (MC) some of these quantities and then find the others using the relations previously found.

3.1.1. NDE AWGN/wPWGN model: Estimator of the variance N_0

With few basic manipulations it is easy to see that

$$\begin{aligned} E[|Y|^2] &= E[|X|^2] + E[Xe^{+j\Theta}N^*] + E[X^*e^{-j\Theta}N] + E[|N|^2] = \\ &= E[|X|^2] + E[X]E[e^{+j\Theta}]E[N^*] + E[X^*]E[e^{-j\Theta}]E[N] + E[|N|^2] = \\ &= E[|X|^2] + E[|N|^2], \end{aligned} \tag{3.1}$$

which means that

$$N_0 = E[|Y|^2] - E[|X|^2]. \tag{3.2}$$

As a consequence, I could estimate via MC the second moment of the received symbol Y and, assuming I know the second moment of the transmitted symbol X (which in my laboratory experiments I do), I could estimate the AWGN variance N_0 as follows:

$$\frac{1}{L}\sum_{i=1}^{L}|Y_i|^2 - E[|X|^2], \tag{3.3}$$

where L is the observation window length.

3.1.2. NDE AWGN/wPWGN model: Estimator of the variance Θ_0

A little bit more complicated is to find a similar correspondence between the PWGN variance Θ_0 and some moment of some function of the transmitted and received symbols X and Y.

Assuming that

$$\Theta_0 \leq \frac{\pi^2}{9} \approx 1.0966, \tag{3.4}$$

so that

$$-\pi = -3\sqrt{\Theta_0} \leq \Theta \leq +3\sqrt{\Theta_0} = +\pi \tag{3.5}$$

99.73% of the times, I can

- neglect the replicas of the Gaussian bell for $k \neq 0$ in (2.10) and only keep the central one. Note that by doing so I am *not* ignoring the case $\Theta = 2\pi, 4\pi, \ldots$ (since when dealing with directional random variables the integrals are anyway always over only one period), rather I am only neglecting the influence of the tails of said replicas ($k \neq 0$) on the central one ($k = 0$);

- extend the integral in the definition of the first moment of $e^{+j\Theta}$ over the whole real axis.

Note that, in general, this assumption is different from the one made in section 2.4 in order to be able to approximate the wPWGN Θ with its Taylor expansion centred at zero. In this work I will always normalise the constellations to have $E_s = 1$, hence this assumption corresponds to a ratio between the energy per transmitted symbol E_s and the PWGN variance Θ_0 larger than roughly -0.4 dB (see section 3.3 for the proper definition).

The first moment of $e^{+j\Theta}$ thus resembles a Fourier transform and I obtain the following result:

$$E[e^{+j\Theta}] \approx \int_{-\pi}^{+\pi} e^{+j\theta} p_\Theta(\theta) d\theta \approx \int_{-\infty}^{+\infty} e^{+j\theta} p_\Theta(\theta) d\theta = \\ = \int_{-\infty}^{+\infty} e^{-j2\pi(-\frac{1}{2\pi})\tau} p_\Theta(\tau) d\tau = P_\Theta(-\frac{1}{2\pi}), \tag{3.6}$$

where $P_\Theta(f)$ is the Fourier transform of $p_\Theta(\tau)$, and the change of variable $\tau = \theta$ has been done to emphasise that (3.6) can be seen as a Fourier transform calculated in $f_0 = -1/(2\pi)$. I used the letter τ because in this work the letter t represents a transmitted transition, see section 2.2. The Fourier transformation has the following properties:

$$e^{-\pi\tau^2} \overset{\mathcal{F}}{\mapsto} e^{-\pi f^2} \text{ and} \tag{3.7}$$

$$af(b\tau) \overset{\mathcal{F}}{\mapsto} a\frac{1}{|b|}F(\frac{f}{b}), \tag{3.8}$$

hence, choosing $f(\tau) = e^{-\pi\tau^2}$ and $a = b = 1/\sqrt{2\pi\Theta_0}$, I obtain

$$P_\Theta(f) = e^{-\pi(\sqrt{2\pi\Theta_0}f)^2} = e^{-2\pi^2\Theta_0 f^2} \tag{3.9}$$

and, finally,

$$E[e^{+j\Theta}] \approx P_\Theta(-\frac{1}{2\pi}) = e^{-2\pi^2\Theta_0(-\frac{1}{2\pi})^2} = e^{-\frac{1}{2}\Theta_0}. \tag{3.10}$$

Now I only need to note that

$$E[\frac{Y}{X}] = E[e^{+j\Theta} + N] = E[e^{+j\Theta}] + E[N] \approx e^{-\frac{1}{2}\Theta_0}, \tag{3.11}$$

which means that

$$\Theta_0 \approx -2\ln(E[\frac{Y}{X}]). \tag{3.12}$$

As a consequence, assuming I know the transmitted symbol X (which in my laboratory experiments I do), I could estimate via MC the first moment of the counter-rotated received symbol Y/X and I could estimate the PWGN variance Θ_0 as follows:

$$-2\ln(\frac{1}{L}\sum_{i=1}^{L}\frac{Y_i}{X_i}). \tag{3.13}$$

The set of the counter-rotated received symbols $Y_i/X_i, i = 1, \ldots, L$, i.e. the counter-rotated received constellation, must not be confused with the counter-rotated and translated (CRT) received constellation of section 2.4. It lies on the point $1+j0$, is symmetric with respect to the real axis $y = 0$ but is *not* symmetric with respect to the axis $x =$

1 because of the wPWGN: in layman's terms, this constellation can be said to have a "bean-like" shape, whereas the CRT constellation of section 2.4 was an ellipse.

For observation windows of *infinite* length, $L \to +\infty$, the argument of the $\ln(\cdot)$ function is a real number between 0 and 1, decreasing as the PWGN variance Θ_0 increases:

$$\frac{1}{L}\sum_{i=1}^{L}\frac{Y_i}{X_i} \to \begin{cases} 1 & \text{when } \Theta_0 \to 0 \text{ and} \\ 0 & \text{when } \Theta_0 \to +\infty. \end{cases} \tag{3.14}$$

When $\Theta_0 \to 0$ there is no PWGN: the counter-rotated received constellation becomes symmetric with respect to the axis $x = 1$, it can be said to lose its "bean-like" shape, the $\ln(\cdot)$ function returns 0 as does the estimator of the PWGN variance Θ_0. On the other hand, when $\Theta_0 \to +\infty$ there is infinite PWGN: the counter-rotated received constellation becomes symmetric with respect to the axis $x = 0$, it can be said to take on a "doughnut-like" shape, the $\ln(\cdot)$ function returns $-\infty$ and the estimator of the PWGN variance Θ_0 returns $+\infty$. In all cases, the "thickness" of its "bean-like" or "doughnut-like" shape depends on the AWGN variance N_0.

3.1.3. NDE AWGN/wPWGN model: Disadvantages

The estimator (3.13) of the PWGN variance Θ_0 has various problems, related to the presence of the $\ln(\cdot)$ function. In fact, the use of observation windows of *finite* length L means that:

- the argument of the $\ln(\cdot)$ function could still have a small complex part, so that precautions should be taken to avoid feeding the $\ln(\cdot)$ function with a complex number;
- the argument of the $\ln(\cdot)$ function could lie outside the interval $(0, 1)$, which would in turn lead to either
  - meaningless results (if the argument of the $\ln(\cdot)$ function is larger than 1, the estimator (3.13) would return a negative PWGN variance Θ_0!) or
  - undefined operations (if the argument of the $\ln(\cdot)$ function is smaller than 0);
- small deviations of the argument of the $\ln(\cdot)$ function from its "true" value (the one it would have for $L \to +\infty$) lead to large variations of its result and, hence, to unreliable estimates of the PWGN variance Θ_0 (recall that the derivative of the $\ln(\cdot)$ function in the interval $(0, 1)$ assumes its largest values);
- small deviations of the argument of the $\ln(\cdot)$ function from the value 1, caused by the AWGN, are wrongly interpreted as caused by the wPWGN.

The estimator (3.13) is, in other words, *unstable* or, better said, might work in theory (with *infinite* observation window length L), but does not in practice (with *finite* observation window length L) and unfortunately I want to apply it to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems.

Moreover, it is hard to judge the performance of both estimators (3.3) and (3.13) (e.g., whether they are *efficient*) because it is virtually impossible to calculate in closed form the Cramér-Rao lower bound (CRLB) for the model (2.8): it would mean calculating the expected value of the first (or second, depending on the definition) derivative of the natural logarithm of the joint probability density function (pdf) of the received symbols $Y_i, i = 1, \ldots, L$ but, as said in section 2.3.2, already the conditional pdf of the received symbol Y conditioned on the transmitted symbol X (2.13) is quite cumbersome ad cannot be solved analytically.

3.1.4. NDE AWGN/wPWGN model: Dependency on the AGC

The estimator (3.3) of the AWGN variance N_0 depends heavily on the performance of the automatic gain control (AGC) stage(s) in the digital signal processing (DSP). In fact, a non-ideal AGC could return each transmitted symbol impaired by AWGN and wPWGN scaled in principle by a different multiplicative factor. These factors would then affect the estimation of the second moment of the received symbol Y in the estimator (3.3) of the AWGN variance N_0 and lead to an over- or underestimated AWGN variance.

The same holds true for the estimator (3.13) of the PWGN variance Θ_0: the multiplicative factors would affect the estimation of the first moment of the counter-rotated received symbol Y/X and increase some of the problems listed before.

A possible solution could be to take into account also these factors and model them as real random variables, most likely *with* memory. However, as explained in chapter 2 (see also figure 2.1), the models lay the foundations for the rest of this work, so a small increase in the complexity at the beginning would translate into a large increase in the complexity at the end, making the treatise most likely unmanageable: the models, the estimators, the theoretical limits of the constellations and possibly the soft-decision (SD) demappers would depend in principle on a series of additional parameters (depending on the statistical description of the multiplicative factors introduced by the non-ideal AGC), while the pdfs as well as the estimators, their variances and their CRLBs would be most likely impossible to obtain in closed forms.

3.2. Estimators based on the ACRT NDE AWGN/PWGN model

Luckily, a second pair of estimators can be developed, based this time on the ACRT NDE AWGN/ PWGN model (2.18), which I also report here for simplicity,

$$Y' = \frac{Y}{X} - 1 = j\Theta + N'.$$

This time I can derive the estimators in the classical Estimation Theory way, by first calculating the CRLB and thus obtaining, automatically, the minimum-variance unbiased (MVU) estimators.

3.2.1. ACRT NDE AWGN/PWGN model: Estimators of the variances N_0 and Θ_0

The first step to calculate the CRLB is to take the $\ln(\cdot)$ of the joint pdf of the CRT received symbols $Y_i', i = 1, \ldots, L$ (2.26) calculated in section 2.4.2:

$$-L\ln(2\pi) - \frac{L}{2}\ln(\frac{1}{2}N_0) - \frac{L}{2}\ln(\Theta_0 + \frac{1}{2}N_0) - \frac{1}{2}\sum_{i=1}^{L}(\frac{\Re(y_i)^2}{\frac{1}{2}N_0} + \frac{\Im(y_i)^2}{\Theta_0 + \frac{1}{2}N_0}). \tag{3.15}$$

Then, since I want to estimate two parameters, I need to calculate the derivatives of (3.15) with respect to N_0 and Θ_0:

$$-\frac{L}{2}\frac{\frac{1}{2}}{\frac{1}{2}N_0} - \frac{L}{2}\frac{\frac{1}{2}}{\Theta_0 + \frac{1}{2}N_0} + \frac{1}{2}\sum_{i=1}^{L}(\frac{\frac{1}{2}\Re(y_i)^2}{(\frac{1}{2}N_0)^2} + \frac{\frac{1}{2}\Im(y_i)^2}{(\Theta_0 + \frac{1}{2}N_0)^2}), \tag{3.16}$$

$$-\frac{L}{2}\frac{1}{\Theta_0 + \frac{1}{2}N_0} + \frac{1}{2}\sum_{i=1}^{L}\frac{\Im(y_i)^2}{(\Theta_0 + \frac{1}{2}N_0)^2}. \tag{3.17}$$

I can rewrite (3.16) and (3.17) organising them in a 2x1 matrix as follows [35, Theorem 3.2]:

$$\begin{bmatrix} \frac{1}{2}\frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2}(\frac{1}{L}\sum_{i=1}^{L}\Im(y_i)^2 - \frac{1}{2}N_0 - \Theta_0) + \frac{1}{4}\frac{\frac{1}{2}L}{(\frac{1}{2}N_0)^2}(\frac{2}{L}\sum_{i=1}^{L}\Re(y_i)^2 - N_0) \\ \frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2}(\frac{1}{L}\sum_{i=1}^{L}\Im(y_i)^2 - \frac{1}{2}N_0 - \Theta_0) \end{bmatrix}. \tag{3.18}$$

This might be rewritten as:

$$\begin{bmatrix} \frac{1}{4}\frac{\frac{1}{2}L}{(\frac{1}{2}N_0)^2} & \frac{1}{2}\frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} \\ 0 & \frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{L}\sum_{i=1}^{L}\Re(y_i)^2 - N_0 \\ \frac{1}{L}\sum_{i=1}^{L}\Im(y_i)^2 - \frac{1}{2}N_0 - \Theta_0 \end{bmatrix}. \tag{3.19}$$

Hence I managed to write the 2x1 matrix (3.18) having as elements the derivative, with respect to N_0 and Θ_0, of the $\ln(\cdot)$ of the joint pdf of the CRT received symbols $Y_i', i = 1, \ldots, L$ in the form [35, Theorem 3.2]

$$\frac{\partial \ln(p_{Y_1',Y_2',\ldots,Y_L'}(y_1, y_2, \ldots, y_L))}{\partial [N_0, \Theta_0]^T} = \boldsymbol{I}([N_0, \Theta_0]^T)(\boldsymbol{g}(Y_1', Y_2', \ldots, Y_L') - [N_0, \Theta_0]^T), \tag{3.20}$$

with

$$\boldsymbol{I}([N_0, \Theta_0]^T) = \begin{bmatrix} \frac{1}{4}\frac{\frac{1}{2}L}{(\frac{1}{2}N_0)^2} & \frac{1}{2}\frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} \\ 0 & \frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} \end{bmatrix} \text{ and} \tag{3.21}$$

$$\boldsymbol{g}(Y_1', Y_2', \ldots, Y_L') = \begin{bmatrix} \frac{2}{L}\sum_{i=1}^{L}\Re(y_i)^2 \\ \frac{1}{L}\sum_{i=1}^{L}\Im(y_i)^2 - \frac{1}{2}N_0 \end{bmatrix}, \tag{3.22}$$

while the inverse of (3.21) is

$$\boldsymbol{I}^{-1}([N_0, \Theta_0]^T) = \frac{1}{\frac{1}{4}\frac{\frac{1}{2}L}{(\frac{1}{2}N_0)^2}\frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2}} \begin{bmatrix} \frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} & -\frac{1}{2}\frac{\frac{1}{2}L}{(\Theta_0+\frac{1}{2}N_0)^2} \\ 0 & \frac{1}{4}\frac{\frac{1}{2}L}{(\frac{1}{2}N_0)^2} \end{bmatrix}. \tag{3.23}$$

This guarantees that

- the estimators [35, Theorem 3.2]

$$\hat{N}_0(Y'_1, Y'_2, \ldots, Y'_L, L) = [\boldsymbol{g}(Y'_1, Y'_2, \ldots, Y'_L)]_{1,1} = \frac{2}{L}\sum_{i=1}^{L} \Re(Y'_i)^2 \text{ and} \tag{3.24}$$

$$\hat{\Theta}_0(Y'_1, Y'_2, \ldots, Y'_L, N_0, L) = [\boldsymbol{g}(Y'_1, Y'_2, \ldots, Y'_L)]_{2,1} = \frac{1}{L}\sum_{i=1}^{L} \Im(Y'_i)^2 - \frac{1}{2}N_0 \tag{3.25}$$

  attain the respective bounds [35, Theorem 3.2] (i.e., they are *efficient*) and are therefore MVU estimators;

- their variances are [35, Theorem 3.2]

$$\sigma^2_{\hat{N}_0}(N_0, L) = [\boldsymbol{I}^{-1}([N_0, \Theta_0]^T)]_{1,1} = \frac{2}{L}(N_0)^2 \text{ and} \tag{3.26}$$

$$\sigma^2_{\hat{\Theta}_0}(N_0, \Theta_0, L) = [\boldsymbol{I}^{-1}([N_0, \Theta_0]^T)]_{2,2} = \frac{2}{L}(\Theta_0 + \frac{1}{2}N_0)^2; \tag{3.27}$$

- their CRLBs are also given by (3.26) and (3.27).

This result agrees with the intuition. Recalling what has been said in section 2.4 about the shape of the CRT received constellation (it is an ellipse, whose minor axis depends on the AWGN variance N_0 and whose major axis depends on the latter and on the PWGN variance Θ_0), the most natural way to estimate the variances N_0 and Θ_0 would probably consists of two steps:

1. to estimate the *second raw moment* of the real part of the CRT received symbols $Y'_i, i = 1, \ldots, L$, which contains *only* the real part of the counter-rotated AWGN N' (since the PWGN Θ is not present). This estimate equals half of the estimated AWGN variance and must therefore multiplied by two;

2. to estimate the *second raw moment* of the imaginary part of the CRT received symbols $Y'_i, i = 1, \ldots, L$, which contains the imaginary part of the counter-rotated AWGN N' *and* the PWGN Θ. This estimate equals the estimated PWGN variance plus half of the estimated AWGN variance, which was obtained at the previous step and has then to be subtracted.

The variances of the estimators $\hat{N}_0$ and $\hat{\Theta}_0$, as well as their CRLB, depend

- linearly on the inverse of the observation window length L: the more the observations, the more precise the estimations;
- quadratically on the variances N_0 and Θ_0: the lower the noises, the more precise the estimations. In particular,
  - the variance of the estimator $\hat{N}_0$ depends only on the AWGN variance N_0, whereas
  - the variance of the estimator $\hat{\Theta}_0$ depends on both the AWGN variance N_0 and the PWGN variance Θ_0.

Sometimes the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems were not exactly symmetrically distributed around the origin; this was especially the case when large amounts of noises were present, since the DSP struggled to converge. I hence applied a slightly modified version of the estimators $\hat{N}_0$ and $\hat{\Theta}_0$ (3.24) and (3.25), which estimates the *variance* (instead of the second raw moment) of the real and of the imaginary part of the CRT received symbols $Y'_i, i = 1, \ldots, L$ and proved even more robust.

3.2.2. ACRT NDE AWGN/PWGN model: Advantages

Not having the $\ln(\cdot)$ function (or any non-linear function, for that matter), the estimators $\hat{N}_0$ and $\hat{\Theta}_0$ (3.24) and (3.25) do not present the problems due to observation windows of finite length L that afflict the estimator of the PWGN variance Θ_0 derived in section 3.1.2.

The estimators $\hat{N}_0$ and $\hat{\Theta}_0$ are, in other words, *stable* or, better said, will work in practice when I will apply them to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems.

Finally, they are efficient, MVU estimators and I was able to calculate their variances and their CRLB in closed form, so I will be able to predict their performance.

3.2.3. ACRT NDE AWGN/PWGN model: Dependency on the AGC

Although less evident, also the estimators $\hat{N}_0$ and $\hat{\Theta}_0$ (3.24) and (3.25) depend on the performance of the AGC stage(s) in the DSP. Even though the CRT received constellation can be "forced" to lie on the origin by substracting its mean value, the multiplicative factors mentioned in section 3.1.4 would still affect the estimation of the second moments of the real and of the imaginary parts of the CRT received symbol Y'.

As explained in section 3.1.4, modelling also these factors would be an elegant solution in theory, but not a viable one in practice. The DSP used in the laboratory experiments presents two *blind* AGCs, a very basic one at the beginning of the cascade and a more advanced one at the end of it. Being blind, they (especially the second) exhibit better performance in presence of smaller amounts of noises, so I assumed that an ideal AGC

is present, at least in these regions, in order to keep the treatise manageable and to propose feasible solutions.

In chapter 4 I will prove this assumption valid for the DSP used in the laboratory experiments.

3.3. SNR, SPNR, OSNR, E_b/N_0, NCG

Rather than on the *absolute* values of the (average) energy per transmitted symbol E_s and the AWGN variance N_0, the performance in terms of bit error rate (BER) of communications systems that can be described with the NDE AWGN model depends on their *ratio*. This quantity, called signal-to-(*additive*-)noise ratio (SNR), is given in dB as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{E_s}{N_0}\right)\ [\mathrm{dB}]. \tag{3.28}$$

Similarly, since with the ACRT NDE AWGN/ PWGN model the PWGN becomes a second additive noise (which, moreover, impairs only the imaginary part of the CRT received symbols), it is helpful to introduce a similar quantity for the ratio between the energy per transmitted symbol E_s and the PWGN variance Θ_0. I call this quantity the signal-to-*phase*-noise ratio (SPNR), which is given in dB as

$$\mathrm{SPNR} = 10\log_{10}\left(\frac{E_s}{\Theta_0}\right)\ [\mathrm{dB}]. \tag{3.29}$$

As said in section 3.1.2, in this work I will always normalise the constellations to have $E_s = 1$. As a consequence, the estimators presented in this chapter, which actually estimate the variances N_0 and Θ_0, can also be said to estimate the two ratios (3.28) and (3.29). In the other chapters I will hence refer to them as "estimators of the SNR and of the SPNR".

In the optical community a third quantity, called *optical* signal-to-noise ratio (OSNR), is customarily used in place of the SNR (3.28), which considers a fixed reference bandwidth for the amplified spontaneous emission (ASE) noise produced by the optical amplifiers. This quantity is given in dB as

$$\mathrm{OSNR} = \mathrm{SNR} + 10\log_{10}\left(\frac{B_s}{B_{\mathrm{ref}}}\right)\ [\mathrm{dB}], \tag{3.30}$$

where B_s is the bandwidth occupied by the transmitted signal and B_{ref} is the reference bandwidth (usually 12.5 GHz, corresponding to 0.1 nm at 1550 nm). Note that hence the performance of an optical communications system in terms of OSNR depends on the occupied bandwidth B_s.

The symbol rate f_s is given by

$$f_s = \frac{R_b}{N_{\mathrm{pol}} m}, \tag{3.31}$$

where R_b is the nominal bit rate, N_{pol} is the number of polarisations used and m is the number of transmitted bits carried by a transmitted symbol X or a transmitted transition T, see (2.3) and (2.30).

In the coding community the (average) energy per transmitted *information bit* E_b is customarily used in place of the energy per transmitted *symbol* E_s, because this quantity allows to compare the performance of *coded* communications system with different code rates R. The two quantities are related as follows:

$$E_b = \frac{E_s}{Rm}. \tag{3.32}$$

The ratio between the energy per transmitted information bit E_b and the AWGN variance N_0 is hence given in dB as

$$10\log_{10}\left(\frac{E_b}{N_0}\right) = \text{SNR} - 10\log_{10}\left(Rm\right) \text{ [dB]}. \tag{3.33}$$

Another quantity often used in the coding community is the net coding gain (NCG), given in dB by the difference between the information bit-to-(*additive*-)noise ratio in dB needed by an uncoded system and that needed by a coded one to achieve a given targeted BER:

$$\begin{aligned}\text{NCG} &= 10\log_{10}\left(\frac{E_b}{N_0}\right)_{\text{unc.}} - 10\log_{10}\left(\frac{E_b}{N_0}\right)_{\text{cod.}} = \\ &= \text{SNR}_{\text{unc.}} - 10\log_{10}\left(m_{\text{unc.}}\right) - \text{SNR}_{\text{cod.}} + 10\log_{10}\left(Rm_{\text{cod.}}\right) \text{ [dB]}.\end{aligned} \tag{3.34}$$

The NCG represents the gain in dB provided by a coded system over an uncoded one. Note that, in principle, one could have $m_{\text{unc.}} \neq m_{\text{cod.}}$. If $m_{\text{unc.}} = m_{\text{cod.}}$, (3.34) reduces to:

$$\text{NCG} = \text{SNR}_{\text{unc.}} - \text{SNR}_{\text{cod.}} + 10\log_{10}\left(R\right) \text{ [dB]}. \tag{3.35}$$

Since this work deals with coding for optical communications systems, it is helpful to relate directly the OSNR to the information bit-to-(*additive*-)noise ratio in dB:

$$\text{OSNR} = 10\log_{10}\left(\frac{E_b}{N_0}\right) + 10\log_{10}\left(Rm\right) + 10\log_{10}\left(\frac{B_s}{B_{\text{ref}}}\right) \text{ [dB]}. \tag{3.36}$$

3.4. Performance of the estimators based on the ACRT NDE AWGN/PWGN model

Let me rewrite the formulas (3.26) and (3.27) for the variances of the estimators $\hat{N}_0$ and $\hat{\Theta}_0$, as well as their CRLB, as a function of the SNR and the SPNR (as well as of the observation window length L), recalling that in this work I will always have $E_s = 1$:

$$\sigma^2_{\hat{N}_0}(\text{SNR}, L) = \frac{2}{L}\left(\frac{1}{\text{SNR}}\right)^2 \text{ and} \tag{3.37}$$

$$\sigma^2_{\hat{\Theta}_0}(\text{SNR}, \text{SPNR}, L) = \frac{2}{L}\left(\frac{1}{\text{SPNR}} + \frac{1}{2}\frac{1}{\text{SNR}}\right)^2. \tag{3.38}$$

Fixed variable	Value
SNR	9.5 dB and 11.0 dB
SPNR	18.0 dB, 21.0 dB and 30.0 dB
L	10^6

Table 3.1.: Variables fixed in figures 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7.

Here the SNR and the SPNR are linear, not in dB.

Since the performance of the estimator $\hat{\Theta}_0$ depends on three variables, it is not possible to show them in one figure; I hence decided that in section 3.4.3 I will show the results in six figures, each time fixing one of the three variables to certain values and parametrizing the curves with another one. When a variable is fixed in one figure, I will highlight it in the others in which it is not (with a thicker curve when it parametrises a curve or with a vertical black dashed line when it is on the x-axis).

Things are easier for the performance of the estimator $\hat{N}_0$, since its variance depends "only" on two variables; in section 3.4.2 I will show it in one figure, parametrizing the curves with one of the two variables.

3.4.1. Values of interest

The decision on which value to fix each variable is motivated as follows:

1. I chose to fix the SNR to two values, 9.5 dB and 11 dB, because they are in the middle of the ranges of the SNR I will estimate in chapter 4 applying the estimator $\hat{N}_0$ to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems. Similarly,

2. I chose to fix the SPNR to three values, 18 dB, 21 dB and 30 dB, because they are in the middle of the ranges of the SPNR I will estimate in chapter 4 applying the estimator $\hat{\Theta}_0$ to the laboratory experiments. Finally,

3. I chose to fix the observation window length L to 10^6, because it is the length of the sequence used in the laboratory experiments.

Note that these values must be considered as *lower limits* of the operating ranges, for in the laboratory experiments I operated the systems in highly non-linear regime in order to stimulate as many non-linearities as possible.

Table 3.1 summarizes which variable I fixed to which value in figures 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7.

3.4.2. ACRT NDE AWGN/PWGN model: Performance of the estimator $\hat{N}_0$

Figure 3.1 shows the variance of the estimator $\hat{N}_0$ as a function of the SNR and parametrised by various values of the observation window length L.

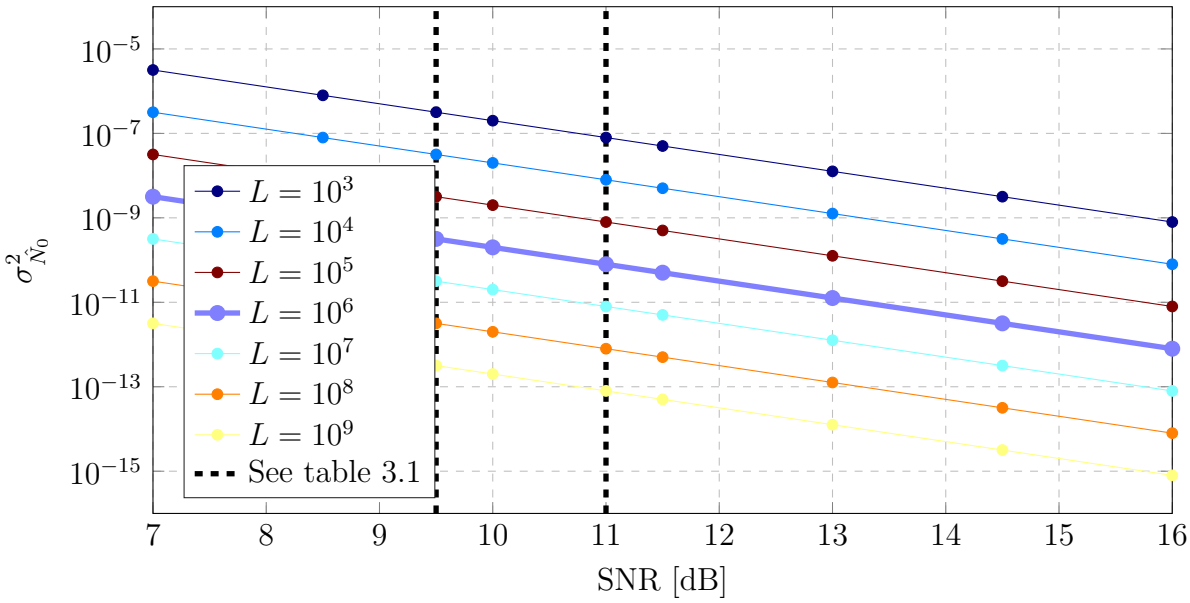

Figure 3.1.: Variance (and CRLB) of the estimator $\hat{N}_0$ as a function of the SNR and parametrised by various values of the observation window length L over the ACRT NDE AWGN/PWGN model.

In the range I am interested in the variance of the estimator $\hat{N}_0$ is in the range 8 $\cdot 10^{-11}$–3 $\cdot 10^{-10}$, whereas N_0 (the value to be estimated) is in the range 8 $\cdot 10^{-2}$–1 $\cdot 10^{-1}$.

3.4.3. ACRT NDE AWGN/PWGN model: Performance of the estimator $\hat{\Theta}_0$

Figures 3.2 and 3.3 show the variance of the estimator $\hat{\Theta}_0$ as a function of the observation window length L, for two fixed value of the SNR = 9.5 dB and 11.0 dB and parametrised by various values of the SPNR.

Figures 3.4, 3.5 and 3.6 show the variance of the estimator $\hat{\Theta}_0$ as a function of the SNR, for three fixed values of the SPNR = 18 dB, 21 dB and 30 dB and parametrised by various values of the observation window length L.

Figure 3.7 shows the variance of the estimator $\hat{\Theta}_0$ as a function of the SPNR, for a fixed value of the observation window length $L = 10^6$ and parametrised by various values of the SNR.

The flattening of the curves in the right parts of figures 3.4, 3.5 and 3.6 shows that increasing the SNR is beneficial only up to a certain point, because even when the AWGN becomes negligible there might still be PWGN. Similarly, the flattening of the curves in the right part of figure 3.7 shows the same with the roles of SNR and SPNR swapped: even when the PWGN becomes negligible there might still be AWGN.

Moreover, in the range I am interested in the variance of the estimator $\hat{\Theta}_0$ is in the range 3 $\cdot 10^{-11}$–1 $\cdot 10^{-9}$, whereas Θ_0 (the value to be estimated) is in the range 1 $\cdot 10^{-3}$–2 $\cdot 10^{-2}$.

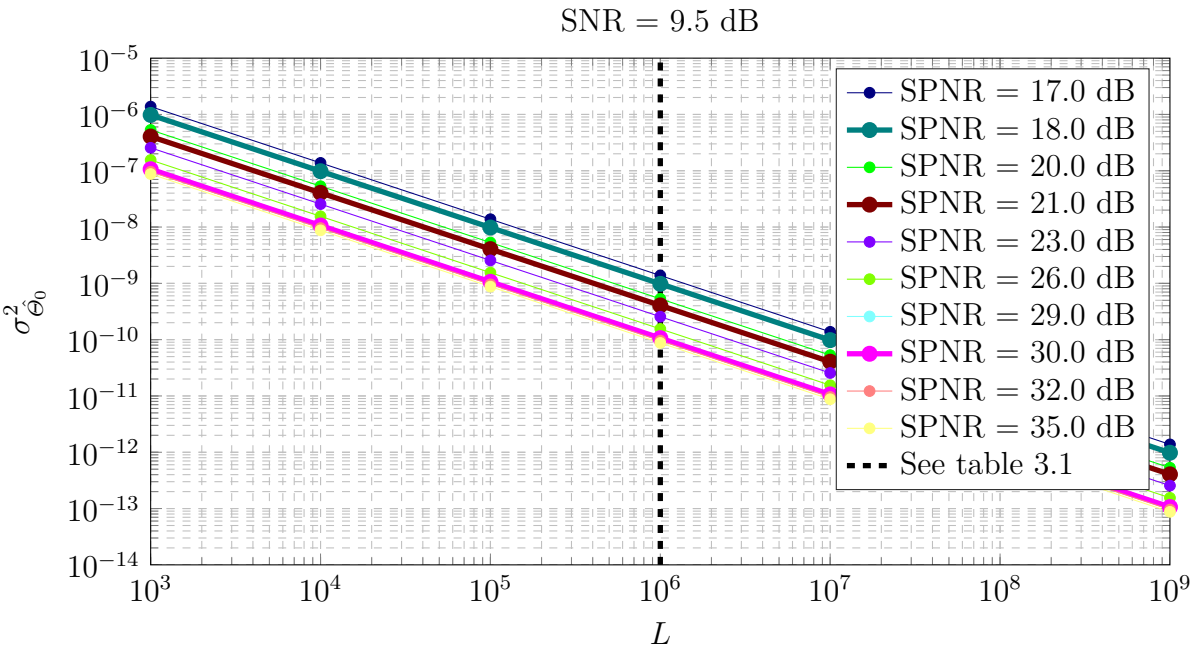

Figure 3.2.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the observation window length L, for a SNR = 9.5 dB and parametrised by various values of the SPNR over the ACRT NDE AWGN/PWGN model.

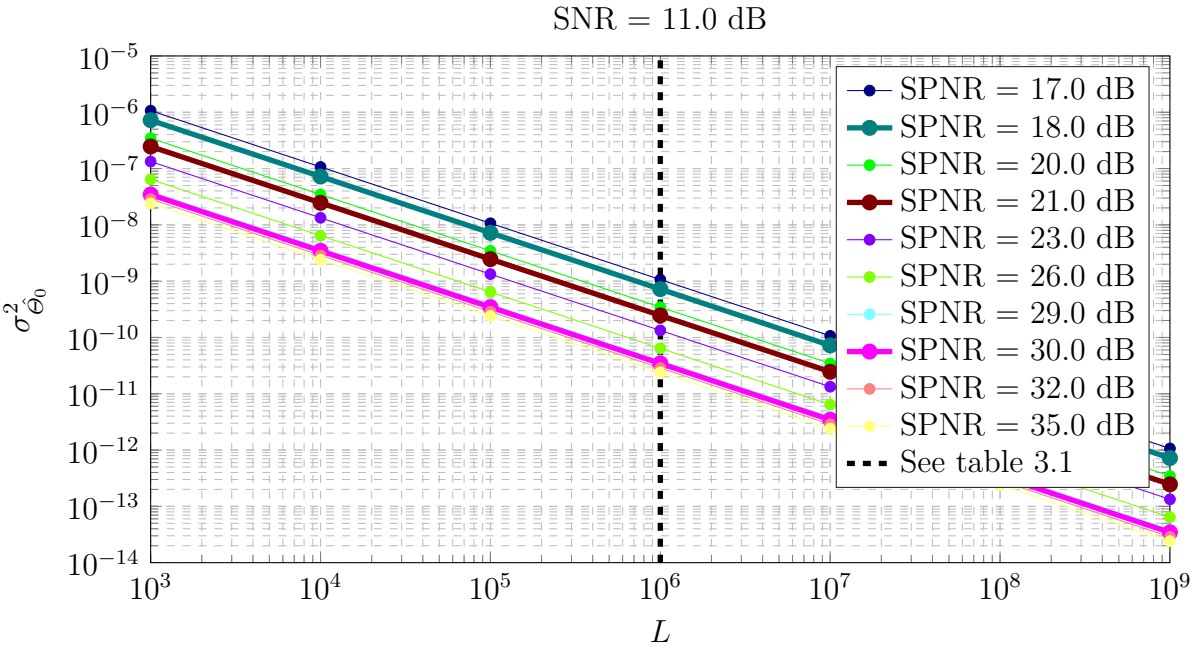

Figure 3.3.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the observation window length L, for a SNR = 11.0 dB and parametrised by various values of the SPNR over the ACRT NDE AWGN/PWGN model.

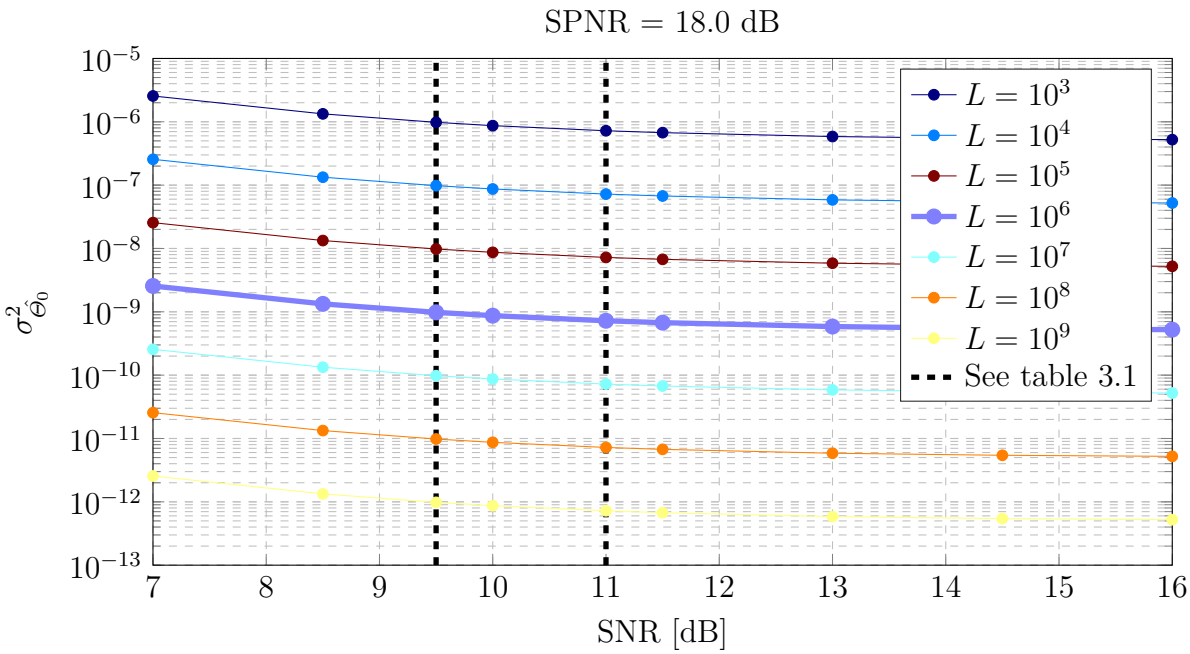

Figure 3.4.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the SNR, for a SPNR = 18.0 dB and parametrised by various values of the observation window length L over the ACRT NDE AWGN/PWGN model.

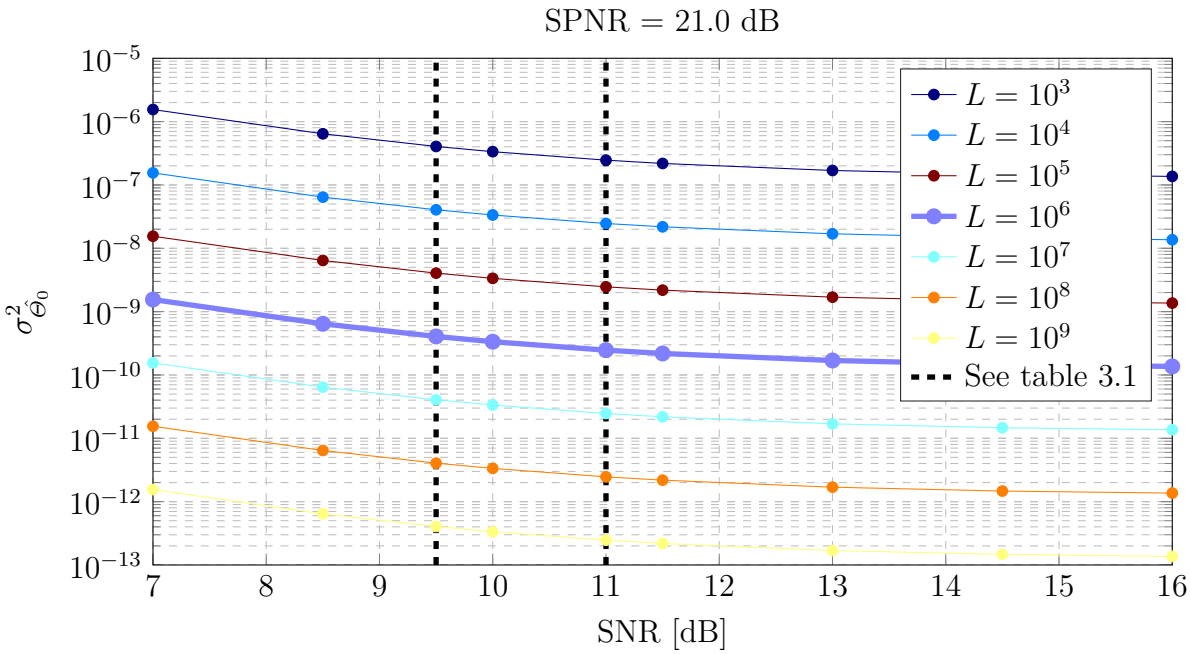

Figure 3.5.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the SNR, for a SPNR = 21.0 dB and parametrised by various values of the observation window length L over the ACRT NDE AWGN/PWGN model.

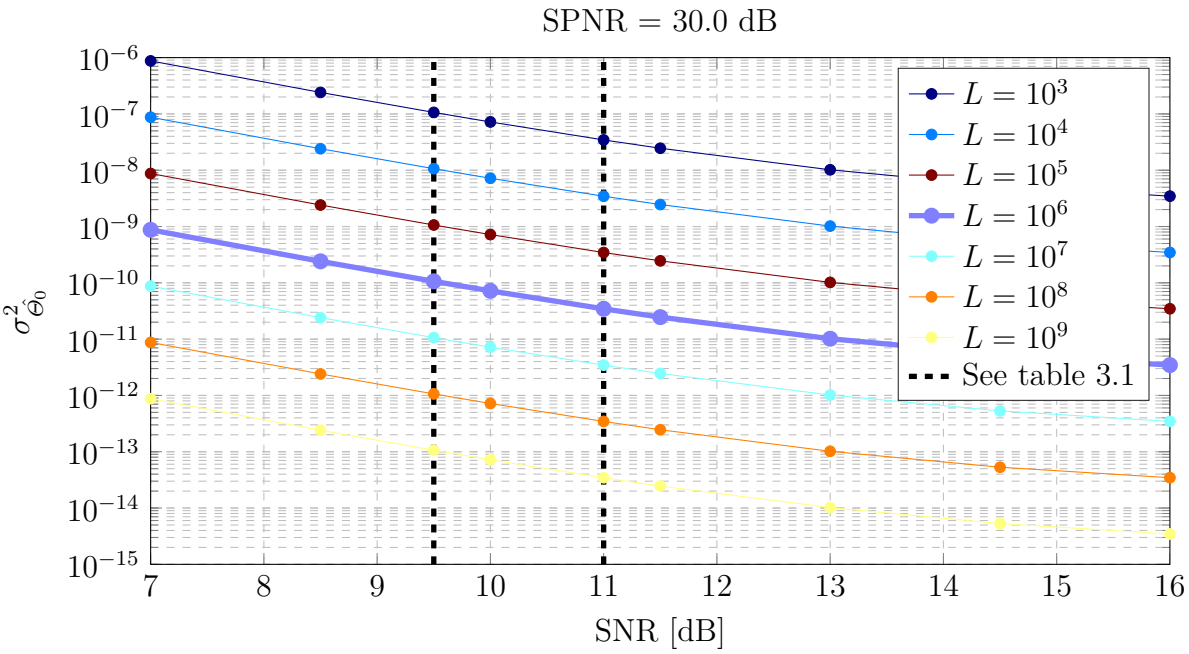

Figure 3.6.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the SNR, for a SPNR = 30.0 dB and parametrised by various values of the observation window length L over the ACRT NDE AWGN/PWGN model.

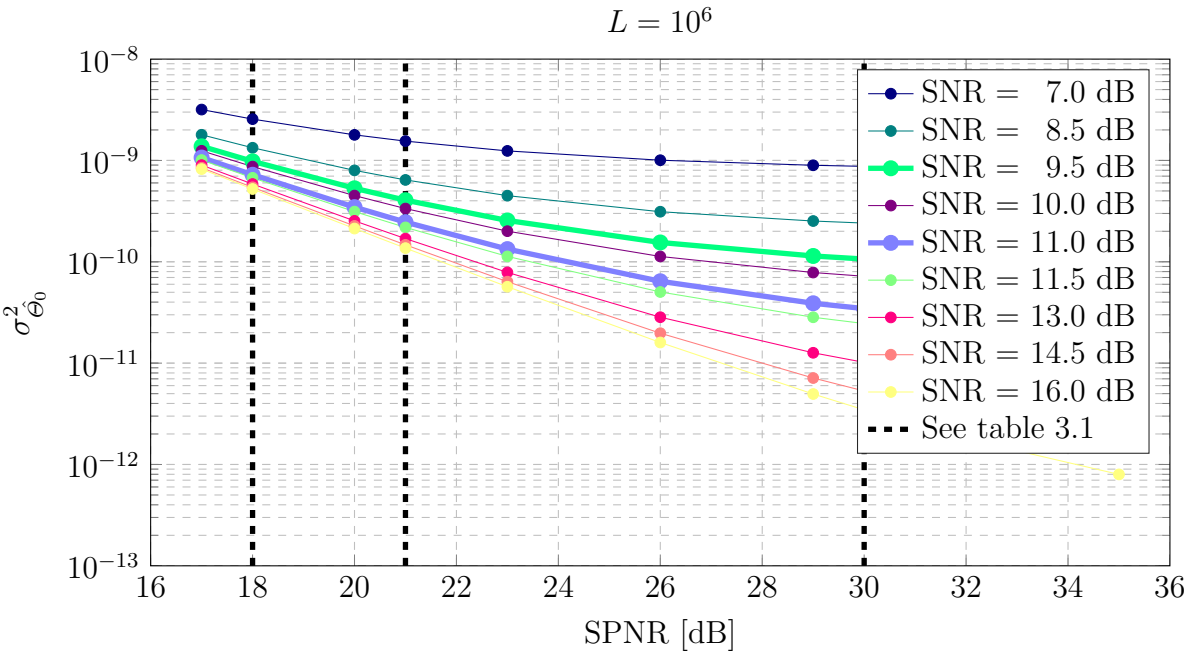

Figure 3.7.: Variance (and CRLB) of the estimator $\hat{\Theta}_0$ as a function of the SPNR, for an observation window length $L = 10^6$ and parametrised by various values of the SNR over the ACRT NDE AWGN/PWGN model.

Model	Formulas	Variance/CRLB
NDE AWGN/wPWGN	$\hat{N}_0 = \frac{1}{L}\sum_{i=1}^{L}\lvert Y_i\rvert^2 - E[\lvert X\rvert^2]$ $\hat{\Theta}_0 = -2\ln(\frac{1}{L}\sum_{i=1}^{L}\frac{Y_i}{X_i})$	unknown
ACRT NDE AWGN/PWGN	$\hat{N}_0 = \frac{2}{L}\sum_{i=1}^{L}\Re(Y_i')^2$ $\hat{\Theta}_0 = \frac{1}{L}\sum_{i=1}^{L}\Im(Y_i')^2 - \frac{1}{2}N_0$	$\sigma^2_{\hat{N}_0} = \frac{2}{L}(\frac{1}{\mathrm{SNR}})^2$ $\sigma^2_{\hat{\Theta}_0} = \frac{2}{L}(\frac{1}{\mathrm{SPNR}} + \frac{1}{2}\frac{1}{\mathrm{SNR}})^2$

Table 3.2.: Estimators presented in this chapter.

3.5. Summary

In this chapter I introduced two pairs of estimators of the two parameters of the models seen in chapter 2, the AWGN variance N_0 and the PWGN variance Θ_0.

The first pair of estimators is based on the NDE AWGN/ wPWGN model and estimates via MC certain quantities which are related to the variances N_0 and Θ_0. Unfortunately, one of them makes use of the non-linear $\ln(\cdot)$ function and is hence particularly sensitive to the imperfections typical of laboratory experiments (i.e., a finite observation window length). Moreover, the performance of these estimators are hard to judge, since a closed form of their variances and their CRLBs could be found.

Hence I developed a second pair of estimators $\hat{N}_0$ and $\hat{\Theta}_0$, this time based on the ACRT NDE AWGN/ PWGN model, which estimate with two aritmetic means directly the variances N_0 and Θ_0. These two new estimators do not make use of any non-linear function and are as such much more robust when applied to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems. Moreover, they are *by construction* efficient, MVU estimators, which means that their variances attain (actually are equal to) their respective CRLBs, for which a closed form could be found.

The performance of the estimators $\hat{N}_0$ and $\hat{\Theta}_0$ depends on the SNR, the SPNR as well as the observation window length. Anticipating some results from chapter 4 regarding the lower limits of the operating ranges of typical wavelength division multiplexing (WDM) systems I was able to prove that the estimators $\hat{N}_0$ and $\hat{\Theta}_0$ have variances that are six to ten orders of magnitude smaller than the values they have to estimate and are, as such, very reliable.

Table 3.2 summarizes the estimators, the models they are based on and their performance.

4. Laboratory Experiments

In this chapter I present some experiments I performed in the laboratories of Nokia Siemens Networks (NSN), nowadays Infinera. I realised a 100G long-haul coherent optical communications system as central channel and surrounded it with either

- eight 100G long-haul coherent optical communications systems or
- eight 10G long-haul incoherent optical communications systems

as neighbouring channels. I operated the central channel and/or the neighbouring channels in highly (actually, as high as possible) non-linear regime, stimulating

- (homogeneous) intra-channel non-linearities (i.e., arising from the propagation of the central channel) and/or
- homogeneous and/or heterogeneous inter-channel non-linearities (i.e., arising from the propagation of the neighbouring channels, modulated as and/or differently from the central one).

Obviously, intra-channel non-linearities can only be homogeneous. The purpose is twofold:

1. to validate the models introduced in chapter 2 and
2. to know in which range the signal-to-(*additive*-)noise ratio (SNR) and most importantly the signal-to-*phase*-noise ratio (SPNR) lie.

After introducing the setups, I will validate the hypothesis of Gaussianity and whiteness made in chapter 2 for the additive and phase noises. Finally, I will apply the estimators presented in chapter 3 to the received symbols resulting from these laboratory experiments to estimate the SNR and the SPNR.

Part of the material presented in this chapter has been published on the IEEE Photonics Technology Letters (PTL) [12], at the Optical Fiber Conference (OFC) [36] as well as at the Signal Processing in Photonic Communications (SPPCom) conference [37, 38].

4.1. The five experiments

I considered two possible modulation formats:

1. polarisation division multiplexing (PDM) quadrature phase-shift keying (QPSK);

2. on-off keying (OOK).

I set the launch power such that the channels were in one of the three following regimes:

1. not present ($\nexists$), launch power = $-\infty$ dBm;
2. in linear (L) regime, launch power = 0 dBm;
3. in non-linear (NL) regime, launch power = 5, 8 and 10 dBm depending on the experiment: each time I chose the highest launch power at which the digital signal processing (DSP) still converged, to be sure to stimulate as many non-linearities as possible.

In principle, this gave me $2 \cdot 3 = 6$ possible "transmitter configurations", and I could choose

1. one of them for the central channel and
2. one of them for the eight neighbouring channels.

for a total of $6 \cdot 6 = 36$ combinations. Clearly, the majority of them are either not meaningful or very similar to (combinations of) others. In particular I did not consider the following cases:

- the central channel is not present;
- the central channel is OOK modulated, since I wanted to propose solutions for 100G long-haul non differentially encoded (NDE) and differentially encoded (DE) coherent optical communications systems;
- the eight neighbouring channels are in linear regime, since in this case almost no non-linear effects can be seen on the central channel;
- the central channel is in linear regime and the eight neighbouring channels are not present, since even in this case there would be almost no non-linearities;
- both the central channel and the eight neighbouring channels are in non-linear regime, with the latters being OOK modulated. This case was unfortunately excluded because of time and resource limitations, not for a lack of interest; however, it would have added little to the investigation, since it is very similar to a combination of other two cases.

The number of neighbouring channels was limited to eight because a higher number would have meant using a much bigger light amplification by stimulated emission of radiation (LASER) bank, with a much more complicated controlling.

Table 4.1 summarizes the experiments I conducted:

1. the *"QPSK-L/QPSK-NL"* experiment, in which

Name	Central channel (regime)	Neighbouring channels (regime)	Total launch power [dBm]	Span length [km]	Non-linearities stimulated
1) QPSK-L/ QPSK-NL	QPSK, 0 dBm (linear)	QPSK, 8 dBm (non-linear)	17.12	380	homogeneous inter-channel
2) QPSK-NL/ ♯	QPSK, 10 dBm (non-linear)	not present	10.00	380	intra-channel
3) QPSK-NL/ QPSK-NL	QPSK, 8 dBm (non-linear)	QPSK, 8 dBm (non-linear)	17.54	380	intra-channel and homogeneous inter-channel
4) QPSK-L/ OOK-NL	QPSK, 0 dBm (linear)	OOK, 5 dBm (non-linear)	14.20	380	heterogeneous inter-channel
5) B2B	QPSK, 0 dBm (linear)	not present	0.00	0	none

Table 4.1.: Experiments conducted in this work.

- the central channel is modulated with a PDM QPSK, in linear regime (launch power = 0.00 dBm);
- the eight neighbouring channels are modulated with a PDM QPSK, in non-linear regime (launch power = 8.00 dBm), generating *homogeneous inter*-channel non-linearities;
- total launch power = $10\log_{10}(1 \cdot 10^{0/10} + 8 \cdot 10^{8/10}) \approx 17.12$ dBm.

This experiment, which helps to isolate the non-linearities arising from the propagation of the sole neighbouring channels, represents a *modern* 100G wavelength division multiplexing (WDM) system in which the neighbouring channels have to travel farther than the central one and have hence a higher launch power. In fact, the longer the distance, the higher the number of optical amplifiers along the way and therefore the lower the *optical* signal-to-noise ratio (OSNR) at the receiver, which must then be increased, usually increasing the launch power at the transmitter;

2. the *"QPSK-NL/♯"* experiment, in which
   - the central channel is modulated with a PDM QPSK, in non-linear regime (launch power = 10.00 dBm), generating (homogeneous) *intra*-channel non-linearities;
   - the eight neighbouring channels are not present;

- total launch power = 10.00 dBm.

This experiment does not represent any practical WDM system but helps to isolate the non-linearities arising from the propagation of the sole central channel;

3. the *"QPSK-NL/QPSK-NL"* experiment, in which
   - the central channel is modulated with a PDM QPSK, in non-linear regime (launch power = 8.00 dBm), generating (homogeneous) *intra*-channel non-linearities;
   - the eight neighbouring channels are modulated with a PDM QPSK, in non-linear regime (launch power = 8.00 dBm), generating *homogeneous inter*-channel non-linearities;
   - total launch power = $10\log_{10}(9 \cdot 10^{8/10}) \approx 17.54$ dBm.

   This experiment represents a *modern* 100G WDM system, in which all channels have the same modulation and the same launch power;

4. the *"QPSK-L/OOK-NL"* experiment, in which
   - the central channel is modulated with a PDM QPSK, in linear regime (launch power = 0.00 dBm);
   - the eight neighbouring channels are modulated with an OOK, in non-linear regime (launch power = 5.00 dBm), generating *heterogeneous inter*-channel non-linearities;
   - total launch power = $10\log_{10}(1 \cdot 10^{0/10} + 8 \cdot 10^{5/10}) \approx 14.20$ dBm.

   This experiment represents an *old* 10G/100G WDM system, in which the central channel originally carrying 10G long-haul incoherent optical communications systems has been replaced with a 100G long-haul coherent optical communications system.

In the following I will refer to the QPSK-L/QPSK-NL experiment, the QPSK-NL/♯ experiment and the QPSK-NL/QPSK-NL experiment as the *"QPSK-only"* experiments for brevity; as I will show, in fact, they will exhibit very similar characteristics.

All four experiments have been conducted transmitting over $4 \cdot 95 = 380$ km of single-mode optical fibre (SMF). The number of spools was limited to four because a higher number would have meant implementing a recirculating loop, with a much more complicated controlling.

In addition to those listed above I also conducted the following experiment:

5. the *"back to back (B2B)"* experiment, in which
   - the central channel is modulated with a PDM QPSK, in linear regime (launch power = 0.00 dBm);
   - no neighbouring channels are present;
   - total launch power = 0.00 dBm.

   This experiment acts as a reference for the performance of the setup itself.

4.2. Experimental setup

The experimental setup was composed of

- either one of two transmitters, depending on the modulation format of the neighbouring channels,
- a transmission link, the same for all the experiments, obviously with the exception of the B2B experiment, and
- a receiver, which was also the same for all the experiments (since it depends on the modulation format of the central channel, which was always PDM QPSK).

4.2.1. Transmitters

Since on the neighbouring channels I wanted to transmit either PDM QPSK or OOK modulated signals, I had to realise two different transmitters.

In both cases the data rate of the central channel was 112 Gbps in order to reproduce the optical transport unit 4 (OTU4) nominal bit rate = 111.809973568 Gbps [39–42, Table 7-1], which in turn corresponds to the optical data unit 4 (ODU4) nominal bit rate of R_b = 104.794445815 Gbps increased by 255/239 because of forward error correction (FEC).

4.2.1.1. Transmitter for the B2B and the three QPSK-only experiments

In order to realise the B2B and the three QPSK-only experiments I prepared a transmitter with one central channel and in the latters eight neighbouring channels, all modulated with PDM QPSK.

Figure 4.1 shows the transmitter used for these four experiments.

In more detail:

- I coupled together
  - five LASERs on the Telecommunication Standardization Sector of International Telecommunication Union (ITU-T) grid (having a 100 GHz spacing) and
  - four LASERs on the ITU-T offset-grid (also having a 100 GHz spacing)

  by using passive couplers, generating 2 LASER combs which I will refer to as the "even" and the "odd" channel groups. These LASERs had a linewidth Δ_f = 100 kHz. Altogether, the channels hence had a 50 GHz spacing. The central channel lay in the middle of the nine channels, with four channels on each side, and was, as such, an odd channel;
- I generated offline a sequence of $\approx 10^6$ QPSK symbols (hence the choice of $L = 10^6$ in figures 3.1 and 3.7), extracted the corresponding I and Q sequences and loaded the latters in a SHF 12100 B pulse pattern generator (PPG) [43], set to run at a 28

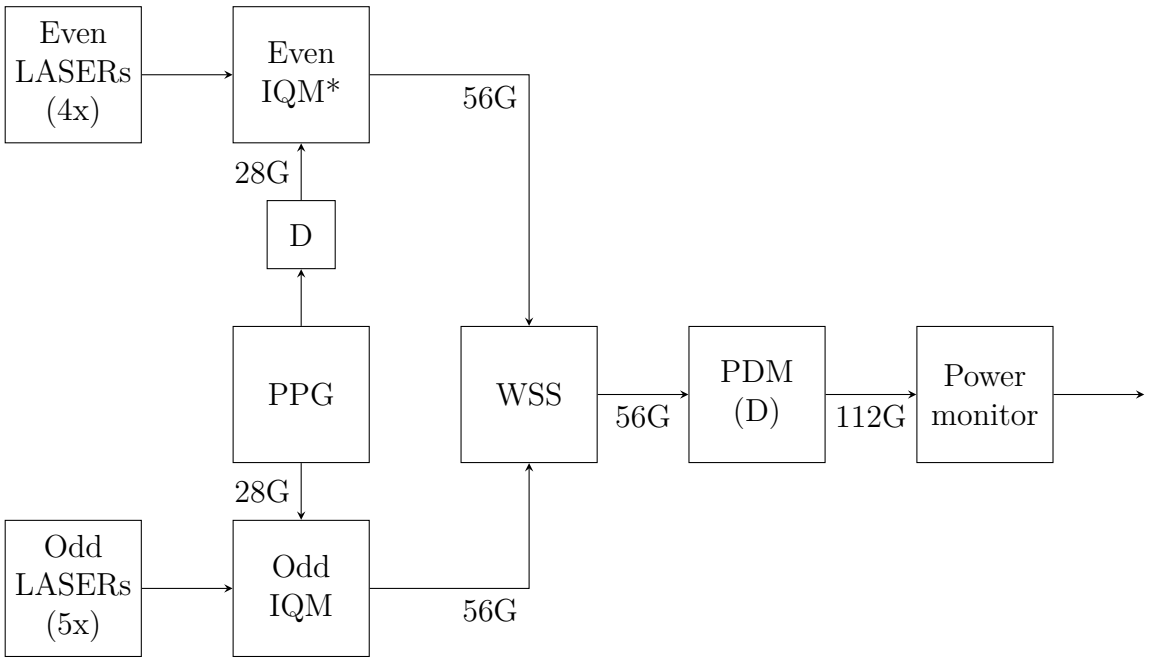

Figure 4.1.: Transmitter used for the three QPSK-only experiments as well as the B2B experiment. The devices marked with an asterisk are not present or have a different position in figure 4.3.

GBd by inserting an electrical 28 GHz clock signal. Unfortunatey, the repetition of longer, user-defined sequences caused some problems, so I was limited to that length.

Figure 4.2 shows the eye diagram at 28 GBaud of the SHF 12100 B PPG [43].

The PPG had two pairs of outputs, so I delayed one of them for decorrelation and

- used them to drive two single-polarisation IQ modulators (IQMs), one modulating the even channel group and the other modulating the odd channel group.

  I hence realised two groups of four and five QPSK modulated channels, each channel carrying data at a rate = $28 \cdot 2 = 56$ Gbps;

- I then multiplexed the even and the odd channel groups using the Waveshaper 4000S wavelength selective switch (WSS) on the 50 GHz ITU-T grid. With the WSS I equalised the channels, regulating the *relative* power of both the central and the neighbouring channels (a 0 or 8 dB difference depending on the experiment, see table 4.1).

  At this point I had one group of nine QPSK modulated channels, each channel still carrying data at a rate = 56 Gbps. In order to build a 100G long-haul coherent optical communications system

- I hence emulated PDM by means of a PDM stage, in which the incoming optical signal is splitted into two equally powered tributaries by means of a polarisation

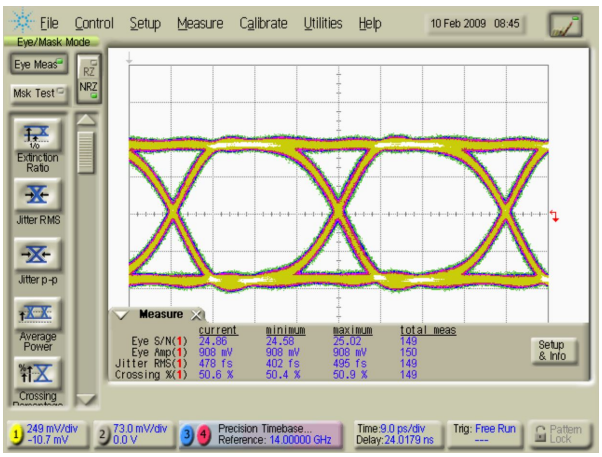

Figure 4.2.: Eye diagram at 28 GBaud of the SHF 12100 B PPG used to realize the PDM QPSK modulated channels for all experiments [43].

beam splitter (PBS), one of them is delayed (by 469, 312 or 229 symbols depending on the experiment), again for decorrelation, and the two are recombined by means of a polarisation beam combiner (PBC).

Thus I obtained one group of nine PDM QPSK modulated channels, each channel carrying data at a rate = $56 \cdot 2 = 112$ Gbps;

- I then pre-amplified the nine channels using an erbium doped fibre amplifier (EDFA) and coupled them into the transmission link. After the EDFA and before the transmission link

- I placed an Eigenlight in-line power monitor, with which I regulated the *absolute* total launch power.

4.2.1.2. Transmitter for the QPSK-L/OOK-NL experiment

In order to realise the QPSK-L/OOK-NL experiment I prepared a transmitter with one central channel, modulated with PDM QPSK, and eight neighbouring channels, modulated with OOK.

Figure 4.3 shows the transmitter used for this fifth experiment.

In more detail and with respect to the transmitter described in section 4.2.1.1:

- I replaced the eight LASERs used for the neighbouring channels (odd and even) with other eight cheaper LASERs. These LASERs had a linewidth $\Delta_f > 100$ kHz. For the OOK modulation format, in fact, the linewidth is less critical than for phase sensitive modulation formats;

- I loaded the I and Q sequences of the same sequence of $\approx 10^6$ QPSK symbols in the same PPG running at a 28 GBaud and

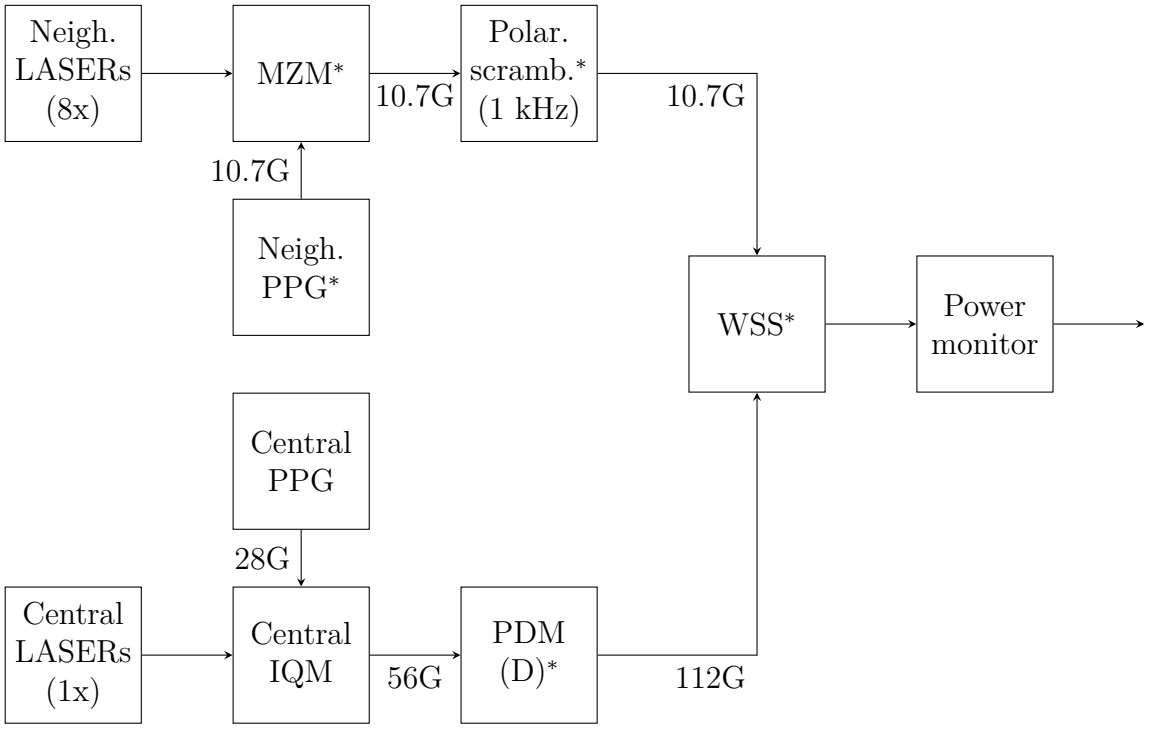

Figure 4.3.: Transmitter used for the QPSK-L/OOK-NL experiment. The devices marked with an asterisk are not present or have a different position in figure 4.1.

- used them to drive one of the two previous single-polarisation IQMs, thus modulating the central channel, now carrying data at a rate = $28 \cdot 2 = 56$ Gbps;
- I then fed the same PDM stage with this signal, with which I emulated PDM obtaining the PDM QPSK modulated central channel, carrying data at a rate = $56 \cdot 2 = 112$ Gbps; in parallel
- I generated a new sequence of OOK symbols, loaded it in another PPG running at a 10.7 GBaud and
- used it to drive a Mach-Zehnder modulator (MZM), thus modulating the neighbouring channels, now carrying data at a rate = 10.7 Gbps.

  The LASER used to generate the neighbouring channels were polarisation aligned, corresponding to worst case launch conditions; in order to test different relative positions between the central and the neighbouring channels
- I placed a polarisation scrambler after the MZM, going through all possible polarisation states within 1 ms (i.e., with a 1 kHz clock);
- I then multiplexed the central and the neighbouring channels using the same WSS, by setting whose attenuations I equalised the *relative* launch powers of both the central and the neighbouring channels (a 5 dB difference, see table 4.1). After the WSS
- I pre-amplified the PDM QPSK modulated central channel, carrying data at a rate = 112 Gbps, and the OOK modulated neighbouring channels, carrying data at a rate = 10.7 Gbps, using the same EDFA and coupled them into the same transmission link; after the EDFA and before the transmission link
- I used again the same EigenLight in-line power monitor to regulate the *absolute* total launch power.

4.2.2. Transmission Link

Figure 4.4 shows the transmission link used in all experiments with the exception of the B2B experiment.

In more detail:

- I used four different 95 km spools of SMF, no dispersion compensating fibre (DCF) was present.

  Table 4.2 reports additional parameters of the optical fibre.

  The overall transmission distance was hence 380 km;
- I placed an EDFA after each span to compensate for the span loss, bringing the total power launched in the next spool back to the values reported in table 4.1.

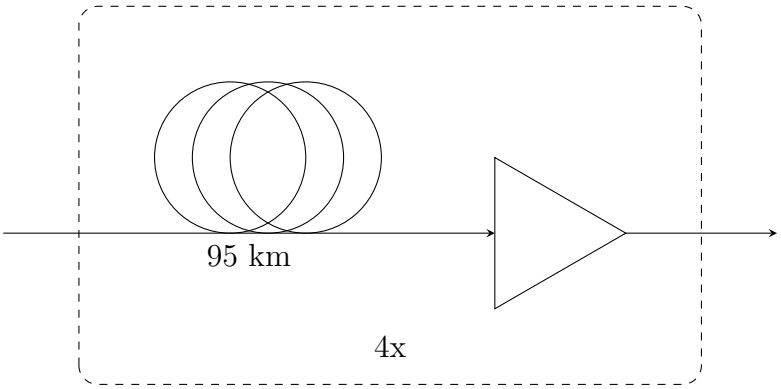

Figure 4.4.: Transmission link used for all but the B2B experiment.

Parameter	Value
Attenuation	0.19 dB/km
Span length	95 km
Span loss (with splices)	19 dB
Chromatic dispersion @1550 nm	16.8 ps/(nm km)
Dispersion slope	0.057 ps/(nm^2 km)
Effective core area	80 µm^2
Non-linear coefficient	1.3 rad/(W km)

Table 4.2.: Average parameters of the SMF used in all experiments.

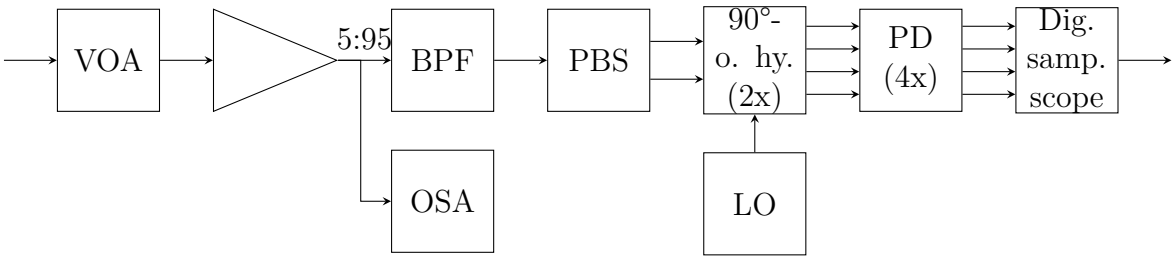

Figure 4.5.: Receiver used for all experiments.

4.2.3. Receiver

Figure 4.5 shows the receiver used for all experiments.

In more detail:

- I attenuated the central channel with a variable optical attenuator (VOA) and then
- I fed an EDFA with it.

  Since amplifiers produce a constant output power, when the power of the central channel is reduced by means of the VOA the amplifier produces more amplified spontaneous emission (ASE) and viceversa. Hence, by manually changing the attenuation of the VOA I could set the OSNR to the desired values. This was, in other words, the noise loading stage (see also section 4.2.3.1). After the EDFA
- I splitted the central channel with a 5:95 splitter and sent 5% of the power to an optical spectrum analyzer (OSA), with which I measured the OSNR set with the VOA;
- I then filtered out the central channel from the remaining 95% of the power with a 50 GHz tunable optical bandpass filter (BPF),
- I splitted it into two orthogonal polarisation components by means of a PBS and
- I combined the outputs of the latter inside two 90°-optical hybrids with a local oscillator (LO) with a 14 dBm power and a linewidth $\Delta_f = 100$ kHz.

  This structure, typical of coherent receivers, is needed to be able to detect both components of the signals of both polarisations, since a photodiode (PD) generates a current which depends on the power of the incoming optical signal and, as such, only on its amplitude;
- I then connected the four outputs of the polarisation diversity 90°-optical hybrid, representing the in-phase and quadrature components of the two polarisations, to four balanced PDs, which detected the four optical signals and translated them into electrical ones. Finally,

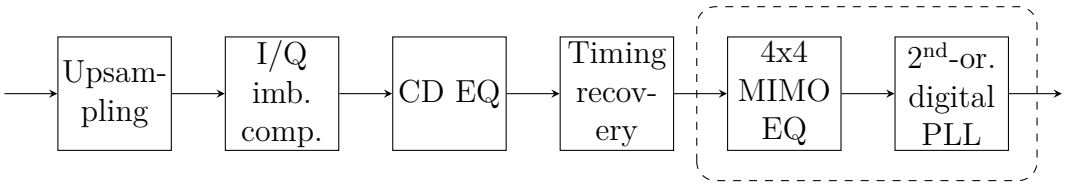

Figure 4.6.: DSP algorithms used for all experiments. The stages framed by a dashed line were performed iteratively.

- I sampled the electrical signals coming from the PDs with a Lecroy LabMaster 945MZi-A 40 GSamples/s digital sampling scope with a 20 GHz electrical bandwidth [44].

4.2.3.1. OSNR measurement

I set the OSA to integrate the power of the optical signal over the reference bandwidth $B_{\text{ref}} = 12.5$ GHz, corresponding to 0.1 nm.

With it I acquired the power at three different points in frequency, once in the middle of the central channel and twice to its sides:

- the value in the middle of the central channel measured the *sum* of the powers of the wanted signal and of the ASE noise *inside* the central channel;
- the values on the sides, instead, measured the power of the *sole* ASE noise *outside* the central channel.

Under the assumption that the noise floor remains more or less constant over all frequencies (or at least it grows or decays linearly with the frequency), I hence estimated

- the power of the ASE noise *inside* the central channel interpolating the values on the sides and
- the power of the wanted signal *inside* the central channel subtracting the latter from the value in the middle.

Finally, I scaled the power of the wanted signal by its real occupied bandwidth $B_s = 28$ GHz, took the ratio of the two powers and obtained the OSNR.

4.2.4. Digital signal processing

Offline I then recombined the sampled in-phase and quadrature components of the two polarisations from section 4.2.3 and post-processed them with a *blind* DSP.

Figure 4.6 shows the DSP algorithms used for all experiments.

In more detail, the cascade was composed of six stages, the last two of which were performed iteratively:

1. *Upsampling* – the signals on the two polarisations were upsampled block-wise from $40/28 \approx 1.43$ samples/symbol to $56/28 = 2$ samples/symbol by means of spline interpolation.

   This stage is needed because the following algorithms require integer or fractional samples/symbol rates. Note that this stage is typically needed in laboratory experiments, where the sampling scopes in general run at fixed sampling rates independent of the symbol rate of the particular laboratory experiment being performed; commercial transponders will most likely not need this stage.

   After this stage, potential direct current (DC) components in the signals on the two polarisations were removed and the powers of the latters were normalised to 1. This normalisation step hence corresponds to a first, very coarse automatic gain control (AGC);

2. *2x I/Q imbalance compensations* – the quadrature components of the signals on the two polarisations were resampled to have their versions temporally aligned with the in-phase ones.

   This stage is needed to compensate for the different lengths of the paths covered by the in-phase and quadrature components of the signals on the two polarisations.

   After this first, manual compensation a second, automatic one was also implemented [45];

3. *Chromatic dispersion (CD) equalisation* – the amount of *accumulated* CD (for in principle the length of the fibre is not known) was estimated with a method [46] derived from the constant modulus algorithm (CMA) algorithm [47]. The fast Fourier transform (FFT) of the signals on the two polarisations was then taken block-wise, multiplied by the inverse of the frequency response representing the estimated accumulated CD and finally the inverse fast Fourier transform (IFFT) of the result was taken. The estimated accumulated CD was hence compensated for *in the frequency domain*. Note that this frequency response has costant amplitude, hence it is safe to take its inverse.

   This stage is needed to compensate for the largest part of the CD which accumulates during propagation in the SMF. Potential *residual* CD was compensated for in the second equalisation stage;

4. *Timing recovery* – the signals on the two polarisations were upsampled block-wise from 2 samples/symbol to 4 samples/symbol, again by means of spline interpolation. The time delays were estimated for each block and averaged over the blocks according to the square timing recovery (STR) algorithm [48]. Finally, the signals were downsampled block-wise from 4 samples/symbol to 2 samples/symbol, again by means of spline interpolation, and resampled block-wise to have their versions with samples at the beginning and in the middle of the symbol period, as needed by the following algorithms.

This stage is needed to compensate for differences in frequency and phase of the transmitter and the receiver clocks, which cause the analog-to-digital converters (ADCs) in the digital sampling scope to sample at random points within a symbol period, not where needed. Note that this stage must come after the CD equalisation one and the upsampling must be at more than 2 samples/symbol, otherwise the spectral component at the symbol rate will not be clearly represented;

5. *Joint equalisation and carrier recovery* – this stage was composed of two substages performed iteratively:
   a) *4x4 multiple-input and multiple-output (MIMO) equalisation* – each in-phase and quadrature component of each signal on the two polarisations was filtered *in the time domain* by four finite impulse response (FIR) filters (different for each component) and then four linear combinations of the results were taken. In total, 16 filters were hence used in a butterfly structure.

      Each filter had 13 taps (uneven, so that the middle tap refers to the center of the central pulse), which means that the filters were able to retrieve information from 3 past symbols and 3 future ones, since the components at this point were sampled with 2 samples/symbol. The choice of the number of taps is always the result of a compromise and, as such, delicate. If the number is too small, meaningful information from past and future symbols that are too far away will be lost; if the number is too large, impairing noise from past and future symbols will be collected.

      The phases of the signals on the two polarisations were then corrected with the values estimated in the previous iteration of the next substage.

      At this point the filters were updated according to various algorithms:

      - for the first blocks (i.e., in the *training* phase) the CMA algorithm was used [47]. Note that this algorithm does not take into account the phase of the signals on the two polarisations, hence it can be used at the beginning, when the frequency offset has not been (or only marginally) corrected yet. Focusing only on the amplitude of the signals on the two polarisations, the CMA algorithm demultiplexed the signals on the two polarisations redistributing the energy of the symbols, taking it from the symbols in the other components or from the neighbouring symbols on the same component and transferring it back to the "original" symbol;
      - for the other blocks (i.e., after the training phase) the decision-directed least mean square (DD-LMS) algorithm was used [49, 50]. Note that this algorithm, instead, takes into account the phase of the signals on the two polarisations, so it can be used only after a while, when the frequency offset has already been (mostly) corrected. Taking into account also the phase of the signals on the two polarisations, the DD-LMS algorithm fine tuned the result of the CMA algorithm.

        Finally, the filters were updated also according to

- one independent component analysis (ICA) algorithm, which achieves blind source separation by avoiding correlation between the resulting output sample sequences or, equivalently, between the off-diagonal converged impulse responses [49, 50]. In fact, the CMA algorithm could force both signals on the two polarisations to the same solution, since it minimised the distance of the signals from a constant, not from a given sequence of symbols.

This stage was needed to compensate for

- potential residual CD left after the CD equalisation stage,
- polarisation mode dispersion (PMD),
- crosstalk between each pair of components,
- bandwidth limitations and
- the random (and hence misaligned) state of polarisation (SOP) of the optical signal entering the receiver. Because of the latter, in fact, the signals detected by the PDs and sampled by the digital sampling scope are actually a linear combination of the transmitted signals on the two polarisations.

This equalisation step hence corresponds to a second, more sophisticated AGC;

b) *2^{nd}-order digital phase-locked loop (PLL)* – a hard-decision (HD) was taken on the samples of the signals on the two polarisations. A *loop filter* was then fed with a function of the samples and the HD decisions and estimated the phase correction terms applied to the signals on the two polarisations before the FIR filters were updated.

Once all the phase correction terms were available, the filter estimated also the frequency offset.

This stage is needed to compensate for the phase and frequency offset between the LASERs used in the transmitter and in the receiver.

More details on the blind DSP algorithms I used can be found in [49–51]. All these algorithms were implemented in MATLAB ®.

4.3. Experimental results

4.3.1. Validation of the statistical models

In order to validate the whiteness and the Gaussianity assumptions I decided to consider the constellation carried on one of the two polarisations and its counter-rotated and translated (CRT) received version taken at an OSNR = 20 dB for the four experiments with the transmission link reported in table 4.1.

Experiment	Est. SNR [dB]	Est. SPNR [dB]
QPSK-L/QPSK-NL	9.67	21.07
QPSK-NL/∄	10.38	21.62
QPSK-NL/QPSK-NL	10.32	20.86
QPSK-L/OOK-NL	12.36	17.75
B2B	10.35	27.91

Table 4.3.: Estimated SNR and SPNR of the shots shown in figure 4.7.

The reason is that I wanted to minimise potential artefacts introduced by the DSP because of too much ASE (the presence of the non-linearities already put the DSP under very stressful conditions where it sometimes failed to converge). In order to have curves in the same ranges, I decided to consider constellations taken instead at an OSNR = 14 dB for the B2B experiment.

Before showing the results, recall what I said in section 2.4, in particular (2.19) and (2.20): according to the approximated, counter-rotated and translated (ACRT) NDE additive white Gaussian noise (AWGN)/ phase white Gaussian noise (PWGN) model,

- the real part of the CRT received symbol equals the real part of the counter-rotated AWGN N', whereas
- its imaginary part equals the sum of the PWGN Θ and the imaginary part of the counter-rotated AWGN N'.

As a consequence, the autocorrelation function of the real part of the CRT received symbol equals the autocorrelation function of the real part of the counter-rotated AWGN,

$$r_{\Re(Y')}(kT_s) = r_{\Re(N')}(kT_s), \quad k \in \mathbb{Z}, \tag{4.1}$$

where $T_s = 1/(28 \cdot 10^9) \approx 35.7$ ps is the symbol period.

Moreover, since the PWGN is assumed to be independent of the counter-rotated AWGN (see section 2.4.1) and they both are assumed to have zero mean, the autocorrelation function of the imaginary part of the CRT received symbol is given by the sum of the autocorrelation function of the PWGN and the autocorrelation function of the imaginary part of the counter-rotated AWGN:

$$r_{\Im(Y')}(kT_s) = r_{\Theta}(kT_s) + r_{\Im(N')}(kT_s), \quad k \in \mathbb{Z}. \tag{4.2}$$

4.3.1.1. Gaussianity

Figure 4.7 shows the *experimental* probability density functions (pdfs) of the real (top) and imaginary (bottom) parts of the CRT received constellations $Y'_i, i = 1, \ldots, L$, together with the corresponding *theoretical* Gaussian pdfs.

Table 4.3 reports the corresponding estimated SNR and estimated SPNR.

The experimental pdfs overlap almost perfectly with the theoretical ones, confirming the Gaussianity assumption. Moreover,

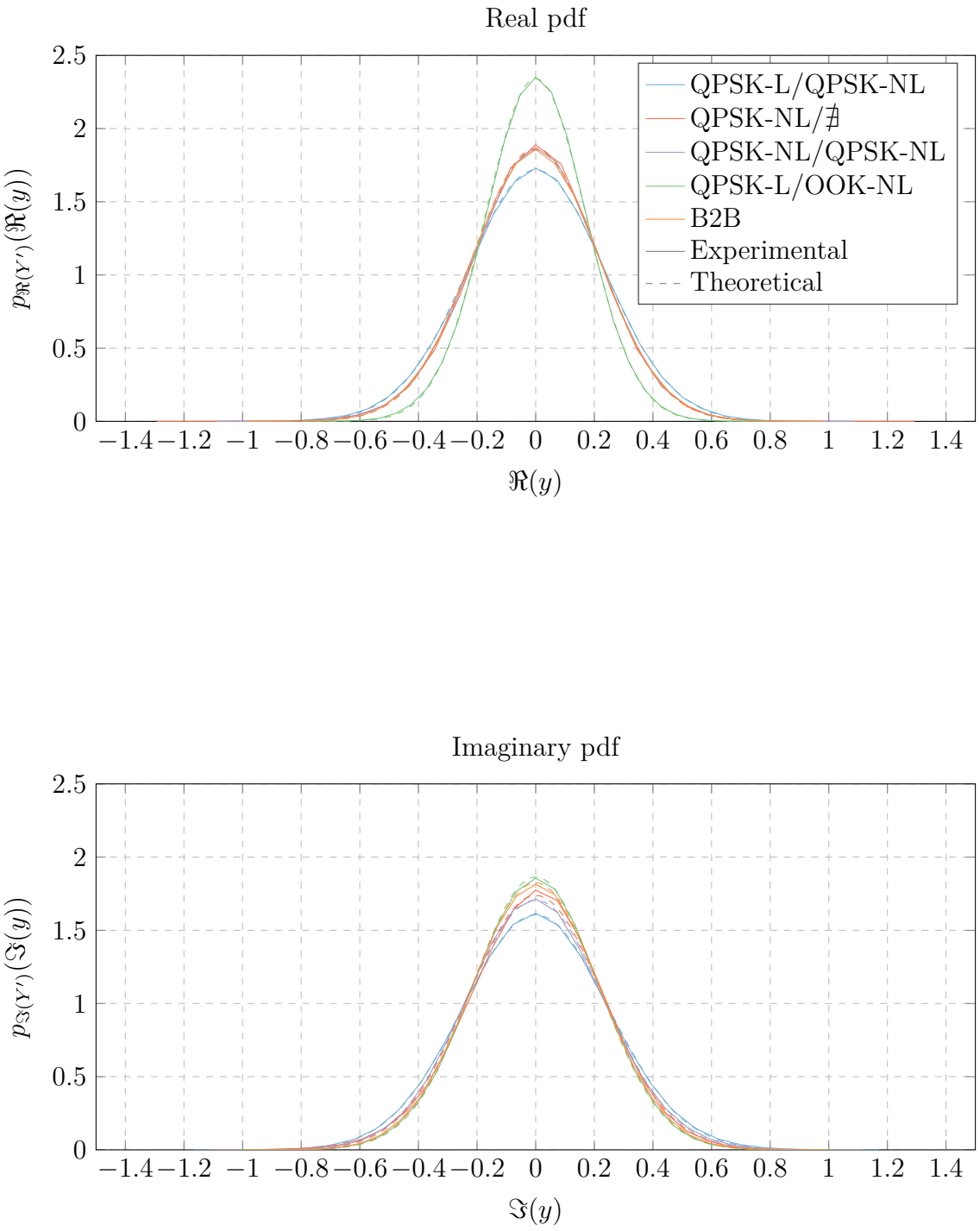

Figure 4.7.: Experimental pdfs (solid) of the real (top) and imaginary (bottom) parts of the CRT received constellations, at an OSNR = 20 dB (14 dB for the B2B experiment). For each curve a *theoretical* Gaussian pdf (dashed) is also reported, with zero mean and variance estimated according to section 3.2.1. The corresponding estimated SNR and SPNR are reported in table 4.3.

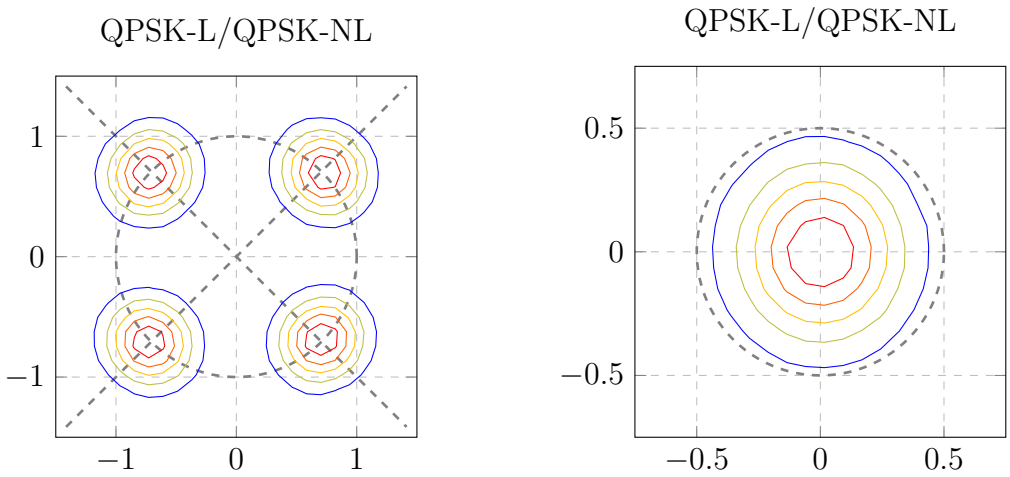

Figure 4.8.: 2D experimental pdfs of the received constellations, before (left) and after (right) being CRT, for the QPSK-L/QPSK-NL experiment, at an OSNR = 20 dB.

- the (experimental and theoretical) "real" and "imaginary" (for brevity) pdfs of the B2B experiment are practically equal, whereas
- the real pdfs of the three QPSK-only experiments have slightly smaller variances and hence are slightly narrower and taller than their imaginary ones. This is due to the presence of PWGN in the imaginary part of the CRT received constellations, which is not present in their real part. This PWGN is due mostly to the intra- and homogeneous inter-channel non-linearities (*after* DSP, to be precise, although I will not repeat it each time, again for brevity) and, to a lesser extent, to the chain of transmitter, receiver and DSP. In fact, as reported in table 4.3, in the three QPSK-only experiments the estimated SPNR is $\approx$ 6–7 dB lower than in the B2B experiment (where only the chain of transmitter, receiver and DSP is present), for approximately the same estimated SNR;
- this phenomenon is even more pronounced for the pdfs of the QPSK-L/OOK-NL experiment, indicating that heterogeneous inter-channel non-linearities contribute more to the PWGN and less to the AWGN than intra- and homogeneous inter-channel ones. As reported in table 4.3, in the OOK experiment the estimated SPNR is $\approx$ 3–4 dB lower than in the three QPSK-only experiments and $\approx$ 10 dB lower than in the B2B experiment, whereas the estimated SNR is $\approx$ 2 dB higher.

Figures 4.8, 4.9, 4.10, 4.11 and 4.12 show the experimental two-dimensional (2D) pdfs of the received constellations, before (left) and after (right) being CRT.

The 2D experimental pdfs of the received constellations after being CRT confirm what figure 4.7 already showed: the 2D pdfs are a series of concentric circles or ellipses, depending on how much PWGN is present. In particular,

- the B2B experiment exhibits the roundest constellations (see figure 4.12),

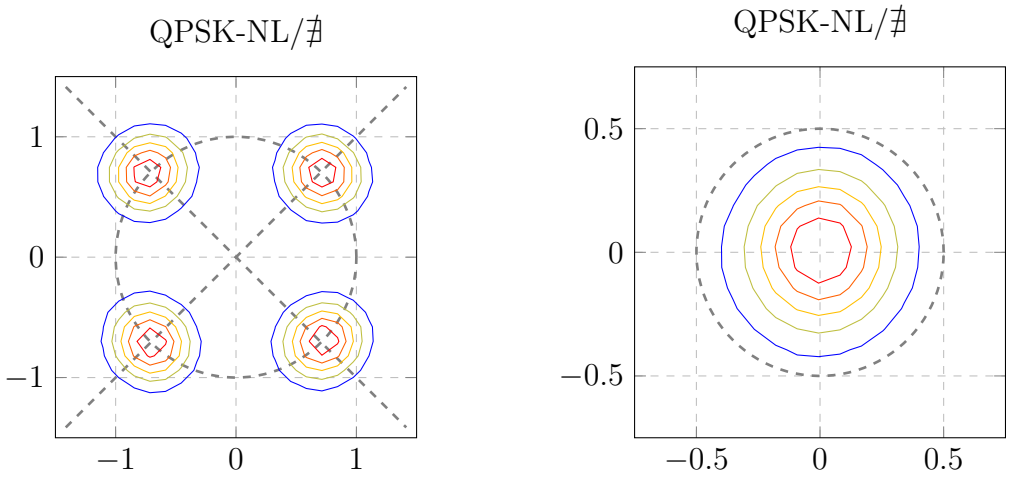

Figure 4.9.: 2D experimental pdfs of the received constellations, before (left) and after (right) being CRT, for the QPSK-NL/♯ experiment, at an OSNR = 20 dB.

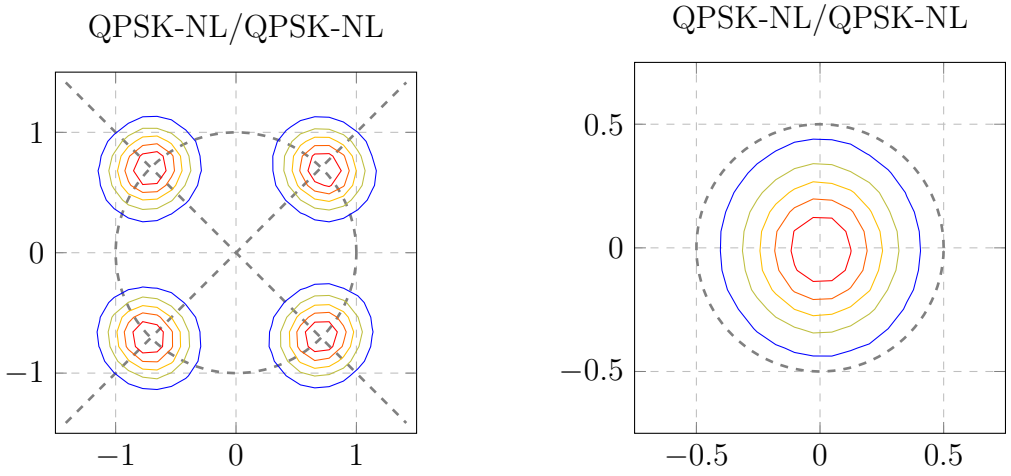

Figure 4.10.: 2D experimental pdfs of the received constellations, before (left) and after (right) being CRT, for the QPSK-NL/QPSK-NL experiment, at an OSNR = 20 dB.

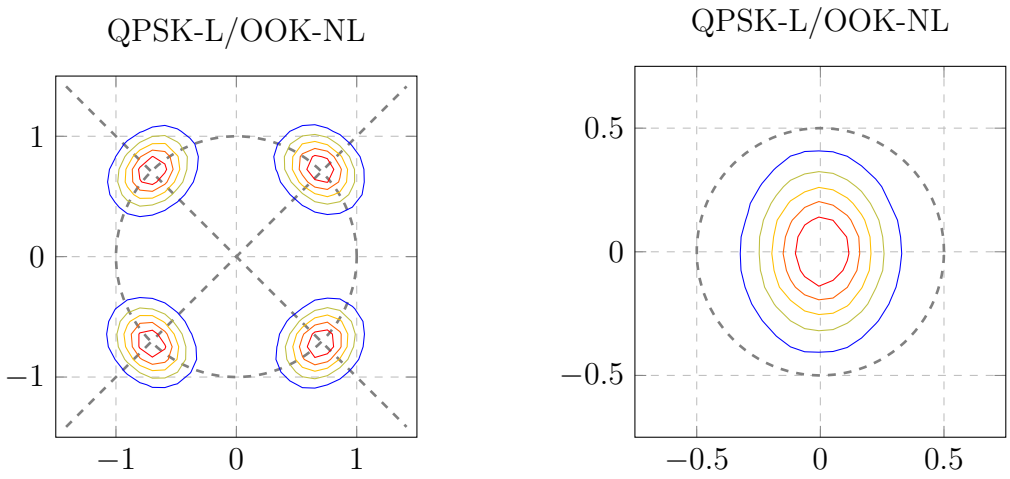

Figure 4.11.: 2D experimental pdfs of the received constellations, before (left) and after (right) being CRT, for the QPSK-L/OOK-NL experiment, at an OSNR = 20 dB.

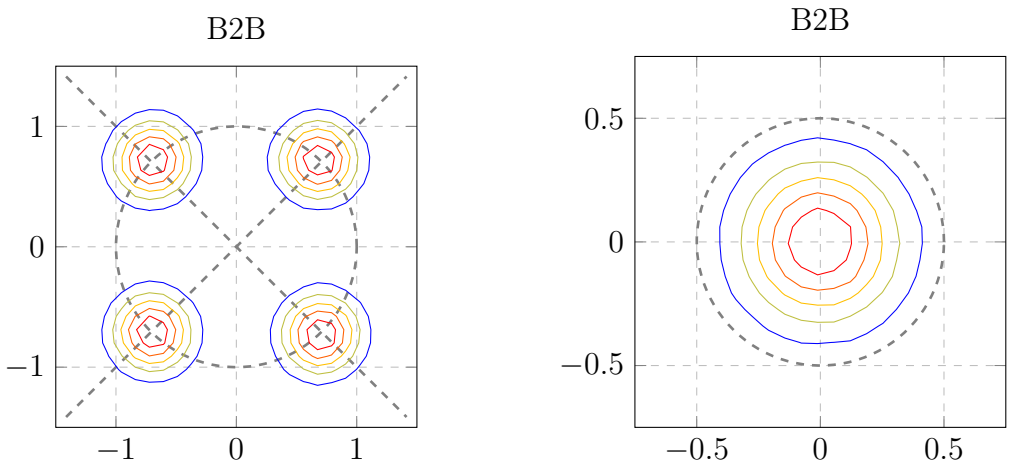

Figure 4.12.: 2D experimental pdfs of the received constellations, before (left) and after (right) being CRT, for the B2B experiment, at an OSNR = 14 dB.

- the QPSK-L/OOK-NL experiment exhibits the least round ones (see figure 4.11) and
- the three QPSK-only experiments are halfway (see figures 4.8, 4.9 and 4.10).

Moreover, the 2D experimental pdfs of the received constellations before being CRT lie on the unit circle, which means that the received constellation is not scaled, confirming the ideality assumption of the AGCs in the DSP.

4.3.1.2. Whiteness

Figure 4.13 shows the *normalised* experimental *autocorrelation* functions of the real (top) and imaginary (middle) parts of the CRT received constellations $Y_i', i = 1, \ldots, L$ as well as their *normalised* experimental *cross-correlation* function (bottom). The normalisation is by the norm in the first two cases and by the squared root of the product of the norms in the last case, so only the autocorrelations will be 1 in the origin [52]. I decided to show the normalised correlations in order to have a meaningful comparison among them. In fact, I am interested into their shapes (rather than their absolute values), whether they are impulsive or not and, if not, after how many symbols they can be considered negligible (the 10% thresholds are indicated with two gray dashed lines in figure 4.13).

The experimental "real" (again for brevity) autocorrelation functions are almost perfectly impulsive, confirming the whiteness assumption for the additive noise (strictly speaking only of its real part).

Some of the experimental "imaginary" (for brevity) autocorrelation functions, instead, are less impulsive, although they all reach their respective half-width at half-maximum (HWHM) already after 1 symbol. In particular,

- the B2B experiment exhibits again a perfectly impulsive experimental imaginary autocorrelation function,
- the three QPSK-only experiments exhibit relatively impulsive ones (they fall below 10% of their maximum value in the origin after 3–5 symbols), whereas
- the QPSK-L/OOK-NL experiment exhibits the least impulsive one (it falls below 10% of its maximum value in the origin after 10 symbols).

Non-linearities and, in particular, heterogeneous inter-channel non-linearities are not white *before* DSP, see section 2.3.1 and references therein. Since the imaginary part of the additive noise can be confidently assumed to be white, figure 4.13 suggests that they manifest also as "non exactly white" phase noise even *after non-ideal* DSP: the latter did not manage to remove *all* memory or, equivalently, left some *residual* correlation. However, as far as the estimation of the SNR and the SPNR is concerned, the effect of this residual correlation proved to be negligible. I tried to insert an interleaver before the transmitted symbols $X_i, i = 1, \ldots, L$ and the corresponding deinterleaver after the received symbols $Y_i, i = 1, \ldots, L$ (see again section 2.2.1 and [26]): the experimental

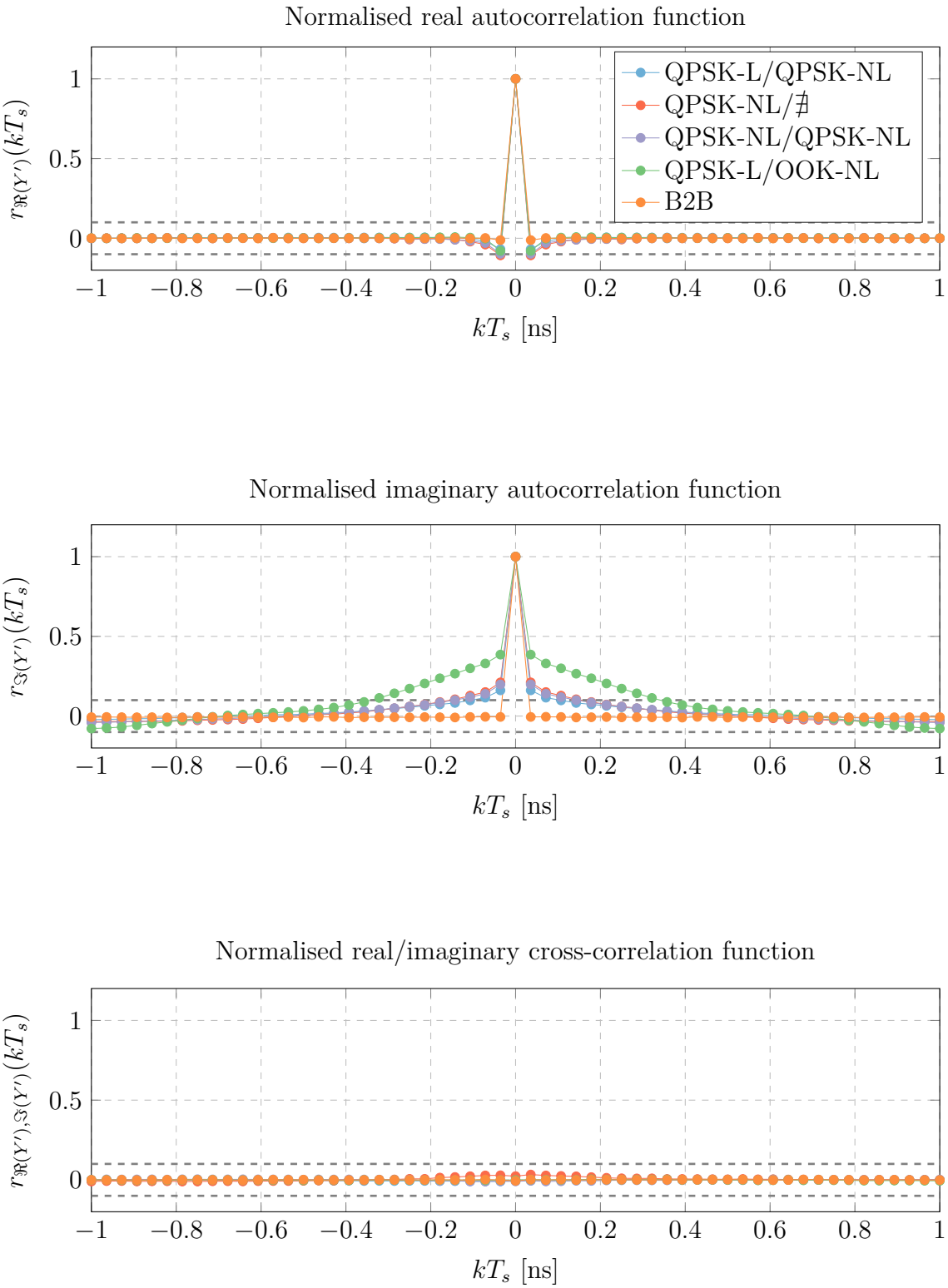

Figure 4.13.: Experimental autocorrelation functions of the real (top) and imaginary (middle) parts of the CRT received constellations as well as their experimental cross-correlation function (bottom), at an OSNR = 20 dB (14 dB for the B2B experiment); $T_s \approx 35.7$ ps is the symbol period, $k \in \mathbb{Z}$.

imaginary autocorrelation function became perfectly impulsive, as expected, but the two estimators of the SNR and of the SPNR returned the same values.

The experimental "real/imaginary" (again for brevity) cross-correlation functions are almost zero everywhere, confirming the uncorrelation assumption for the additive noise and the phase noise (which, being Gaussian, are therefore also independent).

Finally, the fact that the experimental "real" autocorrelation functions are impulsive, the experimental "imaginary" autocorrelation functions are almost impulsive and the experimental "real/imaginary" cross-correlation functions are zero everywhere confirms the independency assumption for the CRT received constellations $Y_i', i = 1, \ldots, L$. In fact, their experimental autocorrelation functions can be written as

$$r_{Y'}(kT_s) = r_{\Re(Y')}(kT_s) + r_{\Im(Y')}(kT_s) + j(r_{\Re(Y'),\Im(Y')}(-kT_s) - r_{\Re(Y'),\Im(Y')}(kT_s)), \quad (4.3)$$

and are therefore almost impulsive, since the first term is impulsive, the second term is almost impulsive and the third term is zero.

Figure 4.14 shows the *normalised* experimental cross-correlation functions of the real and imaginary parts of the CRT *received* constellations $Y_i', i = 1, \ldots, L$ and the real and imaginary parts of the *transmitted* constellation $X_i, i = 1, \ldots, L$.

The experimental "real/real" and "real/imaginary" (for brevity) cross-correlation functions are almost zero everywhere, confirming that the additive noise and the transmitted symbols are uncorrelated. Similarly, the experimental "imaginary/real" and "imaginary/imaginary" (for brevity) cross-correlation functions are also almost zero everywhere, confirming that also the phase noise and the transmitted symbols are uncorrelated.

4.3.2. Estimation of the SNR and SPNR ranges

Figure 4.15 shows the estimated SNR on the y-axis and the estimated SPNR on the x-axis, both obtained applying the estimators presented in section 3.2.1 to the five experiments reported in table 4.1. The OSNR is indirectly given by the intensity of the colours: the higher the OSNR, the darker the colour.

4.3.2.1. Estimated SNR range

Considering both polarisations, the estimated SNR lies

- in the range 8.2–10.5 dB in the QPSK-L/QPSK-NL experiment,
- in the range 7.9–10.5 dB in the QPSK-NL/♯ experiment,
- in the range 7.6–11.1 dB in the QPSK-NL/QPSK-NL experiment,
- in the range 9.4–12.6 dB in the QPSK-L/OOK-NL experiment and
- in the range 7.6–15.2 dB in the B2B experiment.

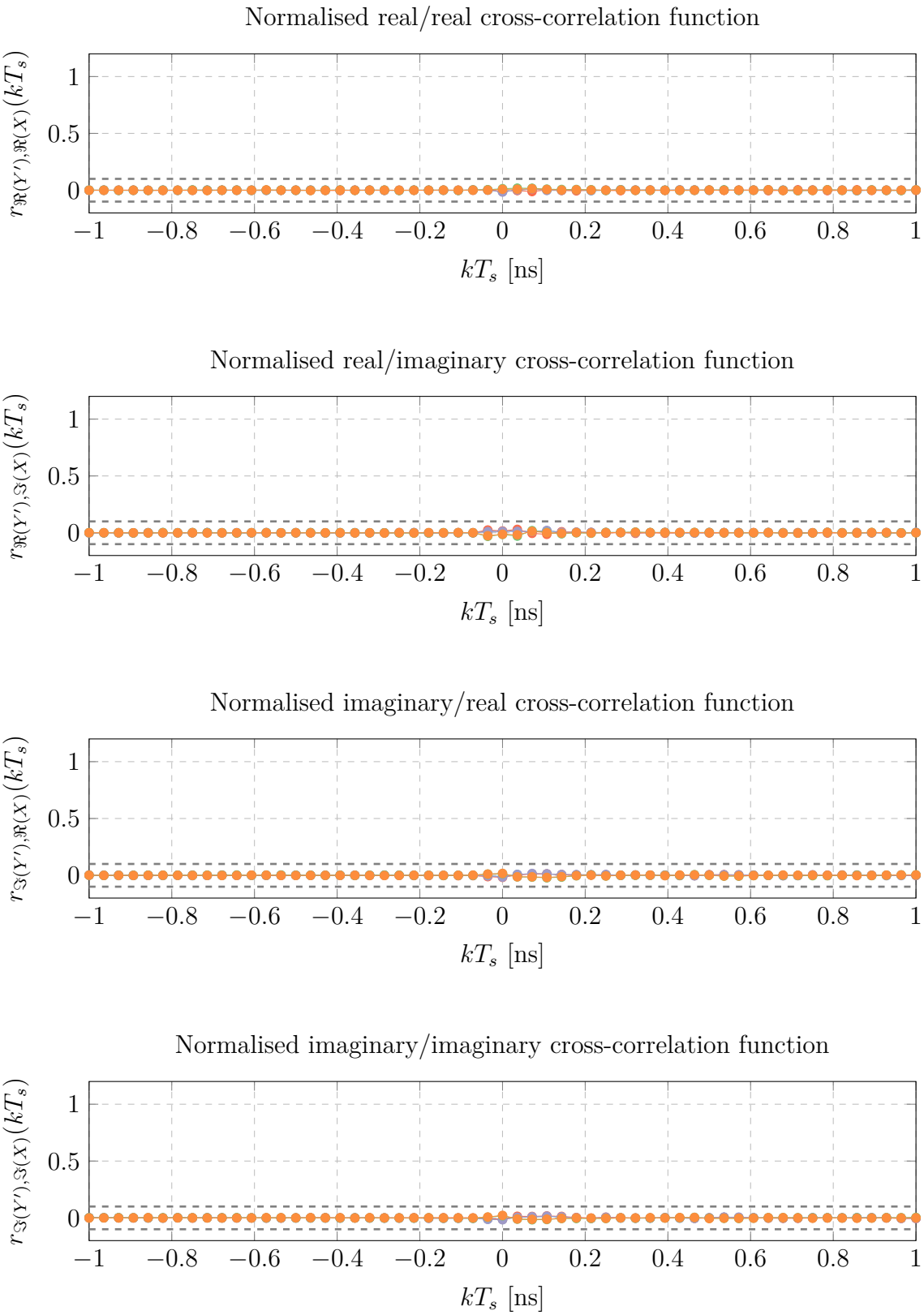

Figure 4.14.: Experimental cross-correlation functions of the real and imaginary parts of the CRT received constellations and the real and imaginary parts of the transmitted constellation, at an OSNR = 20 dB (14 dB for the B2B experiment); $T_s \approx 35.7$ ps is the symbol period, $k \in \mathbb{Z}$. The legends are omitted for space reasons, please refer to figure 4.13.

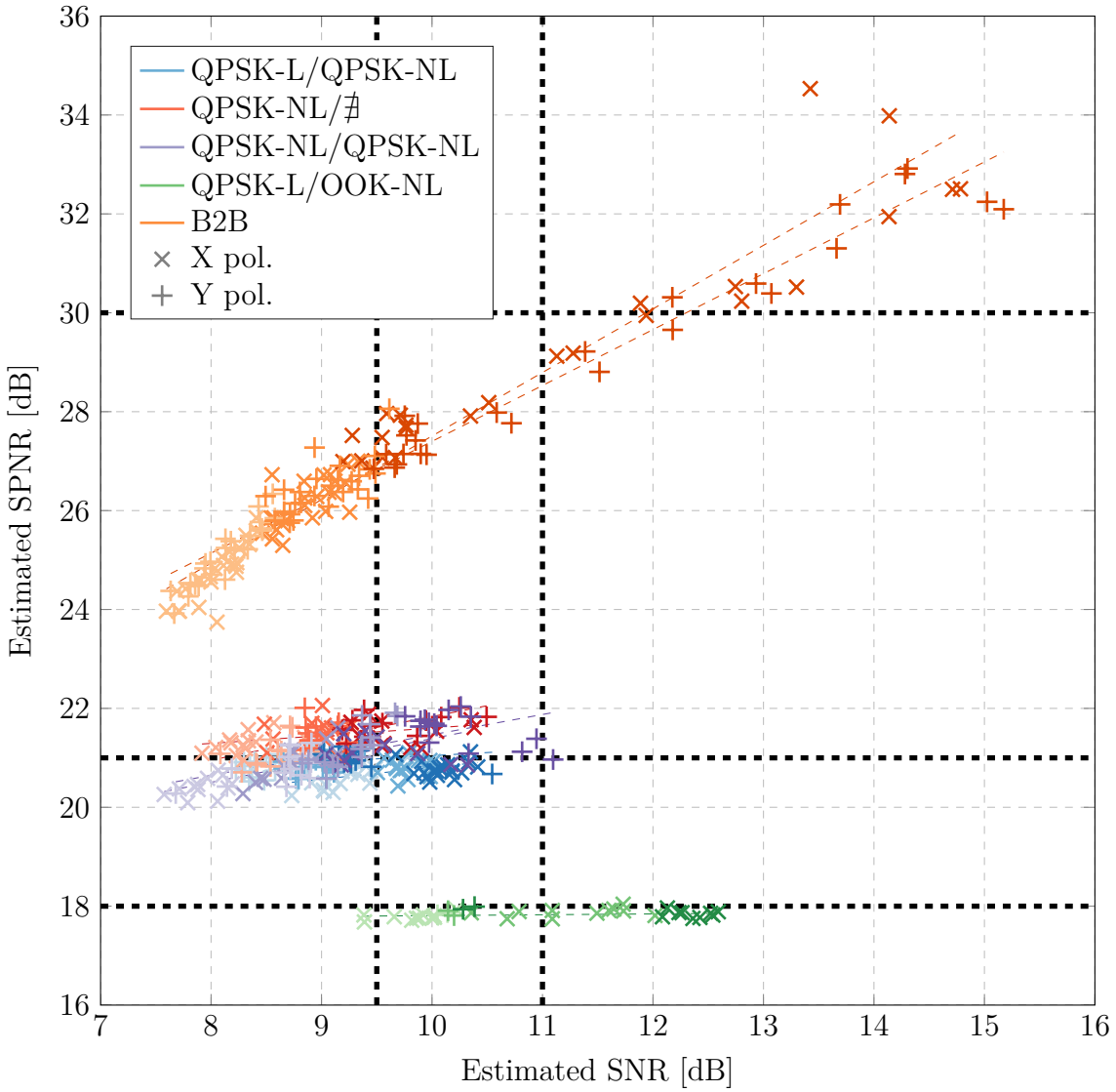

Figure 4.15.: Estimated SPNR vs. estimated SNR (see also table 4.5). Each experiment has a colour whose intensity increases with the OSNR: values corresponding to shots taken at lower OSNRs have a brighter colour, whereas values corresponding to shots taken at higher OSNRs have a darker one. Linear regressions are also shown (dashed lines). The two polarisations are shown separately.

In particular, the estimated SNR increases with the OSNR, but at different rates: faster for the B2B experiments, slower for the four experiments with the transmission link (the estimated SNRs span ranges with different lengths). The reason of this behaviour will be explained in section 4.3.2.3. This is the first part of the results I anticipated in section 3.4.1.

4.3.2.2. Estimation of the setup SNR

As said in section 2.2.1, there are many sources of AWGN, which here I can divide into

1. coming from the setup (i.e., the transmitter, the receiver as well as the DSP),
2. coming from the optical amplifiers (i.e., the ASE) and
3. coming from the optical fibre (i.e., the non-linear effects).

In the B2B experiment the third source is not present and the second can be made negligible (recall that the OSNR was set in the receiver, see section 4.2.3), so this experiment offers a good opportunity to estimate the variance of the first source of AWGN, which I call the *"setup AWGN"*.

To this end I hence took a couple of shots at an OSNR = 36.9 dB, the highest I could reach: the corresponding estimated SNRs on the two polarisations were 19.8, 19.6, 19.7 and 19.7 dB.

Modelling the estimated SNR in the B2B experiment as (in a similar way as in [22–24])

$$\hat{\mathrm{SNR}} = \frac{E_s}{\hat{N}_0} = \frac{E_s}{N_{0,\mathrm{ASE}} + N_{0,\mathrm{setup}}}, \tag{4.4}$$

where $N_{0,\mathrm{ASE}}$ is the *"ASE AWGN"* variance and $N_{0,\mathrm{setup}}$ is the setup AWGN variance, I can solve for $N_{0,\mathrm{setup}}$ obtaining

$$N_{0,\mathrm{setup}} = \frac{E_s}{\hat{\mathrm{SNR}}} - N_{0,\mathrm{ASE}} = \frac{E_s}{\hat{\mathrm{SNR}}} - \frac{E_s}{\mathrm{OSNR}} \cdot \frac{28}{12.5}. \tag{4.5}$$

With $E_s = 1$, $\mathrm{OSNR} = 10^{36.9/10}$ and the four values of estimated SNR reported above and converted to the linear domain, (4.5) leads to values corresponding to *"setup SNRs"* $= 10\log_{10}(E_s/N_{0,\mathrm{setup}}) = 20.0$, 19.8, 19.8 and 19.9 dB, the average being 19.9 dB. It is reasonable that these values are only slightly better than the estimated SNRs: at an OSNR = 36.9 dB the ASE AWGN variance is very small and it can only degrade the setup SNR by a couple of tenths of dB.

4.3.2.3. Estimation of the non-linearities SNR

The reasoning done in section 4.3.2.2 can be extended modelling the estimated SNR as

$$\hat{\mathrm{SNR}} = \frac{E_s}{\hat{N}_0} = \frac{E_s}{N_{0,\mathrm{ASE}} + N_{0,\mathrm{setup}} + N_{0,\mathrm{NL}}}, \tag{4.6}$$

where $N_{0,\mathrm{NL}}$ is the new term representing the *"non-linearities AWGN"* variance. Assuming that

- the setup AWGN variance $N_{0,\text{setup}}$ estimated in section 4.3.2.2 remains more or less constant (which is reasonable for the transmitter and the receiver but not necessarily for the DSP, because the latter is now impaired also by the non-linearities AWGN and might perform worse) and that

- the non-linearities AWGN variance $N_{0,\text{NL}}$ remains constant, since it depends on the launch power which was fixed,

when the OSNR increases only the ASE AWGN variance $N_{0,\text{ASE}}$ decreases. As a result, the estimated SNR does not increase indefinitely, rather it saturates to $E_s/(N_{0,\text{setup}} + N_{0,\text{NL}}) = \text{const}$.

It is hard to tell exactly how much each source contributes to the AWGN. However, a rough attempt can still be done: with an *"ASE SNR"* $\approx$ 16.5 dB (equivalently, an OSNR = SNR + $10\log_{10}(28/12.5)$ = 20.0 dB) and a setup SNR $\approx$ 20.0 dB, it takes

- a *"non-linearities SNR"* $\approx$ 13.3 dB to have an estimated SNR $\approx$ 11.0 dB (in the three QPSK-only experiments) and

- a non-linearities SNR $\approx$ 17.5 dB to have an estimated SNR $\approx$ 13.0 dB (in the QPSK-L/OOK-NL experiment).

While, as said, these numbers should not be taken as exact, they do show a trend: heterogeneous inter-channel non-linearities contribute less to the AWGN than intra- and homogeneous inter-channel non-linearities, for the same OSNR (i.e., for the same ASE, with a constant setup SNRs).

4.3.2.4. Estimated SPNR range

Considering both polarisations, the estimated SPNR lies

- in the range 20.2–21.3 dB in the QPSK-L/QPSK-NL experiment,

- in the range 20.7–22.1 dB in the QPSK-NL/$\nexists$ experiment,

- in the range 20.1–22.0 dB in the QPSK-NL/QPSK-NL experiment,

- in the range 17.7–18.0 dB in the QPSK-L/OOK-NL experiment and

- in the range 23.7–34.5 dB in the B2B experiment.

Even the estimated SPNR increases with the OSNR at different rates for the B2B experiment and the four experiments with the transmission link. The reason of this behaviour will be explained in section 4.3.2.6. This is the second part of the results I anticipated in section 3.4.1.

4.3.2.5. Estimation of the setup SPNR

As said in section 2.3.1, there are many sources of PWGN, which here I can divide into

1. coming from the setup (i.e., the transmitter, the receiver as well as the DSP),
2. coming from the optical fibre (i.e., the non-linear effects).

In the B2B experiment the second source is not present, so this experiment offers again a good opportunity to estimate the variance of the first source of PWGN, which I call the *"setup PWGN"*.

This time I do not need to make use of the shots at an OSNR = 36.9 dB, because I do not need to make the contribution of the optical amplifiers negligible; moreover, figure 4.15 shows that, in the B2B experiment, the *"setup SPNR"* is not constant but rather it increases with the OSNR. I hence decided to consider instead an exemplary estimated SNR = 12 dB, halfway between the values considered in section 4.3.2.3 when estimating the non-linearities SNR and as such of more practical interest. In the B2B experiment, this value corresponds to a setup SPNR $\approx$ 30 dB.

Since the LASERs in the transmitter and in the receiver produce a constant PWGN, the dependence of the estimated SPNR on the OSNR can only be due to the DSP: the lower the estimated SNR, the more troubles the DSP has converging and thus the less phase noise it compensates for (and possibly also the more phase noise it introduces).

4.3.2.5.1. LASER SPNR The shots from the B2B experiment at an OSNR = 36.9 dB, however, can still prove useful. The corresponding estimated SPNRs on the two polarisations were 37.0, 38.9, 35.7 and 35.9 dB, the average being 36.9 dB.

LASER phase noise (before DSP) is usually modelled as a *random walk*, a stochastic process with stationary independent increments. These phase noise increments are modelled with a real, Gaussian random variable, with zero mean and variance $\Theta_{0,\mathrm{LASER}}$ which depends on LASER linewidth Δ_f [53–55]:

$$\Theta_{0,\mathrm{LASER}} = 2\pi\Delta_f T. \tag{4.7}$$

In my laboratory experiments I used 2 LASERs at the transmitter and the receiver with a linewidth $\Delta_f = 100$ kHz, while the symbol period was $T_s = 1/(28{\cdot}10^9) \approx 35.7$ ps. Since the two LASERs were independent, the variances of their phase noise add up leading to a *"LASER SPNR"* $\approx$ 43.5 dB, in the same order of magnitude of the estimated SPNRs in the B2B experiment at an OSNR = 36.9 dB.

4.3.2.6. Estimation of the non-linearities SPNR

Similar to section 4.3.2.3, I can model the estimated SPNR as

$$\hat{\mathrm{SPNR}} = \frac{E_s}{\hat{\Theta}_0} = \frac{E_s}{\Theta_{0,\mathrm{setup}} + \Theta_{0,\mathrm{NL}}}, \tag{4.8}$$

where $\Theta_{0,\mathrm{setup}}$ is the setup PWGN variance and $\Theta_{0,\mathrm{NL}}$ is the *"non-linearities PWGN"* variance. Assuming that

	...SNR [dB]	...SPNR [dB]	Comments
ASE...	16.5	n.r.	at OSNR = 20 dB
setup...	19.9*	30.0	
non-linearities...	13.3*	21.6*	in the three QPSK-only exp.
	17.5*	18.3*	in the QPSK-L/OOK-NL exp.

Table 4.4.: (ASE,) setup and non-linearities SNRs and SPNRs according to (4.6) and (4.8). The values marked with an asterisk are considered constant.

- the non-linearities PWGN variance $\Theta_{0,\mathrm{NL}}$ remains constant, since it depends on the launch power which was fixed,

when the OSNR increases only the setup PWGN variance $\Theta_{0,\mathrm{setup}}$ decreases. As a result, the estimated SPNR does not grow indefinitely, rather it saturates to the *"non-linearities SPNR"* $= E_s/\Theta_{0,\mathrm{NL}} = \mathrm{const}$.

With a setup SPNR $\approx$ 30.0 dB, it takes

- a non-linearities SPNR $\approx$ 21.6 dB to have an estimated SPNR $\approx$ 21.0 dB (in the three QPSK-only experiments) and

- a non-linearities SPNR $\approx$ 18.3 dB to have an estimated SPNR $\approx$ 18.0 dB (in the QPSK-L/OOK-NL experiment).

As in section 4.3.2.3, these numbers do not pretend to be exact but at least show a trend: heterogeneous inter-channel non-linearities contribute more to the PWGN than intra- and homogeneous inter-channel non-linearities, for the same estimated SNR (with a constant setup SPNR), even with a *lower* launch power.

Table 4.4 summarizes these results and those found in sections 4.3.2.2, 4.3.2.3 and 4.3.2.5.

4.3.2.7. Lower limits of operating ranges

The results from sections 4.3.2.1 and 4.3.2.4 can be reformulated in terms of "*lower* limits of operating ranges": based on the laboratory experiments, I expect these limits to be around

- SNR = 11.0 dB $\pm$ 2.0 dB and SPNR = 18 dB, for *old* 10G/100G WDM systems, in which the central channel can be impaired by heterogeneous inter-channel non-linearities (the QPSK-L/OOK-NL experiment), and

- SNR = 9.5 dB $\pm$ 2.5 dB and SPNR = 21 dB $\pm$ 1 dB, for *modern* 100G WDM systems, in which the central channel can be impaired by homogeneous inter-channel non-linearities and/or (homogeneous) intra-channel non-linearities (the three QPSK-only experiments).

Laboratory experiment	Representing	Estim. SNR [dB]	Estim. SPNR [dB]
QPSK-only	modern 100G WDM systems	9.5 ± 2.5	21 ± 1
QPSK-L/OOK-NL	old 10G/100G WDM systems	11.0 ± 2.0	18

Table 4.5.: Estimated SNR and the estimated SPNR as lower limits of operating ranges (see also figure 4.15).

Table 4.5 summarizes these limits.

Note that the adjective "lower" is due to the fact that in the laboratory experiments I purposedly operated the systems in highly non-linear regime, to find these limits; practical system will never be operated in such extreme conditions.

4.4. Summary

In this chapter I introduced five laboratory experiments, four with a transmission link between transmitter and receiver and one in B2B. At the transmitter,

- the central channel, a 100G long-haul coherent optical communications system, was always PDM QPSK modulated with a baud rate = 28 GBaud;
- the eight neighbouring channels were either as the central one or 10G long-haul incoherent optical communications systems, in which case they were OOK modulated with a baud rate = 10.7 GBaud.

The central and/or the neighbouring channels were operated either in linear or in highly non-linear regime, in the latter case stimulating (homogeneous) intra-channel non-linearities and/or homogeneous and/or heterogeneous inter-channel non-linearities. The transmission link was represented by 380 km of SMF without DCF, with an EDFA after every 95 km span to compensate for the span losses. At the receiver, the OSNR was set and the central channel was filtered out, combined with a LO, detected and sampled. Finally, a blind DSP was used.

Analysing the pdfs as well as the auto- and the cross-correlation functions of the real and imaginary parts of the CRT received constellations I showed that

1. the additive and phase noises are, in fact, Gaussian, (almost always) white, uncorrelated with each other (and hence independent) and uncorrelated with the transmitted symbols,

confirming the statistical assumptions made in chapter 2.

Applying the estimators of chapter 3 to the CRT received constellations, I showed that

2. three addends contribute to the AWGN:

- the ASE AWGN (whose variance depends on the OSNR),
- the setup AWGN (whose variance is assumed constant) and
- the non-linearities AWGN (whose variance is assumed constant).

At an OSNR = 20 dB, the latter were responsible for the largest contribution;

3. two addends contribute to the PWGN:
   - the setup PWGN (whose variance depends on the OSNR) and
   - the non-linearities PWGN (whose variance is assumed constant).

   At an estimated SNR = 12 dB, the latter were responsible for the largest contribution;
4. heterogeneous inter-channel non-linearities contribute less to the AWGN than intra- and homogeneous inter-channel non-linearities, for the same OSNR;
5. heterogeneous inter-channel non-linearities contribute more to the PWGN than intra- and homogeneous inter-channel non-linearities, for the same estimated SNR, even with a smaller launch power. Most importantly,
6. even when the system is operated under unrealistic non-linear conditions the estimated SPNR is between 18 and 22 dB.

This latter information in particular will be useful in chapter 5 to choose the alternative constellation and to decide whether the phase noise can be neglected and, if so, to develop solutions based on the much simpler NDE AWGN model and DE AWGN model.

5. Theoretical Limits

In this chapter I present the mutual information (MI) between the input and the output of two of the models presented in chapter 2, namely

- the non differentially encoded (NDE) additive white Gaussian noise (AWGN)/ wrapped phase white Gaussian noise (wPWGN) model and
- the differentially encoded (DE) AWGN/ wPWGN model.

Working with the MI has the nice benefit of removing one variable from the equation, namely the forward error correction (FEC) solution, since the MI assumes an *ideal* code (hence the adjectives *theoretical* and, later on, *potential*). Note that the assumption that the noises and hence the received symbols are continuous corresponds to having an ideal *soft-decision (SD)* code.

As such, the MI depends on the constellation (or the transition constellation) as well as on the signal-to-(*additive*-)noise ratio (SNR) and the signal-to-*phase*-noise ratio (SPNR). I hence calculated (or, better said, estimated) the MI for various SNR, SPNR and, most importantly, nine different constellations for each model.

The purpose is twofold:

1. to understand if and where (i.e., for which values of SNR and SPNR) the phase noise can be neglected. If it can, it means that I can use the simpler NDE AWGN model and DE AWGN model for my proposed solutions, which greatly simplifies the implementation of the SD demapper;
2. to be able to compare the theoretical performance of these constellations and hence to choose two for the proposed solutions, in general over various SNR and SPNR and in particular in the ranges anticipated in section 3.4.1 and individuated in chapter 4 applying the estimators presented in chapter 3 to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems.

I will start with the MI over the NDE AWGN/ wPWGN model, introducing its Monte Carlo (MC) estimation the MI (since it could not be obtained analytically), the constellations considered and, finally, the results. One of the constellations is the quadrature phase-shift keying (QPSK), representing the *traditional* solution; I will hence present the results first in *abolute* terms of the nine constellations and then in *relative* terms as potential gain over the QPSK of the other eight *alternative* constellations.

I will then repeat the same steps for the MI over the DE AWGN/ wPWGN model, highlighting similarities and differences between the two.

Part of the material presented in this chapter has been published on the IEEE Photonics Technology Letters (PTL) [12, 13], at the Optical Fiber Conference (OFC) [14], at the European Conference on Optical Communications (ECOC) [15], at the Signal Processing in Photonic Communications (SPPCom) conference [16] as well as at the International Conference on Transparent Optical Networks (ICTON) [17].

5.1. Theoretical limits of 100G long-haul NDE coherent optical communications systems

5.1.1. NDE AWGN/wPWGN model: MC estimation of the MI

The MI between the transmitted symbol X (the input of the channel) and the received one Y (the output of the channel) can be written as the expectation over the transmitted symbol X and the received symbol Y of the following function [56],

$$\log_2\left(\frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right), \tag{5.1}$$

where at the numerator is the conditional probability density function (pdf) of the received symbol Y conditioned on the transmitted symbol X $p_{Y|X}(y|x)$ and at the denominator is the pdf of the received symbol Y $p_Y(y)$. Note that both pdfs are in this case just functions applied to the transmitted and the received symbols X and Y. This is where the choice of the model comes into play: in this section, since I am using the NDE AWGN/ wPWGN model, I will use (2.13) for the conditional pdf $p_{Y|X}(y|x)$ and calculate the pdf $p_Y(y)$ as indicated at the end of section 2.3.2.

For each constellation, SNR and SPNR

- I generated $N_{\mathrm{MC}} = 20000$ realisations of
  - the transmitted symbols X according to (2.2) for the discrete constellations and according to (2.5) for the Gaussian one;
  - the AWGN N according to (2.6), using the variance N_0 corresponding to the chosen SNR;
  - the wPWGN Θ according to (2.10), using the variance Θ_0 corresponding to the chosen SPNR.

  In section 5.1.4.1 I will show that $N_{\mathrm{MC}} = 20000$ was a good trade-off between the precision of the MC estimation and the execution time, which increases more or less linearly with the size of the constellation, since the numerical integration in (2.13) must be done $|\mathcal{X}|$ times;

- I obtained the corresponding realisations of the received symbols Y according to (2.8);

- I calculated the pdfs in (5.1).

  When implementing (2.10) I noticed that, for the values of SPNR considered in this work (see section 5.1.2), it was sufficient to consider only the Gaussian bell centered around 0 and its two neighbouring replicas centered around $\pm\pi$ (i.e., $-1 \leq k \leq +1$ – see also section 3.1.2).

  The integral in (2.13) cannot be solved analytically, so I resorted to numerical integration [57].

  In order to calculate the pdf $p_Y(y)$ of the Gaussian constellation, I had to numerically integrate also over x [58, 59];

- I obtained the corresponding realisations of the MI I according to (5.1); finally,

- I took an average over the latters, thus estimating the MI via MC:

$$\hat{I} = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} \log_2\left(\frac{p_{Y|X}\left(y_i|x_i\right)}{p_Y(y_i)}\right). \tag{5.2}$$

5.1.2. NDE AWGN/wPWGN model: constellations as well as SNR and SPNR ranges considered

Figure 5.1 shows the nine constellations I decided to consider for 100G long-haul NDE coherent optical communications systems.

The 4-star is sometimes called 2 amplitude- 2 phase-shift keying (2A2P); however I find this name misleading, since the constellation is not given by the Cartesian product of two amplitudes and two phases as is, for instance, the 2 amplitude- 4 phase-shift keying (2A4P), for which I adopt the conventional name. Similarly, the 8-star is sometimes called 8 quadrature amplitude modulation (QAM); this name refers in the literature to constellations of various shape, so I will avoid it.

The choice of these constellations is motivated as follows:

- naturally, I started with the QPSK because it represents the classical approach;

- I decided to include two more constellations with four points, the 4-star and the 4 pulse amplitude modulation (PAM), to show the dependency of the MI not only on the *size* but also on the *shape* of the constellation, since all constellations with four points allocate space for the same amount of redundancy as the QPSK. Intuitively, a constellation with more rings and less points per ring will perform better at high SNR and low SPNR;

- I decided to focus mainly on 8-ary constellations and investigate in particular four constellations with eight points, the 8 phase-shift keying (PSK), the 8-star, the 8PAM and the 2A4P, since I expect the potential gain over the QPSK to increase less than linearly with the constellation size. However, just choosing to move to *any* constellation with eight points is not enough, for even in this case constellations of

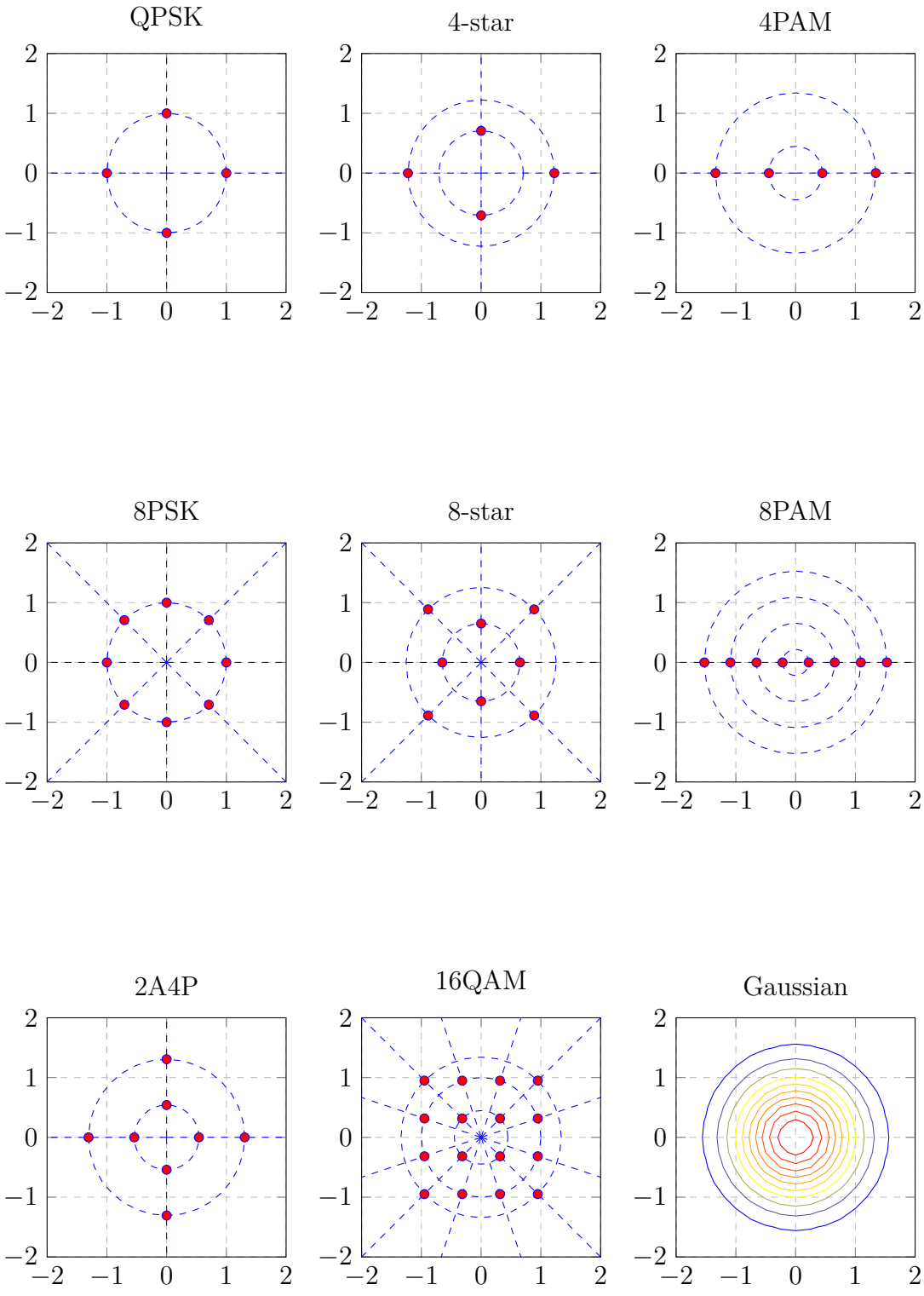

Figure 5.1.: The nine constellations considered for 100G long-haul NDE coherent optical communications systems.

Name	R_2/R_1	R_3/R_2	R_4/R_3	R_1
4-star	$\sqrt{3}$	n.a.	n.a.	≈ 0.707
4PAM	$3/1$	n.a.	n.a.	≈ 0.447
8-star	$(1+\sqrt{3})/\sqrt{2}$	n.a.	n.a.	≈ 0.650
8PAM	$3/1$	$5/3$	$7/5$	≈ 0.218
2A4P	$1+\sqrt{2}$	n.a.	n.a.	≈ 0.541
16QAM	$\sqrt{5}$	$3/\sqrt{5}$	n.a.	≈ 0.447

Table 5.1.: Radii ratios of the constellations with more than one ring considered for 100G long-haul NDE coherent optical communications systems and reported in figure 5.1.

various shapes are expected to perform differently in an AWGN or in a wPWGN dominated system;

- I decided to consider the 16QAM because it is the most representative constellation with 16 points for NDE systems; finally,
- I decided to add also the Gaussian constellation because it is known to achieve the maximum MI over the NDE AWGN model and, although the same has not been rigorously proven yet for the NDE AWGN/ wPWGN model, its circular symmetry, together with its uniform phase distribution, suggests that this might still be the case. In fact, for constant-intensity, constant-envelope, or ring modulation, the capacity-achieving phase distribution is known to be uniform in $[-\pi, +\pi)$ [26].

The radii of the constellations with more than one ring are the classical one chosen for the NDE AWGN model, i.e., those that guarantee that the three closest neighbouring points are equidistant, thus maximizing the MI.

Table 5.1 summarizes these radii ratios.

Ideally, I wanted the wPWGN variance Θ_0 to go from 0 to $\pi^2/36$: in the first case there would be no wPWGN, whereas in the second only a very "unlucky" QPSK point would land on one of its neighbours, since

$$-1.57 \approx -\frac{\pi}{2} = -3\sqrt{\Theta_0} \leq \Theta \leq +3\sqrt{\Theta_0} = +\frac{\pi}{2} \approx +1.57 \tag{5.3}$$

99.73% of the times.

However, dividing by zero is never a good idea when performing numerical integration (recall that the variance is in the denominators in the Gaussian pdf). I therefore took 16 values between 0.01 and 1.51 with a 0.1 step, I divided them by three and I elevated them to the power of two, obtaining 16 values between $\approx 1.1 \cdot 10^{-5}$ and ≈ 0.25, this time with a non-linear step because of the squaring operation. These values corresponds to 16 values of SPNR between ≈ 5.96 dB and ≈ 49.54 dB, also with a non-linear step.

Table 5.2 reports the limits of the intervals in which the random variable Θ would be 99.73% of the times as well as the corresponding variances Θ_0 and SPNRs in dB.

$3\sqrt{\Theta_0}$	Θ_0	SPNR [dB]	$3\sqrt{\Theta_0}$	Θ_0	SPNR [dB]
$0.01 \approx 0.003\pi$	0.00001	49.54*	$0.81 \approx 0.258\pi$	0.07290	11.37
$0.11 \approx 0.035\pi$	0.00134	28.71	$0.91 \approx 0.290\pi$	0.09201	10.36*
$0.21 \approx 0.067\pi$	0.00490	23.10	$1.01 \approx 0.321\pi$	0.11334	9.46
$0.31 \approx 0.099\pi$	0.01068	19.72*	$1.11 \approx 0.353\pi$	0.13690	8.64
$0.41 \approx 0.131\pi$	0.01868	17.29	$1.21 \approx 0.385\pi$	0.16268	7.89*
$0.51 \approx 0.162\pi$	0.02890	15.39	$1.31 \approx 0.417\pi$	0.19068	7.20
$0.61 \approx 0.194\pi$	0.04134	13.84*	$1.41 \approx 0.449\pi$	0.22090	6.56
$0.71 \approx 0.226\pi$	0.05601	12.52	$1.51 \approx 0.481\pi$	0.25334	5.96*

Table 5.2.: Values of SPNR considered for the MI and corresponding limits of the 99.73% interval. The values marked with an asterisk are those reported in the figures 5.2, 5.3, 5.4, 5.7, 5.8 as well as 5.11, 5.12 and 5.14.

The choice of the SNR range was more pragmatic: for each constellation I started with some random values and then iteratively refined the range with a 0.5 dB step, expanding or contracting it in both directions until either the MI was between ≈ 0.5 and $\approx m$ bits per channel use (bpcu), where m is the number of transmitted *encoded* bits carried by a transmitted symbol X, see (2.3), or the curves flattened out (in some cases the SPNR would prevent the MI from reaching $\approx m$ bpcu, even for very high SNR). A notable exception was the Gaussian constellation: given its importance, I looked for SNR values that lead to MI values between ≈ 0 bpcu and ≈ 2.5 bpcu.

5.1.3. NDE AWGN/wPWGN model: values of interest

The MI depends on three variables: the constellation, the SNR and the SPNR. Similar to section 3.4.1, I hence decided that in section 5.1.4 I will show the results in four figures, fixing either thrice the constellation or once the SNR and parametrizing the curves with either the SPNR or the constellations.

Swapping the roles of MI and SNR, the potential gain over the QPSK depends on three variables: the constellation, the targeted MI and the SPNR. Correspondingly, in section 5.1.5 I will show the results in three figures, fixing twice the constellation and parametrizing the curves with the SPNR and fixing once the MI and parametrizing the curves with the constellations.

Contrarily to section 3.4.1, in both cases the figures for a fixed SPNR would have had the constellations on the x-axis, which cannot be ordered as for instance can real numbers; I hence decided to skip this combination. When a variable is fixed in one figure, I will highlight it in the others in which it is not (with a thicker curve when it parametrises a curve or with a vertical or horizontal black dashed line when it is on one of the axes). In addition, I will also highlight the SPNR, although as said I will not fix it.

The decision on which value to fix each variable is motivated as follows:

1. I chose to fix the constellation to three “values”:

Fixed variable	Value
Constellation	QPSK, 8-star, Gaussian
SNR	6.00 dB
MI	1.87 bpcu
SPNR	18.00 dB, 21.00 dB, 30.00 dB and 19.72 dB

Table 5.3.: Variables fixed or highlighted in figures 5.2, 5.3, 5.4, 5.5 and 5.6 as well as in figures 5.7, 5.8 and 5.9.

- the QPSK because it corresponds to the traditional approach;
- the 8-star because, as I will soon show, it is the alternative, among the constellations with 8 points, that offers the highest potential gain over the QPSK;
- the Gaussian constellation because it has certain, unique characteristics, which I will list later on;

2. I chose to fix the SNR to 6 dB, because it is one of the few values, among those for which I estimated the MI, for which *all* constellations, for *all* SPNRs, achieve a MI in the range 1.2–2.4 bpcu, centered roughly around the value chosen for the MI (see next number point);

3. I chose to fix the MI to 1.87 bpcu, because it is the value targeted (per polarisation) by the proposed solutions (I will elaborate on the motivations behind this choice in chapter 6);

4. I chose to highlight three values of the SPNR when it is on the x-axis, 18 dB, 21 dB and 30 dB, because they are the three lower limits of the operating ranges I anticipated in section 3.4.1 and individuated in chapter 4.

   However, I chose to highlight a fourth value of the SPNR when it parametrises a curve, 19.72 dB, because it is the value, among those for which I estimated the MI, closest to these lower limits.

Table 5.3 summarizes which variable I fixed or highlighted to which value in figures 5.2, 5.3, 5.4, 5.5 and 5.6 as well as in figures 5.7, 5.8 and 5.9.

5.1.4. NDE AWGN/wPWGN model: MI

Figures 5.2, 5.3 and 5.4 show the MI of the QPSK, the 8-star and the Gaussian constellation, as a function of the SNR and parametrised by various values of the SPNR.

Figures 5.2 and 5.3 show that the MI of the QPSK and of the 8-star (and the same holds true also for the other discrete constellations) increases monotonically with the SNR (moving from left to right) and the SPNR (moving from bottom to top), but is limited to the *finite* number of bits per symbol m carried by a transmitted symbol. The presence of *moderate* wPWGN can still be compensated by an increased SNR. However,

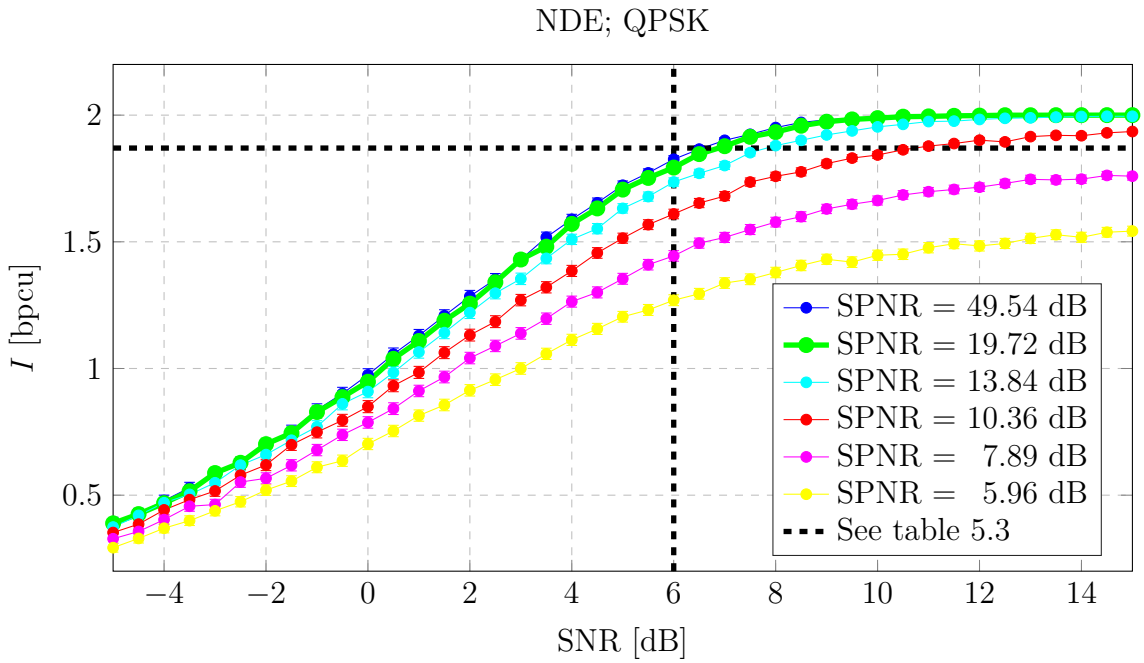

Figure 5.2.: MI of the QPSK as a function of the SNR and parametrised by various values of the SPNR over the NDE AWGN/wPWGN model. Also reported are the confidence intervals, which can be seen more easily in figure 5.6.

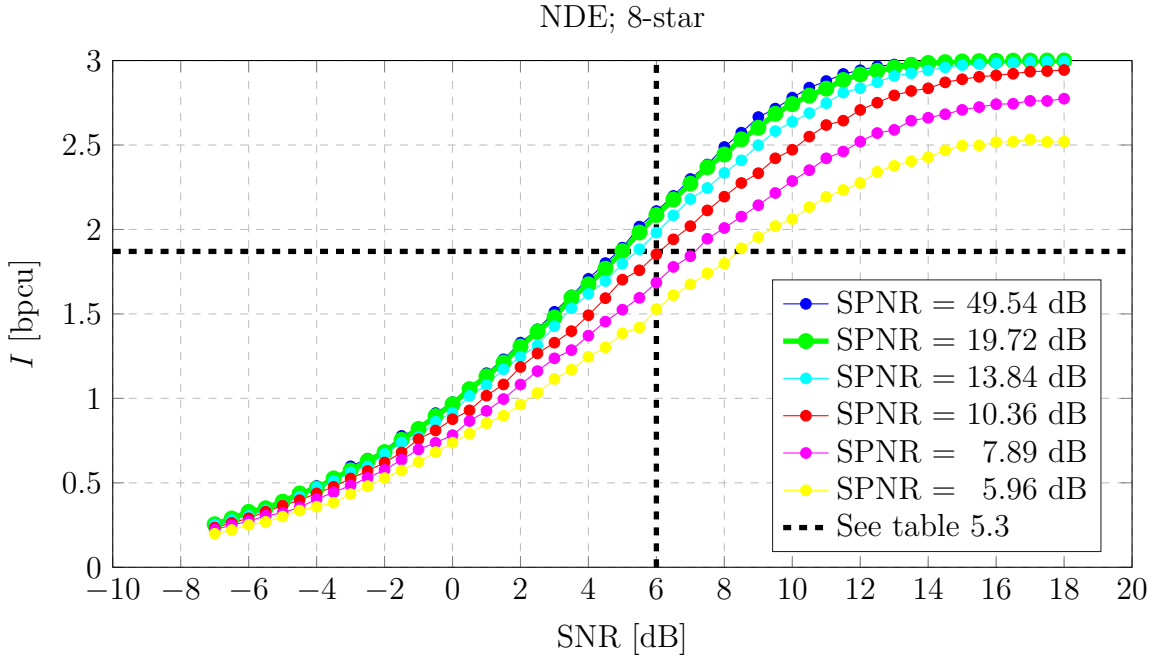

Figure 5.3.: MI of the 8-star as a function of the SNR and parametrised by various values of the SPNR over the NDE AWGN/wPWGN model.

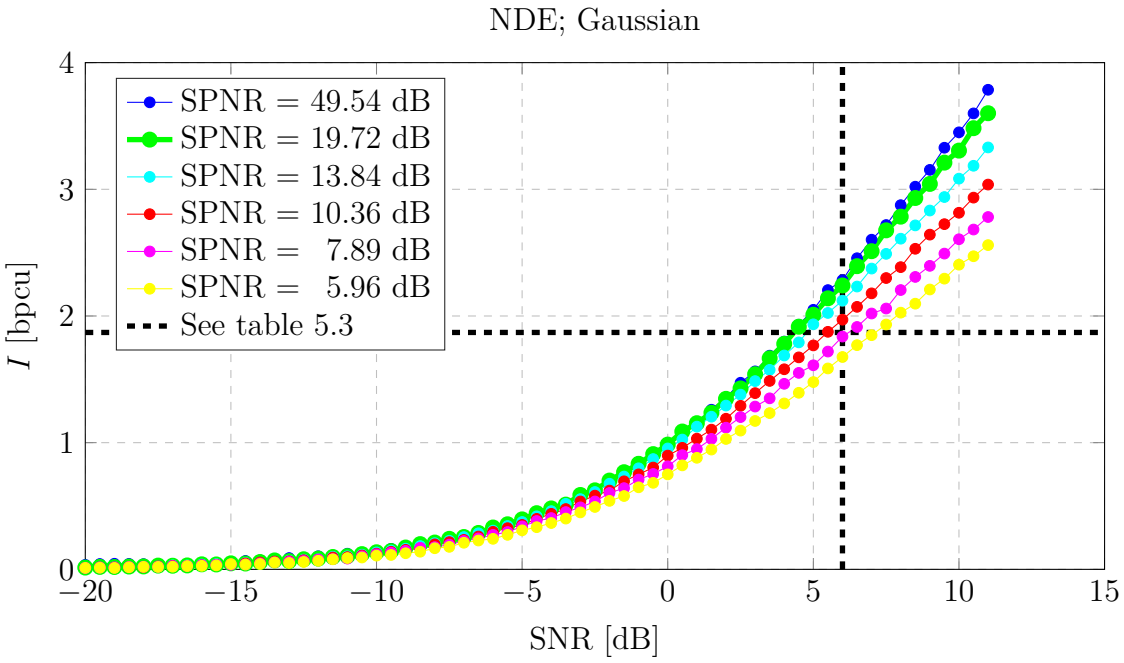

Figure 5.4.: MI of the Gaussian constellation as a function of the SNR and parametrised by various values of the SPNR over the NDE AWGN/wPWGN model.

too much wPWGN can prevent the constellations from reaching values close to the limit $I = m$ bpcu, even for very high SNRs.

Things change in figure 5.4. The Gaussian constellation has infinite cardinality and hence an *infinite* number of bits per symbol m carried by a transmitted symbol. As such its MI

- increases indefinitely for SNR $\to +\infty$ but
- is still limited to the *capacity* of the NDE AWGN model [56],

$$C_{\text{AWGN}} = \log_2(1 + \text{SNR}), \tag{5.4}$$

for SPNR $\to +\infty$.

Also, the presence of wPWGN can *always* be compensated by an increased SNR, at least over the ranges of SNR and SPNR considered here.

Figure 5.5 shows the MI of all constellations, as a function of the SPNR and for a fixed value of the SNR = 6 dB.

Comparing the constellations confirms that what I said for a couple of constellations is actually valid for all of them: the MI increases with the SPNR.

However, this benefit is less pronounced for the PAMs: the 4PAM and the 8PAM are in fact practically insensitive to wPWGN, thanks to the fact that they have only two points per ring. Indeed, they outperform the QPSK for SPNR ≤ 7.5 dB. Also interesting

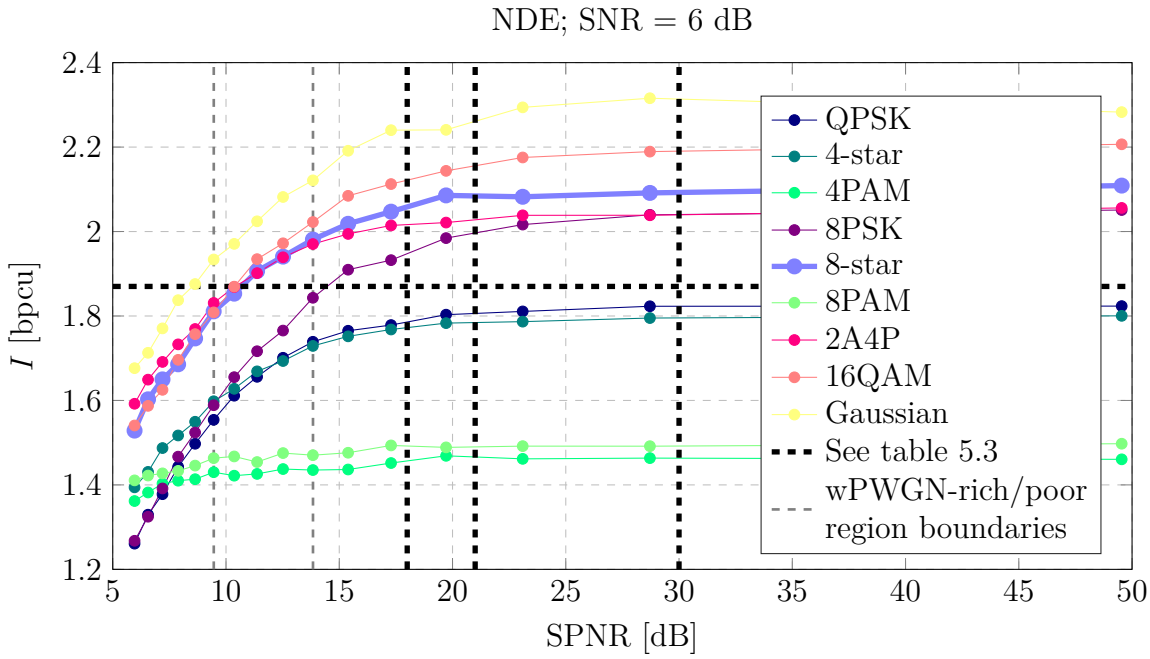

Figure 5.5.: MI of all constellations, as a function of the SPNR and for a SNR = 6 dB over the NDE AWGN/wPWGN model.

to notice is that they achieve almost the same MI, although the 8PAM allows space for one additional encoded bit.

The MI curve of the 16QAM crosses the one of the 2A4P around SPNR = 10.4 dB and overlaps with the one of the 8-star for SPNR $\leq$ 9.5 dB: although it allows space for one additional encoded bit, the 16QAM achieves worse to equal theoretical performance for low SPNRs. This apparently counter-intuitive result can be explained by looking at the 16QAM as a constellation with three rings: one of them has eight points and hence suffers particularly from wPWGN.

A similar reasoning explains why the 8PSK, which achieves the same MI as the 2A4P for SPNR $\geq$ 28.7 dB, has worse performance below this value, up to the point where it has the same performance as the QPSK for SPNR $\leq$ 6.9 dB while allowing space for one additional encoded bit.

The 4-star has slightly worse performance than the QPSK for SPNR $\geq$ 12.5 dB but achieves a higher MI below this value. Again, this is due to the number of points per ring that these two constellations exhibit.

The Gaussian constellation achieves the highest MI at least over the whole range of SPNR and among the constellations considered here, thus partly confirming the assumption made in section 5.1.2.

Among the discrete constellations (i.e., those of interest for practical systems) there is no "clear winner", that is, no single constellation achieves the highest MI over the whole range of SPNR. As said, the MI curve of the 16QAM crosses the one of the 2A4P around SPNR = 10.4 dB:

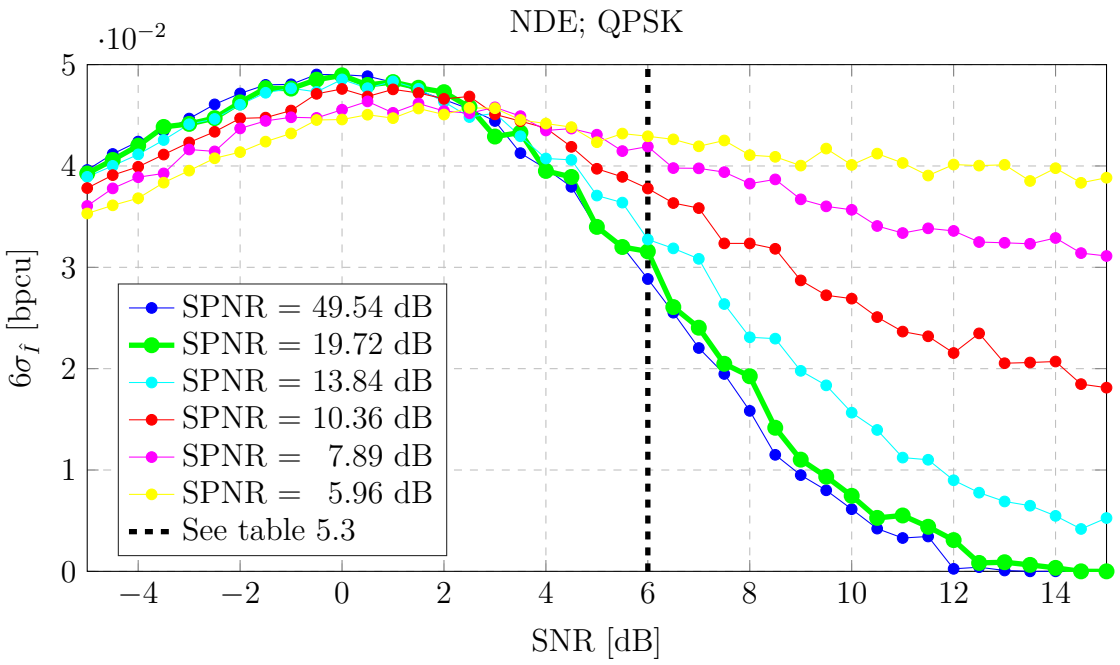

Figure 5.6.: Lengths of the confidence intervals of the estimated MI shown in figure 5.2.

- above this value the 16QAM achieves the highest MI, whereas
- below this value the 2A4P achieves the highest MI.

5.1.4.1. Confidence interval of the MC estimation

The MC estimator (5.2), being a sample mean, tends to a Gaussian distribution for the central limit theorem with variance given by

$$\sigma_{\hat{I}}^2 = \frac{\sigma_I^2}{N_{\mathrm{MC}}}, \tag{5.5}$$

where σ_I^2 is the variance of the realisations of the MI I. Hence, the standard deviation of the MC estimator $\sigma_{\hat{I}}$ decays with $\sqrt{N_{\mathrm{MC}}}$.

Figure 5.2 reports also the confidence intervals of each point, extending from $-3\sigma_{\hat{I}}$ to $+3\sigma_{\hat{I}}$ and containing 99.73% of the realisations. However, since they are hard to see, Figure 5.6 shows directly the length of the confidence intervals 6 $\sigma_{\hat{I}}$ of the MI of the QPSK, as a function of the SNR and parametrised by various values of the SPNR. As it can be seen, they are two orders of magnitude smaller than the values they refer to over the whole range of SNR and SPNR, indicating that the choice $N_{\mathrm{MC}} = 20000$ led to a reliable MC estimation of the MI.

5.1.4.2. NDE AWGN/wPWGN model: wPWGN-rich and wPWGN-poor regions – can wPWGN be neglected?

With the exception of the PAMs, the MI "saturates" (i.e., achieves $\approx 90\%$ of their respective maxima) for 9.46 dB $\leq$ SPNR $\leq$ 13.84 dB (indicated with a vertical gray dashed line in figure 5.5). In other words, I can identify

- a *"wPWGN-rich"* region to the left of this interval, for SPNR $\leq$ 9 dB and
- a *"wPWGN-poor"* region to the right of this interval, for SPNR $\geq$ 14 dB.

Note that these regions have been identified for a SNR = 6 dB. Although not rigorously proven (as said in section 5.1.3, this was one of the few values for which this comparison was possible), it is reasonable to expect the boundaries of the wPWGN-rich and wPWGN-poor regions to shift toward even lower SPNRs for higher SNRs.

In the wPWGN-rich region the effect of wPWGN on the MI must be taken into account; in the wPWGN-poor region, on the other hand, this effect can be neglected.

Recall now the lower limits of the operating ranges anticipated in section 3.4.1 and individuated in chapter 4 (see table 4.5):

- SPNR = 18 dB for old 10G/100G wavelength division multiplexing (WDM) systems and
- SPNR = 21 dB for modern 100G ones.

Both are well inside the wPWGN-poor region. As a consequence, I can use the NDE AWGN model of section 2.2 for the proposed solutions

- without any noticeable loss of precision and, instead,
- with a great simplification of the implementation of the SD demapper, since for this model I have a closed form of the conditional pdfs of the received symbol Y conditioned on the transmitted symbol X, as I showed in section 2.2.2.

5.1.5. NDE AWGN/wPWGN model: potential gain over the QPSK

I define the "potential gain over the QPSK" of a certain constellation as the difference between the SNR needed by the QPSK to achieve a certain MI and the SNR needed by the chosen constellation to achieve the same MI, for the same SPNR. If this quantity is larger than zero, it means that the QPSK needs a higher SNR than the chosen constellation to achieve the same MI, which in turn means that the chosen constellation represents a better alternative. Note that, equivalently, I could have defined the potential gain over the QPSK in terms of E_b/N_0, since the targeted MI is always the same, see (3.33).

Figures 5.7 and 5.8 show the potential gain over the QPSK of the 8-star and the Gaussian constellation, as a function of the targeted MI and parametrised by various values of the SPNR.

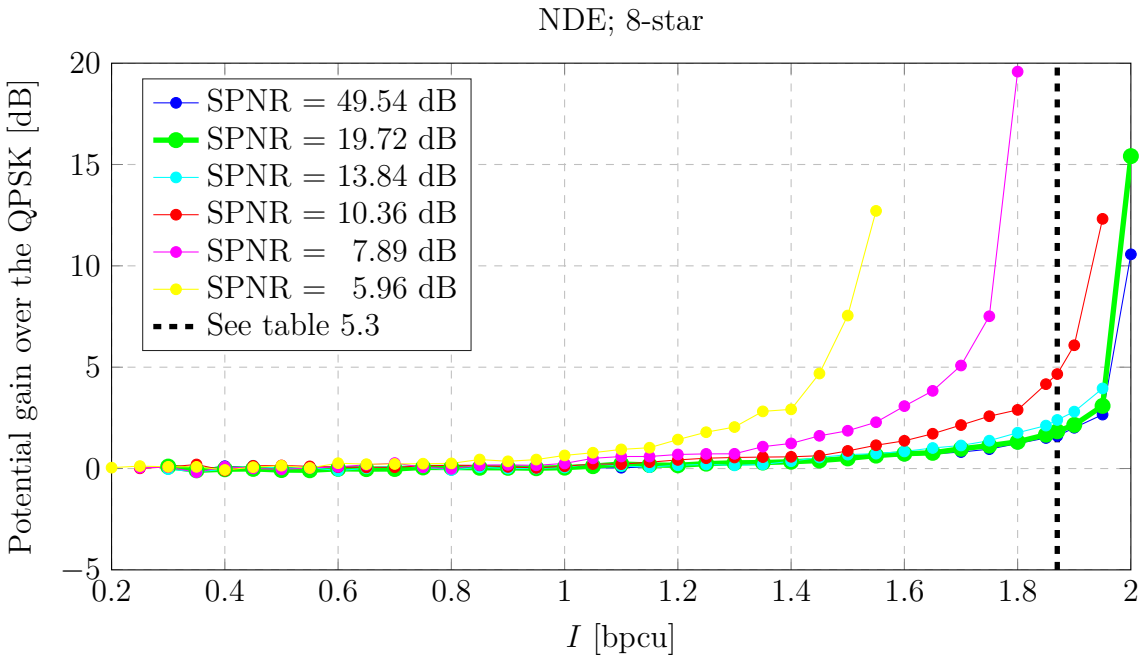

Figure 5.7.: Potential gain over the QPSK of the 8-star as a function of the targeted MI and parametrised by various values of the SPNR over the NDE AWGN/wPWGN model.

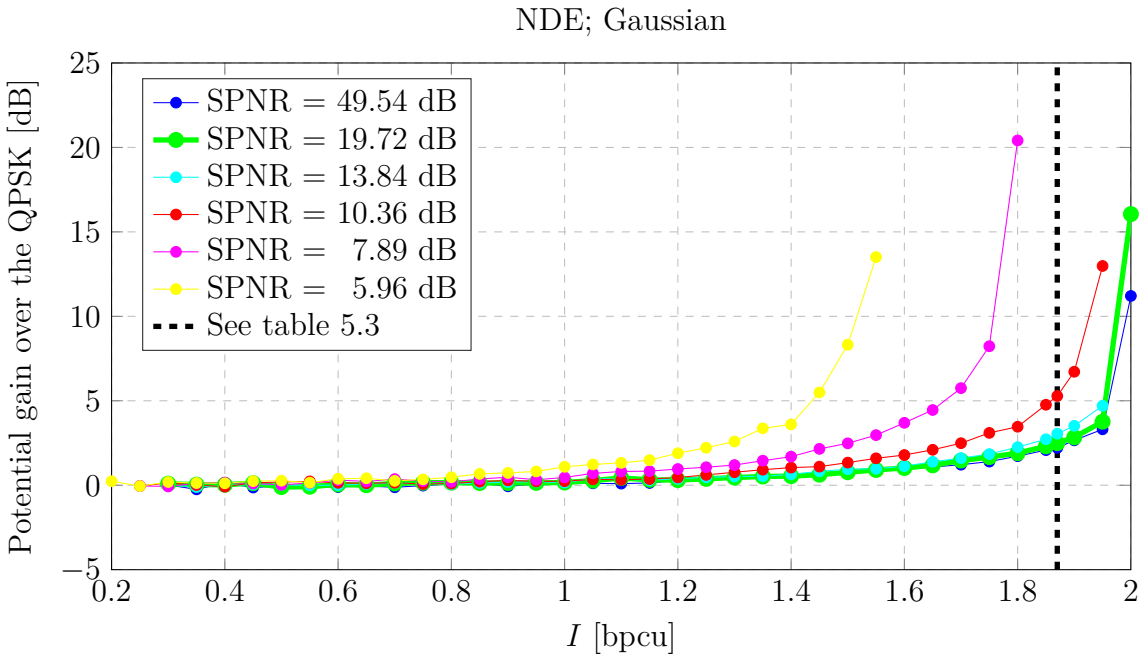

Figure 5.8.: Potential gain over the QPSK of the Gaussian constellation as a function of the targeted MI and parametrised by various values of the SPNR over the NDE AWGN/wPWGN model.

Figure 5.7 shows that the potential gain over the QPSK increases with the targeted MI. This is reasonable: recall that I represents the maximum amount of *information* bits that can be sent, reliably, with a constellation carrying m *encoded* bits and, hence, with a code with code rate $R = I/m$ [56]. This means that

- when $I \to 0$ bpcu, $R \to 0$: both constellations, one with four points (the QPSK) and one with eight points (the 8-star), are strongly (ideally) coded (when $R = 0$ the transmitted bits are only *encoded* bits, no *information* bits are sent!), and indeed in this region the potential gain over the QPSK of the latter is almost zero;
- when $I \to 2$ bpcu, however,
  - $R \to 2/2 = 1$ for a constellation with four points (the QPSK), whereas
  - $R \to 2/3$ for a constellation with eight points (the 8-star):

  the comparison is between an almost *uncoded* constellation with four points (the QPSK) and an still relatively strongly (ideally) coded constellation with eight points (the 8-star), hence in this region the potential gain over the QPSK of the latter is very high.

The potential gain over the QPSK decreases with the SPNR: the QPSK gains more from an increase of the SPNR than the 8-star, especially when it is close to the limit of its MI $I = 2$ bpcu.

Figure 5.8 leads to very similar comments, the only difference being that the absolute values assumed by the potential gain over the QPSK are all slightly higher for the Gaussian constellation, since this constellation achieves the highest MI (or, equivalently, needs the lowest SNR to achieve the same MI), as noticed already in section 5.1.4.

Figure 5.9 shows the potential gain over the QPSK of all constellations, as a function of the SPNR and for a fixed value of the targeted MI $I = 1.87$ bpcu.

For the PAMs the potential gain over the QPSK is negative for SPNR > 10.7 dB and SPNR > 12 dB. This agrees with what I showed in figure 5.5: the PAMs, even the 8PAM with its additional encoded bit, are convenient only in the wPWGN-rich region.

The 4-star has a negative and almost zero potential gain over the QPSK, at least for SPNR > 15.8 dB; it becomes convenient only in the wPWGN-rich region, as I showed already in figure 5.5.

Similarly to figure 5.5, the 16QAM and the 8PSK exhibit a positive potential gain over the QPSK, which is however only slightly higher than, equal to or even lower than the one shown by 2A4P and 8-star.

Finally, the Gaussian constellation offers the highest potential gain over the QPSK.

In conclusion, the constellation with eight points with the highest potential gain over the QPSK is the 8-star, which motivates the choice of this constellation in figures 5.3 and 5.7 and, most importantly, in the proposed solution for 100G long-haul NDE coherent optical communications systems.

A final note on the wPWGN-poor and wPWGN-rich regions. The potential gain over the QPSK decreases with the SPNR and flattens out (achieves $\approx 90\%$ of their respective minima) for 9.46 dB $\leq$ SPNR $\leq$ 17.29 dB; to the right of this interval (roughly in the

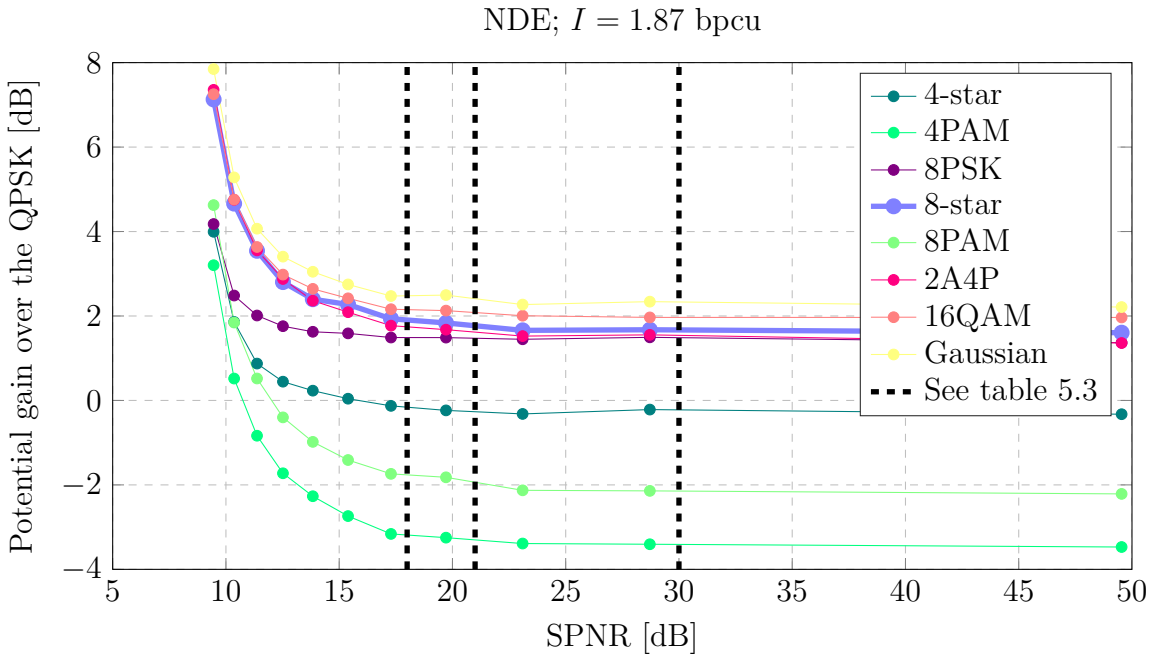

Figure 5.9.: Potential gain over the QPSK of the other constellations, as a function of the SPNR and for a MI $I = 1.87$ bpcu over the NDE AWGN/wPWGN model.

wPWGN-poor region, see section 5.1.4.2) the potential gain over the QPSK practically does not depend on the SPNR and equals the one these constellations exhibit over the NDE AWGN model. Note that, in principle, this could also be due to the wPWGN causing the same penalty to all constellations, so that the potential gain over the QPSK remains constant; however, in figure 5.5 I showed that this is because in the wPWGN-poor region the effect of the wPWGN on the MI can be neglected (see again section 5.1.4).

5.2. Theoretical limits of 100G long-haul DE coherent optical communications systems

In this section I will show the results over the DE AWGN/ wPWGN model; since many concepts are similar I will try to keep the exposition short, highlighting only the parts that changed with respect to section 5.1.

5.2.1. DE AWGN/wPWGN model: MC estimation of the MI

When the communications system is DE I am interested in the MI between the transmitted *transition* T (the input of the channel) and the *pair* of received symbols Y_1, Y_2 (the outputs of the channel). As already explained in section 2.5, this does not mean

that the symbol rate must be doubled, since every received symbol serves in two symbol periods.

In the DE case the MI can be written as the expectation over the transmitted transition T and the received symbols Y_1, Y_2 of the following function [56],

$$\log_2\left(\frac{p_{Y_1,Y_2|T}\left(Y_1, Y_2|T\right)}{p_{Y_1,Y_2}(Y_1, Y_2)}\right), \tag{5.6}$$

where at the numerator is the *joint* conditional pdf of the received symbols Y_1, Y_2 conditioned on the transmitted transition T $p_{Y_1,Y_2|T}\left(y_1, y_2|t\right)$ and at the denominator is the *joint* pdf of the received symbols Y_1, Y_2 $p_{Y_1,Y_2}(y_1, y_2)$. Again, this is where the choice of the model comes into play: in this section, since I am using the DE AWGN/ wPWGN model, I will use (2.44) for the joint conditional pdf $p_{Y_1,Y_2|T}\left(y_1, y_2|t\right)$ and calculate the joint pdf $p_{Y_1,Y_2}(y_1, y_2)$ as indicated at the end of section 2.6.2.

For each constellation, SNR and SPNR value I hence generated $N_{\mathrm{MC}} = 20000$ realisations of the first (previous) transmitted symbol X_1 and of the transmitted transition T according to (2.29), obtained the corresponding second (current) transmitted symbol X_2 according to the DE function (2.28), generated the noise samples N_1, N_2, Θ_1 and Θ_2, obtained the corresponding received symbols Y_1, Y_2 according to (2.8), calculated the pdfs in (5.6), obtained the corresponding realisations of the MI I according to (5.6) and, finally, took an average over the latters.

With respect to what I said in section 5.1.1,

- I did not have to integrate over x, because this time I did not consider the Gaussian constellation;
- the execution time increases more or less quadratically with the size of the constellation, since the numerical integration in (2.44) must be done $2 \cdot |\mathcal{X}|^2$ times.

5.2.2. DE AWGN/wPWGN model: constellations as well as SNR and SPNR ranges considered

Figure 5.10 shows the nine constellations I decided to consider for 100G long-haul DE coherent optical communications systems.

With respect to the constellations considered for NDE systems shown in figure 5.1, I replaced the 16QAM and the Gaussian constellation with two new constellations with 16 points, the 16-star and the 4 amplitude- 4 phase-shift keying (4A4P).

The choice was motivated as follows:

- I decided to remove the 16QAM because only two of the four bits it carries can be used to DE the phase of the transmitted bits [60];
- I decided to remove the the Gaussian constellation because it does not necessarily achieve the maximum MI over the DE AWGN/ wPWGN model anymore (as a matter of facts, it is not even known which constellation achieves the maximum MI over the DE AWGN model [31–34, 61–64]);

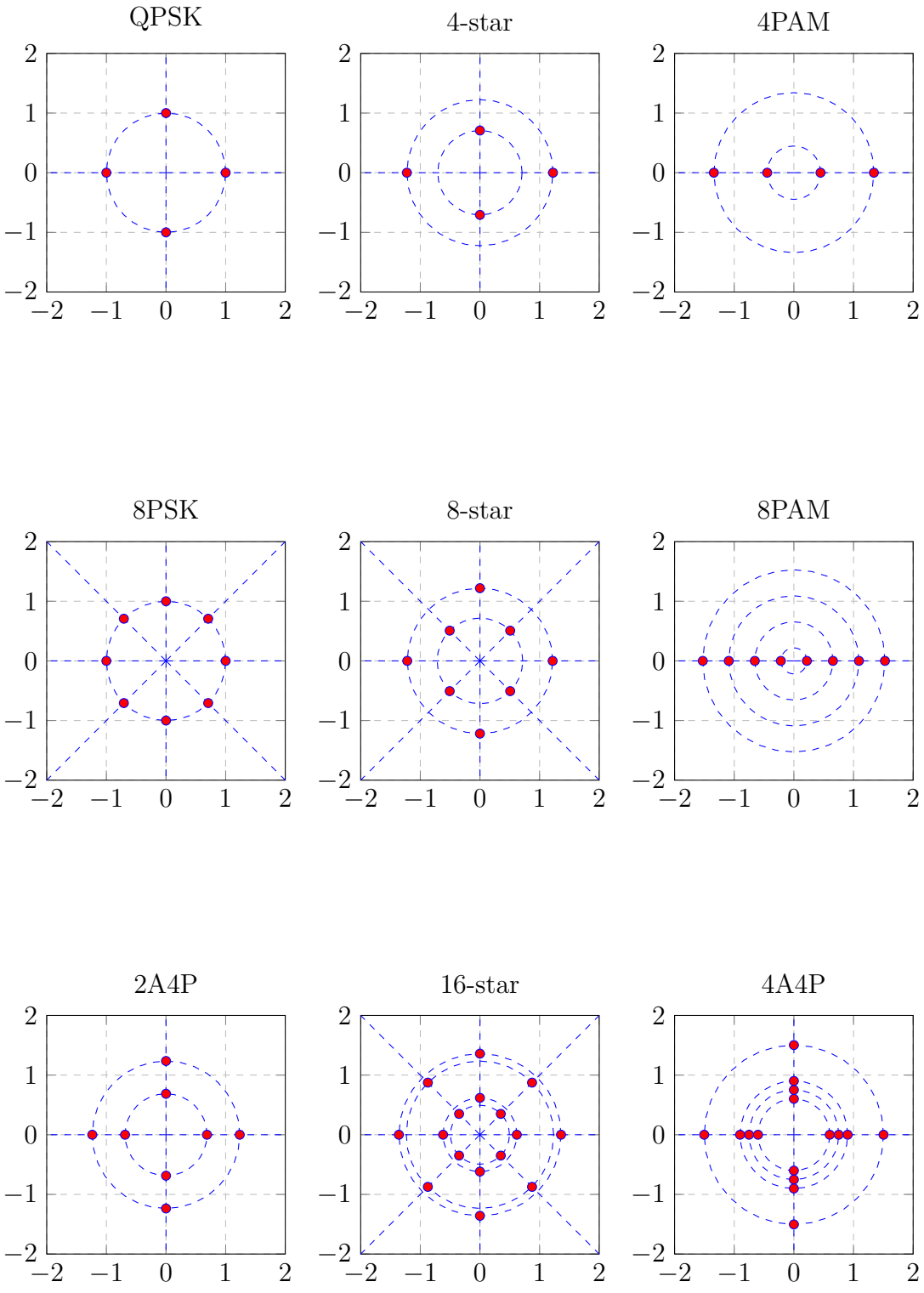

Figure 5.10.: The nine constellations considered for 100G long-haul DE coherent optical communications systems.

Name	R_2/R_1	R_3/R_2	R_4/R_3	R_1
4-star	$\sqrt{3}$	n.a.	n.a.	≈ 0.707
4PAM	3/1	n.a.	n.a.	≈ 0.447
8-star	1.700	n.a.	n.a.	≈ 0.717
8PAM	3/1	5/3	7/5	≈ 0.218
2A4P	1.800	n.a.	n.a.	≈ 0.687
16-star	1.250	2.000	1.100	≈ 0.494
4A4P	1.250	1.200	1.667	≈ 0.601

Table 5.4.: Radii ratios of the constellations with more than one ring considered for 100G long-haul DE coherent optical communications systems and reported in figure 5.10.

- I decided to add the 16-star and the 4A4P because I expect them to perform better than, for instance, the 2 amplitude- 8 phase-shift keying (2A8P) in presence of wPWGN, since they have "only" 4 points on each ring.

The radii of the most promising constellations with more than one ring (8-star, 2A4P, 16-star and 4A4P) have been numerically optimised for the DE AWGN model, i.e. are those that maximise the MI. The others constellations (4-star, 4PAM and 8PAM) have the same radii as in the NDE case.

Table 5.4 summarizes these radii ratios.

I considered the same SPNR range I used for the NDE case and explained in section 5.2.2 (see also table 5.2).

The method behind the choice of the SNR range was also the same, although it led to broader ranges.

5.2.3. DE AWGN/wPWGN model: values of interest

Clearly, even in the DE case the MI and the potential gain over the QPSK depend on the same three variables.

In section 5.2.4 and 5.2.5 I will hence show the results as explained in section 5.1.3, the only difference being that the figures relative to the Gaussian constellation will not be present, since this constellation has not been considered.

The decision on which value to fix each variable is motivated as follows:

1. I chose to fix the constellation to two "values":
   - the QPSK because it corresponds to the traditional approach;
   - the 2A4P because, as I will soon show, it is the alternative, among the constellations with 8 points, that offers the highest potential gain over the QPSK;
2. I chose to fix the SNR to 9 dB, because it is one of the few values, among those for which I estimated the MI, for which *all* constellations, for *all* SPNRs, achieve

Fixed variable	Value
Constellation	QPSK, 2A4P
SNR	9.00 dB
MI	1.87 bpcu
SPNR	18.00 dB, 21.00 dB, 30.00 dB and 19.72 dB

Table 5.5.: Variables fixed or highlighted in figures 5.11, 5.12 and 5.13 as well as in figures 5.14 and 5.15.

a MI in the range 0.8–2.8 bpcu, centered roughly around the value chosen for the MI.

3. I fixed the MI to the same value as in the NDE case;

4. I highlighted the same values of the SPNR as in the NDE case.

Table 5.5 summarizes which variable I fixed or highlighted to which value in figures 5.11, 5.12 and 5.13 as well as in figures 5.14 and 5.15.

5.2.4. DE AWGN/wPWGN model: MI

Figures 5.11 and 5.12 show the MI of the QPSK and the 2A4P as a function of the SNR and parametrised by various values of the SPNR. Also reported in figure 5.11, for comparison, is its MI in the NDE case from figure 5.2 (dashed lines).

Figures 5.11 and 5.12 show behaviours similar to those described in the NDE case, see section 5.1.4.

Even in the DE case the MI increases monotonically with the SNR and the SPNR. The presence of *some* wPWGN can still be compensated by an increased SNR, although *too much* wPWGN can prevent the constellations from reaching values close to the limit $I = m$ bpcu, even for very high SNRs.

The DE encoding causes a SNR penalty, that is, a higher SNR is required to achieve the same MI for the same SPNR. This penalty is not constant, rather it varies with the SNR and the SPNR: at the targeted MI it is around 1.0 dB for high SPNRs and it gets worse for lower SPNRs.

Figure 5.13 shows the MI of all constellations, as a function of the SPNR and for a fixed value of the SNR = 9 dB.

Comparing the constellations it can be seen again that the MI of all constellations increases with the SPNR.

As in the NDE case, the PAMs are practically insensitive to wPWGN: thy achieve almost the same MI and outperform the QPSK for SPNR $\leq$ 11.4 dB.

The MI curve of the 16-star crosses the one of the 2A4P around SPNR = 13.8 dB and almost overlaps with the one of the 8-star around SPNR $\leq$ 12.5 dB, despite its additional encoded bit.

In contrast to the NDE case, the 8PSK achieves always a smaller MI than the 2A4P and, for SPNR $\leq$ 15.5 dB, even than the QPSK while offering one additional encoded

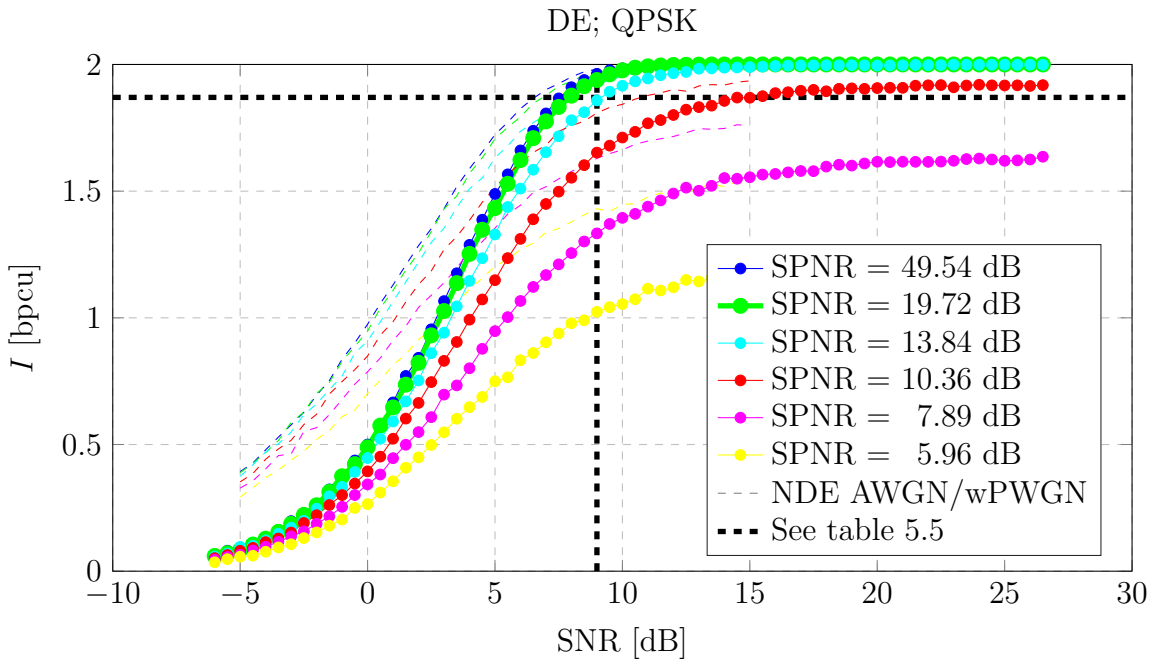

Figure 5.11.: MI of the QPSK as a function of the SNR and parametrised by various values of the SPNR over the DE AWGN/wPWGN model. Also reported is its MI over the NDE AWGN/wPWGN model from figure 5.2 (dashed).

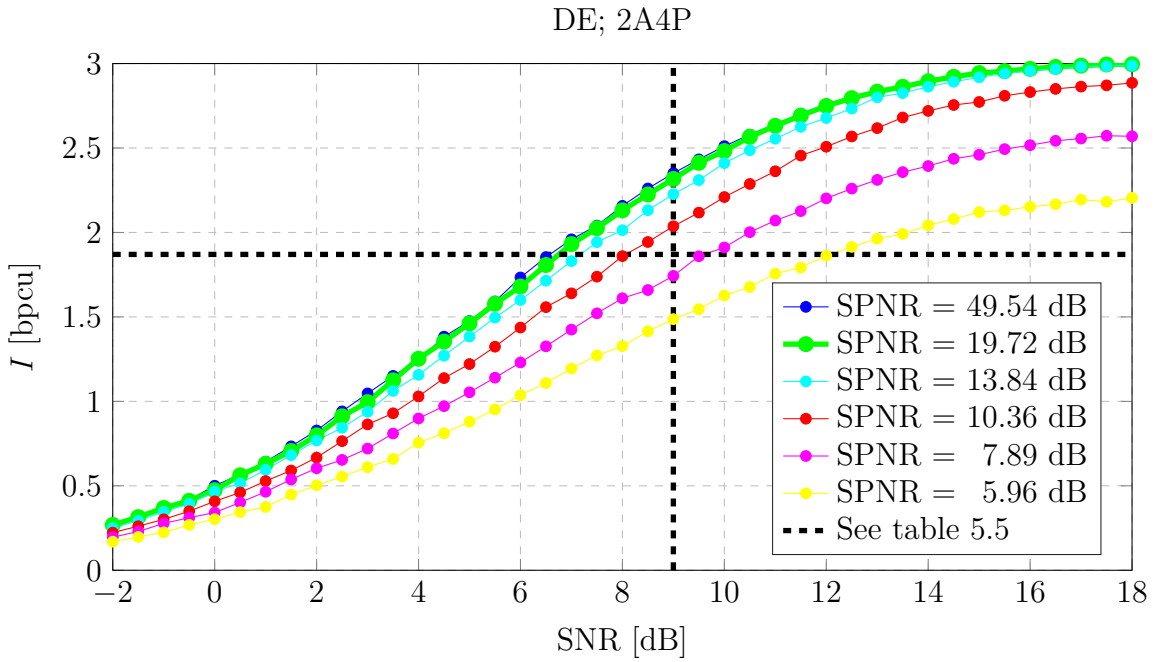

Figure 5.12.: MI of the 2A4P as a function of the SNR and parametrised by various values of the SPNR over the DE AWGN/wPWGN model.

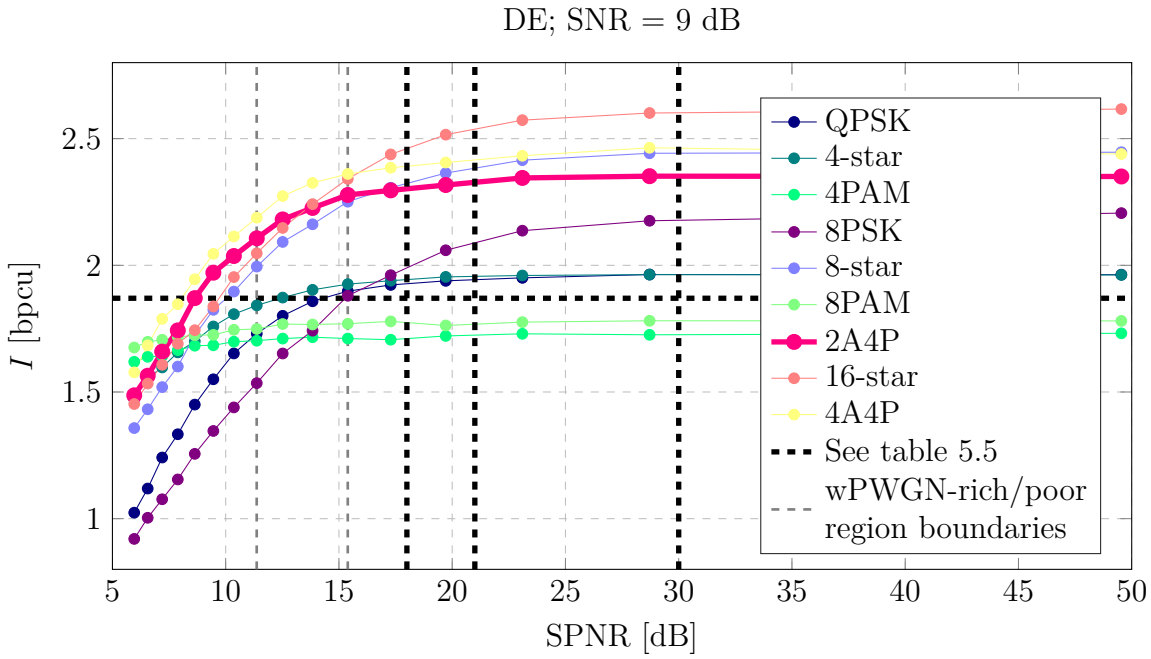

Figure 5.13.: MI of all constellations, as a function of the SPNR and for a SNR = 9 dB over the DE AWGN/wPWGN model.

bit. The reason is that this constellation has eight points on the same ring, thus being particularly sensitive to phase noise when DE.

The 4-star has the same performance as the QPSK for SPNR $\geq$ 23.1 dB and even better below this value, similar to the NDE case.

As in the NDE case, there is no "clear winner". The MI curve of the 16-star crosses the one of the 4A4P around SPNR = 15.5 dB:

- above this value the 16-star achieves the highest MI, whereas
- below this value the 4A4P achieves the highest MI; the exception is
- the extreme region SPNR $\leq$ 6.5 dB, where the 8PAM becomes the best constellation.

5.2.4.1. DE AWGN/wPWGN model: wPWGN-rich and wPWGN-poor regions – can wPWGN be neglected?

The MI "saturates" (i.e., achieves $\approx$ 90% of their respective maxima) for 11.37 dB $\leq$ SPNR $\leq$ 15.39 dB for all constellations but the PAMs (indicated with a vertical gray dashed line in figure 5.13). In other words, even in the DE case I can identify

- a wPWGN-rich region to the left of this interval, for SPNR $\leq$ 11 dB and
- a wPWGN-poor region to the right of this interval, for SPNR $\geq$ 15 dB.

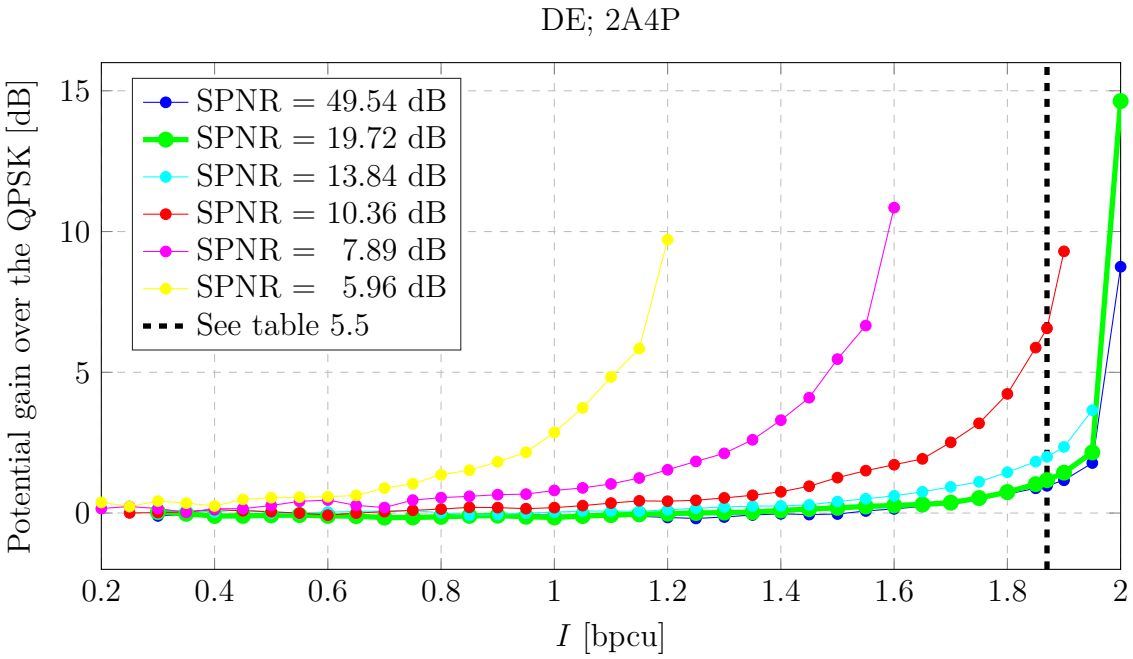

Figure 5.14.: Potential gain over the QPSK of the 2A4P as a function of the targeted MI and parametrised by various values of the SPNR over the DE AWGN/wPWGN model.

Note that these regions have been identified for a SNR = 9 dB. Even in the DE case I expect their boundaries to shift toward even lower SPNRs for higher SNRs.

If a WDM system operated in this region, I could hence neglect the wPWGN and use the DE AWGN model of section 2.5, with the great advantage of having a closed form of the conditional pdfs of the received symbols Y_1, Y_2 conditioned on the transmitted transition T, as I showed in section 2.5.2, which would be useful for the implementation of the SD demapper, without compromising the precision.

This is exactly the case: the lower limits of the operating ranges anticipated in section 3.4.1 and individuated in chapter 4 (see again table 4.5) are well inside this wPWGN-poor region even in the DE case.

5.2.5. DE AWGN/wPWGN model: potential gain over the QPSK

As I did in section 5.1.5 for the NDE case, I find it useful to present also the potential gain over the QPSK of the various alternative constellations.

Figure 5.14 shows the potential gain over the QPSK of the 2A4P, as a function of the targeted MI and parametrised by various values of the SPNR.

Figure 5.14 shows that, even in the DE case, the potential gain over the QPSK increases with the targeted MI and decreases with the SPNR.

Figure 5.15 shows the potential gain over the QPSK of all constellations, as a function of the SPNR and for a fixed value of the targeted MI $I = 1.87$ bpcu.

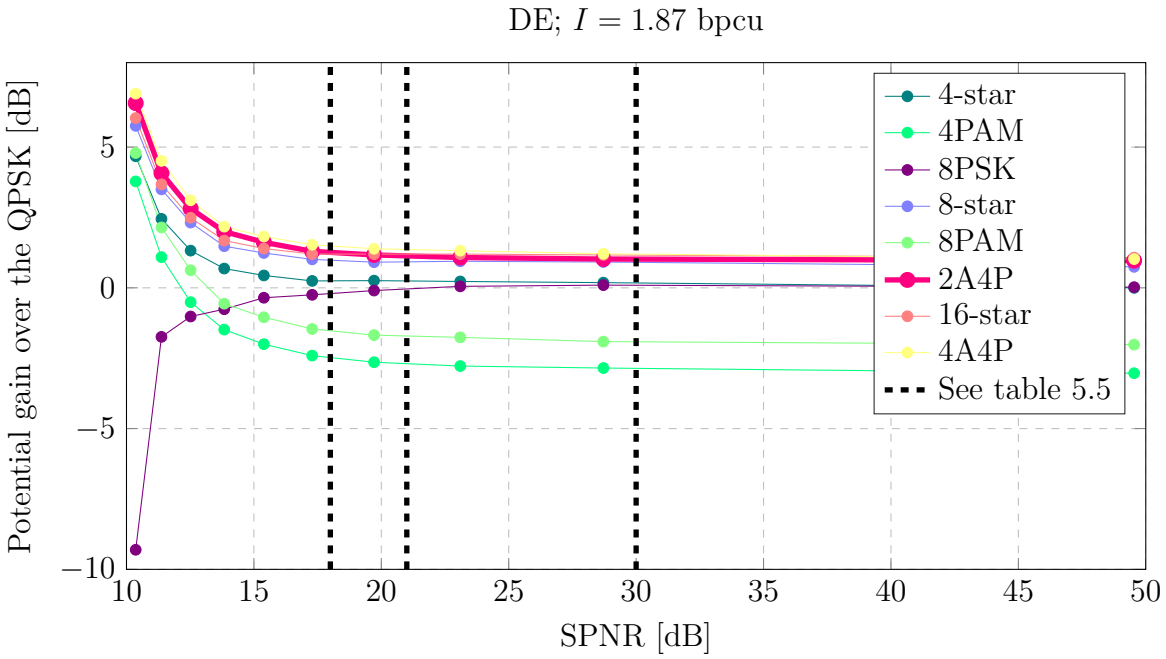

Figure 5.15.: Potential gain over the QPSK of the other constellations, as a function of the SPNR and for a MI $I = 1.87$ bpcu over the DE AWGN/wPWGN model.

For the PAMs the potential gain over the QPSK is negative for SPNR $>$ 12 dB and SPNR $>$ 13 dB: the PAMs, even the 8PAM with its additional encoded bit, are convenient only in the wPWGN-rich region.

The 4-star has always a positive potential gain over the QPSK, which is however $\leq$ 0.25 dB for SPNR $>$ 17.3 dB and becomes significant only in the wPWGN-rich region.

A remarkable difference with figure 5.9 is that, in the DE case, the 8PSK has a *negative or zero* potential gain over the QPSK, at least for $I = 1.87$ bpcu.

The 16-star and the 4A4P with 16 points exhibit a positive potential gain over the QPSK, which is however only slightly higher than the one shown by the 2A4P and the 8-star.

Finally, the 4A4P offers the highest potential gain over the QPSK.

In conclusion, the constellation with eight points with the highest potential gain over the QPSK is the 2A4P, which motivates the choice of this constellation in figures 5.12 and 5.14 and, most importantly, in the proposed solution for 100G long-haul DE coherent optical communications systems.

As in section 5.1.5, a final note on the wPWGN-poor and wPWGN-rich regions. The potential gain over the QPSK decreases with the SPNR and flattens out (achieves $\approx$ 90% of their respective minima) for 15.39 dB $\leq$ SPNR $\leq$ 23.10 dB; to the right of this interval (deep inside the wPWGN-poor region, see section 5.2.4.1) the potential gain over the QPSK practically does not depend on the SPNR and equals the one these constellations exhibit over the DE AWGN model.

Region	NDE AWGN/wPWGN model (SNR = 6 dB)	DE AWGN/wPWGN model (SNR = 9 dB)
wPWGN-rich	SPNR < 9 dB	SPNR < 11 dB
wPWGN-poor	SPNR > 14 dB	SPNR > 15 dB

Table 5.6.: Boundaries of the wPWGN-rich and of the wPWGN-poor regions over the NDE AWGN/wPWGN model and the DE AWGN/wPWGN model for a SNR = 6 dB and 9 dB, respectively.

5.3. Summary

In this chapter I presented the MI achieved by various constellations over the NDE AWGN/ wPWGN model and the DE AWGN/ wPWGN model as a function of the SNR and the SPNR. These quantities represent the *theoretical* (since an *ideal* SD FEC is assumed) limits of the constellations that can be used for 100G long-haul NDE and DE coherent optical communications systems. In both cases one of the considered constellations was the QPSK, which is used in traditional solutions, whereas the others are potential candidates for the proposed solutions. I then showed the same results in relative terms as the *potential* gain over the QPSK of the alternative constellations.

For each model I identified

1. a wPWGN-rich and a wPWGN-poor region, in which the effect of the wPWGN must be taken into account or can be neglected, respectively.

The boundaries of these regions depend on the SNR and I expect them to shift toward lower SPNRs for higher SNRs.

Table 5.6 summarizes the boundaries of these regions.

Recalling the results obtained in chapter 4 applying the estimators presented in chapter 3 to the received symbols resulting from laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems I proved that

2. the SPNR of WDM systems lies well inside the wPWGN-poor region, which means that I can use the NDE AWGN model and the DE AWGN model to implement the SD demapper. This is a great advantage, because for these models the conditional pdfs are known in closed form, which greatly simplifies the implementation of the SD demapper.

Finally, I also showed that, among the discrete constellations with eight points, in the wPWGN-poor region the highest potential gain over the QPSK is achieved by

3. the 8-star over the NDE AWGN/ wPWGN model and

4. the 2A4P over the DE AWGN/ wPWGN model.

Hence, I will use these constellations for the proposed solution for 100G long-haul NDE and DE coherent optical communications systems, respectively.

6. Proposed solutions

The *traditional* solutions for 100G long-haul non differentially encoded (NDE) and differentially encoded (DE) coherent optical communications systems with optical data unit 4 (ODU4) nominal bit rate of $R_b = 104.794445815$ Gbps are based on (see [2] and references therein for an historical overview)

- the quadrature phase-shift keying (QPSK) ($m = 2$ bits/symbol);
- polarisation division multiplexing (PDM) ($N_{\text{pol}} = 2$) and
- various (concatenations of) hard-decision (HD) and/or soft-decision (SD) forward error correction (FEC) schemes, always with
  - $\approx 20\%$ overhead (OH) redundancies, corresponding to
  - code rates $R \approx 0.83$,
  - targeted mutual informations (MIs) $I = Rm \approx 1.67$ bits per channel use (bpcu) per polarisation and
  - a symbol rate $f_s \approx 31.44$ GBd, corresponding to
  - a bandwidth expansion (BE) of $\approx 20\%$ ($f_s \approx 26.20 \rightarrow 31.44$ GBd).

In this final chapter I propose two solutions for 100G long-haul NDE and DE coherent optical communications systems, *alternative* to the traditional ones since they are based on (much) *higher* redundancy allocated mostly in a *larger* constellation and only marginally in a *narrower* occupied bandwidth.

More in particular, the presented solutions make use of

- the 8-ary constellations chosen in chapter 5 ($m = 3$ bits/symbol);
- PDM ($N_{\text{pol}} = 2$) and
- a concatenated HD/ SD FEC scheme with
  - iterative demapping and decoding (IDD) and
  - a 60.04% OH redundancy, corresponding to
  - a code rate $R = 2/3 \cdot 239/255 \approx 0.62$,
  - a targeted MI $I \approx 1.87$ bpcu per polarisation and
  - a symbol rate $f_s \approx 27.95$ GBd, corresponding to
  - a BE = 6.67% ($f_s = 26.20 \rightarrow 27.95$ GBd) as well as

– a constellation expansion (CE) = 50.00% ($m = 2 \rightarrow 3$ bits/symbol)

The performance of an IDD scheme can be predicted by means of the extrinsic information transfer (EXIT) charts, given by the overlap of the transfer characteristics of the two actors iteratively exchanging information, in this case the demapper and the decoder.

I will hence start the chapter by introducing how I estimated the transfer characteristics of

- the demapper of 100G long-haul NDE coherent optical communications systems, which depend on the bit mapping and the signal-to-(*additive*-)noise ratio (SNR) and of
- the decoder, which depend on the number of low-density parity-check (LDPC) iterations.

Note that the number of LDPC iterations N_{LDPC} and the number of IDD iterations N_{IDD} are two different things: the former is the number of times the variable and the check nodes of the LDPC code exchange information, whereas the latter is the number of times the demapper and the decoder exchange information.

I will then present the performance predicted by the EXIT chart of each combination of a bit mapping, a SNR and a number of LDPC iterations. I will also present their net coding gain (NCG) over the uncoded QPSK as well as their approximated complexity in terms of summations per information bit (spib).

I will then repeat the same steps for 100G long-haul DE coherent optical communications systems.

Finally, in order to understand how good these *alternative* solutions are in terms of minimum SNR and complexity, I will compare them with a *traditional* solution realised using the same building blocks with the exception of the constellation and the inner code, which must be different for it to belong to the traditional solutions. Note that this term of comparison (as all traditional solutions) make use of a larger BE and as such have a larger symbol rate and occupy more bandwidth.

Part of the material presented in this chapter has been published at the Optical Fiber Conference (OFC) [14], at the European Conference on Optical Communications (ECOC) [15, 65] as well as at the International Conference on Transparent Optical Networks (ICTON) [17].

6.1. The proposed FEC solution

The proposed FEC solution consists of

- a SD demapper;
- the binary 2/3-rate LDPC code from the digital video broadcasting – satellite – 2$^{\text{nd}}$ generation (DVB-S2) standard [66] as *inner* SD code, whose redundancy is

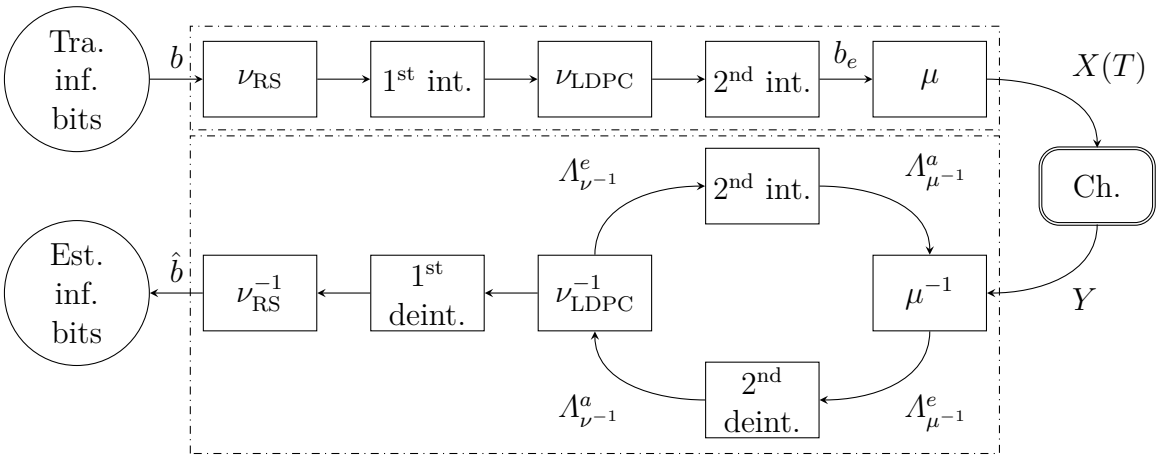

Figure 6.1.: Block diagram of the proposed solutions for 100G long-haul NDE and DE coherent optical communications systems. The part framed by a dash-dotted line is the proposed FEC solution presented in this chapter. The double lined block will be replaced with either the NDE AWGN model or the DE AWGN model, see figure 2.2.

allocated *entirely* in the CE. This code is needed to "quickly" (in terms of SNR) achieve the targeted post-FEC bit error rate (BER) = 10^{-15}.

The demapper and the SD LDPC decoder perform

- IDD, meaning that they iteratively exchange information to refine the log-likelihood ratios (LLRs). This step is needed to achieve the *full* capacity of a *non-binary* constellation employing *binary* codes;
- the non-binary 16-fold interleaved Reed-Solomon (RS) (255, 239) code from the Telecommunication Standardization Sector of International Telecommunication Union (ITU-T) G.975 standard [67] as *outer* HD code, whose redundancy is allocated *entirely* in the BE. This code is needed to correct potential error bursts leaving the IDD.

Figure 6.1 shows a high level block diagram of the proposed solutions. This is what happens at each step:

- the RS encoder ν_{RS} encodes the transmitted information bits b into the transmitted RS encoded bits;
- the first interleaver interleaves the latter into the interleaved transmitted RS encoded bits;
- the LDPC encoder ν_{LDPC} encodes the latter into the transmitted LDPC encoded bits;
- the second interleaver interleaves the latter into the transmitted encoded bits b_e;

- the mapper μ maps the latter to either the transmitted symbols X or the transmitted transitions T, depending on whether the system is NDE or DE, respectively;
- if the system is DE, the channel Ch DE encodes the transmitted transitions T into the transmitted symbols X. It then adds the additive white Gaussian noise (AWGN) to the transmitted symbols X and obtains the received symbols Y;
- the SD demapper μ^{-1} calculates its extrinsic LLRs $\Lambda^e_{\mu^{-1}}$ from the latter and its a priori LLRs $\Lambda^a_{\mu^{-1}}$, which are the extrinsic LLRs $\Lambda^e_{\nu^{-1}}$ of the SD LDPC decoder ν^{-1}_{LDPC} interleaved by a copy of the second interleaver. This is the first half of the IDD stage;
- the SD LDPC decoder ν^{-1}_{LDPC} calculates its extrinsic LLRs $\Lambda^e_{\nu^{-1}}$ from its a priori LLRs $\Lambda^a_{\nu^{-1}}$, which are the extrinsic LLRs $\Lambda^e_{\mu^{-1}}$ of the SD demapper μ^{-1} deinterleaved by the second deinterleaver. This is the second half of the IDD stage.

  After a certain number of IDD iterations the SD LDPC decoder ν^{-1}_{LDPC} takes a HD on its a posteriori LLRs $\Lambda^p_{\nu^{-1}}$ and obtains the interleaved estimated RS encoded bits;
- the first deinterleaver deinterleaves the latter into the estimated RS encoded bits;
- the HD RS decoder ν^{-1}_{RS} decodes the latter into the estimated information bits $\hat{b}$.

In the following I will refer to the SD demapper simply as “demapper” and to the SD LDPC decoder simply as “decoder”, for brevity.

6.1.1. The LDPC code

The decision on the 2/3-rate DVB-S2 LDPC code is in line with the current trends in research [2]. In general, LDPC codes exhibit error correcting performance close to the Shannon limit. Morever, they can be encoded and decoded efficiently with reduced circuit complexity, since they are linear codes and their decoding algorithm is inherently parallel.

These advantages are particularly evident in the LDPC codes from the DVB-S2 standard [66]:

> a powerful FEC system based on LDPC (Low-Density Parity Check) codes concatenated with BCH codes, allowing Quasi-Error-Free operation at about 0.7 dB to 1 dB from the Shannon limit, depending on the transmission mode (AWGN channel, modulation constrained Shannon limit).

The codes exhibit a staircase lower triangular on the right part of the parity-check matrices, resulting in a linear encoding complexity. Their long codewords and, especially in the case of the code for the proposed solutions, their high redundancies lead to increased girths, which dominate the error correction performance in the low post-FEC BER region and hence lead to lower error floors. The drawbacks are increased circuit size and

delays; however, the LDPC codes from the DVB-S2 standard are structured and hence architecture- or hardware-aware codes, whose decoders can be efficiently implemented through high parallelisation.

Finally, the DVB-S2 standard has been developed in 2003 by the digital video broadcasting (DVB) project and ratified in 2005 by European Telecommunications Standards Institute (ETSI), so it has been known for quite some years already and it can be considered a stable, proven code; also, coming from an important standard, I am confident that a great deal of work is being and will be done, aimed at tackling the drawbacks mentioned above.

6.1.2. The RS code

The RS code is used since decades by the optical community and extremely efficient implementations thereof are readily available on the market. I preferred to use a RS code instead of a Bose-Chaudhuri-Hocquenghem (BCH) code (as suggested in the DVB-S2 standard) because [2]:

> RS codes have higher tolerance to error bursts

since they work with *bytes*, not *bits*.

Overall, I intend to prove that excellent results can be achieved by combining available off-the-shelf components: in this sense, the novelty of the proposed solutions does not reside in the single parts, rather in their clever combinations.

6.2. Proposed solution for 100G long-haul NDE coherent optical communications systems: 8-star + DVB-S2-LDPC(64800, 43200) + ITU-T-RS(255, 239)

Table 6.1 shows the minimum SNR required to achieve the targeted MI $I = 1.87$ bpcu per polarisation for the constellations considered in chapter 5 with a symbol rate $f_s = 27.95$ GBd over the NDE AWGN model. The corresponding OSNR can be obtained adding $10\log_{10}(27.95/12.50) \approx 3.5$ dB. The corresponding E_b/N_0 can be obtained subtracting $10\log_{10}(3 \cdot 2/3 \cdot 239/255) \approx 2.7$ dB. Also reported is how much OH they offer space for, together with the corresponding potential gain over the coded QPSK and potential NCG over the uncoded QPSK, which needs a SNR $= 18.00$ dB to achieve a BER $= 10^{-15}$.

The values relative to the 8-star chosen in chapter 5 are marked in bold. Table 6.1 will be useful to assess how close the practical proposed solutions are to their theoretical performance.

Constellation	OH	Min. SNR [dB]	Pot. gain over *coded* QPSK [dB]	Pot. NCG over *uncoded* QPSK [dB]
QPSK	6.69%	6.48	n.r.	+11.22
4-star		6.81	−0.33	+10.90
4PAM		9.95	−3.47	+07.75
8PSK	60.04%	5.12	+1.37	+12.59
8-star		**4.88**	**+1.60**	**+12.83**
8PAM		8.70	−2.21	+09.01
2A4P		5.12	+1.36	+12.58
16QAM	113.39%	4.52	+1.97	+13.19
Gaussian	n.r.	4.27	+2.21	+13.44

Table 6.1.: Minimum SNR (in dB) required to achieve the targeted MI $I = 1.87$ bpcu per polarisation for the constellations considered in chapter 5 with a symbol rate $f_s = 27.95$ GBd, potential gain over the coded QPSK and potential NCG over the uncoded QPSK over the NDE AWGN model. Also reported is the redundancy offered. The constellation chosen in chapter 5 for 100G long-haul NDE coherent optical communications systems is highlighted in bold.

6.2.1. NDE AWGN model: estimation of the EXIT charts

An EXIT chart is the combination of one transfer characteristic of the demapper and one transfer characteristic of the decoder, the latter with the x- and the y-axis swapped.

These transfer characteristics relate

- the MI between the bits and their a priori LLRs $I^a_{\mu^{-1}}$ (for the demapper) or $I^a_{\nu^{-1}}$ (for the decoder), which I will call in the following *"a priori MI"* for brevity, and
- the MI between the bits and their extrinsic LLRs $I^e_{\mu^{-1}}$ (for the demapper) or $I^e_{\nu^{-1}}$ (for the decoder), which I will call in the following *"extrinsic MI"* for brevity.

Note that these MIs must not be confused with the one achieved by a constellation mentioned before. These MIs refer to a completely different channel, whose input is a bit (hence, a "constellation" $\mathcal{X} = \{0, 1\}$ with cardinality $|\mathcal{X}| = 2$ that can carry $m = 1$ bit/symbol) and whose output is a LLR. They can be considered an indicator of the quality of their LLRs.

I estimated the transfer characteristics of the demapper and the decoder following ten Brink's excellent introductory paper [68]; for the interested reader, in appendix A I explain in detail the procedure.

Once I had the transfer characteristics of the demapper and of the decoder, I tried each possible combination and found the minimum SNR and the number of IDD iterations at which the "tunnel", i.e. the region below the transfer characteristic of the demapper and above the swapped transfer characteristic of the decoder, opened up "enough" for the "decoding trajectory", i.e. the polygonal chain going back and forth between the two

transfer characteristics, to achieve the top right part of the EXIT chart (in section 6.2.4.1 I will explain in more detail when I considered a decoding trajectory "successful").

The number of IDD iterations is given by

$$N_{\text{IDD}} = \frac{N_{\text{hops}} - 1}{2}, \tag{6.1}$$

where N_{hops} is the number of hops in the decoding trajectory; this is because one IDD iteration is represented by two "bounces", one on the demapper transfer characteristic (going up – 90° turn – going right) and one on the decoder transfer characteristic (going right – 90° turn – going up); the -1 accounts for the starting point (0,0). The maximum number of IDD iterations was set to 100.

In order not to weigh the treatise down, from now on I will refrain most of the times from explicitly saying whether a quantity is for the demapper or for the decoder, relying on the context and the symbols (μ^{-1} refers to the demapper whereas ν^{-1} refers to the decoder).

6.2.2. NDE AWGN model: 8-star bit mappings as well as SNR and a priori MI ranges considered

I performed an exhaustive search through all possible permutations of the symbols in the constellation (i.e., all possible bit mappings) admitted by an 8-star looking for the "unique" ones as follows:

1. For each bit mapping μ_k, for each symbol X_i in the bit mapping μ_k and for all symbols $X_j, j \neq i$ at minimum *Euclidean* distance from the symbol X_i (there could be more than one) I calculated the *Hamming* distance between X_i and X_j;

2. I estimated the probability density function (pdf) of the *Hamming* distance between X_i and its neighbours counting with an histogram how many times it was 1, 2 or 3;

3. I then averaged the pdf of the *Hamming* distance between X_i and its neighbours over all symbols X_i obtaining an estimate of the average pdf of the *Hamming* distances between two points at minimum Euclidean distance of the bit mapping μ_k. Finally,

4. I found the unique histograms and took one "exemplary" bit mapping for each of them.

Clearly, any other bit mapping could have been taken, as long as there is one representative for each unique histogram: in this sense any further bit mapping can be proved to be equivalent to the previous ones up to bit reordering and/or bit inversion.

An 8-ary constellation as the 8-star admits $8! = 40320$ possible bit mappings; I found 12 distinct bit mappings for the 8-star ($\approx 0.03\%$).

Figure 6.2 shows these 12 distinct bit mappings.

In particular

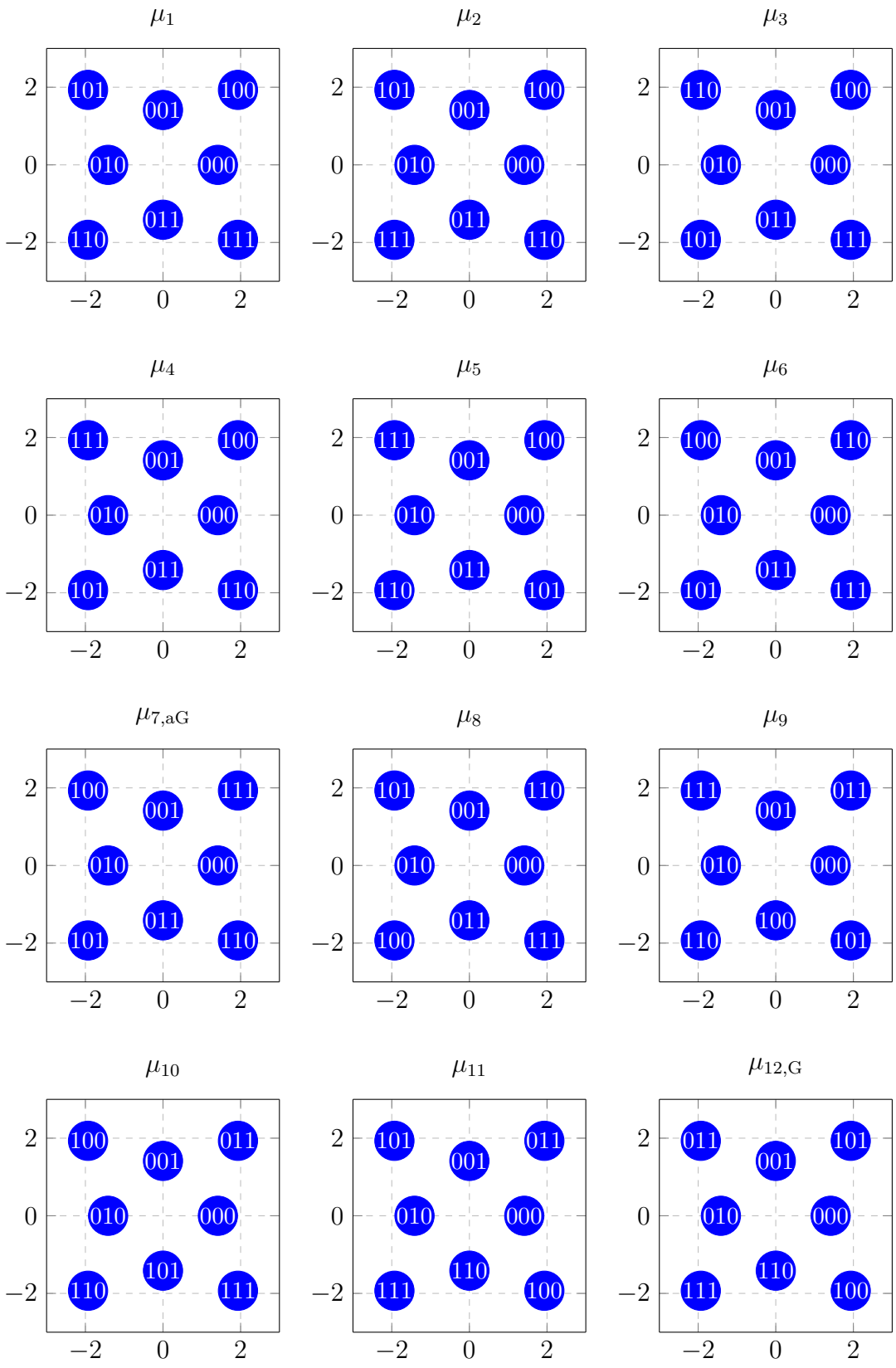

Figure 6.2.: Twelve distinct bit mappings admitted by the 8-star. The bit mapping $\mu_{12,\mathrm{G}}$ is the closest to a Gray-coded bit mapping, the bit mapping $\mu_{7,\mathrm{aG}}$ is the farthest from a Gray-coded bit mapping.

- the bit mapping $\mu_{12,\mathrm{G}}$ is the closest to a Gray-coded bit mapping (the 8-star does not admit a Gray-coded bit mapping), meaning that the average pdf of the Hamming distances between two points at minimum Euclidean distance has the peak around 1; conversely,
- the bit mapping $\mu_{7,\mathrm{aG}}$ is the farthest from a Gray-coded bit mapping (i.e. the anti-Gray-coded bit mapping), meaning that the average pdf of the Hamming distances between two points at minimum Euclidean distance has the peak around 3.

I chose as SNR range an asymmetric interval $[-1, +3]$ around the minimum SNR required for the 8-star to achieve the targeted MI $I = 1.87$ bpcu per polarisation with a linear 0.1 dB step.

The a priori MI is by definition 0 bpcu $< I^a_{\mu^{-1}} < 1$ bpcu; I decided to span this interval with a linear 0.05 bpcu step.

6.2.3. Number of LDPC iterations and a priori MI range considered

The DVB-S2 standard sets a maximum number of LDPC iterations $N_{\mathrm{LDPC}} = 50$ [66]. Expecting the performance to saturate with an increasing number of LDPC iterations, I hence decided to consider for the number of LDPC iterations 10 values between 1 and 50 with a non-linear step: 1, 2, 3, 4, 5, 10, 20, 30, 40 and 50.

Even for the decoder the a priori MI is by definition 0 bpcu $< I^a_{\nu^{-1}} < 1$ bpcu, however this time I spanned this interval with two steps: a first larger linear 0.050 bpcu step over the whole range and a second smaller linear 0.001 bpcu step in a symmetric interval 0.67 $\pm$ 0.2 bpcu around the code rate $R = 2/3$ of the 2/3-rate DVB-S2 LDPC code. The reason is that I expected the extrinsic MI curves of the decoder to change very quickly around the code rate, especially for high numbers of LDPC iterations.

6.2.4. NDE AWGN model: values of interest

The extrinsic MI of the demapper depends on three variables: the SNR, the bit mapping and the a priori MI.

Similar to section 5.1.3, I hence decided that in section 6.2.5 I will show the results in two figures fixing once the SNR and parametrizing the curves with the bit mapping and fixing once the bit mapping and parametrizing the curves with the a priori MI. The figures for a fixed a priori MI would have had the bit mappings on the x-axis, which cannot be ordered as for instance can real numbers; I hence decided to skip this combination. When a variable is fixed in one figure, I will highlight it in the others in which it is not (with a thicker curve when it parametrises a curve or with a vertical black dashed line when it is on the x-axis).

The extrinsic MI of the decoder, instead, depends on two variables: the number of LDPC iterations and the a priori MI.

Hence, in section 6.2.6 I will show the results in one figure, parametrizing the curves with the number of LDPC iterations.

The decision on which value to fix each variable is motivated as follows:

Fixed variable	Value
SNR	4.9 dB
Bit mapping	$\mu_{12,\mathrm{G}}$
A priori MI entering the decoder	2/3 bpcu

Table 6.2.: Variables fixed or highlighted in figures 6.3, 6.4 and 6.5.

1. for the demapper I chose to fix the SNR to 4.9 dB, because it is the theoretical limit of the 8-star over the NDE AWGN model, see table 6.1. The transfer characteristics of the demapper for the values of SNR that give the best match as well as the least complex match with the transfer characteristic of the decoder will be shown in the corresponding EXIT charts;

2. for the demapper I chose to fix the bit mapping to one "value", the $\mu_{12,\mathrm{G}}$, because, as I will soon show, it is the one that gives the best match with the transfer characteristics of the decoder;

3. for the decoder I chose to highlight one value of the a priori MI, 2/3 bpcu, because it corresponds to the code rate of the 2/3-rate DVB-S2 LDPC code.

Table 6.2 summarizes which variable I fixed or highlighted to which value in figures 6.3, 6.4 and 6.5.

6.2.4.1. Successful decoding trajectory

I considered a decoding trajectory successful as soon as the decoder achieved an extrinsic MI, for an a priori MI, such that the corresponding *a posteriori MI* $I^{p}_{\nu^{-1}} \geq 1 - \mathcal{H}(8 \cdot 10^{-5}) \approx 0.9988$ bpcu, where $\mathcal{H}(x)$ is the binary entropy function

$$\mathcal{H}(x) = -x \log_2(x) - (1-x) \log_2(1-x). \tag{6.2}$$

This quantity represents the amount of uncertainty left after the decoder or, equivalently, the amount of uncertainty the 16-fold interleaved ITU-T G.975 RS (255, 239) code can remove, since this code translates a pre-FEC BER = $8 \cdot 10^{-5}$ down to a post-FEC BER = 10^{-15} [69]. Note that the condition must be on the a posteriori MI, since the 16-fold interleaved ITU-T G.975 RS (255, 239) code works on the bits obtained taking a HD on the a posteriori LLRs after the decoder. For the interested reader, in appendix A I explain in detail how I obtained the a posteriori MI from the a priori MI and the extrinsic MI.

6.2.5. NDE AWGN model: transfer characteristics of the demapper

Figure 6.3 shows the extrinsic MI of the 8-star as a function of the a priori MI, for a fixed value of the SNR = 4.9 dB and parametrised by the bit mappings.

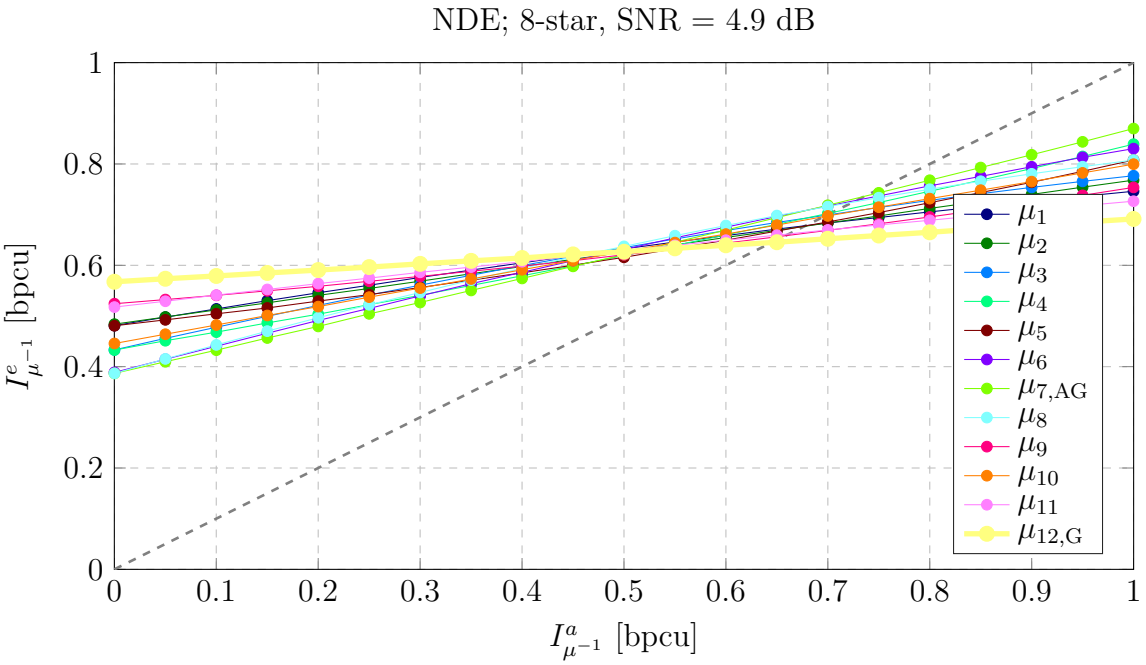

Figure 6.3.: Extrinsic MI of the 8-star as a function of the a priori MI, for a SNR = 4.9 dB and parametrised by the bit mappings over the NDE AWGN model.

In figure 6.3 all 12 lines are visible, which means that the technique described in section 6.2.2 to find the unique bit mappings is correct.

The bit mapping $\mu_{12,\mathrm{G}}$ has the *flattest* curve, exhibiting the *highest* extrinsic MI for $I^a_{\mu^{-1}} = 0$ bpcu and the *lowest* extrinsic MI for $I^a_{\mu^{-1}} = 1$ bpcu. This confirms it as the most "Gray like"-coded bit mapping, as said in section 6.2.2.

The bit mapping $\mu_{7,\mathrm{aG}}$ has the *steepest* curve, exhibiting the *lowest* extrinsic MI for $I^a_{\mu^{-1}} = 0$ bpcu and the *highest* extrinsic MI for $I^a_{\mu^{-1}} = 1$ bpcu. This confirms it as the anti-Gray-coded bit mapping, as said in section 6.2.2.

The extrinsic MI always increases with the a priori MI, as expected: the more reliable the a priori LLRs of the *other* bits input to the demapper, the more reliable the extrinsic LLR of the *one* bit output from the demapper.

The curves cross the bisector $I^e_{\mu^{-1}} = I^a_{\mu^{-1}}$ around $I^a_{\mu^{-1}} = 0.70$ bpcu, being above it for roughly $I^a_{\mu^{-1}} \leq 0.65$ bpcu and below it for roughly $I^a_{\mu^{-1}} \geq 0.75$ bpcu. When the curves are below the bisector it means that the quality of the extrinsic LLRs output by the demapper is lower than that of the a priori LLRs input to it. For a higher SNR the curves moves upward and the region where they cross the bisector moves to the right.

Figure 6.4 shows the extrinsic MI of the 8-star as a function of the SNR, for the bit mapping $\mu_{12,\mathrm{G}}$ and parametrised by various values of the a priori MI.

Figure 6.4 shows that the extrinsic MI increases monotonically with the SNR: the less impaired by AWGN the symbols input to the demapper, the more reliable the extrinsic LLRs output from the demapper. The extrinsic MI, however, cannot increase indefinitely with the SNR: although not seen here, the curves will all collapse to $I^e_{\mu^{-1}} = 1$ bpcu for

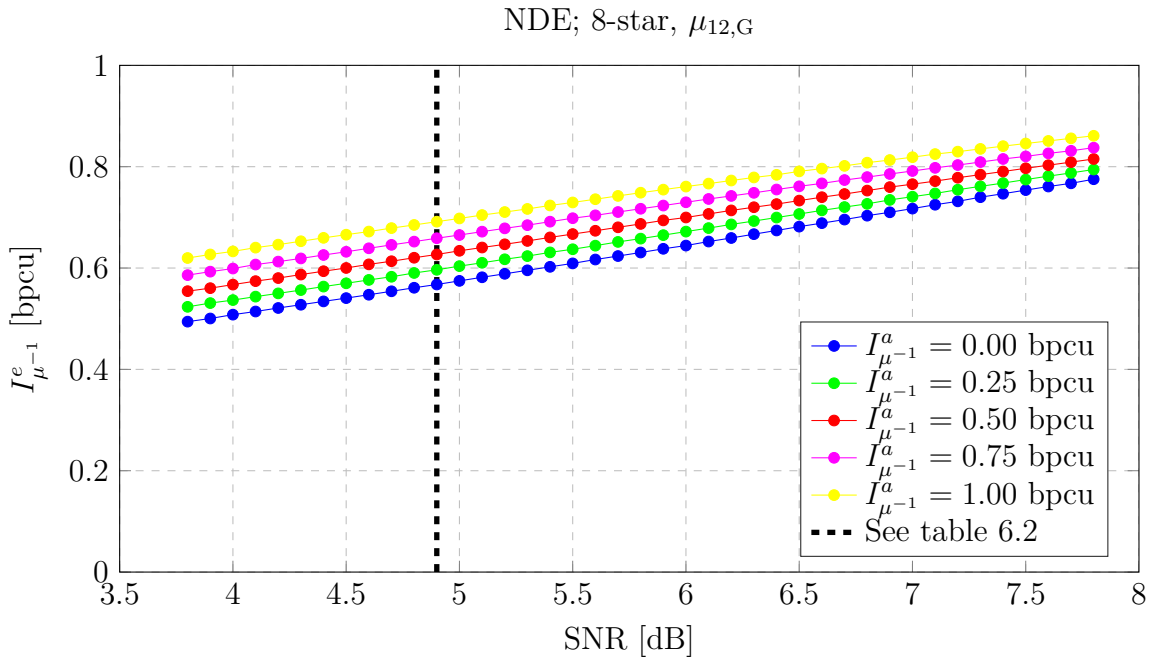

Figure 6.4.: Extrinsic MI of the 8-star as a function of the SNR, for the $\mu_{12,G}$ bit mapping and parametrised by various values of the a priori MI over the NDE AWGN model.

SNR $\to +\infty$.

The extrinsic MI increases also with the a priori MI, as already shown in figure 6.3.

6.2.6. Transfer characteristics of the decoder

Figure 6.5 shows the extrinsic MI of the 2/3-rate LDPC code from the DVB-S2 standard [66] as a function of the a priori MI and parametrised by the number of LDPC iterations.

Figure 6.5 shows that even in this case the extrinsic MI always increases with the a priori MI: the more reliable the a priori LLRs of the *other* bits input to the decoder, the more reliable the extrinsic LLR of the *one* bit output from the decoder. Also, the extrinsic MI increases with the number of LDPC iterations: the decoder refines the extrinsic LLRs at each iteration.

The curves cross the bisector $I^e_{\nu^{-1}} = I^a_{\nu^{-1}}$, being below it for roughly $I^a_{\nu^{-1}} \leq 0.70$ bpcu and above it for roughly $I^a_{\nu^{-1}} \geq 0.85$ bpcu (with the exception of the curve for $N_{\mathrm{LDPC}} = 1$, which is always below the bisector). Even in this case, when the curves are below the bisector it means that the quality of the extrinsic LLRs output by the decoder is lower than that of the a priori LLRs input to it.

A comparison with figure 6.3 shows that this time the curves are much steeper. This can be explained considering also the bit mapping as a very weak code introducing very little memory (i.e., a SD code with code rate $R = 1$ and code words length $n = 3$

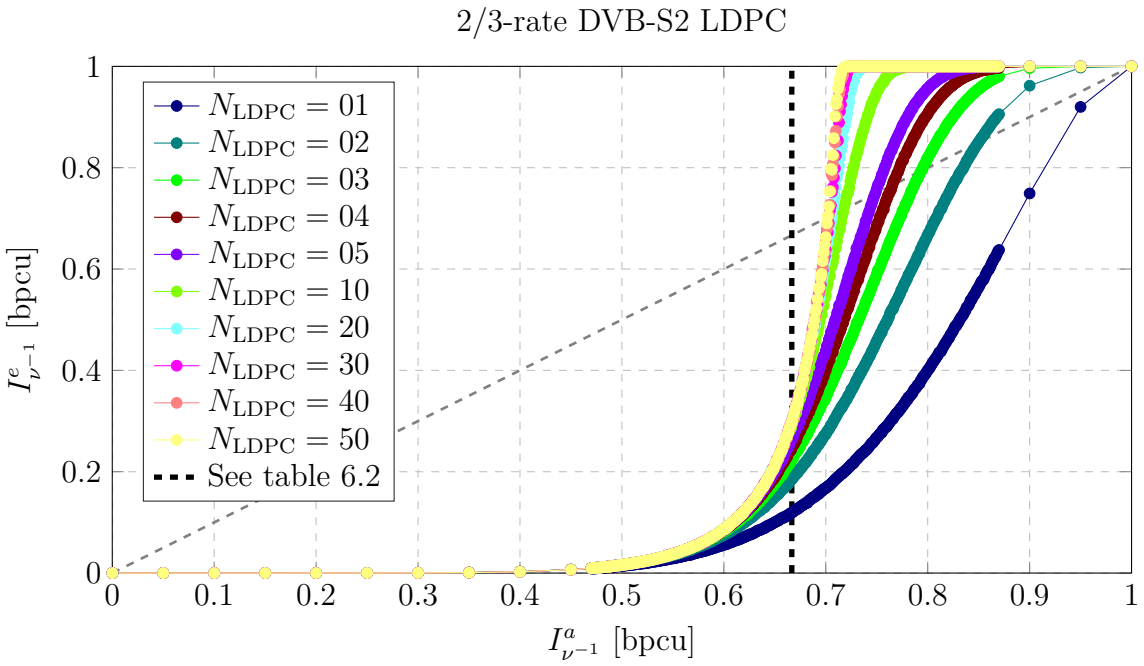

Figure 6.5.: Extrinsic MI of the 2/3-rate DVB-S2 LDPC code as a function of the a priori MI and parametrised by the number of LDPC iterations.

bits) and the demapper as the decoder of such code. In general, the more the memory introduced by a code, the steeper the transfer characteristic of its decoder [68, 70–73].

Increasing the number of LDPC iterations, the a priori MI at which

- $I^e_{\nu^{-1}} \approx 1$ bpcu is achieved decreases, but
- $I^e_{\nu^{-1}} \neq 0$ bpcu is achieved (i.e., the transfer characteristic "raises" from the x-axis) remains the same, $I^a_{\nu^{-1}} \approx 0.35$ bpcu.

In other words, the transfer characteristic of the decoder tends to the indicator function of the subset $I^a_{\nu^{-1}} > R$,

$$I^e_{\nu^{-1}} \to 1_{I^a_{\nu^{-1}} > R}(I^a_{\nu^{-1}}), \qquad \text{for } N_{\mathrm{LDPC}} \to +\infty, \tag{6.3}$$

more on its right region than on its left region. However, it does not reach it: the first a priori MI for which the decoder reaches $I^e_{\nu^{-1}} \approx 1$ bpcu, with $N_{\mathrm{LDPC}} = 50$, is $I^a_{\nu^{-1}} = 0.72$ bpcu, slightly larger than the code rate. This will lead to some penalty with respect to the theoretical limit represented by the MI.

For $I^a_{\nu^{-1}} \geq 0.72$ bpcu, the higher the a priori MI the less LDPC iterations are needed to achieve $I^e_{\nu^{-1}} \approx 1$ bpcu.

6.2.7. NDE AWGN model: estimation of the minimum SNR required

Table 6.3 shows, for each combination of a demapper transfer characteristic and a decoder transfer characteristic, the minimum SNR (in dB) and the corresponding number of IDD iterations (in parenthesis) at which the decoding trajectory is successful, i.e. the decoder achieved the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu (see section 6.2.4.1).

Table 6.4 shows the same results but in terms of NCG over the uncoded QPSK.

Certain combinations do not achieve the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for any SNR considered, which can be up to 3 dB more than those reported in table 6.1. Considering even higher values of SNR would have been possible, but it would have meant considering systems of no practical interest.

In no case a single IDD iteration is enough, not even with the most "Gray like"-coded bit mapping: in other words, a bit-interleaved coded modulation (BICM) scheme [74] would require a SNR > 7.8 dB (the biggest among the values considered here) to achieve the targeted a posteriori MI.

The "best" combination (i.e., requiring the minimum SNR and achieving the largest NCG over the uncoded QPSK) is given by

- $\mu_{12,\mathrm{G}}$ and
- $N_{\mathrm{LDPC}} = 20$: it achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for
- a SNR = 5.9 dB (1.0 dB from the theoretical limit of the 8-star and 0.6 dB *beyond* the theoretical limit of the QPSK, see table 6.1), corresponding to
- a NCG over the uncoded QPSK = 11.8 dB, with
- $N_{\mathrm{IDD}} = 18$.

There are actually three more combinations that achieve the targeted a posteriori MI for the same SNR = 5.9 dB, given by the same bit mapping, $N_{\mathrm{LDPC}} = 30$, 40 and 50 and $N_{\mathrm{IDD}} = 15$: however, I consider the first the best combination because, as I will show in section 6.2.8, the overall complexity of the proposed solution is dominated by the complexity of the decoder and hence goes with $N_{\mathrm{IDD}} N_{\mathrm{LDPC}}$.

The "least complex" combination is given by

- $\mu_{12,\mathrm{G}}$ and
- $N_{\mathrm{LDPC}} = 4$: it achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for
- a SNR = 7.8 dB (2.9 dB from the theoretical limit of the 8-star and 1.3 dB from the theoretical limit of the QPSK, see table 6.1), corresponding to
- a NCG over the uncoded QPSK = 9.9 dB, with
- $N_{\mathrm{IDD}} = 3$.

		N_{LDPC}									
		1	2	3	4	5	10	20	30	40	50
μ	1	n.a.	n.a.	n.a.	7.1 (06)	6.6 (09)	6.3 (23)	6.3 (20)	6.3 (20)	6.3 (20)	6.3 (20)
	2	n.a.	n.a.	7.5 (05)	6.7 (10)	6.5 (13)	6.4 (14)	6.4 (12)	6.4 (12)	6.4 (12)	6.4 (12)
	3	n.a.	n.a.	7.5 (05)	6.7 (16)	6.6 (49)	6.6 (18)	6.6 (17)	6.6 (17)	6.6 (18)	6.6 (17)
	4	n.a.	7.8 (07)	6.9 (21)	6.8 (21)	6.8 (14)	6.7 (35)	6.7 (31)	6.7 (29)	6.7 (44)	6.7 (34)
	5	n.a.	n.a.	7.0 (09)	6.6 (23)	6.6 (12)	6.5 (14)	6.5 (12)	6.5 (12)	6.5 (12)	6.5 (11)
	6	n.a.	7.8 (07)	7.0 (36)	7.0 (14)	7.0 (12)	6.9 (26)	6.9 (25)	6.9 (24)	6.9 (25)	6.9 (25)
	7,aG	n.a.	7.3 (22)	7.1 (19)	7.1 (11)	7.0 (19)	7.0 (13)	7.0 (12)	7.0 (12)	7.0 (12)	7.0 (12)
	8	n.a.	n.a.	7.0 (18)	6.9 (29)	6.9 (18)	6.9 (14)	6.9 (13)	6.9 (13)	6.9 (13)	6.9 (13)
	9	n.a.	n.a.	7.8 (05)	7.0 (06)	6.5 (10)	6.2 (17)	6.2 (13)	6.2 (13)	6.2 (12)	6.2 (12)
	10	n.a.	n.a.	7.0 (10)	6.7 (19)	6.7 (13)	6.6 (31)	6.6 (28)	6.6 (27)	6.6 (35)	6.6 (30)
	11	n.a.	n.a.	n.a.	7.4 (05)	6.9 (06)	6.1 (22)	6.1 (16)	6.1 (16)	6.1 (16)	6.1 (16)
	12,G	n.a.	n.a.	n.a.	*7.8* (03)	7.4 (03)	6.3 (06)	**5.9** (18)	5.9 (15)	5.9 (15)	5.9 (15)

Table 6.3.: Minimum SNR (in dB) and number of IDD iterations (in parenthesis) required to achieve the targeted a posteriori MI $I^{p}_{\nu^{-1}} = 0.9988$ bpcu for each combination of a bit mapping of the 8-star and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the NDE AWGN model. The best combination is highlighted in bold; the least complex combination is highlighted in italics. Their EXIT charts are shown in figures 6.6 and 6.7, respectively.

		N_{LDPC}									
		1	2	3	4	5	10	20	30	40	50
μ	1	n.a.	n.a.	n.a.	10.6	11.1	11.4	11.4	11.4	11.4	11.4
	2	n.a.	n.a.	10.2	11.0	11.2	11.3	11.3	11.3	11.3	11.3
	3	n.a.	n.a.	10.2	11.0	11.1	11.1	11.1	11.1	11.1	11.1
	4	n.a.	9.9	10.8	10.9	10.9	11.0	11.0	11.0	11.0	11.0
	5	n.a.	n.a.	10.7	11.1	11.1	11.2	11.2	11.2	11.2	11.2
	6	n.a.	9.9	10.7	10.7	10.7	10.8	10.8	10.8	10.8	10.8
	7,aG	n.a.	10.4	10.6	10.6	10.7	10.7	10.7	10.7	10.7	10.7
	8	n.a.	n.a.	10.7	10.8	10.8	10.8	10.8	10.8	10.8	10.8
	9	n.a.	n.a.	9.9	10.7	11.2	11.5	11.5	11.5	11.5	11.5
	10	n.a.	n.a.	10.7	11.0	11.0	11.1	11.1	11.1	11.1	11.1
	11	n.a.	n.a.	n.a.	10.3	10.8	11.6	11.6	11.6	11.6	11.6
	12,G	n.a.	n.a.	n.a.	*9.9*	10.3	11.4	**11.8**	11.8	11.8	11.8

Table 6.4.: NCG over the uncoded QPSK (in dB) of each combination of a bit mapping of the 8-star and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the NDE AWGN model. Bold and italics highlightings are the same as in table 6.3.

It is interesting to notice that the most complex combination, given by μ_4 and $N_{\text{LDPC}} = 40$ with $N_{\text{IDD}} = 44$, actually needs a 0.8 dB higher SNR than the best one.

Figures 6.6 and 6.7 show the EXIT charts of the best and the least complex combinations.

In figure 6.6, corresponding to the best combination, the transfer characteristics of the demapper and of the decoder match very well, at least in the central region. On the left region the mismatch is due to the transfer characteristic of the decoder "rising too slowly", see figure 6.5, or alternatively to the transfer characteristic of the demapper "being too straight". Things are better on the right region, where the two transfer characteristics are closer and the area between them is smaller.

This mismatch is responsible for the 1.0 dB capacity loss between the values reported in tables 6.1 and 6.3. The area between the two transfer characteristics, in fact, relates to the capacity loss of the combination, as shown by Ashikhmin rigorously at least for the binary erasure channel (BEC) [75, 76]. Unfortunately, no bit mapping exhibit a transfer characteristic which bends more downwards for $I^a_{\mu^{-1}} \leq 0.25$ bpcu, they are all quite straight and differ only in their "slope": a bit mapping with a steeper transfer characteristic would reduce the gap on the left region of the EXIT chart, but it would increase the gap on its right region. Similarly, no number of LDPC iterations exhibit a more squared transfer characteristic around $I^a_{\nu^{-1}} \approx 0.60$ bpcu.

In figure 6.7, corresponding to the least complex combination, the transfer characteristics of the demapper and the decoder are very far. This does allow the IDD stage to achieve the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu with very few IDD iterations, but it is also responsible for the remarkable 2.9 dB capacity loss.

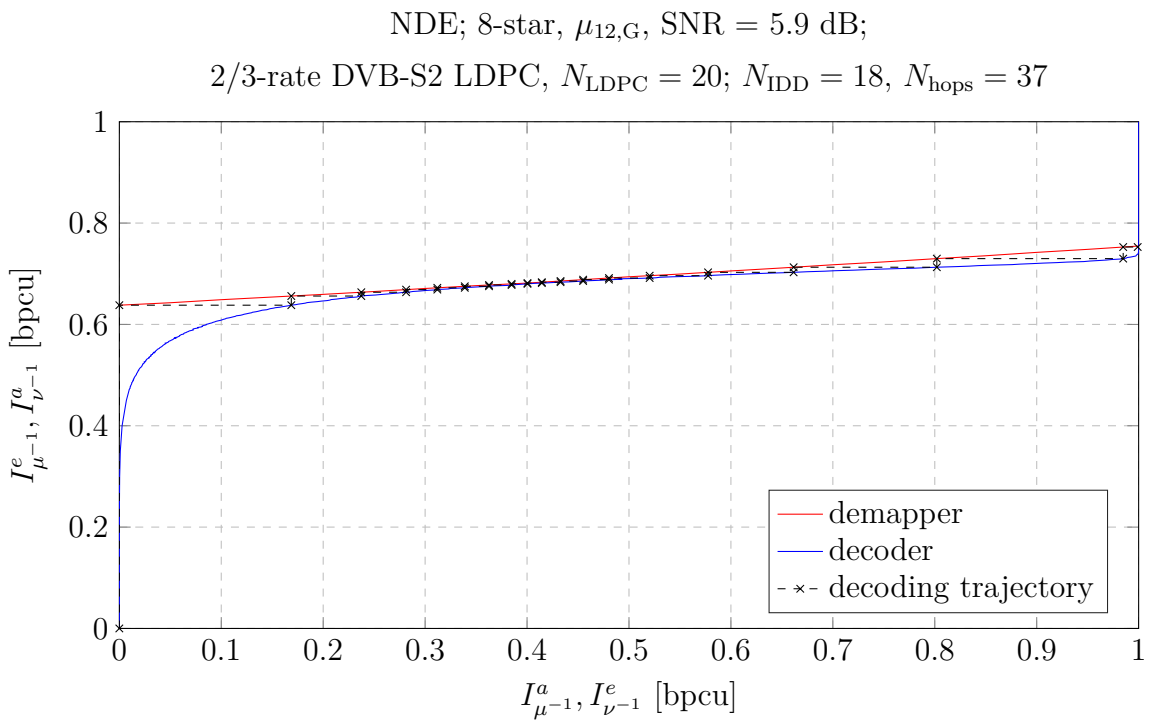

Figure 6.6.: EXIT chart of the best combination over the NDE AWGN model.

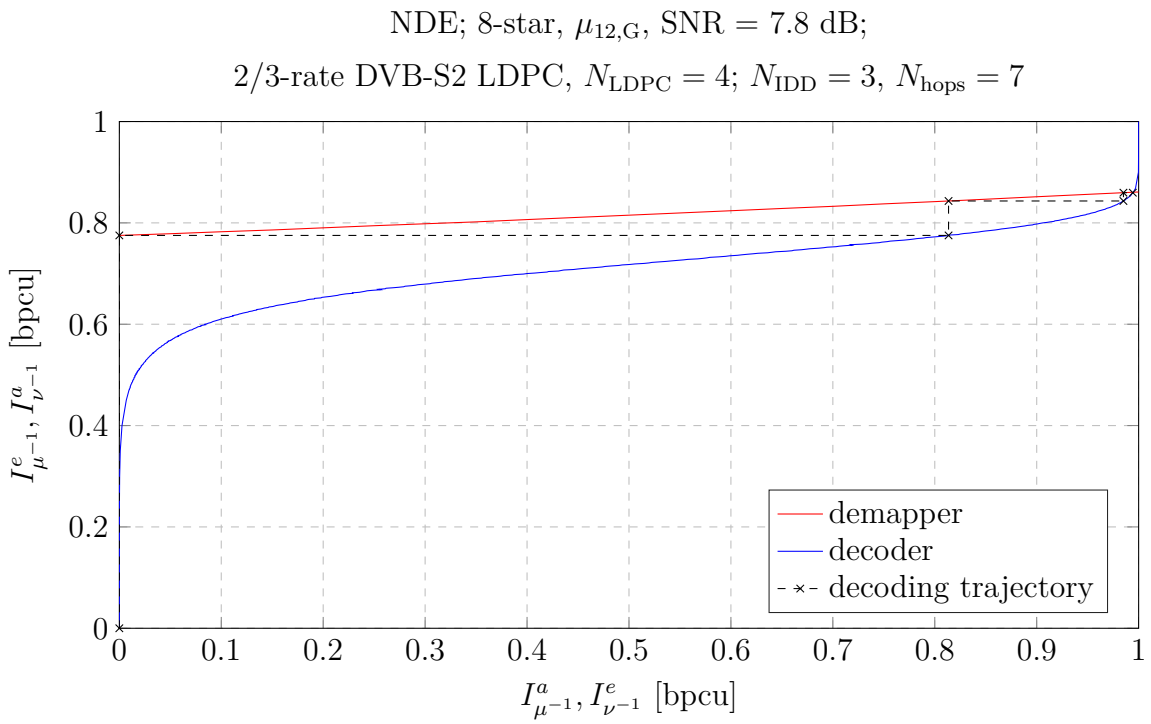

Figure 6.7.: EXIT chart of the least complex combination over the NDE AWGN model.

6.2.8. NDE AWGN model: estimation of the complexity

In order to evaluate the complexity of the IDD stage I calculated the number of summations needed

- by the demapper to implement (A.6), (A.7) and (A.8) (see appendix A) as well as
- by the decoder to implement the SD LDPC decoding algorithm [77, 78]

under the following assumptions:

1. the two's complement representation is used;
2. the max-log approximation (MLA) is used, according to which [79]

$$\sum_i a_i \approx \max_i\{a_i\}. \tag{6.4}$$

   Note that, whereas in practice a high performing system would resort to more accurate approximations, the MLA arithmetic provides a simple criterion to compare the relative overall complexity of different systems, since it requires only summations;

3. the min-sum algorithm (MSA) is used, according to which [79]

$$\prod_i \tanh(a_i) \approx \tanh(\min_i\{a_i\}); \tag{6.5}$$

4. the complexity of the operations "max/min between N quantities" is $N-1$ summations. For instance, in order to find the maximum between the three numbers a, b and c, one could:
   a) calculate $a-b$ (first summation).
      If $a-b>0$ keep a,
      else keep b,
      call d the number kept;
   b) calculate $c-d$ (second summation).
      If $c-d>0$ c is the maximum,
      else d is the maximum;
5. the complexity of the operation "absolute value of a real quantity" is 1 summation, since in the two's complement representation changing the sign means changing all bits and then adding 1;
6. the complexity of the product of sign operations can be neglected, since it can be implemented with a binary exclusive OR (XOR) [80].

6.2.9. NDE AWGN model: complexity of the demapper at each IDD iteration

Applying the MLA (6.4) and exploiting the fact that the exponential is a strictly monotonically increasing function, so that I can write

$$\max_i\{e^{a_i}\} = e^{\max_i\{a_i\}}, \tag{6.6}$$

I obtained the following approximated versions of the extrinsic LLR of the bit i $\Lambda^{e,i}_{\mu^{-1}}$ (the exact versions are reported in appendix A):

$$\begin{aligned}\Lambda^{e,2}_{\mu^{-1}} \approx &+ \max\{-d(000) + \Lambda^{a,1}_{\mu^{-1}} + \Lambda^{a,0}_{\mu^{-1}}; -d(001) + \Lambda^{a,1}_{\mu^{-1}}; -d(010) + \Lambda^{a,0}_{\mu^{-1}}; -d(011)\} + \\ &- \max\{-d(100) + \Lambda^{a,1}_{\mu^{-1}} + \Lambda^{a,0}_{\mu^{-1}}; -d(101) + \Lambda^{a,1}_{\mu^{-1}}; -d(110) + \Lambda^{a,0}_{\mu^{-1}}; -d(111)\},\end{aligned} \tag{6.7}$$

$$\begin{aligned}\Lambda^{e,1}_{\mu^{-1}} \approx &+ \max\{-d(000) + \Lambda^{a,2}_{\mu^{-1}} + \Lambda^{a,0}_{\mu^{-1}}; -d(001) + \Lambda^{a,2}_{\mu^{-1}}; -d(100) + \Lambda^{a,0}_{\mu^{-1}}; -d(101)\} + \\ &- \max\{-d(010) + \Lambda^{a,2}_{\mu^{-1}} + \Lambda^{a,0}_{\mu^{-1}}; -d(011) + \Lambda^{a,2}_{\mu^{-1}}; -d(110) + \Lambda^{a,0}_{\mu^{-1}}; -d(111)\},\end{aligned} \tag{6.8}$$

$$\begin{aligned}\Lambda^{e,0}_{\mu^{-1}} \approx &+ \max\{-d(000) + \Lambda^{a,2}_{\mu^{-1}} + \Lambda^{a,1}_{\mu^{-1}}; -d(010) + \Lambda^{a,2}_{\mu^{-1}}; -d(100) + \Lambda^{a,1}_{\mu^{-1}}; -d(110)\} + \\ &- \max\{-d(001) + \Lambda^{a,2}_{\mu^{-1}} + \Lambda^{a,1}_{\mu^{-1}}; -d(011) + \Lambda^{a,2}_{\mu^{-1}}; -d(101) + \Lambda^{a,1}_{\mu^{-1}}; -d(111)\},\end{aligned} \tag{6.9}$$

where I call

$$d(b_2b_1b_0) \triangleq \frac{|y - \mu_k(b_2, b_1, b_0)|^2}{N_0} \tag{6.10}$$

for space reasons.

The normalised squared Euclidean distance (6.10) between the received symbol y and the transmitted symbol to which the bit mapping μ_k maps the bits b_2, b_1, b_0 is the term carrying the information coming from the channel. Although in this term operations other than summations are indeed present, the complexity of its calculation can be neglected, for this term has to be calculated only once and then it remains constant throughout the IDD iterations (what changes at each IDD iteration are the a priori LLRs). Note that these approximated formulas are only used to calculate the complexity of the demapper, whereas I used the exact formulas to calculate the transfer characteristics and the EXIT charts.

Looking at (6.7), (6.8) and (6.9) some general rules can be inferred:

1. each max is between 2^{m-1} numbers;
2. in each max the a priori LLR of the *one* bit whose extrinsic LLR is being calculated is never added;
3. in each max the a priori LLRs of the *other* bits whose extrinsic LLRs are not being calculated are added when the corresponding bits are 0 in the quantity (6.10).

Hence, in each max the number of a priori LLRs being added equals

$$\sum_{i=0}^{m-1} \binom{m-1}{i} i = (m-1)2^{m-2}, \tag{6.11}$$

because I need

- 0 summations for the $\binom{m-1}{0} = 1$ combination with 0 bits out of $m-1$ being 0,
- 1 summation for each of the $\binom{m-1}{1} = m-1$ combinations with 1 bit out of $m-1$ being 0,
- 2 summations for each of the $\binom{m-1}{2} = (m-1)(m-2)/2$ combinations with 2 bits out of $m-1$ being 0,
- ...,
- $m-1$ summations for the $\binom{m-1}{m-1} = 1$ combination with $m-1$ bits out of $m-1$ being 0.

In (6.11) I used the known identity

$$\sum_{k=0}^{n} k\binom{n}{k} = n2^{n-1}, \tag{6.12}$$

with $k = i$ and $n = m-1$.

Overall I need

- (6.11) summations, because each a priori LLR entails a summation,
- $2^{m-1} - 1$ summations for the complexity of the max,
- twice as many summations because there are two max's and, finally,
- 1 additional summation for the difference between the two max's.

The number of summations needed by the demapper to calculate the extrinsic LLR of one *encoded* bit at *each* IDD iteration is hence given by:

$$2((m-1)2^{m-2} + 2^{m-1} - 1) + 1 = 2^{m-1}(m+1) - 1. \tag{6.13}$$

6.2.9.1. NDE AWGN model: complexity of the demapper per information bit at each IDD iteration

However, I am interested in the number of summations needed by the demapper to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration, which means that I have to rescale (6.13) by the inverse of the code rate, obtaining

$$K'_{\mu^{-1}} = \frac{1}{R}(2^{m-1}(m+1) - 1). \tag{6.14}$$

Since

- the chosen constellation has $m = 3$ bits/symbol and
- the 2/3-rate DVB-S2 LDPC code has code rate $R = 2/3$,

the complexity of the demapper of the proposed solution at *each* IDD iteration is

$$K'_{\mu^{-1}} = 22.5 \text{ spib}.$$

6.2.10. Complexity of the decoder at each IDD iteration

In a SD LDPC decoder, the variable and the check nodes iteratively exchange information to refine the LLRs. Hence, to estimate the complexity of the decoder it is helpful to estimate the complexity of the two steps separately and then add them up.

6.2.10.1. Complexity of the message from one variable node to one check node

Without the need to invoke any approximation, the message sent from a variable node V_i to *one* check node C_j is given by [79]:

$$M_{V_i \to C_j} = \Lambda^{a,i}_{\nu^{-1}} + \sum_{j' \neq j} M_{C_{j'} \to V_i},$$

where $M_{C_{j'} \to V_i}$ are the messages sent from the *other* check nodes $C_{j'}$ to the variable node V_i.

The number of these messages clearly equals the number of the *other* check nodes $C_{j'}$ connected to the variable node V_i. One way to know the latter is to use the degree distribution *from the edge perspective*:

$$\sum_{\ell=1}^{\lambda_{\max}} \lambda_\ell (\ell - 1), \tag{6.15}$$

where the quantity λ_ℓ indicates the fraction of edges connected to a variable node of degree ℓ, which can be obtained from the code matrix [79], and the -1 excludes the edge between the variable node V_i and the check node C_j on which the message currently being calculated will travel.

Overall I need

- (6.15) -1 summations, because between n elements there are $n - 1$ summations, and
- 1 additional summation for the a priori LLR,

so that in the end the *average* number of summations needed by the decoder to calculate one message sent from a variable node V_i to a check node C_j at *each* LDPC iteration and at *each* IDD iteration is also given by (6.15).

6.2.10.2. Complexity of the message from one check node to one variable node

Applying the MSA (6.5), the message travelling in the other direction, i.e. the one sent from a check node C_j to *one* variable node V_i, is given by [79]:

$$M_{C_j \to V_i} \approx (\prod_{i' \neq i} \text{sign}(M_{V_{i'} \to C_j})) \min_{i' \neq i} \{|M_{V_{i'} \to C_j}|\},$$

where $M_{V_{i'} \to C_j}$ are the messages sent from the *other* variable nodes $V_{i'}$ to the check node C_j.

As before, the number of these messages equals the number of the *other* variable nodes $V_{i'}$ connected to the check node C_j, and the latter can be calculated with the degree distribution *from the edge perspective*:

$$\sum_{\ell=1}^{\rho_{\max}} \rho_\ell (\ell - 1), \tag{6.16}$$

where the quantity ρ_ℓ indicates the fraction of edges connected to a check node of degree ℓ, which as before can be obtained from the code matrix [79], and as before the -1 excludes the edge between the check node C_j and the variable node V_i on which the message currently being calculated will travel.

Overall I need

- (6.16) summations for the absolute values, and
- (6.16) -1 summations for the complexity of the min,

which means that the *average* number of summations needed by the decoder to calculate one message sent from a check node C_j to a variable node V_i at *each* LDPC iteration and at *each* IDD iteration is given by

$$\sum_{\ell=1}^{\rho_{\max}} \rho_\ell (\ell - 1) + (\sum_{\ell=1}^{\rho_{\max}} \rho_\ell (\ell - 1) - 1) = \sum_{\ell=1}^{\rho_{\max}} \rho_\ell (2\ell - 3). \tag{6.17}$$

6.2.10.3. Complexity of all messages

The sum of (6.15) and (6.17) gives the *average* number of summations needed by the decoder to calculate the *two* messages travelling on *one* edge in *both* directions at *each* LDPC iteration and at *each* IDD iteration.

Multiplying it by the number of edges N_e I hence obtain the number of summations needed by the decoder to calculate the extrinsic LLRs of n *encoded* bits (i.e., of one code word) at *each* LDPC iteration and at *each* IDD iteration:

$$N_\text{e}(\sum_{\ell_1=1}^{\lambda_{\max}} \lambda_{\ell_1} (\ell_1 - 1) + \sum_{\ell_2=1}^{\rho_{\max}} \rho_{\ell_2} (2\ell_2 - 3)). \tag{6.18}$$

ℓ	λ_ℓ	ρ_ℓ
1	$1/215999 \approx 4.6297 \cdot 10^{-6}$	0
2	$43198/215999 \approx 1.9999 \cdot 10^{-1}$	0
3	$116640/215999 \approx 5.4000 \cdot 10^{-1}$	0
...	...	...
9	0	$9/215999 \approx 4.1667 \cdot 10^{-5}$
10	0	$215990/215999 \approx 9.9996 \cdot 10^{-1}$
...	...	...
13	$56160/215999 \approx 2.6000 \cdot 10^{-1}$	0
Total	$215999/215999 = 1$	$215999/215999 = 1$

Table 6.5.: Degree distribution *from the edge perspective* of the 2/3-rate DVB-S2 LDPC code, indicating the fraction of edges connected to a variable (λ_ℓ) or to a check (ρ_ℓ) node of degree ℓ. The graph has $N_\mathrm{e} = 215999$ edges.

6.2.10.4. Complexity of the decoder per information bit at each IDD iteration

Dividing (6.18) by the information word length k I obtain the number of summations needed by the decoder to calculate the extrinsic LLR of one *information* bit at *each* LDPC iteration and at *each* IDD iteration:

$$K''_{\nu^{-1}} = \frac{N_\mathrm{e}}{k}\left(\sum_{\ell_1=1}^{\lambda_{\max}} \lambda_{\ell_1}(\ell_1 - 1) + \sum_{\ell_2=1}^{\rho_{\max}} \rho_{\ell_2}(2\ell_2 - 3)\right). \tag{6.19}$$

Finally, multiplying (6.19) by the number of LDPC iterations N_{LDPC} I have the number of summations needed by the decoder to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration:

$$K'_{\nu^{-1}} = N_{\mathrm{LDPC}} K''_{\nu^{-1}} = N_{\mathrm{LDPC}} \frac{N_\mathrm{e}}{k}\left(\sum_{\ell_1=1}^{\lambda_{\max}} \lambda_{\ell_1}(\ell_1 - 1) + \sum_{\ell_2=1}^{\rho_{\max}} \rho_{\ell_2}(2\ell_2 - 3)\right). \tag{6.20}$$

Table 6.5 shows the degree distribution *from the edge perspective* of the 2/3-rate LDPC code from the DVB-S2 standard [66].

Since the 2/3-rate DVB-S2 LDPC code has

- the degree distribution *from the edge perspective* shown in table 6.5,
- $N_\mathrm{e} = 215999$ edges and
- $k = 43200$ information bits,

the complexity of the decoder of the proposed solution at *each* IDD iteration is

$$K'_{\nu^{-1}} \approx 107.0 N_{\mathrm{LDPC}} \text{ spib}.$$

Recall now that the complexity of the demapper of the proposed solution at *each* IDD iteration is $K'_{\mu^{-1}} = 22.5$ spib (see section 6.2.9.1): as anticipated in section 6.2.7, the complexity of the demapper is indeed negligible when compared to the complexity of the decoder, since the latter is approximately a factor 4.8 N_{LDPC} that of the former.

6.2.11. NDE AWGN model: complexity of the whole IDD stage

The sum of (6.14) and (6.20) represents the number of summations needed by the whole IDD scheme (demapper and decoder) to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration:

$$\begin{aligned} K' &= K'_{\mu^{-1}} + K'_{\nu^{-1}} = \\ &= \frac{1}{R}(2^{m-1}(m+1)-1) + N_{\mathrm{LDPC}}\frac{N_{\mathrm{e}}}{k}(\sum_{\ell_1=1}^{\lambda_{\max}} \lambda_{\ell_1}(\ell_1-1) + \sum_{\ell_2=1}^{\rho_{\max}} \rho_{\ell_2}(2\ell_2-3)). \end{aligned} \tag{6.21}$$

Finally, if I multiply (6.21) by the number of IDD iterations N_{IDD} I get the number of summations needed by the whole IDD scheme to calculate the extrinsic LLR of one *information* bit:

$$\begin{aligned} K =& N_{\mathrm{IDD}}K' = \\ =& N_{\mathrm{IDD}}(\frac{1}{R}(2^{m-1}(m+1)-1) + N_{\mathrm{LDPC}}\frac{N_{\mathrm{e}}}{k}(\sum_{\ell_1=1}^{\lambda_{\max}} \lambda_{\ell_1}(\ell_1-1) + \sum_{\ell_2=1}^{\rho_{\max}} \rho_{\ell_2}(2\ell_2-3))). \end{aligned} \tag{6.22}$$

With the numbers reported in section 6.2.9.1 and section 6.2.10.4 the overall complexity of the proposed solution is

$$K \approx N_{\mathrm{IDD}}(22.5 + 107.0N_{\mathrm{LDPC}}) \text{ spib.}$$

Table 6.6 shows the overall complexity of each combination of a demapper transfer characteristic and a decoder transfer characteristic.

The best combination has an overall complexity $K \approx 38.92$ kspib. The least complex combination has an overall complexity $K \approx 1.35$ kspib. The most complex combination has an overall complexity $K \approx 189.31$ kspib.

6.2.12. NDE AWGN model: 2A4P instead of 8-star?

The best combination achieves the targeted a posteriori MI for SNR = 5.9 dB, but as I showed in figure 6.6 it has an EXIT chart in which the transfer characteristics of the demapper and of the decoder match do not match very well on the sides. This mismatch is responsible for the 1.0 dB gap to the theoretical limit of the 8-star.

One could then ask, what if I use another constellation? Maybe there is another 8-ary constellation, with a higher theoretical limit, but with a combination that achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu at a lower SNR. In the end I am interested in what can be practically achieved, more than in what the theoretical limits are.

I therefore went through the same procedure as described so far with a 2 amplitude-4 phase-shift keying (2A4P) which, as I will show in more detail in section 6.3, admits 18 distinct bit mappings. The results was that the "best" combination of a 2A4P bit mapping and the 2/3-rate DVB-S2 LDPC code

		N_{LDPC}									
		1	2	3	4	5	10	20	30	40	50
μ	1	n.a.	n.a.	n.a.	0.003	0.005	0.025	0.043	0.065	0.086	0.107
	2	n.a.	n.a.	0.002	0.005	0.007	0.015	0.026	0.039	0.052	0.064
	3	n.a.	n.a.	0.002	0.007	0.027	0.020	0.037	0.055	0.077	0.091
	4	n.a.	0.002	0.007	0.009	0.008	0.038	0.067	0.094	0.189	0.183
	5	n.a.	n.a.	0.003	0.010	0.007	0.015	0.026	0.039	0.052	0.059
	6	n.a.	0.002	0.012	0.006	0.007	0.028	0.054	0.078	0.108	0.134
	7,aG	n.a.	0.005	0.007	0.005	0.011	0.014	0.026	0.039	0.052	0.064
	8	n.a.	n.a.	0.006	0.013	0.010	0.015	0.028	0.042	0.056	0.070
	9	n.a.	n.a.	0.002	0.003	0.006	0.019	0.028	0.042	0.052	0.064
	10	n.a.	n.a.	0.003	0.009	0.007	0.034	0.061	0.087	0.151	0.161
	11	n.a.	n.a.	n.a.	0.002	0.003	0.024	0.035	0.052	0.069	0.086
	12,G	n.a.	n.a.	n.a.	*0.001*	0.002	0.007	**0.039**	0.048	0.065	0.081

Table 6.6.: Overall complexity (in Mspib) required to achieve the targeted a posteriori MI $I^{p}_{\nu^{-1}} = 0.9988$ bpcu for each combination of a bit mapping of the 8-star and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the NDE AWGN model. Bold and italics highlightings are the same as in table 6.3.

- achieves the targeted a posteriori MI for SNR = 6.2 dB, which is 1.1 dB away from the theoretical limit of the 2A4P (see table 6.1) and, most importantly, 0.3 dB more than the best combination identified in section 6.2.7;
- has a similar overall complexity, since it requires $N_{\text{LDPC}} = 30$ LDPC iterations and $N_{\text{IDD}} = 12$ IDD iterations.

In other words, the best combination found in section 6.2.7 is to be preferred.

6.3. Proposed solution for 100G long-haul DE coherent optical communications systems: 2A4P + DVB-S2-LDPC(64800, 43200) + ITU-T-RS(255, 239)

In this section I will show the results relative to 100G long-haul DE coherent optical communications systems; since many concepts are similar to those introduced in section 6.2 I will keep the exposition short, highlighting only the parts that changed.

Table 6.7 shows the minimum SNR required to achieve the targeted MI $I = 1.87$ bpcu per polarisation for the constellations considered in chapter 5 with a symbol rate $f_s = 27.95$ GBd over the DE AWGN model. Also reported is how much OH they offer space for, together with the corresponding potential gain over the coded QPSK and

Constellation	OH	Min. SNR [dB]	Pot. gain over *coded* QPSK [dB]	Pot. NCG over *uncoded* QPSK [dB]
QPSK	6.69%	7.54	n.r.	+10.25
4-star		7.55	0.00	+10.25
4PAM		10.58	−3.03	+07.22
8PSK	60.04%	7.52	+0.02	+10.28
8-star		6.80	+0.75	+11.00
8PAM		9.56	−2.02	+08.24
2A4P		**6.58**	**+0.97**	**+11.22**
16-star	113.39%	6.47	+1.07	+11.33
4A4P		6.52	+1.02	+11.27

Table 6.7.: Minimum SNR (in dB) required to achieve the targeted MI $I = 1.87$ bpcu per polarisation for the constellations considered in chapter 5 with a symbol rate $f_s = 27.95$ GBd, potential gain over the coded QPSK and potential NCG over the uncoded QPSK over the DE AWGN model. Also reported is the redundancy offered. The constellation chosen in chapter 5 for 100G long-haul DE coherent optical communications systems is highlighted in bold.

potential NCG over the uncoded QPSK, which needs a SNR = 18.09 dB to achieve a BER = 10^{-15}. The values relative to the 2A4P chosen in chapter 5 are marked in bold. Table 6.1 will be useful to assess how close the practical proposed solutions are to their theoretical performance.

6.3.1. DE AWGN model: estimation of the EXIT charts

I estimated the transfer characteristics of the demapper again following [68]. The formulas for the extrinsic LLRs are very similar to those in the NDE case, the only difference is the term which carries the information coming from the channel, since now the channel model is different. For the interested reader, in appendix A I explain in detail the procedure.

Obviously, the transfer characteristics of the decoder as well as the procedure to find the minimum SNR and the number of IDD iterations required to have a successful decoding trajectory remained the same as in the NDE case.

6.3.2. DE AWGN model: 2A4P bit mappings as well as SNR and a priori MI ranges considered

Using the same exhaustive search described in 6.2.2 I found 18 distinct bit mappings for the 2A4P, out of the 8! = 40320 possible bit mappings it admits ($\approx 0.04\%$).

Figure 6.8 shows these 18 distinct bit mappings.

In particular,

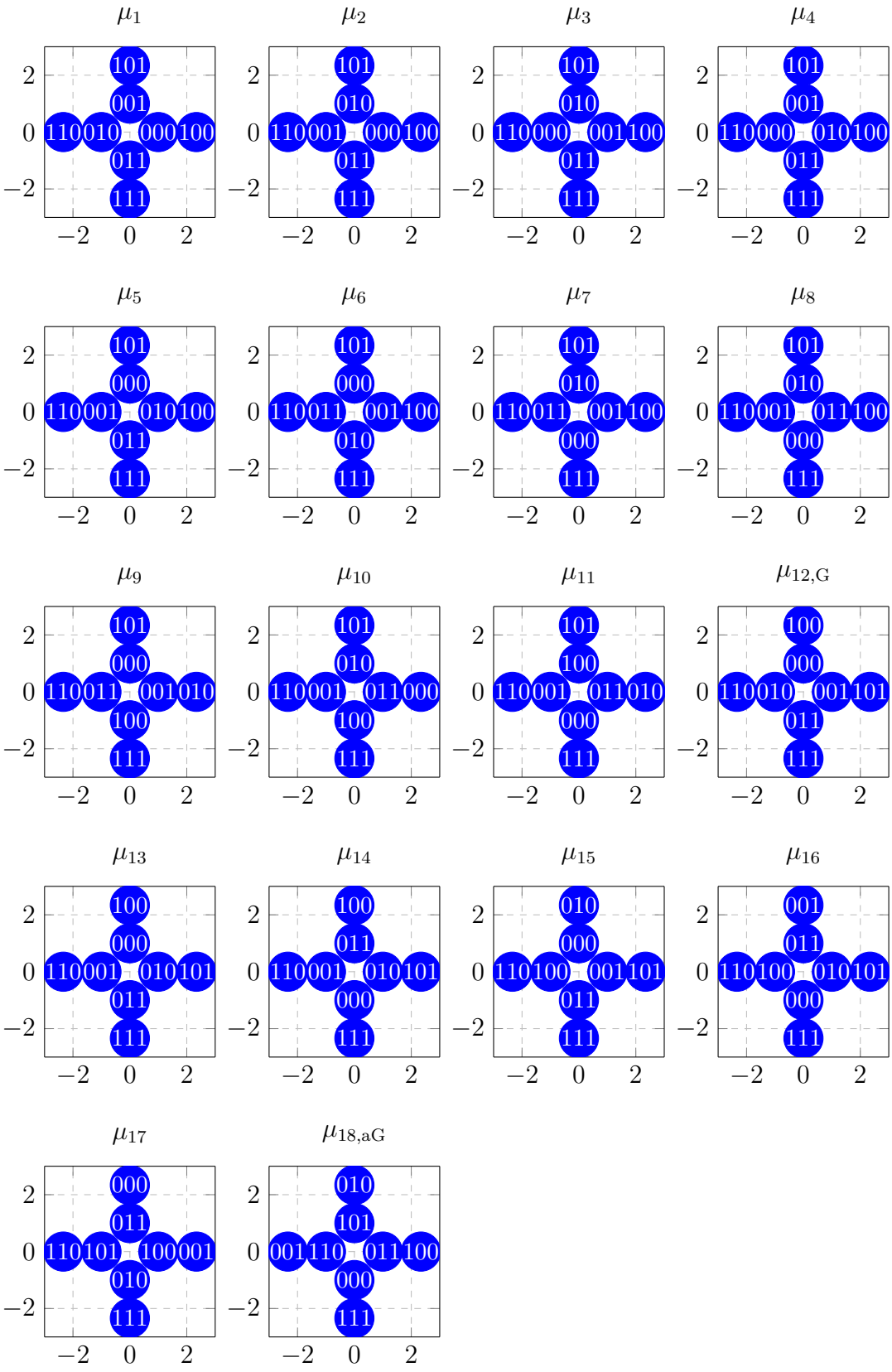

Figure 6.8.: Eighteen distinct bit mappings admitted by the 2A4P. The bit mapping $\mu_{12,\mathrm{G}}$ is the Gray-coded bit mapping, the bit mapping $\mu_{18,\mathrm{aG}}$ is the farthest from a Gray-coded bit mapping.

Fixed variable	Value
SNR	6.6 dB
Bit mapping	μ_{15}

Table 6.8.: Variables fixed or highlighted in figures 6.9 and 6.10.

- the bit mapping $\mu_{12,\mathrm{G}}$ is the Gray-coded bit mapping (contrarily to the 8-star, the 2A4P does admit one); conversely,
- the bit mapping $\mu_{18,\mathrm{aG}}$ is the farthest from a Gray-coded bit mapping (i.e. the anti-Gray-coded bit mapping).

I chose the SNR range again asymmetrically around the minimum SNR required for the 2A4P to achieve the targeted MI $I = 1.87$ bpcu per polarisation. I considered the same range for the a priori MI I used for the NDE case.

6.3.3. DE AWGN model: values of interest

Clearly, even in the DE case the extrinsic MI of the demapper depends on the same three variables.

In section 6.3.4 I will hence show the results as explained in section 6.2.4.

The decision on which value to fix each variable is motivated as follows:

1. I chose to fix the SNR to 6.6 dB, because it is the theoretical limit of the 2A4P over the DE AWGN model, see table 6.7. The transfer characteristics of the demapper for the value of SNR that give the best match as well as the least complex match with the transfer characteristic of the decoder will be shown in the corresponding EXIT charts;
2. I chose to fix the bit mapping to one "value", the μ_{15}, because, as I will soon show, it is the one that gives the best match with the transfer characteristics of the decoder.

Table 6.8 summarizes which variable I fixed or highlighted to which value in figures 6.9 and 6.10.

6.3.4. DE AWGN model: transfer characteristics of the demapper

Figure 6.9 shows the extrinsic MI of the 2A4P as a function of the a priori MI, for a fixed value of the SNR = 6.6 dB and parametrised by the bit mappings.

In figure 6.9 all 18 lines are visible, which further confirms the correctness of the technique described in section 6.2.2 to find the unique bit mappings.

The bit mapping $\mu_{12,\mathrm{G}}$ has a literally *flat* curve, exhibiting the *same* extrinsic MI for $I^a_{\mu^{-1}} = 0$ bpcu and for $I^a_{\mu^{-1}} = 1$ bpcu. This confirms it as the Gray-coded bit mapping, as said in section 6.3.2.

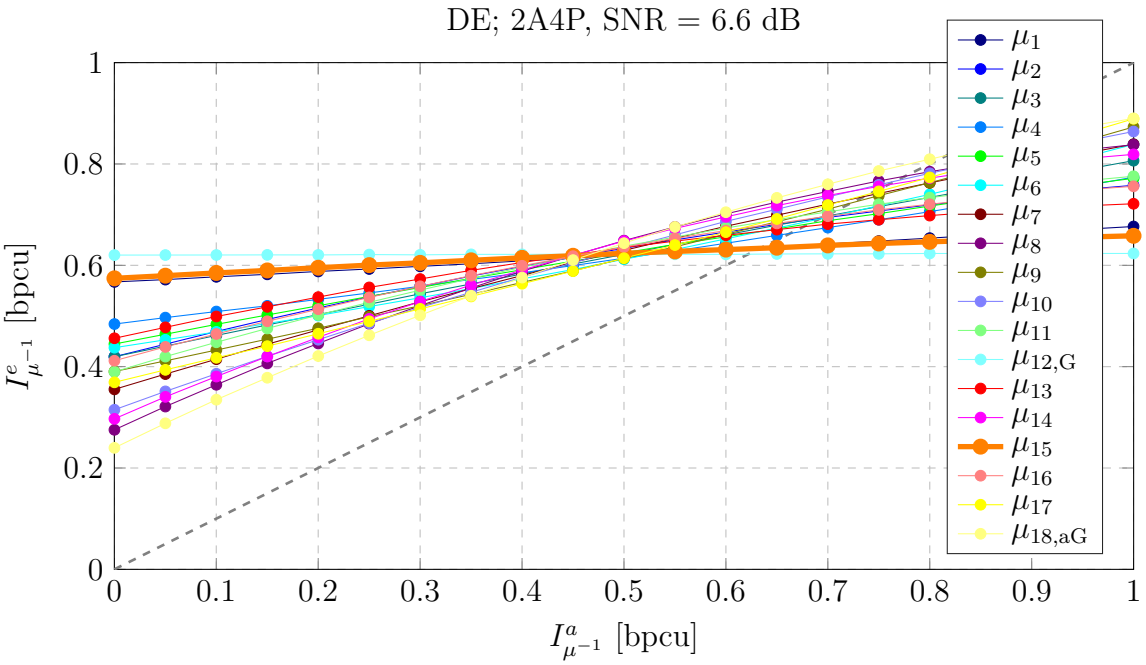

Figure 6.9.: Extrinsic MI of the 2A4P as a function of the a priori MI, for a SNR = 6.6 dB and parametrised by the bit mappings over the DE AWGN model.

The bit mapping $\mu_{18,aG}$ has the *steepest* curve, which confirms it as the anti-Gray-coded bit mapping, as said in section 6.3.2.

Similarly to figure 6.3, the extrinsic MI always increases with the a priori MI, as expected (with the exception of the bit mapping $\mu_{12,G}$).

Similarly to figure 6.3, the curves cross the bisector $I^e_{\mu^{-1}} = I^a_{\mu^{-1}}$ around $I^a_{\mu^{-1}} = 0.70$ bpcu, being above it for roughly $I^a_{\mu^{-1}} \leq 0.6$ bpcu and below it for roughly $I^a_{\mu^{-1}} \geq 0.8$ bpcu.

Figure 6.10 shows the extrinsic MI of the 2A4P as a function of the SNR, for the bit mapping μ_{15} and parametrised by various values of the a priori MI.

Similarly to figure 6.4, the extrinsic MI increases with the SNR, tending to $I^e_{\mu^{-1}} = 1$ bpcu for SNR $\to +\infty$, and increases also with the a priori MI.

6.3.5. DE AWGN model: estimation of the minimum SNR required

Table 6.9 shows, for each combination of a demapper transfer characteristic and a decoder transfer characteristic, the minimum SNR (in dB) and the corresponding number of IDD iterations (in parenthesis) at which the decoder achieved the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu.

Table 6.10 shows the same results but in terms of NCG over the uncoded QPSK.

Even in the DE case, certain combinations do not achieve the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for any value of the SNR I considered. In particular, the bit mapping $\mu_{18,aG}$ never makes it.

μ	N_{LDPC} = 1	2	3	4	5	10	20	30	40	50
1	n.a.	n.a.	n.a.	n.a.	9.6 (03)	8.3 (04)	7.6 (32)	7.6 (18)	7.6 (17)	7.6 (16)
2	n.a.	n.a.	9.4 (05)	8.5 (17)	8.5 (13)	8.5 (10)	8.4 (64)	8.5 (10)	8.5 (10)	8.5 (10)
3	n.a.	n.a.	8.7 (18)	8.6 (20)	8.6 (14)	8.6 (11)	8.6 (10)	8.6 (10)	8.6 (10)	8.6 (10)
4	n.a.	n.a.	9.3 (05)	8.4 (11)	8.2 (14)	8.1 (14)	8.1 (12)	8.1 (12)	8.1 (12)	8.1 (12)
5	n.a.	n.a.	8.9 (09)	8.4 (27)	8.4 (16)	8.4 (11)	8.4 (10)	8.4 (10)	8.4 (10)	8.4 (10)
6	n.a.	9.0 (32)	8.8 (16)	8.7 (15)	8.6 (29)	8.6 (14)	8.6 (13)	8.6 (13)	8.6 (14)	8.6 (13)
7	n.a.	9.3 (25)	9.2 (15)	9.1 (24)	9.1 (18)	9.1 (15)	9.1 (14)	9.1 (14)	9.1 (14)	9.1 (14)
8	n.a.	n.a.	n.a.	9.6 (30)	9.6 (22)	9.6 (20)	9.6 (19)	9.6 (19)	9.6 (20)	9.6 (19)
9	n.a.	9.2 (17)	8.9 (44)	8.9 (14)	8.9 (11)	8.8 (20)	8.8 (19)	8.8 (19)	8.8 (19)	8.8 (20)
10	n.a.	9.5 (17)	9.3 (38)	9.3 (17)	9.3 (14)	9.3 (12)	9.3 (12)	9.3 (12)	9.3 (12)	9.3 (12)
11	n.a.	n.a.	9.1 (07)	8.7 (14)	8.7 (12)	8.6 (27)	8.6 (27)	8.6 (28)	8.6 (28)	8.6 (29)
12,G	n.a.	n.a.	n.a.	n.a.	n.a.	*8.9* (01)	8.3 (01)	8.1 (01)	8.0 (02)	8.0 (01)
13	n.a.	n.a.	n.a.	9.0 (06)	8.6 (06)	8.2 (13)	8.2 (12)	8.2 (12)	8.2 (12)	8.2 (12)
14	n.a.	9.6 (27)	9.5 (23)	9.5 (15)	9.5 (14)	9.5 (12)	9.5 (12)	9.5 (12)	9.5 (12)	9.5 (12)
15	n.a.	n.a.	n.a.	n.a.	n.a.	8.6 (02)	7.9 (04)	7.7 (06)	**7.6** (10)	7.6 (09)
16	n.a.	n.a.	9.5 (05)	8.7 (09)	8.5 (14)	8.4 (65)	8.4 (62)	8.5 (10)	8.5 (10)	8.5 (10)
17	n.a.	9.2 (28)	9.0 (37)	9.0 (14)	9.0 (11)	8.9 (25)	8.9 (24)	8.9 (23)	8.9 (25)	8.9 (24)
18,aG	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 6.9.: Minimum SNR (in dB) and number of IDD iterations (in parenthesis) required to achieve the targeted a posteriori MI $I^{p}_{\nu^{-1}} = 0.9988$ bpcu for each combination of a bit mapping of the 2A4P and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the DE AWGN model. The best combination is highlighted in bold; the least complex combination is highlighted in italics. Their EXIT charts are shown in figures 6.11 and 6.12, respectively.

		N_{LDPC}									
		1	2	3	4	5	10	20	30	40	50
μ	1	n.a.	n.a.	n.a.	n.a.	8.2	9.5	10.2	10.2	10.2	10.2
	2	n.a.	n.a.	8.4	9.3	9.3	9.3	9.4	9.3	9.3	9.3
	3	n.a.	n.a.	9.1	9.2	9.2	9.2	9.2	9.2	9.2	9.2
	4	n.a.	n.a.	8.5	9.4	9.6	9.7	9.7	9.7	9.7	9.7
	5	n.a.	n.a.	8.9	9.4	9.4	9.4	9.4	9.4	9.4	9.4
	6	n.a.	8.8	9.0	9.1	9.2	9.2	9.2	9.2	9.2	9.2
	7	n.a.	8.5	8.6	8.7	8.7	8.7	8.7	8.7	8.7	8.7
	8	n.a.	n.a.	n.a.	8.2	8.2	8.2	8.2	8.2	8.2	8.2
	9	n.a.	8.6	8.9	8.9	8.9	9.0	9.0	9.0	9.0	9.0
	10	n.a.	8.3	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5
	11	n.a.	n.a.	8.7	9.1	9.1	9.2	9.2	9.2	9.2	9.2
	12,G	n.a.	n.a.	n.a.	n.a.	n.a.	*8.9*	9.5	9.7	9.8	9.8
	13	n.a.	n.a.	n.a.	8.8	9.2	9.6	9.6	9.6	9.6	9.6
	14	n.a.	8.2	8.3	8.3	8.3	8.3	8.3	8.3	8.3	8.3
	15	n.a.	n.a.	n.a.	n.a.	n.a.	9.2	9.9	10.1	**10.2**	10.2
	16	n.a.	n.a.	8.3	9.1	9.3	9.4	9.4	9.3	9.3	9.3
	17	n.a.	8.6	8.8	8.8	8.8	8.9	8.9	8.9	8.9	8.9
	18,aG	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 6.10.: NCG over the uncoded QPSK (in dB) of each combination of a bit mapping of the 2A4P and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the DE AWGN model. Bold and italics highlightings are the same as in table 6.9.

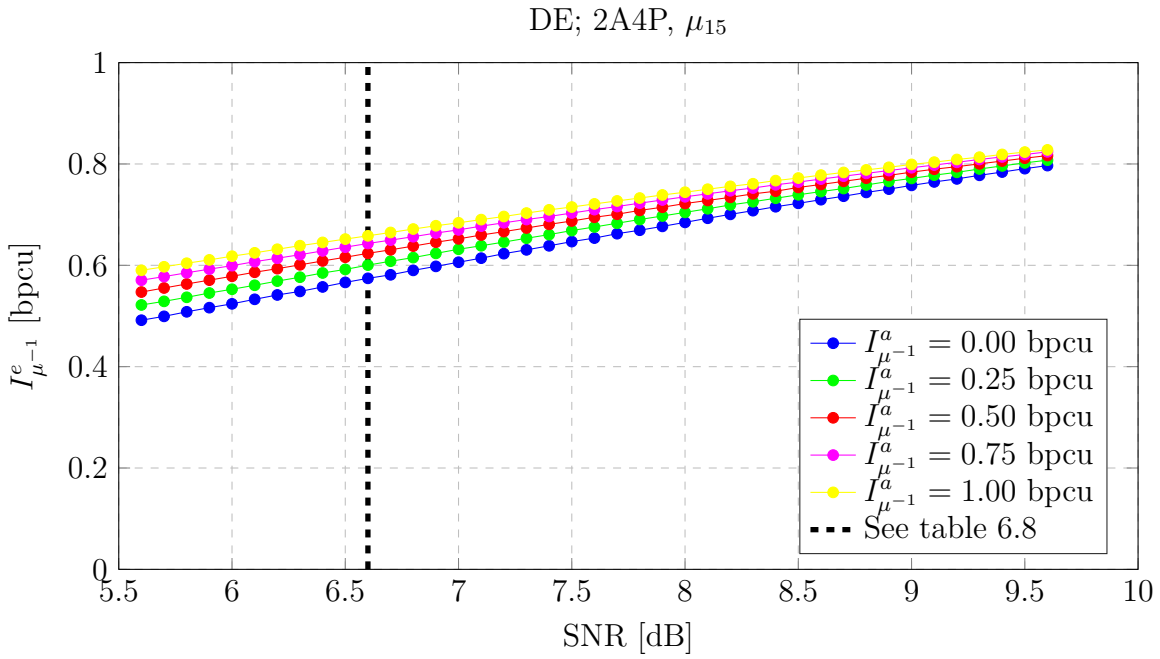

Figure 6.10.: Extrinsic MI of the 2A4P as a function of the SNR, for the μ_{15} bit mapping and parametrised by various values of the a priori MI over the DE AWGN model.

Contrarily to the NDE case, there are some combinations that required only a single IDD iteration, given by the bit mapping $\mu_{12,\mathrm{G}}$ and $N_{\mathrm{LDPC}} = 10$, 20, 30 and 50. A BICM scheme [74] would require a SNR = 8.9 dB for $N_{\mathrm{LDPC}} = 10$ and a SNR = 8.0 dB for $N_{\mathrm{LDPC}} = 50$ to achieve the targeted a posteriori MI.

The best combination is given by

- μ_{15} and
- $N_{\mathrm{LDPC}} = 40$: it achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for
- a SNR = 7.6 dB (1.0 dB from the theoretical limit of the 2A4P and 0.1 dB from the theoretical limit of the QPSK, see table 6.7), corresponding to
- a NCG over the uncoded QPSK = 10.2 dB, with
- $N_{\mathrm{IDD}} = 10$.

As in the NDE case, there are actually five more combinations that achieve the targeted a posteriori MI for SNR = 7.6 dB, one with the same bit mapping and four with the bit mapping μ_1, which has a transfer characteristic very similar to the one of the former (see again figure 6.9). Even in this case, I consider as the best combination the one with the lowest complexity, i.e. the lowest $N_{\mathrm{IDD}} N_{\mathrm{LDPC}}$.

The least complex combination is given by

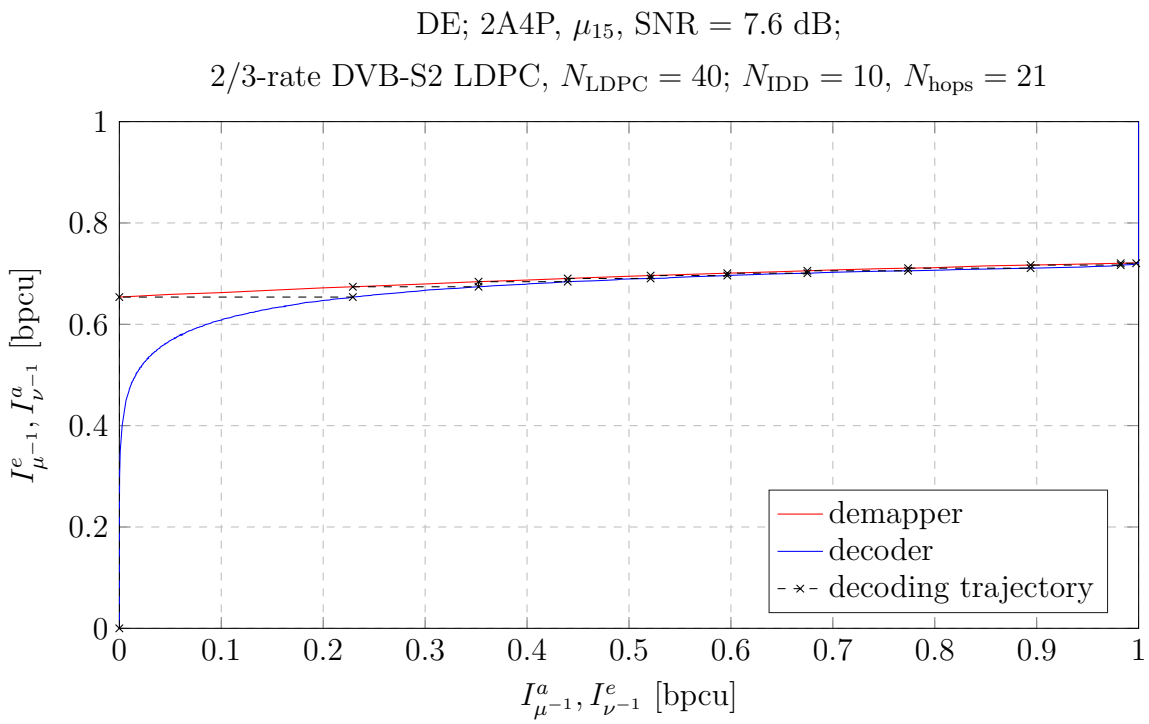

Figure 6.11.: EXIT chart of the best combination over the DE AWGN model.

- $\mu_{12,\text{G}}$ and
- $N_{\text{LDPC}} = 10$ (i.e., the BICM scheme with the lowest number of LDPC iterations): it achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for
- a SNR = 8.9 dB (2.3 dB from the theoretical limit of the 2A4P and 1.4 dB from the theoretical limit of the QPSK, see table 6.7), corresponding to
- a NCG over the uncoded QPSK = 8.9 dB, with
- $N_{\text{IDD}} = 1$.

It is interesting to notice that the most complex combination, given by μ_{11} and $N_{\text{LDPC}} = 50$ with $N_{\text{IDD}} = 29$, actually needs a 1.0 dB higher SNR than the best one.

Figures 6.11 and 6.12 show the EXIT charts of the best and the least complex combinations.

In figure 6.11, corresponding to the best combination, the transfer characteristics of the demapper and of the decoder match impressively well, at least in the central and right region. As in the NDE case (see figure 6.6), the mismatch on the left region is responsible for the 1.0 dB capacity loss between the values reported in tables 6.7 and 6.9 [75, 76]. The fact that the capacity loss is equal to the one suffered by the best combination in the NDE case is due to the fact that the two mappings have very similar

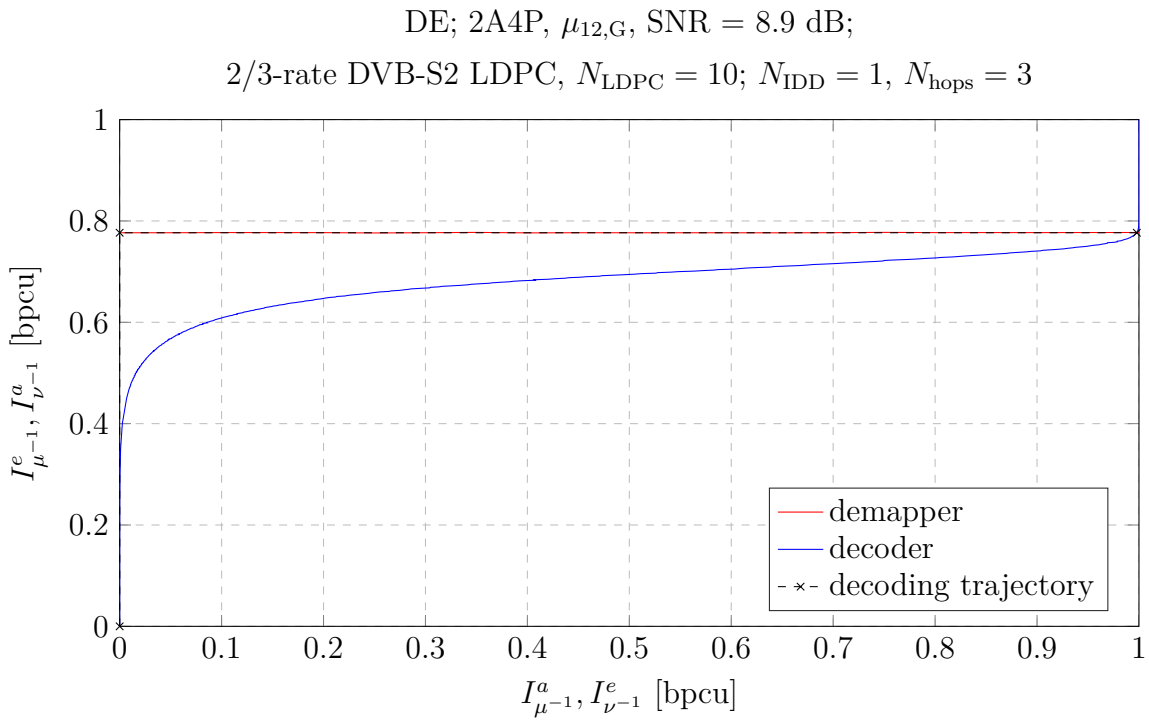

Figure 6.12.: EXIT chart of the least complex combination over the DE AWGN model.

transfer characteristics, hence the area between them and the transfer characteristic of the decoder is also very similar.

In figure 6.12, corresponding to the least complex combination, the big distance between the transfer characteristics of the demapper and the decoder ensures that the IDD stage requires only one IDD iteration to achieve the targeted a posteriori MI $I^p_{\nu^{-1}}$ = 0.9988 bpcu, but is also responsible for the 2.3 dB capacity loss.

6.3.6. DE AWGN model: estimation of the complexity

I evaluated the complexity of the IDD stage in the DE case under the same assumptions as in section 6.2.8.

Obviously, the number of summations needed by the decoder to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration is still given by (6.20).

6.3.7. DE AWGN model: complexity of the demapper at each IDD iteration

As I explained in section 6.3.1, the only thing that changes in the formulas for the extrinsic LLRs with respect to the NDE case is the term carrying the information coming from the channel.

As a consequence, the approximated versions of the extrinsic LLRs (6.7), (6.8) and

(6.9) are still valid once the normalised Euclidean distance between the received symbol y and the transmitted symbol to which the bit mapping μ_k maps the bits b_2, b_1, b_0 (6.10) is replaced by

$$d(b_2 b_1 b_0) \triangleq \min_{x_1 \in \mathcal{X}} \{ \frac{|y_1 - x_1|^2}{N_0} + \frac{|y_2 - \varepsilon(\mu_k(b_2, b_1, b_0), x_1)|^2}{N_0} \}, \tag{6.23}$$

the smallest of eight *sums* given by

- the normalised squared Euclidean distance between the first (previous) received symbol y_1 and the first (previous) transmitted symbol x_1 and
- the normalised squared Euclidean distance between the second (current) received symbol y_2 and the second (current) transmitted symbol, given by the DE encoding of the first (previous) transmitted symbol x_1 and the transmitted transition to which the bit mapping μ_k maps the bits b_2, b_1, b_0.

The minimum is due to the sum over the first (previous) transmitted symbol x_1 in (2.36). As in the NDE case, this terms does require operations other than summations, but the complexity of its calculation can be neglected because it must be calculated only once.

In the end, the number of summations needed by the demapper to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration is still given by (6.14).

6.3.8. DE AWGN model: complexity of the whole IDD stage

As a consequence, the number of summations needed by the whole IDD scheme to calculate the extrinsic LLR of one *information* bit at *each* IDD iteration is still given by (6.22).

Table 6.11 shows the overall complexity of each combination of a demapper transfer characteristic and a decoder transfer characteristic.

The best combination has an overall complexity $K \approx 43.02$ kspib. The least complex combination, i.e. the BICM combination, has an overall complexity $K \approx 1.09$ kspib. The most complex combination has an overall complexity $K \approx 155.80$ kspib,

6.3.9. DE AWGN model: 8-star instead of 2A4P?

The best combination achieves the targeted a posteriori MI for SNR = 7.6 dB, with a 1.0 dB gap to the theoretical limit of the 2A4P.

Even in the DE case it is worth exploring the possibility of using another 8-ary constellation, with a higher theoretical limit but hopefully with a combination that achieves the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu for a SNR < 7.6 dB.

Unfortunately, the best combination of an 8-star bit mapping and the 2/3-rate DVB-S2 LDPC code

- achieves the targeted a posteriori MI for SNR = 7.8 dB, which is 1.0 dB away from the theoretical limit of the 8-star (see table 6.7) and, most importantly, 0.2 dB more than the best combination identified in section 6.3.5;

		N_{LDPC}									
		1	2	3	4	5	10	20	30	40	50
μ	1	n.a.	n.a.	n.a.	n.a.	0.002	0.004	0.069	0.058	0.073	0.086
	2	n.a.	n.a.	0.002	0.008	0.007	0.011	0.138	0.032	0.043	0.054
	3	n.a.	n.a.	0.006	0.009	0.008	0.012	0.022	0.032	0.043	0.054
	4	n.a.	n.a.	0.002	0.005	0.008	0.015	0.026	0.039	0.052	0.064
	5	n.a.	n.a.	0.003	0.012	0.009	0.012	0.022	0.032	0.043	0.054
	6	n.a.	0.008	0.005	0.007	0.016	0.015	0.028	0.042	0.060	0.070
	7	n.a.	0.006	0.005	0.011	0.010	0.016	0.030	0.045	0.060	0.075
	8	n.a.	n.a.	n.a.	0.014	0.012	0.022	0.041	0.061	0.086	0.102
	9	n.a.	0.004	0.015	0.006	0.006	0.022	0.041	0.061	0.082	0.107
	10	n.a.	0.004	0.013	0.008	0.008	0.013	0.026	0.039	0.052	0.064
	11	n.a.	n.a.	0.002	0.006	0.007	0.029	0.058	0.091	0.120	0.156
	12,G	n.a.	n.a.	n.a.	n.a.	n.a.	*0.001*	0.002	0.003	0.009	0.005
	13	n.a.	n.a.	n.a.	0.003	0.003	0.014	0.026	0.039	0.052	0.064
	14	n.a.	0.006	0.008	0.007	0.008	0.013	0.026	0.039	0.052	0.064
	15	n.a.	n.a.	n.a.	n.a.	n.a.	0.002	0.009	0.019	**0.043**	0.048
	16	n.a.	n.a.	0.002	0.004	0.008	0.071	0.134	0.032	0.043	0.054
	17	n.a.	0.007	0.013	0.006	0.006	0.027	0.052	0.074	0.108	0.129
	18,aG	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 6.11.: Overall complexity (in Mspib) required to achieve the targeted a posteriori MI $I^{p}_{\nu^{-1}} = 0.9988$ bpcu for each combination of a bit mapping of the 2A4P and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code over the DE AWGN model. Bold and italics highlightings are the same as in table 6.9.

- has a smaller overall complexity, since it requires $N_{\mathrm{LDPC}} = 10$ LDPC iterations and $N_{\mathrm{IDD}} = 13$ IDD iterations.

In other words, the best combination found in section 6.3.5 is to be preferred.

6.4. Traditional solution for 100G long-haul NDE and DE coherent optical communications systems: (D)QPSK + DVB-S2-LDPC(64800, 57600) + ITU-T-RS(255, 239)

The traditional solution with which I will compare my proposed solutions make use of

- the QPSK ($m = 2$ bits/symbol);
- PDM ($N_{\mathrm{pol}} = 2$) and
- a concatenated HD/ SD FEC scheme with
  - BICM and
  - a 20.03% OH redundancy, corresponding to
  - a code rate $R = 8/9 \cdot 239/255 \approx 0.83$,
  - a targeted MI $I \approx 1.67$ bpcu per polarisation and
  - a symbol rate $f_s \approx 31.45$ GBd, corresponding to
  - a BE = 20.03% ($f_s = 26.20 \rightarrow 31.45$ GBd).

Its FEC scheme consists of

- a SD demapper;
- a code from the same standard, i.e. the binary 8/9-rate LDPC code from the DVB-S2 standard [66] as *inner* SD code, whose redundancy is allocated in the BE. This standard admits a total of 11 code rates: the closest alternatives 5/6 and 9/10 would have led to 28.03% and 18.55% OH redundancies, respectively.

  The demapper and the SD LDPC decoder are part of

- a BICM scheme (i.e., they perform IDD with $N_{\mathrm{IDD}} = 1$). This step, also suggested by the standard, is enough to achieve the *full* capacity of a QPSK, since the latter admits a Gray-coded bit mapping [74];
- the non-binary 16-fold interleaved RS (255, 239) code from the ITU-T G.975 standard [67] as *outer* HD code, whose redundancy is also allocated in the BE.

	Traditional	Proposed (NDE/DE)
constellation	QPSK	8-star/2A4P
m [bits/symbol]	2	3
N_{pol}	2	2
PDM	yes	yes
bit mapping	Gray-coded	$\mu_1, \ldots, \mu_{12,\text{G}}$/ $\mu_1, \ldots, \mu_{18,\text{aG}}$
inner SD code	8/9-rate DVB-S2 LDPC [66]	2/3-rate DVB-S2 LDPC [66]
N_{LDPC}	50	1, ..., 5, 10, 20, 30, 40, 50
BICM/IDD	BICM	IDD
N_{IDD}	1	1, ..., 100
outer HD code	RS(255, 239) [67]	RS(255, 239) [67]
overall redundancy	20.03%	60.04%
overall R	$8/9 \cdot 239/255 \approx 0.83$	$2/3 \cdot 239/255 \approx 0.62$
targeted MI [bpcu per pol.]	$2 \cdot 8/9 \cdot 239/255 \approx \mathbf{1.67}$	$3 \cdot 2/3 \cdot 239/255 \approx \mathbf{1.87}$
BE	**20.03%**	**6.69%**
CE	0%	50%
f_s [GBd]	**31.45**	**27.95**

Table 6.12.: Comparison of traditional and proposed solutions. The differences in targeted MI, BE and symbol rate f_s are highlighted in bold.

Table 6.12 summarizes and compares the parameters of the traditional solution described in this section and the proposed solutions presented in sections 6.2 and 6.3.

The traditional solution targets a *smaller* MI and makes use of *more* BE and a *larger* symbol rate f_s, thus occupying a *wider* bandwidth, as highlighted in bold. Nonetheless, I decided to use this solution as term of comparison because 6.69% SD-FEC is rather uncommon (practical solutions adopt simpler HD-FEC codes).

Since the occupied bandwidth is different, the theoretical limits and the NCGs must be recalculated. A QPSK requires a minimum SNR = 4.57 dB to achieve the targeted MI $I = 1.67$ bpcu per polarisation with a symbol rate $f_s = 31.45$ GBd over the NDE AWGN model. The corresponding potential NCG over the uncoded QPSK = 12.65 dB.

Over the DE AWGN model, instead, it requires a minimum SNR = 6.06 dB, which corresponds to a potential NCG over the uncoded QPSK = 11.25 dB.

6.4.1. Estimation of the minimum SNR required

For the traditional solution, since the bit mapping, the number of LDPC iterations and the number of IDD iterations were fixed, running two simulations was easier than re-doing the whole procedure with the EXIT charts.

In this case I hence looked directly for the minimum SNR required to achieve a post-FEC BER = $8 \cdot 10^{-5}$, which corresponds to the criterion presented in section 6.2.4.1; for the interested reader, in appendix A I explain in detail the procedure.

Figure 6.13 shows the pre-FEC BER and the post-FEC BER as a function of the SNR

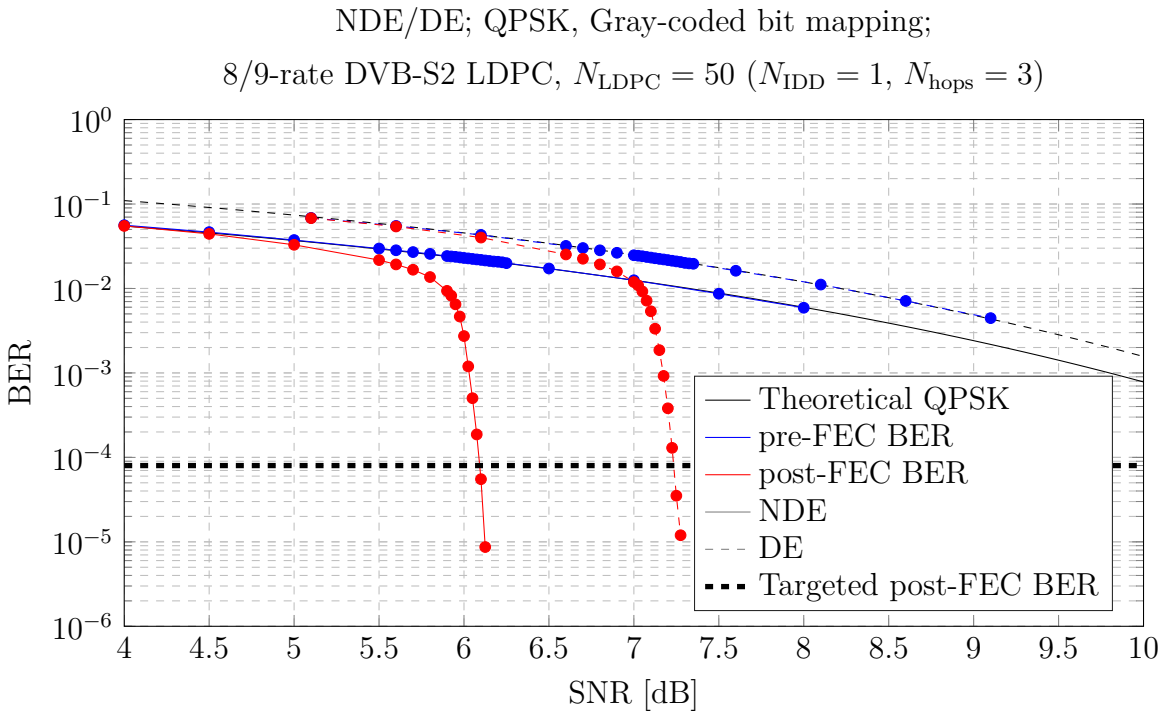

Figure 6.13.: Pre- and post-FEC BER as a function of the SNR of the traditional solution over both models.

of the traditional solution over the NDE AWGN model and over the DE AWGN model.

In the NDE case the traditional solution achieves the targeted post-FEC BER = 8 $\cdot 10^{-5}$ at

- a SNR = 6.1 dB (1.5 dB from the theoretical limit of the QPSK, see section 6.4), corresponding to
- a NCG over the uncoded QPSK = 11.1 dB;

in the DE case, instead, it achieves the targeted post-FEC BER = 8 $\cdot 10^{-5}$ at

- a SNR = 7.2 dB (1.1 dB from the theoretical limit of the QPSK, see section 6.4), corresponding to
- a NCG over the uncoded QPSK = 10.1 dB.

6.4.2. Estimation of the complexity

Using (6.14) with

- $m = 2$ bits/symbol and
- $R = 8/9$

ℓ	λ_ℓ	ρ_ℓ
1	$1/194399 \approx 5.1441 \cdot 10^{-6}$	0
2	$14398/194399 \approx 7.4064 \cdot 10^{-2}$	0
3	$151200/194399 \approx 7.7778 \cdot 10^{-1}$	0
4	$28800/194399 \approx 1.4815 \cdot 10^{-1}$	0
...	...	...
26	0	$26/194399 \approx 1.3375 \cdot 10^{-4}$
27	0	$194373/194399 \approx 9.9987 \cdot 10^{-1}$
Total	$194399/194399 = 1$	$194399/194399 = 1$

Table 6.13.: Degree distribution *from the edge perspective* of the 8/9-rate DVB-S2 LDPC code, indicating the fraction of edges connected to a variable (λ_ℓ) or to a check (ρ_ℓ) node of degree ℓ. The graph has $N_\mathrm{e} = 194399$ edges.

results in a complexity of the demapper of the traditional solution of

$$K'_{\mu^{-1},\mathrm{trad.}} \approx 5.6 \text{ spib},$$

which is a factor 0.3x that of the proposed one (see section 6.2.9.1). The demapper of a quaternary constellation needs a *lower* number of summations at *each* IDD iteration than the demapper of an 8-ary constellation because (6.7), (6.8) and (6.9) reduce to the difference between two max operations between two values, of which only one is a sum, and only the a priori LLR of the other bit must be considered.

Table 6.13 shows the degree distribution *from the edge perspective* of the 8/9-rate LDPC code from the DVB-S2 standard [66].

Using (6.20) with

- the degree distribution *from the edge perspective* shown in table 6.13,
- $N_\mathrm{e} = 194399$ and
- $k = 57600$ information bits,

results in a complexity of the decoder of the traditional solution of

$$K'_{\nu^{-1},\mathrm{trad.}} \approx 179.1 \cdot 50 \text{ spib},$$

which is a factor $1.67 \cdot 50/N_\mathrm{LDPC}$ that of the proposed one (see section 6.2.10.4). The decoder of the 8/9-rate DVB-S2 LDPC code needs a *higher* number of summations at *each* IDD iteration than the decoder of the 2/3-rate DVB-S2 LDPC code because its degree distribution *from the edge perspective* has non-zero values for higher ℓ (compare tables 6.5 and 6.13).

Finally, using (6.21) or, equivalently, (6.22) with $N_\mathrm{IDD} = 1$, results in an overall complexity of the traditional solution of

$$K_\mathrm{trad.} = K'_\mathrm{trad.} \approx 5.6 + 179.1 \cdot 50 \approx 8.96 \text{ kspib}.$$

6.4.3. NDE AWGN model: comparison

Figure 6.14 shows graphically the information reported numerically in table 6.6, "rescaled" by the complexity of the traditional solution, that is,

- on the y-axis is the *ratio* (since it is linear) between the overall complexity of the various combinations of the proposed solution and the overall complexity of the traditional solution, whereas
- on the x-axis is the *difference* (since it is already in dB) between the minimum $10\log_{10}(E_b/N_0)$ needed by the traditional solution to achieve the targeted post-FEC BER $= 8\cdot10^{-5}$ and the minimum $10\log_{10}(E_b/N_0)$ needed by the various combinations of the proposed solution to achieve the targeted a posteriori MI $I^p_{\nu^{-1}} = 0.9988$ bpcu.

  This number corresponds to the opposite of the difference between the NCGs over the uncoded QPSK of the traditional solution and of the various combinations of the proposed solution, as well as to the difference between their minimum *optical* signal-to-noise ratios (OSNRs).

The traditional solution clearly corresponds to the point $(0, 1)$, which is also indicated with a vertical and an horizontal line; interesting solutions are in the bottom right region, where the minimum E_b/N_0 needed by the various combinations is lower than that of the traditional solution and the complexity is lower than that of the traditional solution.

The best combination (▼)

- *gains* $6.1 - 10\log_{10}(2\cdot 8/9\cdot 239/255) - 5.9 + 10\log_{10}(3\cdot 2/3\cdot 239/255) \approx 0.7$ dB over the traditional solution with
- a complexity $\approx$ 4.3x that of the traditional solution.

The least complex combination (●)

- *loses* $6.1 - 10\log_{10}(2\cdot 8/9\cdot 239/255) - 7.8 + 10\log_{10}(3\cdot 2/3\cdot 239/255) \approx -1.2$ dB to the traditional solution with
- a complexity $\approx$ 0.2x that of the traditional solution.

Finally, there are also some combinations in the bottom right part of figure 6.14, which offer marginal gains up to 0.3 dB for a slightly smaller complexity.

6.4.3.1. NDE AWGN model: How to use the lower SNR and the narrower occupied bandwidth

In order to see how the additional 0.7 dB of SNR gained by the best combination could be used, let me recall the famous formula for the optical power budget [50] in dB

$$\text{OSNR} = 58 + P_{\text{s,dBm}} - \alpha L_{\text{span}} - N_{\text{spans,dB}} - F_N \text{ [dB]}, \tag{6.24}$$

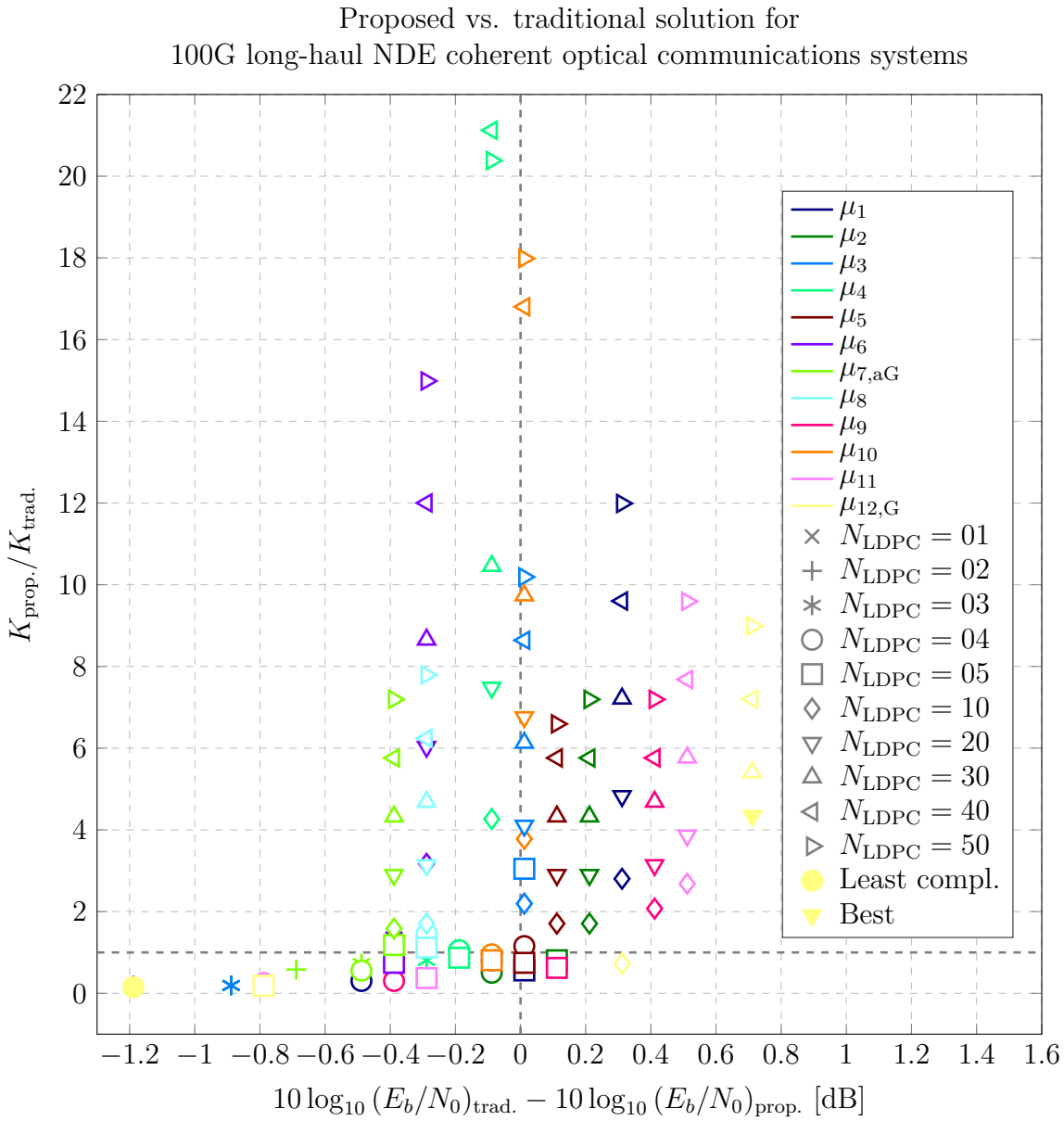

Figure 6.14.: Comparison of each combination of a bit mapping of the 8-star and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code with the traditional solution over the NDE AWGN model. The best and the least complex combinations highlighted in table 6.6 have here full markers.

Loss [dB]	Reach decrease	Loss [dB]	Reach decrease	Gain [dB]	Reach increase
−1.9	−35%	−0.9	−19%	+0.1	+2%
−1.8	−34%	−0.8	−17%	+0.2	+5%
−1.7	−32%	−0.7	−15%	+0.3	+7%
−1.6	−31%	−0.6	−13%	+0.4	+10%
−1.5	−29%	−0.5	−11%	+0.5	+12%
−1.4	−28%	−0.4	−9%	+0.6	+15%
−1.3	−26%	−0.3	−7%	+0.7	+17%
−1.2	−24%	−0.2	−5%		
−1.1	−22%	−0.1	−2%		
−1.0	−21%	0.0	0%		

Table 6.14.: dB to linear conversion.

where $P_{\mathrm{s,dBm}}$ is the signal power leaving the transmitter (in dBm), α is the fibre attenuation coefficient, L_{span} is the length of a span, $N_{\mathrm{spans,dB}} = 10\log_{10}(N_{\mathrm{spans}})$ is the number of spans "expressed in dB", F_N is the noise figure of an erbium doped fibre amplifier (EDFA) and a reference bandwidth $B_{\mathrm{ref}} = 12.5$ GHz is assumed.

If the transmitter, the optical fibres and the amplifiers remain equal, in the linear domain (6.24) implies that

$$\mathrm{OSNR} \cdot N_{\mathrm{spans}} = \text{const.}, \tag{6.25}$$

which means that

- any decrease in the OSNR corresponds to an equal increase in the number of spans N_{spans}, i.e. in the reach.

  Table 6.14 shows a conversion from the dB to the linear domain, which can help understand which reach increase a gain corresponds to.

  A 0.7 dB gain, such as the one offered by the proposed solution, correspond to a reach increase of 17%.

Alternatively,

- the transmitted power $P_{\mathrm{s,dBm}}$ can be reduced, thereby potentially reducing nonlinearities and hence increasing the SNR (i.e., a virtuous circle could be started);
- cheaper optical fibres with higher attenuation coefficients α can be used;
- fewer amplifiers can be spaced further apart increasing the length of a span L_{span}, again increasing the SNR;
- cheaper amplifiers with a higher noise figure F_N can be adopted;

Last but not least, since the proposed solution needs a symbol rate and occupies a bandwidth which are 0.9x those of the traditional solution,

- the overall data rate of the whole wavelength division multiplexing (WDM) system can be increased by 10%, because the carriers can be packed closer together.

In general,

- the requirements of the electrical and optical components in terms of SNR and/or symbol rate and hence occupied bandwidth can be relaxed.

This, however, comes to a cost, since the complexity is also increased by a factor 4.3x; according to the Moore's law [81,82], this corresponds to a time span of 3–4 years, which means that, in principle, the proposed solution should be feasible by the time this work is published.

The least complex combination can be regarded to as a way to reduce the complexity, rather than to increase the reach (or any of the other benefits listed above): if complexity is an issue and one is willing to accept a reach reduced by one fourth (-1.2 dB correspond to $\approx -0.24\%$, see table 6.14) or, equivalently, to compensate somewhere else for this loss, the complexity can be knocked down by 80% by switching to the proposed solution.

6.4.4. DE AWGN model: comparison

Figure 6.15 shows graphically the information reported numerically in table 6.11, "rescaled" by the complexity of the traditional solution as explained in section 6.4.3.

The best combination (◀)

- *gains* $7.2 - 10\log_{10}(2 \cdot 8/9 \cdot 239/255) - 7.6 + 10\log_{10}(3 \cdot 2/3 \cdot 239/255) \approx 0.1$ dB over the traditional solution with
- a complexity ≈ 4.8x that of the traditional solution.

The least complex combination (♦)

- *loses* $7.2 - 10\log_{10}(2 \cdot 8/9 \cdot 239/255) - 8.9 + 10\log_{10}(3 \cdot 2/3 \cdot 239/255) \approx -1.2$ dB to the traditional solution with
- a complexity ≈ 0.1x that of the traditional solution.

Unfortunately, there are no combinations in the bottom right region of figure 6.15.

6.4.4.1. DE AWGN model: How to use the narrower occupied bandwidth

Contrarily to the NDE case, the gain offered by the proposed solution has shrunk dramatically, to the point where moving to a larger constellation becomes questionable.

However, the two solutions target different MIs or, equivalently, have different symbol rates and hence occupy different bandwidths. In this sense, the best combination can still be regarded to as a way to reduce the symbol rate, narrow the occupied bandwidth

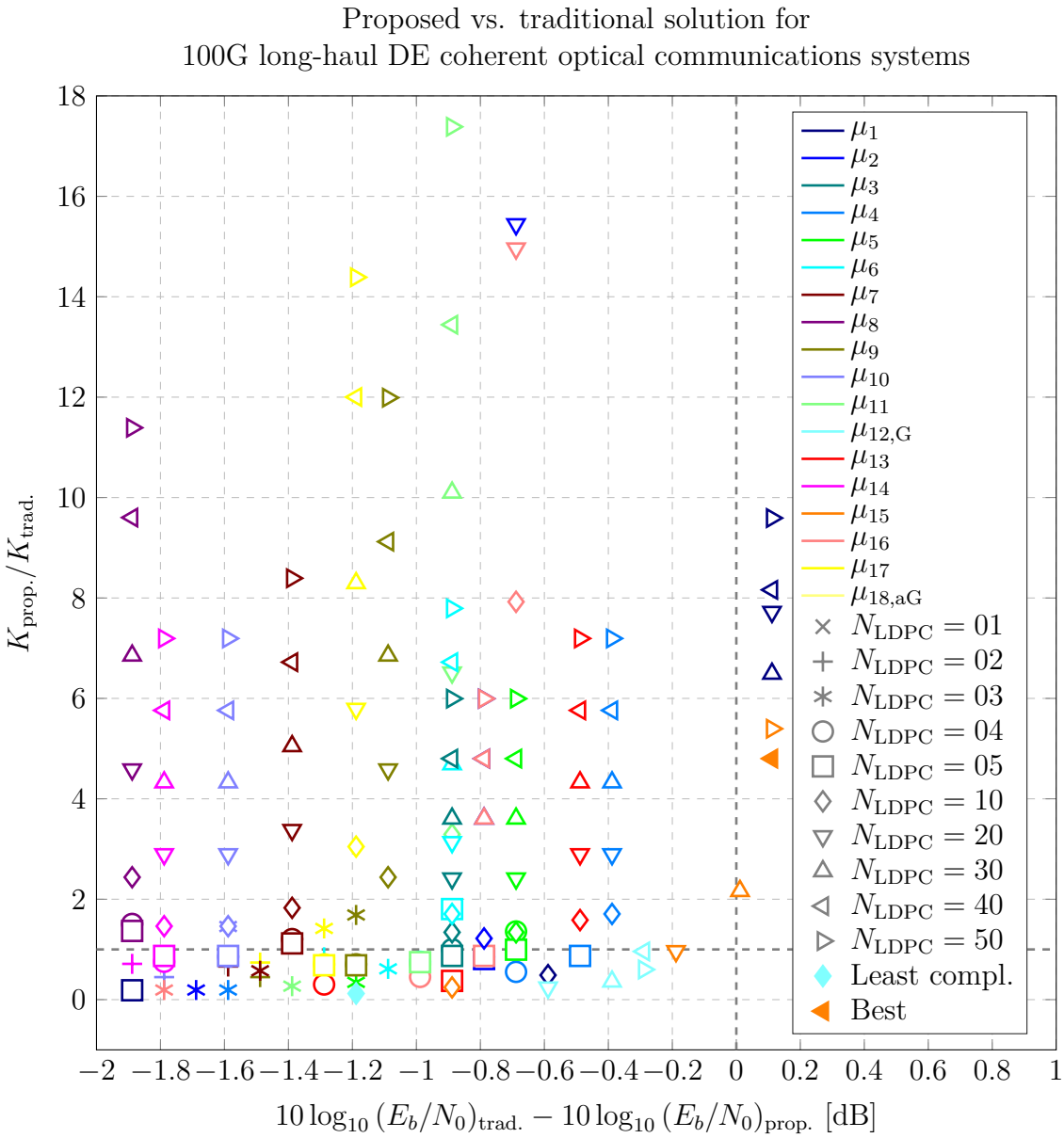

Figure 6.15.: Comparison of each combination of a bit mapping of the 2A4P and a number of LDPC iterations of the 2/3-rate DVB-S2 LDPC code with the traditional solution over the DE AWGN model. The best and the least complex combinations highlighted in table 6.11 have here full markers.

and increase the overall data rate of the whole WDM system, while achieving the same performance, rather than to increase the reach (or any of the other benefits mentioned in section 6.4.3.1).

Similar to section 6.4.3, the least complex combination can be regarded to as a way to reduce the complexity, rather than to increase the reach: if the performance can be reduced by 1.2 dB (or this loss can be somehow compensated for), the complexity can be knocked down by 90%.

6.5. Summary

Traditional solutions for 100G long-haul NDE and DE coherent optical communications systems are usually based on QPSK, PDM and concatenated HD/ SD FEC schemes, whose 20% OH redundancy is allocated in the bandwidth; traditional solutions hence target a MI $I = 1.67$ bpcu per polarisation and have a symbol rate $f_s = 31.44$ GBd.

In this chapter, instead, I proposed some alternative solutions based on the 8-ary constellations chosen in chapter 5, PDM and a much stronger concatenated HD/ SD FEC scheme, given by the 2/3-rate DVB-S2 LDPC code and the 16-fold interleaved ITU-T G.975 RS (255, 239) code, whose 60.04% OH redundancy is allocated mostly in the constellation and only marginally in the bandwidth; the proposed solutions target a *higher* MI $I = 1.87$ bpcu per polarisation, have a *smaller* symbol rate $f_s = 27.95$ GBd and hence occupy a *narrower* bandwidth.

The choice to use 8-ary constellations brings with it a series of additional challenges. First of all, I employed IDD between the demapper and the decoder to achieve the full capacity of a *non-binary* constellation employing *binary* codes. I estimated the transfer characteristics of the demapper and the decoder and resorted on EXIT charts to predict the performance of all combinations of a transfer characteristic of the demapper, which depends on the bit mapping (12 for the 8-star, 18 for the 2A4P) and on the SNR, and a transfer characteristic of the decoder, which depends on the number of LDPC iterations (10 different values). Finally, assuming approximated versions of the demapper and the decoder, I estimated the complexity of all combinations.

All results obtained so far were *absolute*. To put them in perspective, I considered as term of comparison also a traditional solution reusing as many building blocks as possible, changing only the constellation (a QPSK with Gray-coded bit mapping) and the inner code (the 8/9-rate DVB-S2 LDPC code) as well as removing the IDD stage (since it was not needed); this traditional solution has a 20.03% OH redundancy allocated in the bandwidth, targets a MI $I = 1.67$ bpcu per polarisation and has a symbol rate $f_s = 31.45$ GBd. I estimated its performance (this time by means of a direct simulation) and its complexity.

In the NDE case, the best combination is given by

- the twelfth bit mapping of figure 6.2 (the most “Gray like”-coded bit mapping) and
- 20 LDPC iterations; it achieves “quasi error-free” communication at

- a SNR = 5.9 dB (NCG over uncoded QPSK = 11.8 dB) after
- 18 IDD iterations, requiring
- ≈ 38.92 kspib.

This solution

- is still 1.0 dB away from the theoretical performance of a 8-star over the NDE AWGN model, but
- is already 0.6 dB *beyond* the theoretical performance of a QPSK over the NDE AWGN model; compared to the traditional solution, it
- gains 0.7 dB and
- has a 4.3x higher complexity (feasible within 3–4 years, according to the Moore's law),
- with a 0.9x smaller symbol rate and hence narrower occupied bandwidth.

The 0.7 dB gained by the solution proposed could be used to increase the reach by 17% or, alternatively, to reduce the transmitted power (thereby potentially increasing the SNR), to deploy longer spans of optical fibres with higher attenuations, to use fewer amplifiers with higher noise figures or to increase the overall data rate of the whole WDM system. The 0.9x narrower occupied bandwidth could be used to increase the overall data rate of the whole WDM system by 10%. Alternatively, the least complex combination loses 1.2 dB but has a 0.2x lower complexity.

In the DE case the best combination is given by

- the fifteenth bit mapping of figure 6.8 and
- 40 LDPC iterations; it achieves "quasi error-free" communication at
- a SNR = 7.6 dB (NCG over uncoded QPSK = 10.2 dB) after
- 10 IDD iterations, requiring
- ≈ 43.02 kspib.

This solution

- is still 1.0 dB away from the theoretical performance of a 2A4P over the DE AWGN model, but
- is *only* 0.1 dB away from the theoretical performance of a QPSK over the DE AWGN model; compared to the traditional solution, it
- gains 0.1 dB and

- has a 4.8x higher complexity,
- with a 0.9x smaller symbol rate and hence narrower occupied bandwidth.

The 0.1 dB gain translates to a negligible increase in reach; however, this combination can still be used to increase the overall data rate of the whole WDM system by 10%. Alternatively, the least complex combination loses 1.2 dB but has a 0.1x lower complexity.

7. Conclusions

The research presented in this work was performed between February 2010 and January 2015, when the first 100G long-haul coherent optical communications systems were coming to market. All these systems follow the same approach: quadrature phase-shift keying (QPSK), polarisation division multiplexing (PDM) and forward error correction (FEC) redundancy allocated in the bandwidth, a technique referred to in this work as bandwidth expansion (BE). Since the occupied bandwidth cannot be increased indefinitely, the general consensus was to use FEC solutions with a maximum of 20% redundancy.

However, once the constellation and the redundancy (as well as the channel model) are fixed, reliable transmission of information can be achieved only above a certain minimum signal-to-(*additive*-)noise ratio (SNR). This limit cannot be beaten by improving the electrical and optical components (the approach followed in the early days of optical communications), but only by changing the constellation. This is the core of Ungerböck's proposal, to use "larger than strictly needed" constellations and to allocate there the FEC redundancy, a technique referred to in this work as constellation expansion (CE). The "bet" is that the stronger codes *more than compensate* for the increased minimum required SNR needed by the larger constellations, leading to systems that overall achieve the targeted bit error rate (BER) for lower *optical* signal-to-noise ratios (OSNRs).

The solutions for 100G long-haul non differentially encoded (NDE) and differentially encoded (DE) coherent optical communications systems proposed in this work are based both on BE and on CE, since they use

- larger 8-ary constellations (the next step after QPSK),
- PDM and
- a (much) stronger FEC solution, whose redundancy is allocated
  - mostly in the constellation and only
  - marginally in the bandwidth and

  whose higher complexity is still (or will soon become) reasonable.

In particular,

- the redundancy allocated in the CE comes from a very hardware-aware standardised iterative soft-decision (SD) low-density parity-check (LDPC) code, the 2/3-rate LDPC code from the digital video broadcasting – satellite – 2$^{\text{nd}}$ generation (DVB-S2) standard [66], which is needed to achieve quickly the targeted BER;

- the redundancy allocated in the BE comes from a very well-known classic algebraic hard-decision (HD) Reed-Solomon (RS) code, the 16-fold interleaved RS (255, 239) code from the Telecommunication Standardization Sector of International Telecommunication Union (ITU-T) G.975 standard [67], which is needed to clean up potential error-floors of the first one.

The proposed solution is hence a concatenated SD/ HD FEC scheme with an overall code rate $R = 2/3 \cdot 239/255 \approx 0.62$, corresponding to a 60.04% overhead (OH) and a symbol rate $f_s = 27.95$ GBd.

In order to maximise the performance of the proposed solutions, certain steps have been taken:

- the need for more accurate channel models was discussed, which take into account also the (residual) phase noise beside the additive one.

  Phase noise is a known problem already in other kinds of communications (wired, such as communications over twisted pair or coaxial cable, as well as wireless, such as satellite or mobile communications). However, optical communications have to cope with larger amounts of phase noise, mostly due to the larger linewidths of their oscillators, the light amplification by stimulated emission of radiations (LASERs), and to the non-linear effects, which are always present but become detrimental once the launch power grows beyond a certain threshold.

  In chapter 5, it was shown that the phase noise can reduce considerably the mutual information (MI) that QPSK and other constellations can achieve over the NDE additive white Gaussian noise (AWGN)/ wrapped phase white Gaussian noise (wPWGN) model and the DE AWGN/ wPWGN model. It was also shown that, however, its effect on the MI can be neglected for signal-to-*phase*-noise ratios (SPNRs) > 15 dB.

  In chapter 4, using the estimators presented in chapter 3, it was shown that in 100G long-haul NDE and DE coherent optical communications systems representing old and modern wavelength division multiplexing (WDM) system, even when operated in highly non-linear regimes and hence dominated by

  – (homogeneous) intra-channel and/or

  – homo- or heterogeneous inter-channel non-linearities,

  the SPNR is always above this threshold. All the more must this then be the case for practical 100G long-haul NDE and DE coherent optical communications systems operating in realistic regimes. The conclusion was hence that the phase noise can be neglected and that channel models considering only additive noise can still be used;

- the 8-ary constellations that achieved the targeted MI for the lowest SNR or, equivalently, the highest potential gain over the QPSK, were used.

  In chapter 5, it was shown that these constellations are

  - the 8-star over the NDE AWGN model and
  - the 2 amplitude- 4 phase-shift keying (2A4P) over the DE AWGN model;

- iterative demapping and decoding (IDD) was employed to achieve the full capacity of the chosen *8-ary* constellations with a *binary* LDPC code.

  In chapter 6, all bit mappings admitted by the two chosen constellations were identified. The performance of all "bit mapping-number of LDPC iterations" combinations were investigated by means of extrinsic information transfer (EXIT) charts. It was shown that the best combinations were:

  - the 8-star with $\mu_{12,\mathrm{G}}$ and $N_{\mathrm{LDPC}} = 20$ over the NDE AWGN model, achieving "quasi error-free" communication at an OSNR = 9.4 dB, corresponding to a net coding gain (NCG) over the uncoded QPSK = 11.8 dB, and
  - the 2A4P with μ_{15} and $N_{\mathrm{LDPC}} = 40$ over the DE AWGN model, achieving "quasi error-free" communication at an OSNR = 11.1 dB, corresponding to a NCG over the uncoded QPSK = 10.2 dB.

A solution following the traditional approach was also considered as a term of comparison, useful to gain some insight into the performance and into the complexity of the proposed solutions.

In chapter 6, this solution was obtained reusing as many building blocks as possible and consisted of QPSK, PDM, 8/9-rate DVB-S2 LDPC code and 16-fold interleaved ITU-T G.975 RS (255, 239) code, with a higher overall code rate $R = 8/9 \cdot 239/255 \approx 0.83$, corresponding to a lower 20.03% OH, a *larger* symbol rate $f_s = 31.45$ GBd and hence most of all a *wider* occupied bandwidth. All bit mapping-number of LDPC iterations combinations were compared with this solution. It was shown that to achieve "quasi error-free" communication,

- the best combination over the NDE AWGN model needs a 0.7 dB lower SNR for a 4.3x higher complexity and
- the best combination over the DE AWGN model needs a 0.1 dB lower SNR for a 4.8x higher complexity.

On the other side of the spectrum are the least complex combinations, which in both cases require a 1.2 dB higher SNR but for a 0.2x or 0.1x lower complexity.

7.1. Outlook

Being one of the first works in its direction (at least to the best of my knowledge), this work has only scratched the surface of what could be done. Also, by definition FEC is at the two very ends of a communications system and many things can happen in between that can have an impact on it. In this section, I will briefly touch upon some parts of this work that I would have liked to study in even more depth, as well as some directions that I wish research will look into in the future.

7.1.1. Other amounts of CE and BE, other codes

In this work, Ungerböck's idea has been applied to 100G long-haul NDE and DE coherent optical communications systems and two specific solutions have been proposed with certain amounts of CE and BE. However, his idea is very general and nothing prohibits to consider larger amounts of CE and/or BE, for instance

- a 16-ary constellation in lieu of the 8-star or the 2A4P (corresponding to a 100% CE) and/or

- a symbol rate $f_s = 31.44$ GBd in lieu of 27.95 GBd (corresponding to a 20% BE).

Of the two, the second idea sounds more promising. The first idea, which would allow space for an even stronger inner code with 100% redundancy, would probably result in only marginally better performance (as shown in chapter 5, at least the 16-ary constellations considered in this work have only few tenths of dB higher potential gain over the QPSK), but also in an even higher complexity. Moreover, all bit mappings of the chosen 16-ary constellations should be identified and the performance of all bit mapping-number of LDPC iterations combinations should be investigated, which might become unfeasible.

The second idea, which would allow space for a slightly stronger outer code with 20% redundancy, would probably be able to translate an even higher pre-FEC BER down to the same targeted post-FEC BER. Note that in this case all performance metrics (theoretical limits, potential gain over the QPSK, potential NCG over the uncoded QPSK) should be recalculated, as the occupied bandwidth and the targeted MI would be different.

Alternatively, the approach presented in this work (50% CE, 6.69% BE) could be maintained, but the building blocks could be different. The search among the bit mappings of the two chosen constellations is indeed exhausted, but other codes with the same OHs could be used. The 2/3-rate LDPC code from the DVB-S2 standard [66] represented a good compromise between performance and complexity, but

- inner codes with transfer characteristic that better match those of the SD demappers could achieve the targeted a posteriori MI for lower SNRs, however possibly for higher complexities: if the two transfer characteristics are very close, many bounces and hence more IDD iterations are needed. On the other hand,

- more advanced outer codes could manage with a lower targeted a posteriori MI; this in turn could mean that less IDD iterations could be needed, reducing the complexity of the IDD scheme. Finally,

- inner codes with even sparser parity-check matrix could exhibit lower complexity, however possibly with worse performance.

7.1.2. Data-aided DSP, larger iterative loops

Digital signal processing (DSP) is a huge field by itself and fine-tuning it is an art form: countless Ph.D. theses have been (and will be) written on these topics, so by no means this work pretends to have covered it all. Since DSP is part of the channel seen by the SD demapper and FEC decoder, I would like to see studies on the effects of the former on the latter in presence of non-linear effects.

In the laboratory experiments, I optimised the blind DSP algorithms to minimise the BER. This resulted in Gaussianly distributed counter-rotated and translated (CRT) received constellations with more or less (depending on the experiment) impulsive auto- and cross-correlation functions of their real and imaginary parts. It would be interesting to study whether the *residual* correlations can be removed with a different optimisation of the same DSP algorithms or even different DSP algorithms.

In the same experiments the blind DSP struggled (and sometimes even failed) to converge at low OSNRs. Unfortunately, this is exactly the direction in which FEC "pushes" a communications system: reduce the minimum SNR (equivalently, increase the pre-FEC BER) at which the targeted post-FEC BER is achieved. If the DSP cannot converge at such low SNRs, it becomes the performance limiting factor.

Convergence at low SNRs can probably be improved by further optimizing the DSP algorithms, but at a certain point the limits of blind DSP will be reached. Beyond these limits, it will be necessary to switch to *data-aided DSP*, in which few pilot symbols (which can also be seen as "redundancy", for which space must then also be made) are introduced at regular intervals in the transmitted symbols. Data-aided DSP is more robust against cycle-slips, which in turn can make DE encoding as well as techniques such as multisymbol detection (MSD) [63] – which could have been adopted to close (or at least narrow) the gap between NDE and DE systems [83] – superfluous; in other words, only 100G long-haul NDE coherent optical communications systems would be needed, which are those that gain the most from CE. The laboratory experiments presented in this work could thus be redone, this time with a data-aided DSP, and from a comparison of the two important insights could definitely be gained: for instance, I expect the contribution of the DSP to the "setup SNR" and the "setup SPNR" to be reduced.

Finally, another possibility is to increase the scope of the iterative scheme, including also the DSP (or at least part of it) thus creating an iterative DSP, demapping and decoding (IDDD) scheme. Starting from the end of the chain of DSP algorithms and going backwards, assuming that some DSP algorithms can be adapted to accept some kind of *a priori* information on the *symbols*, these algorithms could be fed with the (appropriately adapted) *extrinsic* information on the *bits* from the SD decoder of the inner code. If transfer characteristics similar to those of the demapper and the decoder presented in this work can be obtained for these algorithms, the performance of the IDDD scheme could be predicted by means of an extended "tridimensional" version of the EXIT charts. Assuming for instance that only one DSP algorithm takes part in the IDDD scheme,

- the DSP algorithm would translate an a priori MI I_1 into an extrinsic MI I_2,

- the demapper would translate an a priori MI I_2 into an extrinsic MI I_3,
- the decoder would translate an a priori MI I_3 into an extrinsic MI I_4,
- the DSP algorithm would translate an a priori MI I_4 into an extrinsic MI I_5 and so on.

For the DSP algorithms used in this work, this would mean starting with the joint equalisation and carrier recovery stage. This could improve the compensation of the LASER phase noise, possibly to the point where cycle-slips can be avoided; DE encoding would then become superfluous, with the same advantages mentioned above when talking about data-aided DSP. Clearly, the price to pay would be the usual increase of complexity and latency.

7.1.3. Experiments with larger constellations, more channels and wider occupied bandwidths

So far, I presented some possible theoretical directions. However, this work contains also a practical part, the laboratory experiments representing 100G long-haul NDE and DE coherent optical communications systems with eight neighbouring channels operating in highly non-linear regime and thus impaired by (homogeneous) intra-channel non-linearities and/or homogeneous (100G) or heterogeneous (10G) inter-channel non-linearities. With these experiments, I investigated how non-linear effects caused by the propagation of

- QPSK modulated channels employing PDM and occupying a bandwidth $B_s \approx 28$ GHz and
- on-off keying (OOK) modulated channels occupying a bandwidth $B_s \approx 10$ GHz

manifest in the channel "seen by the SD demapper and FEC decoder".

However, to keep up with the constantly growing domand for more data, larger constellations with various shapes as well as more channels in parallel with wider occupied bandwidths will be needed (not to store the redundancy, but to increase the bit rate); to increase reach, higher launch powers, longer transmission links and more optical amplifiers will be needed. The question then arises, how does all this reflect in the channel seen by the SD demapper and FEC decoder?

Qualitatively, the variance of the LASER phase noise increments depends linearly on the symbol period, so if the latter is reduced because each channel has a wider occupied bandwidth, the former will decrease. On the other hand, non-linear effects depend on the total launch power, so if the latter is increased because a higher number of channels with the same occupied bandwidth or the same number of channels with a wider occupied bandwidth are transmitted or, finally, because some points of the constellation have higher power than others, the former will increase. The interaction between the non-linear effects and the chromatic dispersion (CD) should also be taken into account, when the channels occupy a wider bandwidth and are spaced further apart. Finally,

more optical amplifiers means more additive noise. In all these directions a quantitative analysis would be very interesting.

Note that, in case a constellation with more than one ring is used, the approximated, counter-rotated and translated (ACRT) NDE AWGN/ phase white Gaussian noise (PWGN) model must be adapted, because the transmitted symbol does not have unitary amplitude anymore and hence the counter-rotation operation will change the variance of the random variable modelling the additive noise.

Also, it might be worth extending this channel model including *residual* I/Q imbalances in order to make the estimators of the AWGN and the PWGN variances based on it even more robust. In fact, the underlying assumption of the counter-rotation and translation operation is that the DSP fully compensates for potential I/Q imbalances such as the phase difference between the in-phase and quadrature components not being exactly 90° and/or the two components undergoing slightly different amplification factors along their paths.

If, instead, *residual* I/Q imbalances are present, the in-phase and quadrature components are mixed together and the received constellation resembles a "stretched" QPSK, looking like either a diamond or a rectangle. Correspondingly, the CRT received constellation looks like an ellipse "stretched" along one and/or both of its axes; the estimators will misinterpret this shape and end up overestimating the additive and phase noise variances or, equivalently, underestimating the SNR and the SPNR. Note that correctly estimating the SNR is important not only for the laboratory experiments like those presented in this work or suggested above, but also for practical optical communications systems, since the corresponding estimated AWGN variance is needed to calculate the extrinsic log-likelihood ratios (LLRs).

7.1.4. CE and BE for 100G+ long-haul coherent optical communications systems

The hybrid approach based on CE and BE presented in this work for 100G long-haul NDE and DE coherent optical communications systems can be applied also to 100G+ long-haul coherent optical communications systems, especially if they are realised increasing the symbol rate and hence the occupied bandwidth and not using larger constellations. In fact, as shown, the benefit of moving from a constellation with m symbols to a constellation with $m+1$ symbols is maximised for low values of m (in this work $m = 2$).

An *uncoded* 200G long-haul coherent optical communications system employing PDM can be realised with either

- a constellation with $m = 4$ bits/symbol with a symbol rate $f_s = 25$ GBd or
- a constellation with $m = 2$ bits/symbol with a symbol rate $f_s = 50$ GBd,

depending on the reach requirements. In the first case the spectral efficiency (SE) is doubled but the reach is (in theory) reduced by a factor 0.25x; in the second case, instead, the SE is the same and the reach is (in theory) the same [84].

II-VI Incorporated, for instance, announced in 2020 a coherent optical transceiver which, when operated at 200G, has a reach of 2000 km [85] and [86]

> uses 64-GBaud QPSK modulation, which enables longer reach than alternatives based on 32-GBaud 16QAM modulation.

To transmit twice the optical data unit 4 (ODU4) nominal bit rate of $R_b \approx 104.79$ Gbps an uncoded system would need a symbol rate $f_s = 52.4$ GBd: this solution is hence using a BE = 22.14%.

It has, in other words, parameters very similar to those of the traditional solution used in this work as term of comparison, with the exception of the symbol rate and hence of the occupied bandwidth. Hence, assuming that their system is NDE and that their FEC solution (not disclosed) has performance similar to the FEC scheme I used for the traditional solution, replacing the QPSK with a 8-star would guarantee the same advantages as for the solutions I proposed in this work: a 0.7 dB lower OSNR and a 0.9x smaller symbol rate and hence narrower occupied bandwidth, for a 4.3x higher complexity.

Finally, a 200G long-haul coherent optical communications system can also be realised pairing two 100G long-haul coherent optical communications systems very close to each other in frequency, a technique known as "dual carrier". Even in this case CE and BE can be used separately on each 100G system, with the smaller symbol rate and hence narrower occupied bandwidth becoming a particularly interesting feature.

The next step is 400G long-haul coherent optical communications system. Even when they employ PDM, these systems will most likely be realised with a constellation with $m = 4$ bits/symbol, most likely a 16 quadrature amplitude modulation (QAM), and when *uncoded* a symbol rate $f_s = 50$ GBd. This is the way chosen by II-VI Incorporated for the same transceiver when operated at 400G [86] and by Acacia [87], among others. In this case CE becomes less attractive, for the reasons mentioned above.

A. Estimation of the transfer characteristics

The idea is to

1. choose a certain a priori MI $I^a_{\mu^{-1}}$ or $I^a_{\nu^{-1}}$,
2. generate N_{bits} encoded bits $B_{e,i}$,
3. *generate* (not calculate!) their a priori LLRs $\Lambda^{a,i}_{\mu^{-1}}$ or $\Lambda^{a,i}_{\nu^{-1}}$ *knowing already* that the corresponding a priori MI will equal the value $I^a_{\mu^{-1}}$ or $I^a_{\nu^{-1}}$ chosen at step 1,
4. calculate the extrinsic LLRs $\Lambda^{e,i}_{\mu^{-1}}$ or $\Lambda^{e,i}_{\nu^{-1}}$ from the a priori LLRs $\Lambda^{a,i}_{\mu^{-1}}$ or $\Lambda^{a,i}_{\nu^{-1}}$ (in two different ways for the demapper and the decoder), and finally
5. calculate the extrinsic MI $I^e_{\mu^{-1}}$ or $I^a_{\nu^{-1}}$ corresponding to the extrinsic LLRs $\Lambda^{e,i}_{\mu^{-1}}$ or $\Lambda^{e,i}_{\nu^{-1}}$.

A.1. Generate a priori LLRs corresponding to a given a priori MI

For the sake of readability in the following I will only refer to the demapper, leaving to the reader's goodwill the task to substitute $\Lambda^{a,i}_{\mu^{-1}}$ with $\Lambda^{a,i}_{\nu^{-1}}$ to adapt the formulas to the decoder.

Clearly the problem is the step 3, since normally one

1. *calculates*, not generates, the LLRs and
2. calculates the MI *from* the LLRs (and the bits), not generates the LLRs from the MI.

Luckily, ten Brink noted that [68, 70]

> the extrinsic information LLRs $\Lambda^e_{\nu^{-1}}$ (i.e. $\Lambda^a_{\mu^{-1}}$) as fed back from the outer decoder are almost Gaussian distributed. Additionally, large interleavers keep the a priori LLRs $\Lambda^a_{\mu^{-1}}$ fairly uncorrelated over many iterations.

Hence he suggested to generate the a priori LLRs $\Lambda^{a,i}_{\mu^{-1}}$ according to a normal distribution with mean $1/2 \cdot B'_e \sigma^2_{\Lambda^a_{\mu^{-1}}}$ and variance $\sigma^2_{\Lambda^a_{\mu^{-1}}}$, solving the first problem. Note that B'_e is the encoded bit B_e scaled and translated to be either -1 or $+1$,

$$B'_e = 2B_e - 1 = \begin{cases} -1 & \text{, if } B_e = 0 \\ +1 & \text{, if } B_e = 1, \end{cases} \tag{A.1}$$

since the LLRs must be symmetrically distributed around 0. Note also that mean and variance are related, which means that the distribution of the a priori LLR $\Lambda^a_{\mu^{-1}}$ is described by only *one* parameter; this will prove useful in the following. Hence

$$-3\sigma_{\Lambda^a_{\mu^{-1}}} - \frac{1}{2}\sigma^2_{\Lambda^a_{\mu^{-1}}} \leq \Lambda^a_{\mu^{-1}} \leq +3\sigma_{\Lambda^a_{\mu^{-1}}} + \frac{1}{2}\sigma^2_{\Lambda^a_{\mu^{-1}}}$$

99.73% of the times.

As for the second problem, under the assumption that the a priori LLR $\Lambda^a_{\mu^{-1}}$ is Gaussianly distributed, he knows its conditional probability density function (pdf) conditioned on B'_e,

$$p_{\Lambda^a_{\mu^{-1}}|B'_e}(\lambda^a_{\mu^{-1}}|b) = \frac{1}{\sqrt{2\pi\sigma^2_{\Lambda^a_{\mu^{-1}}}}} e^{-\frac{1}{2}\frac{\left(\lambda^a_{\mu^{-1}} - \frac{1}{2}b\sigma^2_{\Lambda^a_{\mu^{-1}}}\right)^2}{\sigma^2_{\Lambda^a_{\mu^{-1}}}}}. \tag{A.2}$$

Multiplying (A.2) by the pdf of B'_e, assumed to be uniformly distributed,

$$p_{B'_e}(b) = \frac{1}{2}, \quad b \in \{-1, +1\}, \tag{A.3}$$

he obtains the joint pdf of the a priori LLR and the bit $p_{\Lambda^a_{\mu^{-1}}, B'_e}(\lambda^a_{\mu^{-1}}, b)$, and summing the latter over $b \in \{-1, +1\}$ he gets the pdf of the a priori LLR $p_{\Lambda^a_{\mu^{-1}}}(\lambda^a_{\mu^{-1}})$.

As usual, the a priori MI $I^a_{\mu^{-1}}$ can be written as the expectation over B'_e and the a priori LLR $\Lambda^a_{\mu^{-1}}$ of the following function,

$$\log_2\left(\frac{p_{\Lambda^a_{\mu^{-1}}|B'_e}(\Lambda^a_{\mu^{-1}}|B'_e)}{p_{\Lambda^a_{\mu^{-1}}}(\Lambda^a_{\mu^{-1}})}\right). \tag{A.4}$$

In [68] ten Brink calls this quantity $J(\sigma^2_{\Lambda^a_{\mu^{-1}}})$,

$$I^a_{\mu^{-1}} \triangleq J(\sigma^2_{\Lambda^a_{\mu^{-1}}}) = \sum_{b\in\{-1,+1\}} \int_{-\infty}^{+\infty} p_{\Lambda^a_{\mu^{-1}}, B'_e}(\lambda^a_{\mu^{-1}}, b) \log_2\left(\frac{p_{\Lambda^a_{\mu^{-1}}|B'_e}(\lambda^a_{\mu^{-1}}|b)}{p_{\Lambda^a_{\mu^{-1}}}(\lambda^a_{\mu^{-1}})}\right) d\lambda^a_{\mu^{-1}}, \tag{A.5}$$

but does not say how to calculate it; this time I used the definition itself, numerically integrating (A.5) [57] instead of estimating it via Monte Carlo (MC), as I did in chapter

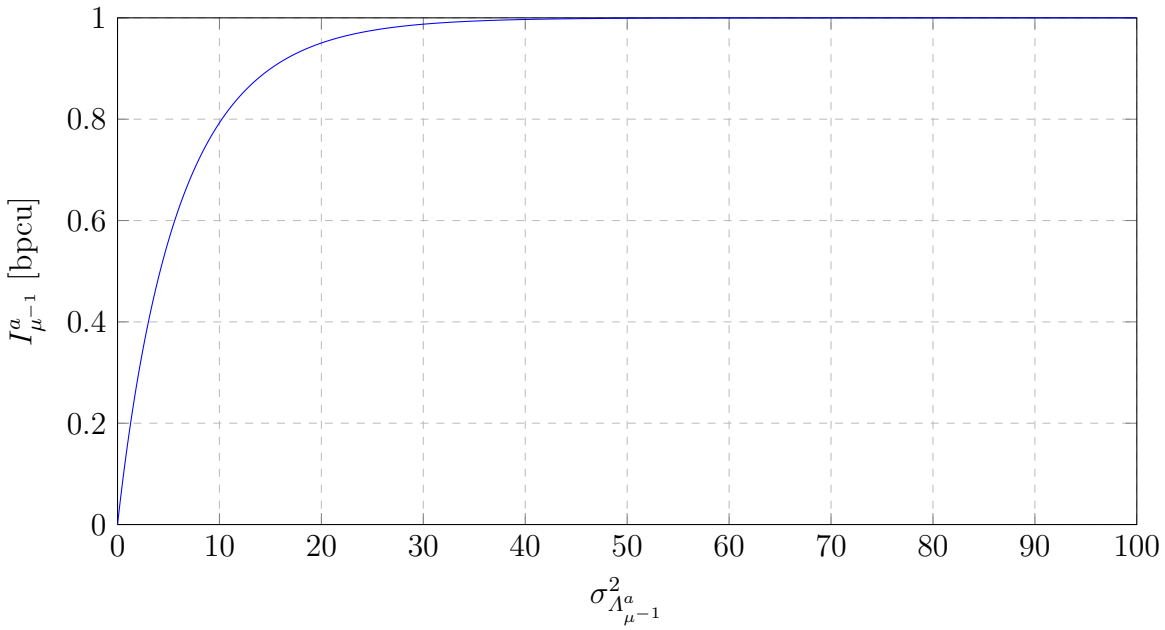

Figure A.1.: A priori MI as a function of the variance of the a priori LLRs.

5, since I had a closed form of the three pdfs $p_{\Lambda^a_{\mu^{-1}}|B'_e}(\lambda^a_{\mu^{-1}}|b)$, $p_{\Lambda^a_{\mu^{-1}},B'_e}(\lambda^a_{\mu^{-1}},b)$ and $p_{\Lambda^a_{\mu^{-1}}}(\lambda^a_{\mu^{-1}})$. An approximated version is also available [79,88]; nonetheless, I decided to calculate it to gain more insight in the topic and because, once I calculated and saved it, I could store it and look it up when needed.

Figure A.1 shows the a priori MI $I^a_{\mu^{-1}}$ as a function of the a priori LLR variance $\sigma^2_{\Lambda^a_{\mu^{-1}}}$. As it can be seen the function is monotonically increasing and thus reversible, so it can be used swapping the x- and the y-axis to know which variance $\sigma^2_{\Lambda^a_{\mu^{-1}}}$ the a priori LLR $\Lambda^a_{\mu^{-1}}$ must have in order to have a certain a priori MI $I^a_{\mu^{-1}}$, solving the second problem.

A.2. Calculate the extrinsic LLRs from the a priori LLRs

Once I had a way to generate a priori LLRs $\Lambda^{a,i}_{\mu^{-1}}$ or $\Lambda^{a,i}_{\nu^{-1}}$ corresponding to a given a priori MI I fed them to the demapper or to the decoder and obtained the extrinsic LLRs $\Lambda^{e,i}_{\mu^{-1}}$ or $\Lambda^{e,i}_{\nu^{-1}}$.

A.2.1. Demapper for the NDE AWGN model

For each bit mapping μ_k and each SNR:

1. I generated N_{bits} encoded bits $B_{e,i}$ according to (A.3);
2. I generated their a priori LLRs $\Lambda^{a,i}_{\mu^{-1}}$ according to a normal distribution with mean $1/2 \cdot B'_e\sigma^2_{\Lambda^a_{\mu^{-1}}}$ and variance $\sigma^2_{\Lambda^a_{\mu^{-1}}}$ obtained interpolating figure A.1;

3. I mapped the encoded bits $B_{e,i}$ to $N_{\text{bits}}/3$ transmitted symbols X_j with the chosen bit mapping μ_k;

4. I generated the AWGN samples N_j according to (2.6), with the AWGN variance corresponding to the chosen SNR;

5. I obtained the received symbols Y_j according to (2.1);

6. I calculated the extrinsic LLRs of the three bits adapting the formulas reported in [89, 90] to the case of an 8-ary constellation:

$$\Lambda_{\mu^{-1}}^{e,2} = \ln\left(\frac{p(000)e^{\Lambda_{\mu^{-1}}^{a,1}}e^{\Lambda_{\mu^{-1}}^{a,0}} + p(001)e^{\Lambda_{\mu^{-1}}^{a,1}} + p(010)e^{\Lambda_{\mu^{-1}}^{a,0}} + p(011)}{p(100)e^{\Lambda_{\mu^{-1}}^{a,1}}e^{\Lambda_{\mu^{-1}}^{a,0}} + p(101)e^{\Lambda_{\mu^{-1}}^{a,1}} + p(110)e^{\Lambda_{\mu^{-1}}^{a,0}} + p(111)}\right), \tag{A.6}$$

$$\Lambda_{\mu^{-1}}^{e,1} = \ln\left(\frac{p(000)e^{\Lambda_{\mu^{-1}}^{a,2}}e^{\Lambda_{\mu^{-1}}^{a,0}} + p(001)e^{\Lambda_{\mu^{-1}}^{a,2}} + p(100)e^{\Lambda_{\mu^{-1}}^{a,0}} + p(101)}{p(010)e^{\Lambda_{\mu^{-1}}^{a,2}}e^{\Lambda_{\mu^{-1}}^{a,0}} + p(011)e^{\Lambda_{\mu^{-1}}^{a,2}} + p(110)e^{\Lambda_{\mu^{-1}}^{a,0}} + p(111)}\right), \tag{A.7}$$

$$\Lambda_{\mu^{-1}}^{e,0} = \ln\left(\frac{p(000)e^{\Lambda_{\mu^{-1}}^{a,2}}e^{\Lambda_{\mu^{-1}}^{a,1}} + p(010)e^{\Lambda_{\mu^{-1}}^{a,2}} + p(100)e^{\Lambda_{\mu^{-1}}^{a,1}} + p(110)}{p(001)e^{\Lambda_{\mu^{-1}}^{a,2}}e^{\Lambda_{\mu^{-1}}^{a,1}} + p(011)e^{\Lambda_{\mu^{-1}}^{a,2}} + p(101)e^{\Lambda_{\mu^{-1}}^{a,1}} + p(111)}\right), \tag{A.8}$$

where I call

$$p(b_2b_1b_0) \triangleq p_{Y|X}\left(y|\mu_k(b_2,b_1,b_0)\right) \tag{A.9}$$

for space reasons. The quantity $p_{Y|X}\left(y|\mu_k(b_2,b_1,b_0)\right)$, clearly given by (2.7), represents the information coming from the channel, i.e. the probability that the symbol y has been received *given that* the symbol to which the bit mapping μ_k maps the bits b_2, b_1, b_0 has been transmitted. Note that, in order to calculate the extrinsic MI of *one* bit, only the a priori MIs of the *other* bits are used;

7. I generated $N_{\text{bits}} = 9 \cdot 10^6$ bits, corresponding to $3 \cdot 10^6$ symbols.

A.2.2. Decoder

The procedure to obtain the extrinsic LLRs for the decoder was similar to the one described in the previous section.

In particular, I skipped the steps from 3 to 5 and I replaced the step 6, obtaining for each number of LDPC iterations N_{LDPC} the extrinsic LLRs of the n bits $\Lambda_{\nu^{-1}}^{e}$ using the built-in functions of the Communications Toolbox of MATLAB ®. These functions [77]

> performs LDPC decoding using the belief-passing or message-passing algorithm, implemented as the log-domain sum-product algorithm

and are optimised for speed, since they can be run on the graphics processing unit (GPU) [78]. I generated $N_{\text{bits}} \approx 9 \cdot 10^6$ bits, corresponding to 139 code words.

A.2.3. Demapper for the DE AWGN model

The procedure to calculate the extrinsic LLRs of the demapper over the DE AWGN model was very similar to the one presented in the previous section, the only differences being that

- I now mapped the N_{bits} generated encoded bits $B_{e,i}$ to $N_{\text{bits}}/3$ transmitted *transitions* T_j with the chosen bit mapping μ_k;
- I generated the first (previous) transmitted symbol X_1 according to (2.2);
- I obtained the following transmitted symbols according to the DE function (2.28);
- I calculated the extrinsic LLRs of the three bits adapting (A.6), (A.7) and (A.8) to the DE case. It turned out that (A.6), (A.7) and (A.8) are still valid once the information coming from the channel (A.9) (where the dependency on the chosen bit mapping μ_k is present) is adapted as follows:

$$p(b_2 b_1 b_0) \triangleq p_{Y_1,Y_2|T}(y_1, y_2|\mu_k(b_2, b_1, b_0)). \tag{A.10}$$

  The quantity $p_{Y_1,Y_2|T}(y_1, y_2|\mu_k(b_2, b_1, b_0))$, clearly given by (2.36), represents the information coming from the channel, i.e. the probability that the symbols y_1, y_2 have been received *given that* the transition to which the bit mapping μ_k maps the bits b_2, b_1, b_0 has been transmitted.

A.3. Calculate the extrinsic MI corresponding to given extrinsic LLRs

Once I had the extrinsic LLRs $\Lambda^{e,i}_{\mu^{-1}}$ or $\Lambda^{e,i}_{\nu^{-1}}$ I could have estimated their variance and interpolated again figure A.1 to obtain the corresponding extrinsic MI; however, I preferred to use the definition itself, in order not to have to rely on the assumption that the extrinsic LLR is Gaussianly distributed.

In fact, even the extrinsic MI $I^e_{\mu^{-1}}$ can be written as expectation over B'_e and the extrinsic LLR $\Lambda^e_{\mu^{-1}}$ of a function similar to (A.4), this time with the conditional pdf of the extrinsic LLR conditioned on the bit $p_{\Lambda^e_{\mu^{-1}}|B'_e}(\lambda^e_{\mu^{-1}}\big|b)$, the joint pdf of the extrinsic LLR and the bit $p_{\Lambda^e_{\mu^{-1}},B'_e}(\lambda^e_{\mu^{-1}}, b)$ and the pdf of the extrinsic LLR $p_{\Lambda^e_{\mu^{-1}}}(\lambda^e_{\mu^{-1}})$.

Since this time no closed form for these three pdfs is known, I estimated the first one by opportunely scaling (such that the integral of the pdf over the real axis would be 1) the histograms of the extrinsic LLR $\Lambda^e_{\mu^{-1}}$ relative to $B'_e = 0$ and those relative to $B'_e = 1$; the histograms had 10^3 bins.

Multiplying it by (A.3) I obtained the second one and summing the latter over $b \in \{-1, +1\}$ I got the third one.

Finally, I approximated the integral $\int_{-\infty}^{+\infty} d\lambda^e_{\mu^{-1}}$ from the definition of expectation with a sum.

A.4. Calculate the a posteriori MI from the a priori MI and the extrinsic MI

The a posteriori LLRs are given by the sum of the a priori LLRs and the extrinsic LLRs,

$$\Lambda_{\nu^{-1}}^{p,i} = \Lambda_{\nu^{-1}}^{a,i} + \Lambda_{\nu^{-1}}^{e,i}. \tag{A.11}$$

Under the assumption that both the a priori LLR and the extrinsic LLR are Gaussianly distributed and uncorrelated [68,70], also the a posteriori LLR is Gaussianly distributed and uncorrelated, with variance given by the sum of the variances of the a priori LLR and the extrinsic LLR:

$$\sigma^2_{\Lambda^p_{\nu^{-1}}} = \sigma^2_{\Lambda^a_{\nu^{-1}}} + \sigma^2_{\Lambda^e_{\nu^{-1}}}. \tag{A.12}$$

Hence, interpolating as usual figure A.1 I can calculate:

$$I^p_{\nu^{-1}} = J(J^{-1}(I^a_{\nu^{-1}}) + J^{-1}(I^e_{\nu^{-1}})). \tag{A.13}$$

A.5. Estimate the BER vs. SNR curves of the traditional solution

For each SNR:

1. I generated $N_{\text{bits}} \cdot 8/9$ information bits according to (A.3);
2. I encoded the information bits into N_{bits} encoded bits $B_{e,i}$ with the built-in functions of the Communications Toolbox of MATLAB ® [91];
3. I mapped the encoded bits $B_{e,i}$ to either $N_{\text{bits}}/2$ symbols X_j (in the NDE case) or $N_{\text{bits}}/2$ transitions T_j (in the DE case) with the Gray-coded bit mapping. In the DE case I also generated the first (previous) transmitted symbol X_1 according to (2.2) and I obtained the following transmitted symbols according to the DE function (2.28);
4. I generated the AWGN samples N_j according to (2.6), with the AWGN variance corresponding to the chosen SNR;
5. I obtained the received symbols Y_j according to (2.1);
6. I calculated the a posteriori LLRs $\Lambda_{\mu^{-1}}^{p,i}$ according to [89,90] and setting the a priori LLRs $\Lambda_{\mu^{-1}}^{a,i}$ to zero;
7. I took the sign of the a posteriori LLRs $\Lambda_{\mu^{-1}}^{p,i}$ and I obtained the estimated encoded bits $\hat{B_{e,i}}$;

8. I decoded the a posteriori LLRs $\Lambda_{\mu^{-1}}^{p,i}$ with the built-in functions of the Communications Toolbox of MATLAB ® [77, 78] and I obtained the estimated information bits;

9. I estimated the pre-FEC BER and the post-FEC BER;

10. in order to speed up the simulation, I generated a number of information bits $N_{\text{bits}} \cdot 8/9$ which depended on the expected post-FEC BER, ranging from $N_{\text{cw}}k = 22 \cdot 57600 \approx 10^6$ to $N_{\text{cw}}k = 2200 \cdot 57600 \approx 126 \cdot 10^6$, where N_{cw} is the number of code words. Correspondingly, the number of generated encoded bits N_{bits} ranged from $N_{\text{cw}}n = 22 \cdot 64800 \approx 10^6$ to $N_{\text{cw}}n = 2200 \cdot 64800 \approx 143 \cdot 10^6$.

Acknowledgements

This thesis covers 14 years, 2 months and 22 days of my life (excluding the end date). As such, the list of people I met while working on it (in full or part time) is quite long.

I would like to thank my university supervisor, Prof. Lankl. He helped me keep my work on academic standards, did not give up on me while time went by and led me to my defense even after his retirement. I would also like to thank Prof. Hanik, who agreed to be my second reviewer.

I would like to thank my industry mentor, Dr. Calabrò. He patiently taught me many things, indulged me when I had many questions (often only mildly related to my topic) and helped me even after he left the company. With him, I would like to thank Dr. Jansen, who encouraged me to go to the lab, as well as Dr. Kuschnerov, who developed and implemented the digital signal processing (DSP) algorithms I used for the lab experiments.

Among the many students I crossed paths during those 5 full time years, I obviously would like to thank Vincent. We shared a desk in the office (and sometimes even the lab), a love for things done well and the pursuit of truth behind the smoke and mirrors, the same healthy do-not-take-yourself-too-seriously attitude, many Schnitzel and Helle in Munich and a common passion for football. Beside this, he helped me set up the lab for the experiments and always pushed me to do my best (in his own peculiar way). Vos, here's to you!

With him, I would like to thank also our leader Mohammad, as well as Adriana, Beril and Susmita, who did their Ph.D.'s at NSN in the same years. We had fun during lunches, played Kicker in the office and freaked out together before conference deadlines.

I also had an office at the university, so I would like to thank Christian, Daniel, Kittipong, Michael, Nora and Vito. Beside our lunches when I was there (lunches are important!), they helped me survive the bureaucracy of our university without speaking a single word of German, not an easy task.

I would like to thank Dr. Erseghe, Dr. Laurenti and Prof. Vangelista from the university of Padova in Italy, where I studied and worked after my degree, as well as Dr. Talli and Prof. Townsend from the Tyndall Institute in Ireland, where I did my master thesis. They all encouraged me to do a Ph.D., convincing me that I might have it in me. I would also like to thank Alberto (Zeta) for refreshing my memory (he too in his own peculiar way) on the basics of Communications Theory over and over (and over) again. God knows I needed that while writing this thesis.

I would like to thank my parents Rosa and Roberto, as well as all the friends, flatmates, bosses, colleagues and even strangers that listened to me ranting about my Ph.D., let me take some time off to write (various versions of) this thesis and encouraged me not to give up. Among my friends in Germany, I would like to mention Don Diego, who I met

in Munich while he was studying Philosophy. I will always treasure our conversations in front of many Hendln and Augustiner, where he revived my interest in humanities. With him, I would like to recall the other two philosophers in my life, Aldo and Bernardo (with their completely opposite ways of pursuing a Ph.D.), my long-time flatmates Thomas and Cristian (Zucca), as well as my cards fellows Giorgio, Luca and Tommaso (Pans). At the same time, I could not forget my friends in Italy, in particular Alvise (Dado) and Giampaolo (Gp), to whom I still owe a trip to Iceland.

Last but not least, I would like to thank my girlfriend Daniela. She was not there at the beginning but she was there at the end, when I needed it most. She supported me, listened patiently to the rehearsals of my presentation and helped me value what I was achieving.

What a journey.

Bibliography

[1] "Quotes falsely attributed to Winston Churchill," https://winstonchurchill.org/resources/quotes/quotes-falsely-attributed/, accessed: 11/05/2024.

[2] F. Chang, K. Onohara, and T. Mizuochi, "Forward error correction for 100G transport networks," *IEEE Commun. Mag.*, vol. 48, pp. S48–S55, Mar. 2010.

[3] E. M. Ip and J. M. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *J. Lightw. Technol.*, vol. 28, pp. 502–519, Feb. 2010.

[4] "Fiber-optic technologies > a brief history of fiber-optic communications," Apr. 2004. [Online]. Available: https://www.ciscopress.com/articles/article.asp?p=170740

[5] M. v. Sivers, "A brief history of optical communication," Feb. 2021. [Online]. Available: https://hackaday.com/2021/02/18/a-brief-history-of-optical-communication/

[6] C. Wilson, "Verizon switches on 100G in Europe," Dec. 2009. [Online]. Available: https://www.lightreading.com/ethernet-ip/verizon-switches-on-100g-in-europe/d/d-id/673196

[7] "Verizon, Nortel light first commercial 100G wave," Dec. 2009. [Online]. Available: https://www.fiercewireless.com/europe/verizon-nortel-light-first-commercial-100g-wave

[8] "Nortel unveils industry's first commercially available 100G optical solution," Dec. 2009. [Online]. Available: https://www.fiberopticsonline.com/doc/nortel-unveils-industrys-first-commercially-0001

[9] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, pp. 55–67, Jan. 1982.

[10] ——, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, pp. 5–11, Feb. 1987.

[11] ——, "Trellis-coded modulation with redundant signal sets Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, pp. 12–21, Feb. 1987.

[12] P. Leoni, S. Calabrò, and B. Lankl, "Constellation expansion for 100G transmission," *IEEE Photon. Technol. Lett.*, vol. 25, pp. 1904–1907, Oct. 2013.

[13] ——, "Constellation expansion for differentially encoded 100G transmission," *IEEE Photon. Technol. Lett.*, vol. 26, pp. 1142–1145, Jun. 2014.

[14] P. Leoni, V. A. J. M. Sleiffer, S. Calabrò, and B. Lankl, "Constellation expansion and multi-symbol detection for differentially encoded 100G systems," in *Proc. OFC'14*, San Francisco, CA, Mar. 23–26, 2014, paper W2A.15.

[15] ——, "Constellation expansion and iterative demapping and decoding for 100G systems," in *Proc. ECOC'13*, London, UK, Sep. 22–26, 2013, paper Tu.1.E.1.

[16] P. Leoni, S. Calabrò, B. Lankl, and B. Spinnler, "On the joint optimization of modulation and channel coding for high data-rate optical communication systems," in *Proc. SPPCom'11*, Toronto, Canada, Jun. 12–15, 2011, paper SPTuA3.

[17] P. Leoni, V. A. J. M. Sleiffer, S. Calabrò, and B. Lankl, "Constellation expansion: Theoretical limits and practical examples (invited)," in *Proc. ICTON'14*, Graz, Austria, Jul. 6–10, 2014, paper Tu.D1.1.

[18] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L. D. Coelho, and N. Hanik, "Reduced complexity digital back-propagation methods for optical communication systems," *J. Lightw. Technol.*, vol. 32, pp. 1351–1362, Apr. 2014.

[19] K.-P. Ho, "Probability density of nonlinear phase noise," *J. Opt. Soc. Am. B*, vol. 20, pp. 1875–1879, Sep. 2003.

[20] ——, "Asymptotic probability density of nonlinear phase noise," *Optics Lett.*, vol. 28, no. 15, pp. 1350–1352, Aug. 2003.

[21] H. Kim and A. H. Gnauck, "Experimental investigation of the performance limitation of DPSK systems due to nonlinear phase noise," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 320–322, Feb. 2003.

[22] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," *IEEE Photon. Technol. Lett.*, vol. 23, pp. 742–744, Jun. 2011.

[23] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, pp. 1524–1539, May 2012.

[24] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, pp. 3857–3879, Dec. 2012.

[25] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. Waltham, Massachusetts, US: Academic Press, 2001.

[26] B. Goebel, R. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent gaussian channels with applications to fiber optics," *IEEE Trans. Inf. Theory*, vol. 57, pp. 5720–5736, Sep. 2011.

[27] K.-P. Ho, "Error probability of DPSK signals with cross-phase modulation induced nonlinear phase noise," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, pp. 421–427, Mar./Apr. 2004.

[28] T.-K. Chiang, N. Kagi, M. E. Marhic, and L. G. Kazovsky, "Cross-phase modulation in fiber links with multiple optical amplifiers and dispersion compensators," *J. Lightw. Technol.*, vol. 14, pp. 249–260, Mar. 1996.

[29] W. Yan, Z. Tao, L. Li, L. Liu, S. Oda, T. Hoshida, and J. C. Rasmussen, "A linear model for nonlinear phase noise induced by cross-phase modulation," in *Proc. OFC'09*, San Diego, CA, Mar. 22–26, 2009, paper OTuD5.

[30] Z. Tao, W. Yan, S. Oda, T. Hoshida, and J. C. Rasmussen, "A simplified model for nonlinear cross-phase modulation in hybrid optical coherent system," *Opt. Express*, vol. 17, no. 16, pp. 13 860–13 868, Aug. 3, 2009.

[31] L. H. Lampe, R. F. Fischer, S. Calabrò, and S. H. Müller-Weinfurtner, "Coded modulation for DPSK on fading channels," in *Proc. GLOBECOM'99*, Rio de Janeiro, Brazil, Dec. 5– 9, 1999.

[32] R. F. H. Fischer, S. Calabrò, L. H. Lampe, and S. H. Müller-Weinfurtner, "Performance of coded modulation employing differential encoding over rayleigh fading channels," *Electron. Lett.*, vol. 35, pp. 122–123, Jan. I999.

[33] L. H. Lampe, R. F. Fischer, and S. Calabrò, "Channel capacity of fading channels for differentially encoded transmission," *Electron. Lett.*, vol. 35, pp. 192–194, Feb. 1999.

[34] R. F. Fischer, L. H. Lampe, and S. Calabrò, "Differential encoding strategies for transmission over fading channels," *AEÜ Int. J. Electron. Commun.*, vol. 54, pp. 59–67, Jan. 2000.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Upper Saddle River, New Jersey, US: Prentice Hall, 1993.

[36] P. Leoni, V. A. J. M. Sleiffer, S. Calabrò, V. Veljanovski, M. Kuschnerov, S. Jansen, and B. Lankl, "Impact of interleaving on SD-FEC operating in highly non-linear XPM-limited regime," in *Proc. OFC'13*, Anaheim, CA, Mar. 17–21, 2013, paper OW1E.6.

[37] P. Leoni, V. A. J. M. Sleiffer, S. Calabrò, M. Kuschnerov, S. L. Jansen, B. Spinnler, and B. Lankl, "On the performance of a soft decision FEC scheme operating in highly non-linear regime," in *Proc. SPPCom'12*, Colorado Springs, CO, Jun. 19–21, 2012, paper SpTu3A.6.

[38] P. Leoni, V. A. J. M. Sleiffer, S. Calabrò, V. Veljanovski, M. Kuschnerov, and B. Lankl, "Pre FEC error burstiness for 100G systems operating in heavily XPM-limited regime," in *Proc. SPPCom'13*, Rio Grande, PR, Jul. 14–17, 2013, paper SPT5D.4.

[39] *Interfaces for the optical transport network*, ITU-T Recommendation G.709/Y.1331, Jun. 2016.

[40] *Interfaces for the optical transport network - Amendment 1*, ITU-T Recommendation G.709/Y.1331, Nov. 2016.

[41] *Interfaces for the optical transport network - Corrigendum 1*, ITU-T Recommendation G.709/Y.1331, Aug. 2017.

[42] *Interfaces for the optical transport network - Amendment 2*, ITU-T Recommendation G.709/Y.1331, Jun. 2018.

[43] "Datasheet SHF 12100 B 50 Gbps bit pattern generator," https://www.shf-communication.com/wp-content/uploads/datasheets/datasheet_shf_12100_b.pdf, accessed: 18/05/2022.

[44] V. A. J. M. Sleiffer, Y. Jung, V. Veljanovski, R. G. H. van Uden, M. Kuschnerov, H. Chen, B. Inan, L. G. Nielsen, Y. Sun, D. J. Richardson, S. U. Alam, F. Poletti, J. K. Sahu, A. Dhar, A. M. J. Koonen, B. Corbett, R. Winfield, A. D. Ellis, and H. de Waardt, "73.7 Tb/s (96 x 3 x 256-gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA," *Opt. Express*, vol. 20, no. 26, pp. B428–B438, Dec. 10, 2012.

[45] L. Anttila, M. Valkama, and M. Renfors, "Blind moment estimation techniques for I/Q imbalance compensation in quadrature receivers," in *Proc. PIMRC'06*, Helsinki, Finland, Sep. 11–14, 2006, pp. 1–5.

[46] M. Kuschnerov, F. N. Hauske, K. Piyawanno, B. Spinnler, A. Napoli, and B. Lankl, "Adaptive chromatic dispersion equalization for non-dispersion managed coherent systems," in *Proc. OFC'09*, San Diego, CA, Mar. 22–26, 2009, paper OMT1.

[47] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.

[48] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, pp. 605–612, May 1988.

[49] M. Kuschnerov, "Signal processing for coherent optic receivers," Ph.D. dissertation, Universität der Bundeswehr München, München, Germany, 2011.

[50] V. A. J. M. Sleiffer, "Towards petabit per second optical long-haul transmission links using space-division multiplexing technology," Ph.D. dissertation, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, Jun. 17, 2014.

[51] M. Kuschnerov, F. N. Hauske, K. Piyawanno, B. Spinnler, M. S. Alfiad, A. Napoli, and B. Lankl, "DSP for coherent single-carrier receivers," *J. Lightw. Technol.*, vol. 27, pp. 3614–, Aug. 2009.

[52] "xcorr," https://uk.mathworks.com/help/matlab/ref/xcorr.html, accessed: 12/06/2023.

[53] M. G. Taylor, "Phase estimation methods for optical coherent detection using digital signal processing," *J. Lightw. Technol.*, vol. 27, pp. 901–914, Apr. 2009.

[54] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *J. Lightw. Technol.*, vol. 27, pp. 989–999, Apr. 2009.

[55] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *J. Lightw. Technol.*, vol. 30, pp. 1480–1486, May 2012.

[56] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, New Jersey, US: John Wiley & Sons, 2006.

[57] "quadgk," https://uk.mathworks.com/help/matlab/ref/quadgk.html, accessed: 27/04/2019.

[58] "quad," https://uk.mathworks.com/help/matlab/ref/quad.html, accessed: 27/04/2019.

[59] "dblquad," https://uk.mathworks.com/help/matlab/ref/dblquad.html, accessed: 27/04/2019.

[60] *2400 bits per second duplex modem using the frequency division technique standardized for use on the general switched telephone network and on point-to-point 2-wire leased telephone-type circuits*, ITU-T Recommendation V.22 bis, Nov. 1988.

[61] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, pp. 300–308, Mar. 1990.

[62] D. Divsalar, M. K. Simon, and M. Shahshahani, "The performance of trellis-coded MDPSK with multiple symbol detection," *IEEE Trans. Commun.*, vol. 38, pp. 1391–1403, Sep. 1990.

[63] D. Divsalar and M. K. Simon, "Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels - metrics and performance," *IEEE Trans. Commun.*, vol. 42, pp. 76–89, Jan. 1994.

[64] M. Peleg and S. Shamai (Shitz), "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, pp. 603–609, May 1998.

[65] P. Leoni, S. Calabrò, V. A. J. M. Sleiffer, and B. Lankl, "Implementation complexity of constellation expansion," in *Proc. ECOC'14*, Cannes, France, Sep. 21–25, 2014, paper Tu.1.3.6.

[66] *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)*, ETSI Recommendation EN 302 307, Aug. 2009.

[67] *Forward error correction for submarine systems*, ITU-T Recommendation G.975, Oct. 2000.

[68] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEÜ Int. J. Electron. Commun.*, vol. 54, pp. 389–398, Dec. 2000.

[69] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 gb/s OTN," *J. Lightw. Technol.*, vol. 30, pp. 110–117, Jan. 2012.

[70] S. ten Brink, "Design of serially concatenated codes based on iterative decoding convergence," in *Proc. ISTC'00*, Brest, France, Sep. 04–07, 2000.

[71] ——, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, pp. 806–808, May 1999.

[72] ——, "Editor's correction: Convergence of iterative decoding," *Electron. Lett.*, vol. 35, pp. 1117–1119, Jun. 1999.

[73] ——, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[74] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.

[75] A. Ashikhmin, G. Kramer, and S. ten Brink, "Code rate and the area under extrinsic information transfer curves," in *Proc. ISIT'02*, Lausanne, Switzerland, Jun. 30– 5, 2002.

[76] ——, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2657–2673, Nov. 2004.

[77] "comm.LDPCDecoder," https://uk.mathworks.com/help/comm/ref/comm.ldpcdecoder-system-object.html, accessed: 16/02/2019.

[78] "comm.gpu.LDPCDecoder," https://uk.mathworks.com/help/comm/ref/comm.gpu.ldpcdecoder-system-object.html, accessed: 16/02/2019.

[79] S. J. Johnson, *Iterative Error Correction: Turbo, Low-Density Parity-Check and Repeat-Accumulate Codes*, 1st ed. Cambridge, UK: Cambridge University Press, 2010.

[80] F. Zarkeshvari and A. H. Banihashemi, "On implementation of min-sum algorithm for decoding low-density parity-check (LDPC) codes," in *Proc. GLOBECOM'02*, Taipei, Taiwan, Nov. 17–21, 2002.

[81] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, Apr. 1965.

[82] ——, "Cramming more components onto integrated circuits," *Proc. IEEE*, vol. 86, pp. 82–85, Jan. 1998.

[83] S. L. Howard and C. Schlegel, "Differential turbo-coded modulation with APP channel estimation," *IEEE Trans. Commun.*, vol. 54, pp. 1397–1406, Aug. 2006.

[84] "The role of higher baud rates in evolving coherent transport," 2019. [Online]. Available: https://www.infinera.com/white-paper/the-role-of-higher-baud-rates-in-evolving-coherent-transport/

[85] "400G CFP2-DCO digital coherent optics transceiver." [Online]. Available: https://ii-vi.com/product/400g-cfp2-dco-digital-coherent-optics-transceiver/

[86] S. Hardy, "II-VI sampling 400G CFP2-DCO coherent optical transceiver," Mar. 2020. [Online]. Available: https://www.lightwaveonline.com/optical-tech/transmission/article/14169190/iivi-sampling-400g-cfp2dco-coherent-optical-transceiver

[87] E. Park, "400zr: Accessible 400g for edge dcis and beyond," Jul. 2019. [Online]. Available: https://acacia-inc.com/acacia-resources/400zr-accessible-400g-for-edge-dcis-and-beyond/

[88] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, pp. 670–678, Apr. 2004.

[89] S. ten Brink, J. Speidel, and R. Yan, "Iterative demapping and decoding for multi-level modulation," in *Proc. GLOBECOM'98*, Sydney, Australia, Nov. 8–12, 1998.

[90] ——, "Iterative demapping for QPSK modulation," *Electron. Lett.*, vol. 34, pp. 1459–1460, Jul. 1998.

[91] "comm.LDPCEncoder," https://uk.mathworks.com/help/comm/ref/comm.ldpcencoder-system-object.html, accessed: 30/03/2019.

[92] H. Chen, V. A. J. M. Sleiffer, F. Huijskens, R. van Uden, C. Okonkwo, P. Leoni, M. Kuschnerov, L. Gruner-Nielsen, Y. Sun, H. de Waardt, and T. Koonen, "Employing prism-based three-spot mode couplers for high capacity MDM/WDM transmission," *IEEE Photon. Technol. Lett.*, vol. 25, pp. 2474–2477, Dec. 2013.

[93] V. A. J. M. Sleiffer, Y. Jung, P. Leoni, M. Kuschnerov, N. Wheeler, N. Baddela, R. van Uden, C. Okonkwo, J. Hayes, J. Wooler, E. Fokoua, R. Slavik, F. Poletti, M. Petrovich, V. Veljanovski, S. Alam, D. Richardson, and H. de Waardt, "30.7 tb/s (96x320 gb/s) DP-32QAM transmission over 19-cell photonic band gap fiber," in *Proc. OFC'13*, Anaheim, CA, Mar. 17–21, 2013, paper OWI1.5.

[94] V. A. J. M. Sleiffer, P. Leoni, Y. Jung, H. Chen, M. Kuschnerov, S. U. Alam, M. Petrovich, F. Poletti, N. Wheeler, N. Baddela, J. Hayes, E. Numkam, D. J. Richardson, L. Grüner-Nielsen, Y. Sun, and H. de Waardt, "Ultra-high capacity transmission with few-mode silica and hollow-core photonic bandgap fibers," in *Proc. OFC'14*, San Francisco, CA, Mar. 23–26, 2014, paper Tu2J.3.

[95] V. A. J. M. Sleiffer, P. Leoni, Y. Jung, J. Surof, M. Kuschnerov, V. Veljanovski, D. Richardson, S. Alam, L. Grüner-Nielsen, Y. Sun, B. Corbett, R. Winfield, S. Calabrò, B. Sommerkorn-Krombholz, H. von Kirchbauer, and H. de Waardt, "20 x 960 gb/s MDM-DP-32QAM transmission over 60km FMF with inline MM-EDFA," in *Proc. ECOC'13*, London, UK, Sep. 22–26, 2013, paper We.2.D.2.

[96] V. A. J. M. Sleiffer, H. Chen, Y. Jung, P. Leoni, M. Kuschnerov, A. Simperler, H. Fabian, H. Schuh, F. Kub, D. J. Richardson, S. U. Alam, L. Grüner-Nielsen, Y. Sun, A. M. J. Koonen, and H. de Waardt, "Field demonstration of mode-division multiplexing upgrade scenarios on commercial networks," *Opt. Express*, vol. 21, no. 25, pp. 31 036–31 046, Dec. 16, 2013.

[97] V. A. J. M. Sleiffer, P. Leoni, Y. Jung, J. Surof, M. Kuschnerov, V. Veljanovski, S. U. Alam, D. J. Richardson, L. Grüner-Nielsen, Y. Sun, B. Corbett, R. Winfield, S. Calabrò, and H. de Waardt, "20 Œ 960-Gb/s space-division-multiplexed 32QAM transmission over 60 km few-mode fiber," *Opt. Express*, vol. 22, no. 1, pp. 749–755, Jan. 13, 2014.

[98] V. A. J. M. Sleiffer, Y. Jung, P. Leoni, M. Kuschnerov, V. Veljanovski, N. V. Wheeler, N. Baddela, J. R. Hayes, J. Wooler, E. Numkam, R. Slavik, F. Poletti, M. N. Petrovich, S. U. Alam, D. J. Richardson, and H. de Waardt, "A first glance at coherent optical transmission using photonic bandgap fiber as a transmission medium," in *Proc. Photonics Society Summer Topical Meeting'13*, Waikoloa, HI, Jul. 8–10, 2013.

[99] V. A. J. M. Sleiffer, H. Chen, Y. Jung, P. Leoni, M. Kuschnerov, H. Fabian, H. Schuh, F. Kub, D. J. Richardson, S. U. Alam, L. Grüner-Nielsen, Y. Sun, A. D. Ellis, A. M. J. Koonen, and H. de Waardt, "A field trial of mode-division

multiplexing," in *Proc. Photonics Society Summer Topical Meeting'14*, Montreal, QC, Jul. 14–16, 2014.

[100] V. A. J. M. Sleiffer, Y. Jung, P. Leoni, M. Kuschnerov, R. G. H. van Uden, V. Veljanovski, L. Grüner-Nielsen, Y. Sun, D. J. Richardson, S. U. Alam, F. Poletti, B. Corbett, R. Winfield, and H. de Waardt, "High capacity multi-mode transmission systems using higher-order modulation formats," in *Proc. OECC'13*, Kyoto, Japan, 30 June-4 July, 2013.

[101] S. Adhikari, S. L. Jansen, M. Alfiad, B. Inan, V. A. J. M. Sleiffer, A. Lobato, P. Leoni, and W. Rosenkranz, "Self-coherent optical OFDM: An interesting alternative to direct or coherent detection," in *Proc. ICTON'11*, Stockholm, Sweden, Jun. 26–30, 2011.

List of Acronyms

2A2P 2 amplitude- 2 phase-shift keying

2A4P 2 amplitude- 4 phase-shift keying

2A8P 2 amplitude- 8 phase-shift keying

2D two-dimensional

4A4P 4 amplitude- 4 phase-shift keying

ACRT approximated, counter-rotated and translated

ADC analog-to-digital converter

AGC automatic gain control

ASE amplified spontaneous emission

AWGN additive white Gaussian noise

B2B back to back

BCH Bose-Chaudhuri-Hocquenghem

BE bandwidth expansion

BEC binary erasure channel

BER bit error rate

BICM bit-interleaved coded modulation

bpcu bits per channel use

BPF bandpass filter

CD chromatic dispersion

CE constellation expansion

CMA constant modulus algorithm

CR carrier recovery

CRLB Cramér-Rao lower bound

CRT counter-rotated and translated

DAC digital-to-analog converter

DBP digital back-propagation

DC direct current

DCF dispersion compensating fibre

DD-LMS decision-directed least mean square

DE differentially encoded

DSP digital signal processing

DVB digital video broadcasting

DVB-S2 digital video broadcasting – satellite – 2$^{\text{nd}}$ generation

ECOC European Conference on Optical Communications

EDFA erbium doped fibre amplifier

ETSI European Telecommunications Standards Institute

EXIT extrinsic information transfer

FEC forward error correction

FFT fast Fourier transform

FIR finite impulse response

GPU graphics processing unit

HD hard-decision

HWHM half-width at half-maximum

ICA independent component analysis

ICTON International Conference on Transparent Optical Networks

IDD iterative demapping and decoding

IDDD iterative DSP, demapping and decoding

IFFT inverse fast Fourier transform

IQM IQ modulator

ISI intersymbol interference

ITU-T Telecommunication Standardization Sector of International Telecommunication Union

LASER light amplification by stimulated emission of radiation

LDPC low-density parity-check

LLR log-likelihood ratio

LO local oscillator

MC Monte Carlo

MI mutual information

MIMO multiple-input and multiple-output

MLA max-log approximation

MSA min-sum algorithm

MSD multisymbol detection

MVU minimum-variance unbiased

MZM Mach-Zehnder modulator

NCG net coding gain

NDE non differentially encoded

NSN Nokia Siemens Networks

ODU4 optical data unit 4

OFC Optical Fiber Conference

OH overhead

OOK on-off keying

OSA optical spectrum analyzer

OSNR *optical* signal-to-noise ratio

OTU4 optical transport unit 4

PAM pulse amplitude modulation

PBC polarisation beam combiner

PBS polarisation beam splitter

PD photodiode

pdf probability density function

PDM polarisation division multiplexing

PLL phase-locked loop

PMD polarisation mode dispersion

PPG pulse pattern generator

PSK phase-shift keying

PTL IEEE Photonics Technology Letters

PWGN phase white Gaussian noise

QAM quadrature amplitude modulation

QPSK quadrature phase-shift keying

ROADM reconfigurable optical add-drop multiplexer

RS Reed-Solomon

SD soft-decision

SE spectral efficiency

SMF single-mode optical fibre

SNR signal-to-(*additive*-)noise ratio

SOP state of polarisation

spib summations per information bit

SPNR signal-to-*phase*-noise ratio

SPPCom Signal Processing in Photonic Communications

STR square timing recovery

VLSI very large scale integration

VOA variable optical attenuator

WDM wavelength division multiplexing

wPWGN wrapped phase white Gaussian noise

WSS wavelength selective switch

XOR exclusive OR

XPM cross-phase modulation

List of Symbols

α Fibre attenuation coefficient

B Random variable modelling the bit

B' Random variable modelling the scaled and translated bit

B_e Random variable modelling the *encoded* bit

B_{ref} Reference bandwidth

B_s Bandwidth occupied by the signal

C_j j-th check node in the graph of the LDPC code

Δ_f LASER linewidth

E_b (average) energy per transmitted *information* bit

ε DE function

E_s (average) energy per transmitted symbol

F_N Noise figure of an amplifier

f_s Symbol rate

$\mathcal{H}$ Binary entropy function

I Mutual information

$I^a_{\mu^{-1}}$ Mutual information between the bits and their a priori LLRs entering the demapper

$I^a_{\nu^{-1}}$ Mutual information between the bits and their a priori LLRs entering the decoder

$I^e_{\mu^{-1}}$ Mutual information between the bits and their extrinsic LLRs leaving the demapper

$I^e_{\nu^{-1}}$ Mutual information between the bits and their extrinsic LLRs leaving the decoder

$I^p_{\nu^{-1}}$ Mutual information between the bits and their a posteriori LLRs leaving the decoder

J Function relating the variance of the a priori LLRs and the mutual information between the bits and their a priori LLRs

j Imaginary unit

K Complexity of the IDD scheme

$K'_{\mu^{-1}}$ Complexity of the SD demapper at *each* IDD iteration

K' Complexity of the IDD scheme at *each* IDD iteration

$K'_{\nu^{-1}}$ Complexity of the SD decoder at *each* IDD iteration

$K''_{\nu^{-1}}$ Complexity of the SD decoder at *each* LDPC iteration and at *each* IDD iteration

L Length of the observation window of the estimators

$\Lambda^a_{\mu^{-1}}$ A priori LLRs entering the demapper

$\Lambda^a_{\nu^{-1}}$ A priori LLRs entering the decoder

$\Lambda^e_{\mu^{-1}}$ Extrinsic LLRs leaving the demapper

$\Lambda^e_{\nu^{-1}}$ Extrinsic LLRs leaving the decoder

λ_ℓ Fraction of edges in the graph of the LDPC code connected to a variable node of degree ℓ

$\Lambda^p_{\nu^{-1}}$ A posteriori LLRs leaving the decoder

L_{span} Length of a span

m Number of transmitted bits carried by a transmitted symbol or a transmitted transition

$M_{C_j \to V_i}$ Message sent from the check node C_j to the variable node V_i

μ_k k-th bit mapping

μ^{-1} Demapper

$M_{V_i \to C_j}$ Message sent from the variable node V_i to the check node C_j

N Random variable modelling the AWGN

N_0 Variance of the random variable modelling the AWGN

$N_{0,\text{ASE}}$ Variance of the random variable modelling the ASE AWGN

$N_{0,\text{NL}}$ Variance of the random variable modelling the non-linearities AWGN

$N_{0,\text{setup}}$ Variance of the random variable modelling the setup AWGN

N' Random variable modelling the counter-rotated AWGN

N_{bits} Number of bits

N_{cw} Number of code words

N_{e} Number of edges in the graph of the LDPC code

N_{hops} Number of hops in the decoding trajectory

N_{IDD} Number of IDD iterations

N_{LDPC} Number of LDPC iterations

N_{MC} Number of realisations for the MC estimation

N_{MSD} Number of received symbols used for MSD

N_{pol} Number of polarisations used for PDM

N_{spans} Number of spans

ν^{-1} Decoder

P_{s} Launch power of the signal

R Code rate

R_b Nominal bit rate

ρ_ℓ Fraction of edges in the graph of the LDPC code connected to a check node of degree ℓ

$\sigma^2_{\Lambda^a_{\mu^{-1}}}$ Variance of the random variable modelling the a priori LLRs entering the demapper

$\sigma^2_{\Lambda^a_{\nu^{-1}}}$ Variance of the random variable modelling the a priori LLRs entering the decoder

$\sigma^2_{\Lambda^e_{\nu^{-1}}}$ Variance of the random variable modelling the extrinsic LLRs leaving the decoder

$\sigma^2_{\Lambda^p_{\nu^{-1}}}$ Variance of the random variable modelling the a posteriori LLRs leaving the decoder

$\sigma^2_{\hat{I}}$ Variance of the mutual information estimator

σ^2_I Variance of the realisations of the mutual information

T Random variable modelling the transmitted transition

$\mathcal{T}$ Transition constellation

Θ Random variable modelling the (wrapped) PWGN

Θ_0 Variance of the random variable modelling the (wrapped) PWGN

$\Theta_{0,\mathrm{LASER}}$ Variance of the random variable modelling the LASER (wrapped) PWGN

$\Theta_{0,\mathrm{NL}}$ Variance of the random variable modelling the non-linearities (wrapped) PWGN

$\Theta_{0,\mathrm{setup}}$ Variance of the random variable modelling the setup (wrapped) PWGN

T_s Symbol period

V_i i-th variable node in the graph of the LDPC code

X Random variable modelling the transmitted symbol

$\mathcal{X}$ Constellation

Y Random variable modelling the received symbol

Y' Random variable modelling the counter-rotated and translated received symbol

List of Figures

List of Tables