Applications of Artificial Intelligence to Control and Analyze the Performance of Fiber-Optic Transmission Systems

Xiaoyan Ye

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAT048

Thèse de doctorat

Applications of Artificial Intelligence to Control and Analyze the Performance of Fiber-Optic Transmission Systems

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

École doctorale n°626 Ecole doctorale de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Réseaux, Informations et Communications

Thèse présentée et soutenue à Palaiseau, le 13/12/2023, par

XIAOYAN YE

Composition du Jury :

Prof. Catherine LEPERS Professeure, Télécom SudParis, Institut Polytechnique de Paris, France	Présidente/Examinatrice
Dr. Paolo SERENA Professeur associé, University of Parma, Italy	Rapporteur
Prof. Lilin YI Professeur, Shanghai Jiao Tong University, China	Rapporteur
Prof. Alan Pak Tao LAU Professeur, Hong-Kong Polytechnic University, China	Examinateur
Prof. Yves JAOUËN Professeur, Télécom Paris, Institut Polytechnique de Paris, France	Directeur de thèse
Dr. Élie AWWAD Maître de Conférences, Télécom Paris, Institut Polytechnique de Paris, France	Co-encadrant de thèse

Invité :

Dr. Amirhossein GHAZISAEIDI Chercheur, Nokia Bell Labs, France	Co-encadrant de thèse

To my sister ***Amandine***

for her unconditional support in all my decisions.

Acknowledgements

Embarking on this doctoral odyssey has been both an extraordinary intellectual adventure and a profound journey of personal growth. It is with immense gratitude that I acknowledge the remarkable individuals who have been guiding lights along this path.

First and foremost, I would like to extend my heartfelt appreciation to all the members of the jury who generously dedicated their time to review my thesis. Your expertise, detailed suggestions and insightful comments have been invaluable in refining this final document.

To Elie and Yves, my academic supervisors and directors, your patience, kindness, and support have been instrumental throughout this journey. Your guidance during challenging periods has been a source of strength, and I would like to express my deepest thanks to you for being so caring.

I am deeply grateful to Amir for the supervision of this thesis and for your mentorship. The unspoken understanding and your encouragement have been the propelling when the challenges seemed insurmountable. None of those achievements would have been possible be possible without your exceptional expertise and our fruitful meetings and mutual perseverance. I consider myself fortunate to have walked this research road under your guidance. I would like to express my gratitude to my colleagues of Nokia Bell Labs, specially to Haik and Sylvain for their unlimited support particularly during these challenging times during the pandemic. It is a true privilege to conduct research within this esteemed team. I reserve a special acknowledgment to my manager Jeremie for proposing to me an internship opportunity and then this CIFRE thesis contract and for kindly sharing your insights and advice. Beyond the professional collaborations, my journey has been enriched with valuable experiences with my esteemed colleagues. it has been a pleasure to share these moments with you, Isaia, Henrique, Guillaume, Samar, Daniel and many others—whether it is our coffee chats or our informal gatherings outside of work.

A profound thank you to my friends and family. Each encouraging word has had a profound impact on me. I am particularly grateful for the special support and attention received during the thesis writing period from individuals such as Amandine, Amy, Qiuyi, Nickson, Alix, Sterenn, Yang, Lina, Roya, Amber, Esther, my friends from Lycée Henri IV and and countless others who have played invaluable roles in my journey.

A special note of appreciation goes to my friends Lei, Zhongzheng, Ruiqiao and Xiaohui whose camaraderie and mutual support as fellow Ph.D. students have been invaluable.

I would like to express my sincere gratitude to my former colleagues, Aymeric, Dylan and Kaoutar who played a significant role in shaping my decision to pursue a Ph.D. in this domain and within this team and have supported me during internships and the early stages of my thesis. I would like also to express my gratitude to Gabriel who kindly has accepted my application for a one-month internship which marked the beginning of my journey in the field of optical communication.

A heartfelt acknowledgment also goes to Nokia Bell Labs, Télécom Paris and Institut Polytechnique de Paris for providing a warm and stimulating research environment.

以梦为马，以风为缰，不忘初心，不负韶华

Résumé en Français

Au cours des dernières décennies, les demandes de trafic internet n'ont cessé d'augmenter. Aujourd'hui, les nouveaux usages comme le cloud computing, les jeux en ligne, le streaming vidéo, l'IoT, etc. nécessitent une augmentation incessante de la capacité. Le nombre total des usagers mobiles va croitre de 5.1 en 2018 à 5.7 milliards d'ici la fin de 2023.

Les systèmes de communication par fibre optique constituent les fondements des systèmes de communication mondiaux actuels, des les systèmes sous-marins transocéaniques à longue distance aux réseaux terrestres. Pour répondre aux besoins d'une plus grande capacité, ces systèmes de communication optique ont bien évolué, de l'utilisation d'amplificateurs optiques et de la détection cohérente à la mise en œuvre d'algorithmes avancés de traitement des signaux numériques (DSP). Pour accroître toujours plus la capacité d'câble de fibres, qui est le produit entre le nombre de fibres par câble, l'efficacité spectrale (exprimée en bit/s/Hz) et la largeur de bande du système (exprimée en Hz), la recherche sur les schémas à bande ultra-large (UWB) et la conception de réseaux à faible marge sont fondamentaux.

Diverses solutions ont été proposées pour réaliser l'amplification UWB, certaines d'entre elles avec des configurations hybrides utilisant des amplificateurs Raman et des amplificateurs à fibre dopée à l'erbium (EDFA) ou des amplificateurs optiques à semi-conducteurs (SOA) pour leur flexibilité, leur grande largeur de bande et la réduction du bruit d'émission spontanée amplifiée (ASE) et des distorsions non linéaires. Étant donné que l'adaptation du profil de gain spectral des amplificateurs optiques est essentielle pour les schémas de transmission UWB, plusieurs méthodes ont été proposées pour prédire de manière efficace le gain Raman dans ces schémas d'amplification ou pour trouver la configuration optimale des pompes Raman. Bien que la physique de la diffusion Raman soit bien connue et modélisée, donc le gain peut être prédit en résolvant un système d'équations différentielles ordinaires non linéaires régissant l'évolution du profil de puissance au cours de la propagation, mais malheureusement cette méthode prend beaucoup de temps et sa complexité augmente avec l'extension de la largeur de bande du spectre et le nombre de pompes. De plus, la tâche de conception inverse, c'est-à-dire la recherche de la configuration optimale des pompes, peut être encore plus complexe et plus longue avec ces méthodes numériques ou itératives.

D'autre part, l'apprentissage automatique (ML) s'est révélé très prometteur pour de nombreux secteurs et a été largement utilisé dans la communauté de la recherche scientifique, notamment dans le domaine des communica-

tions optiques, en particulier lorsqu'il n'y a pas de réponse directe et simple dérivée des lois physiques. La ML a démontré sa capacité à réduire la complexité des calculs et/ou à augmenter la précision d'une méthode classique.

Un autre défi à relever est la conception de systèmes à faible marge. Pour tirer pleinement parti de la capacité d'un réseau optique, il est essentiel de disposer d'un outil d'estimation de la qualité de transmissionn (QoT) fiable et précis. En outre, dans le contexte des réseaux optiques flexibles, le système doit réagir en temps réel à tout changement. En effet, l'établissement ou la reconfiguration d'une connexion nécessite une prédiction précise et rapide. Cela implique la bonne connaissance et la bonne maîtrise de tous les types de bruit présents dans le lien de transmission. En effet, la performance des systèmes de transmission par fibre optique est sévèrement affectée par l'atténuation de la fibre, le bruit ASE, la dispersion chromatique (CD), la dispersion des modes de polarisation (PMD), le bruit de phase du laser (PN) et les effets d'interférence non linéaire (NLI) due à l'effet Kerr. Grâce au DSP sophistié déjà mis en oeuvre, les dégradations linéaires peuvent être bien compensées, mais la mise en œuvre de l'algorithme de compensation des effets non linéaires reste un défi. Il est donc fondamental de fournir rapidement une estimation fiable de l'impact des dégradations non linéaires et des autres pénalités résiduelles.

Parmi toutes les pénalités qui contribuent au rapport signal-bruit (RSB) du signal transmis, le bruit ASE et le bruit de l'émetteur-récepteur (TRX) sont faciles à caractériser, mais l'estimation de l'impact du bruit non linéaire et du bruit de phase renforcé électroniquement (EEPN) est plus délicate.

En effet, dans nos réseaux complexes actuels, une estimation correcte de la variance du NLI pour les systèmes WDM longue distance nécessite des simulations informatiques importantes avec la méthode de Fourier à pas divisé (SSFM) et un grand nombre de paramètres. Le nombre de paramètres augmente avec le nombre de canaux WDM et le nombre de tronçons de fibres.

Par conséquent, des modèles analytiques ont été proposés et le temps de calcul a été considérablement réduit tout en atteignant une précision notable. Toutefois, le temps de calcul reste malheureusement trop long pour des applications en temps réel, car les modèles analytiques sont basés sur le calcul des intégrales multidimensionnels et des modèles approchés plus rapides induisent un manque de généralisation et de précision.

En ce qui concerne l'estimation du bruit EEPN, qui augmente avec la largeur de raie du laser, le débit symbol et la dispersion accumulée (qui augmente donc avec la distance), il existe un modèle analytique, mais il semble trop idéal et ne tient pas compte de l'impact du DSP. Des méthodes expérimentales ont été proposées, mais elles nécessitent une mesure fine du bruit laser avec un équipement ou un montage avancés et une caractérisation des impacts du TRX.

La motivation de ce travail est donc d'explorer différentes méthodes basées sur les données, y compris des modèles ML, pour améliorer la performance du système de transmission optique en répondant au problème complexe de conception des amplificateurs UWB ou pour fournir un outil d'estimation de performance précis pour l'estimation de la variance du NLI et de la variance de l'EEPN.

Plan du manuscrit et mes contributions

Durant cette thèse, nous avons travailler sur de nouvelles solutions pour relever les défis posés par les systèmes de transmission UWB et à grand débit sur de longues distances. Nous nous concentrons sur les méthodes d'estimation pour deux dégradations complexes dans les systèmes de transmission optique ainsi que sur un outil efficace pour aider à la conception d'amplificateurs UWB.

Le manuscrit est organisé comme suit.

Dans le deuxième chapitre, nous passons en revue les notions fondamentales des systèmes de communication optique et une brève introduction aux techniques de ML axées sur les réseaux neuronaux artificiels (ANN) et nous discutons de l'impact de la qualité de l'ensemble des données d'apprentissage sur les performances des cadres basés sur les réseaux neuronaux artificiels (ANN). Nous nous concentrons sur la description et la modélisation des effets linéaires et non linéaires survenant dans les canaux de communication optique et présentons l'émetteur et le récepteur cohérents à double polarization ainsi que les algorithmes DSP pour compenser les distorsions du signal dues à sa propagation à travers la fibre optique.

Le troisième chapitre présente un modèle assisté par ML pour les amplificateurs Raman. En utilisant des données expérimentales, deux modèles sont formés : un modèle génératif pour prédire le profil de perte global (y compris le profil de gain de l'amplificateur, l'atténuation et les effets de diffusion Raman stimulée à l'intérieur du canal) et un modèle inverse pour prédire la configuration de pompe appropriée compte tenu d'un profil de perte désiré. Nous démontrons les capacités de la solution basée sur l'ANN pour résoudre le problème complexe de conception inverse des amplificateurs Raman hybrides UWB.

Une nouvelle méthode d'estimation de la qualité de transmission basée sur la ML fournissant une estimation ultra-rapide et très précise de la variance NLI dans les systèmes de multiplexage par répartition en longueur d'onde (WDM), appelée KerrNet, est décrite en détail dans le quatrième chapitre. L'outil KerrNet proposé est évalué par des simulations pour des liaisons homogènes et hétérogènes arbitraires dans le cadre général de schémas de chargement WDM aléatoires, de débits de symboles, de formats de modulation et de profils de gain-perte résiduels arbitraires par portée, et ses capacités sont démontrées.

Le cinquième chapitre contient une description de la nouvelle méthode de caractérisation proposée pour un autre bruit supplémentaire, résultant du DSP et du bruit de phase du laser : l'EEPN, qui ne peut plus être négligé pour les systèmes de transmission longue distance à taux de symbole élevé. Cette solution est basée sur une méthode phénoménologique avec des données expérimentales et une estimation des paramètres et ne nécessite pas une connaissance a priori des caractéristiques du bruit de phase du laser ou une caractérisation du bruit de l'émetteur-récepteur avec des mesures "back-to-back".

La dernière partie de ce manuscrit présente les conclusions de ce travail et les perspectives futures.

Contents

Glossary

ADC analog-to-digital converter

ANN artificial neural networks

ASE amplified spontaneous emission

AWGN additive white Gaussian noise

B2B back-to-back

BER bit-error rate

BPS blind phase search

CD chromatic dispersion

CDF cumulative density function

CMA constant modulus algorithm

COI channel of interest

DAC digital-to-analog converter

DCF dispersion compensation fiber

Demux demultiplexer

DGD differential group delay

DM dispersion-managed

DSP digital signal processing

DU dispersion un-managed

EDFA erbium-doped fiber amplifiers

EEPN electronically enhanced phase noise

EGN model enhanced Gaussian noise model

FANN feed-forward artificial neural network

FANN feed-forward artificial neuron network

FEC forward error correction

FFT fast Fourier transforms

GAWBS guided acoustic wave Brillouin scattering

GFF gain flattering filter

GN model Gaussian noise model

GT ground truth

GVD group velocity dispersion

I-SRS intra-channel stimulated Raman scattering

IQ-modulator in-phase and quadrature modulator

ISI inter-symbol interference

LO local oscillator

LR learning rate

MAE mean absolute error

MC integration Monte Carlo integration

MCF multi-core fibers

MIMO multiple-input multiple-output

ML machine learning

MLP multi-layer perceptron

MMF multi-mode fiber

MSE mean square error

Mux multiplexer

MZM Mach-Zehnder modulator

NF noise figure

NLI nonlinear interference

NLSE nonlinear Schrödinger equation

NLT nonlinear threshold

ODE ordinary differential equations

OOK on-off keying

optical signal noise ratio OSNR

PBC polarization beam coupler

PCS probabilistic constellation shaping

PDF probability density function

PDL polarization-dependent loss

PDM polarization division multiplexing

PMD polarization mode dispersion

PN phase noise

QAM quadrature amplitude modulation

QoT quality of transmission

QPSK phase shift keying

RC raised-cosine

ReLU rectified linear unit

RF radio-frequency

RMSE root mean square error

ROADM reconfigurable optical add-drop multiplexers

RP regular perturbation

RRC root-raised-cosine

SCI self-channel interference

SDM space division multiplexed

SMF single mode fiber

SNR signal-to-noise ratio

SOA semiconductor optical amplifiers

SpS samples per symbol

SRS stimulated Raman scattering

SSFM split step Fourier method

SSMF standard single mode fiber

tanh hyperbolic tangent

TD-RP time-domain regular perturbation

TRX transceiver

UWB ultra-wideband

WDM wavelength-division multiplexing

XCI cross-channel interference

List of figures

Introduction

In the past decades, the internet traffic demands kept increasing. Nowadays, new usages like cloud computing, online gaming, video streaming, IoT, etc., require a constant increase of the capacity. The total number of global mobile subscribers will grow from 5.1 billion in 2018 to 5.7 billion by 2023 [1].

Optical fiber communications constitute the backbone of the current worldwide communication systems, from long-haul transoceanic submarine systems to terrestrial networks. Optical communication systems have evolved to meet the needs for higher capacity, from the use of optical amplifiers and coherent detection to the implementation of advanced digital signal processing (DSP) algorithms. To further increase the capacity of an optical fiber cable, which is the product between the number of fibers per cable, the spectral efficiency of the transmitted modulation format (expressed in bit/s/Hz) and the system bandwidth (expressed in Hz), the works on ultra-wideband (UWB) schemes [2, 3, 4, 5] are fundamental. Moreover, to fully exploit the capacity of existing and future networks, low margin network design tools are indispensable.

Various solutions have been proposed to achieve UWB amplification, some of them employ hybrid configurations, using Raman amplifiers with erbium-doped fiber amplifiers (EDFA) or with semiconductor optical amplifiers (SOA), for their flexibility, large bandwidth, reduced amplified spontaneous emission (ASE) noise and little nonlinear distortions. Since tailoring the spectral gain profile of optical amplifiers is essential for UWB transmission schemes, several methods have been proposed to correctly predict the Raman amplifier gain profile in these amplification schemes or to find the optimal pumps configuration to achieve a desired target gain profile. Although the physics of the Raman scattering effects is well-known and modelled, so the Raman amplifier gain can be predicted by solving a system of nonlinear ordinary differential equations governing the power evolution of the signal and the pumps during the propagation [6], but unfortunately this numerical solver method is time-consuming, with an increase of complexity with the extension of the spectrum bandwidth and the total number of pump lasers. Moreover, the inverse design task, i.e. finding the optimal configuration of the pumps (frequency and power) for a gain profile we want to achieve, can be even more complex and time-consuming with classical numerical methods or iterative methods.

On the other hand, machine learning (ML) is revealed to be of great potential for many sectors, and has been used widely in the scientific research community including optical communication [7, 8, 9], in particular when there is no direct simple answer derived from the laws of physics. ML has also shown its ability to reduce the computation

complexity and/or increase the accuracy of a classical method.

Another challenge to face is the design of low-margin systems, to completely take advantage of the capacity of an optical network. Thus, a reliable and accurate quality of transmission (QoT) tool is fundamental. Besides, in the context of flexible optical networks, the system needs to react in real–time to any change. Indeed, the establishment or reconfiguration of a connection requires accurate and ultra-fast prediction. That implies the good knowledge and the good mastering of all types of noise in the transmission link. Indeed, the performance optical fiber transmission systems is severely affected by fiber attenuation, ASE noise, chromatic dispersion (CD), polarization mode dispersion (PMD), laser phase noise (PN) and nonlinear interference (NLI) effects. Thanks to the already-implemented advanced DSP, the linear impairments can be well compensated, however compensation algorithms for non-linear effects impose more challenges from an implementation point of view. Thus, it is fundamental to provide a highly-reliable and ultra-fast estimation of the impact of nonlinear impairments and other residual penalties on the QoT. Among all penalties that contribute to the signal-to-noise ratio (SNR) of the transmitted signal, ASE noise and transceiver (TRX) noise are easy to be characterized, but the estimation of the impacts of NLI and electronically enhanced phase noise (EEPN) is more delicate.

Indeed, in current complex optical networks, the estimation of NLI effects for long-haul wavelength-division multiplexing (WDM) systems involves a large amount of parameters and the number of parameters increase with the number of WDM channels and number of fiber spans and a proper estimation of NLI variance requires high computational simulations with the split step Fourier method (SSFM). Consequently, analytical models were proposed [10, 11, 12, 13] and the computation time has been tremendously reduced while achieving a notable accuracy. However, unfortunately, the computation time with exact analytical models is still too long for real-time applications since the computation requires operations with multi-dimensional integrals, while faster closed-form models are less accurate and rarely generalizable.

As for the estimation of the EEPN noise, which increases with the laser linewidth, the symbol rate and the accumulated dispersion (so it increases with transmission distance), an analytical model exists [14] but lacks accuracy and it does not take into account the impact of DSP. Experimental methods have been proposed, yet they require an fine measurement of the laser linewidth with advanced equipment or a built-in setup and a fine TRX characterization.

The motivation of this work is therefore to explore different ML-based or data-driven methods to provide an accurate performance estimation tool for NLI variance and EEPN variance or to enhance the performance of optical transmission systems by solving the complex design problem of UWB amplifiers.

Thesis outline and contributions

This PhD thesis was carried out through an industrial grant (CIFRE N^{o} 2020/0921) between Nokia Bell Labs and Télécom Paris.

In this thesis, we investigate new approaches to meet challenges that arise with UWB and high-speed long-haul transmission systems. We focus on estimation methods for two complex impairments in optical transmission systems as well as an efficient tool to assist the design of UWB amplification.

The thesis is organized as follows.

In Chapter 1, we first review fundamental notions of optical communication systems. Then we provide a brief introduction to ML techniques focused on artificial neural networks (ANN) and we discuss impact of the training dataset quality on ANN-based frameworks' performance. We focus on the description and modelling of the linear and nonlinear effects arising in optical communication and present the dual-polarization coherent transmitter and receiver as well as the DSP algorithms to compensate the signal distortions due to its propagation through the optical fiber.

In Chapter 2, two ML-aided models for Raman amplifier are presented. Using experimental data, two models are trained : a generative model to predict the overall loss profile (including amplifier gain profile, attenuation and intra-channel stimulated Raman scattering effects) and an inverse model to predict the appropriate pump configuration given a desired loss profile. We demonstrate the capabilities of the ANN-based solution to solve the complex inverse design problem of UWB hybrid Raman amplifiers.

A new ML-based QoT method providing an ultra-fast and highly accurate estimation of the NLI variance in WDM systems, called *KerrNet* is detailed in Chapter 3. The proposed KerrNet tool is evaluated through simulations for both homogeneous and arbitrary heterogeneous links in the general setting of random WDM loading schemes, symbol rates, modulation formats and arbitrary per-span residual gain/loss profiles and demonstrated its high accuracy and high speed capabilities.

Chapter 4 contains description of the proposed new characterization method for another additional noise, resulting from the DSP and the laser's phase noise : the EEPN which cannot be neglected anymore for long-haul transmission systems at high symbol rate. This solution is based on a phenomenological method with experimental data and data-driven models' parameter estimation and does not require a priori knowledge of the laser phase noise characteristics or a characterisation of the transceiver noise with back-to-back measurements.

The final part of this thesis presents the conclusions of this work and future perspectives.

Chapter 1

State of the art of modern fiber optical communication systems

In this chapter, we first present the basic concepts of a coherent optical fiber system in Section 1.1 to provide the context of our research work. We review the modelling of propagation effects in fiber optics and different impairments, as well as classical DSP methods to mitigate the signal distortions, and finally different noise contributions in the performance evaluation using the expression of the signal-to-noise ratio (SNR).

Then, in Section 1.2 we introduce fundamental notions of ML algorithm with a particular attention on feed-forward ANN (FANN) architectures and the importance of well-prepared dataset.

1.1 Optical fiber coherent communication systems

1.1.1 Snapshot of current systems

In digital communications, information is coded at the transmitter by a sequence of bits (sequence of binary digits composed of "0" and "1"). Then the coded information pass through a physical channel (free-space, fiber optics, copper wires, etc.) and the goal is to be able to recover the information at the receiver.

An overview of the digital communication system through a fiber optic channel is presented in Fig. 1.1. The digital source of information is coded with a source encoder in a sequence of bits then transformed into a sequence of symbols $\mathbf{X} = \{x_1, x_2, ..., x_N\}$ with a coded modulation encoder including a forward error correction (FEC) to add redundancy for a more robust and reliable transmission. Then the modulation mapper transforms the bits into a constellation (also called modulation format). The coded symbols $\mathbf{X}$ with the chosen constellation are transmitted via the optical communication channel. First, a digital signal processing (DSP) step is performed (including pulse-shaping, digital compensation, etc.) to create an appropriate signal waveform for achieving a better transmission

result. Then using a high-resolution digital-to-analog converter (DAC), the symbols $\mathbf{X}$ are converted to a radio-frequency (RF) signal that will be used to generate the optical signal leveraging an electro-optical modulator and a source laser. The resulting optical signal is sent to the fiber and at the output of the optical fiber transmission link, the received optical signal is converted into a RF signal with an optical coherent receiver, then converted into a digital signal with an analog-to-digital converter (ADC). The coherent receiver enables advanced modulation format such as quadrature amplitude modulation (QAM) format which exploits both amplitude and phase domains of optical signal to encode the information data. The DSP techniques (cf subsection 1.1.7 paragraph ii) are applied to compensate different impairments due to the propagation inside the optical fiber to reconstruct the received symbols $\mathbf{Y}$. The demapper, that constitutes the first part of the coded modulation decoder, recovers the bits information from the constellation of the recovered received symbols sequence. Then the FEC decoder corrects the errors using the redundancy relationships with the bits to achieve a better performance. Finally, the recovered bits are decompressed by source decoding (inverse of source encoding) and send to the final destination.

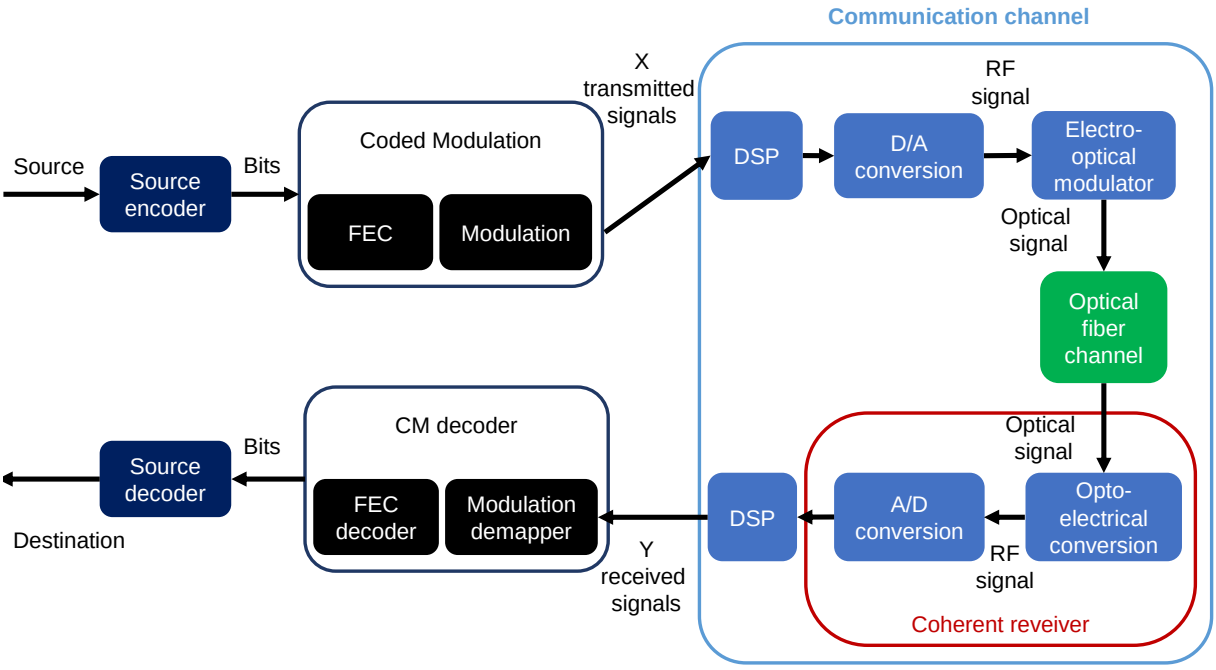


FIGURE 1.1 – Diagram of a coherent optical fiber communication systems

1.1.2 Multiplexing

To enhance the transmission capacity, multiple multiplexing methods can be used. A brief description of three commonly used multiplexing techniques is presented.

i Wavelength division multiplexing

A popular multiplexing technique is wavelength division multiplexing (WDM), developed since 1992. WDM scheme can transmit in parallel N channels in the same fiber, one per optical wavelength. First, each source information is coded as described in the section above (see 1.1.1) with a laser at carrier wavelength λ_{TX} and a dedicated transmitter. Then all the modulated optical signal transmitting at different wavelength are combined together with a multiplexer (Mux) before sending to the optical link. At the end of the transmission link, the optical signal is divided into N signals at different carrier wavelength λ_n with a demultiplexer (Demux) then detected with a receiver separately. A generic representation of a WDM point-to point long haul optical communication scheme, with a homogeneous multi-span link configuration,commonly used for submarine links, is shown in Fig. 1.2.

The WDM system combined with multiple reconfigurable optical add-drop multiplexers (ROADM) are key components of our currently deployed meshed networks. In Fig. 1.3, an illustration of a mesh network is represented and it shows the flexibility of using such a system. Indeed, at each nodes, where ROADMs are placed, it is possible to add or to drop one or several channels that either going to a different destination or coming from a distinct destination.

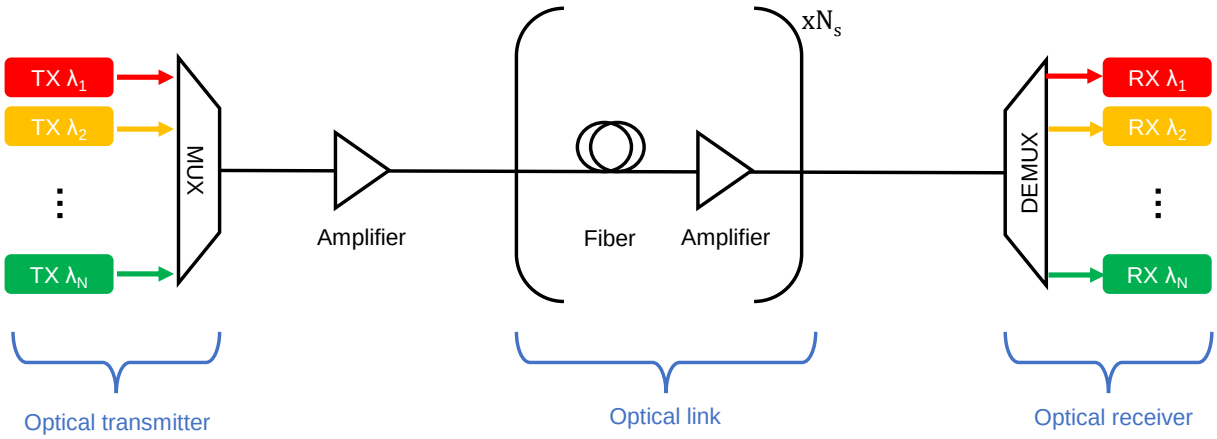


FIGURE 1.2 – Representation of a WDM point-to-point long haul optical communication scheme with a homogeneous link

ii Ultra-wide band transmission

Conventionally, WDM optical transmission system was first performed in the C-band because of the emergence of EDFAs within this bandwidth and the low power attenuation in this optical band. But over years, with the saturation of the capacity of C-band systems, research projects to extend the transmission bandwidth by transmitting altogether with other bands, such as S+C band, C+L band, S+C+L band, etc., are have extensively conducted because an UWB system could provide a much larger available transmission bandwidth, thus achieving a higher overall capacity [15, 16].

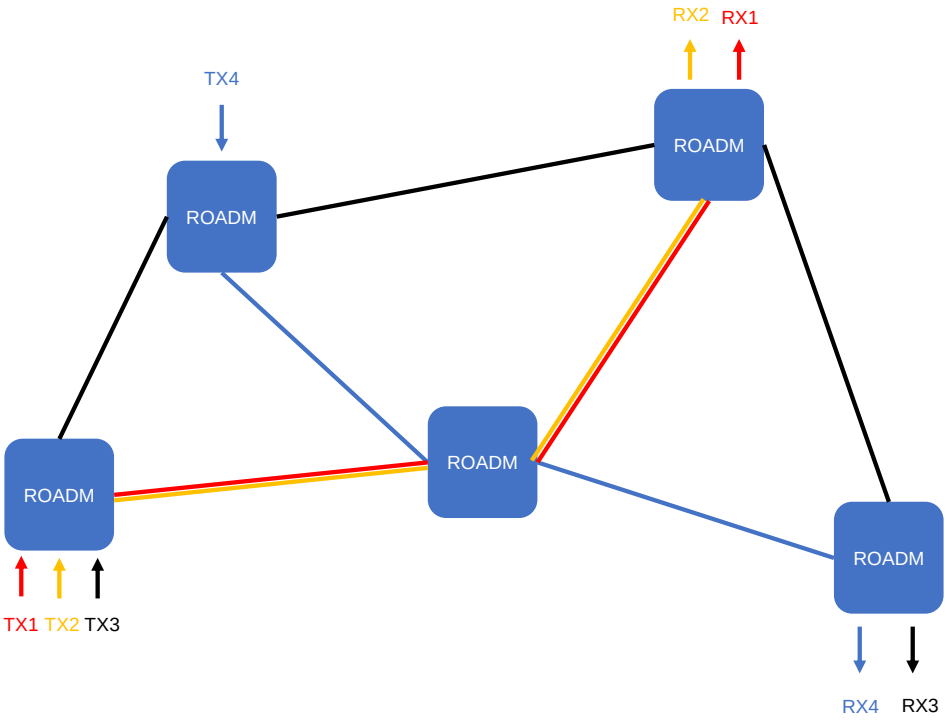


FIGURE 1.3 – Representation of a meshed network

iii Polarization division multiplexing

Polarization is another interesting property of the optical signal that can be used to increase the capacity by a factor of two by using a polarization division multiplexing (PDM) scheme. It consists in using a dual-polarization system with two orthogonal polarizations to transmit the information as depicted in Fig. 1.4 where a single polarization optical signal and a dual orthogonal polarization signal are represented.

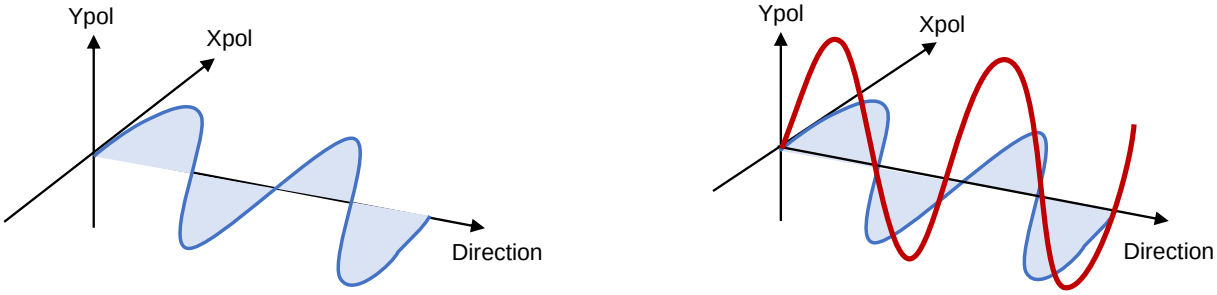


FIGURE 1.4 – Representation of (a) a single polarization and (b) a dual-polarization signal

iv Space division multiplexing

Using a coherent system (with I and Q components of the optical signal field) with two orthogonal polarizations and multiplexed by the frequency in a ultra-wide band configuration can achieve a very-high data rate in single

mode fiber (SMF). Fig. 1.5 illustrates the schematic representation of an SMF and its index distribution. Another solution that has been intensively investigated since 2010 to further increase the capacity is to take the advantage of the spatial dimension in space division multiplexed (SDM) systems. In this configuration, data streams can be transmitted over different transmission paths of the same fiber. Moreover WDM and SDM can be combined to achieve a remarkable increase in system capacity : in such case, multiplexing of N wavelengths and M parallel transmission paths provides $N \times M$ independent data channels.

First attempts to achieve SDM scheme were by packaging several SMFs together to create a fiber bundle as illustrated in Fig. 1.6 (a). Fiber bundles were already adopted in our current optical infrastructure but due to its larger dimension, other solutions with a higher space efficiency have been studied.

Another idea is to build fibers using a different physical architecture. Several single mode cores are packed inside the fiber coating to make multi core fibers (MCF) so the spatial multiplexing can be achieved over different cores (cf Fig. 1.6). The cores inside MCF can be either coupled (coupling between co-propagating signals) or uncoupled. A major drawback of MCF is the inter-core cross-talk causing interference between co-propagating signals of different adjacent cores so many ongoing research projects are conducted to minimize this cross-talk.

Multi-mode fiber (MMF) can also be very efficient in short reach systems to design SDM scheme by exploiting the transverse modes of an optical source transmitting at a given frequency and polarization. To enable multi modes propagation of the optical signal, MMF's core dimension is much higher than SMF (Fig. 1.6 (c) and Fig. 1.5). In long distance transmission, the main obstacles are the modal dispersion, modal interference and high differential mode group delay of co-propagation modes which leads to the implementation of computational heavy DSP algorithms.

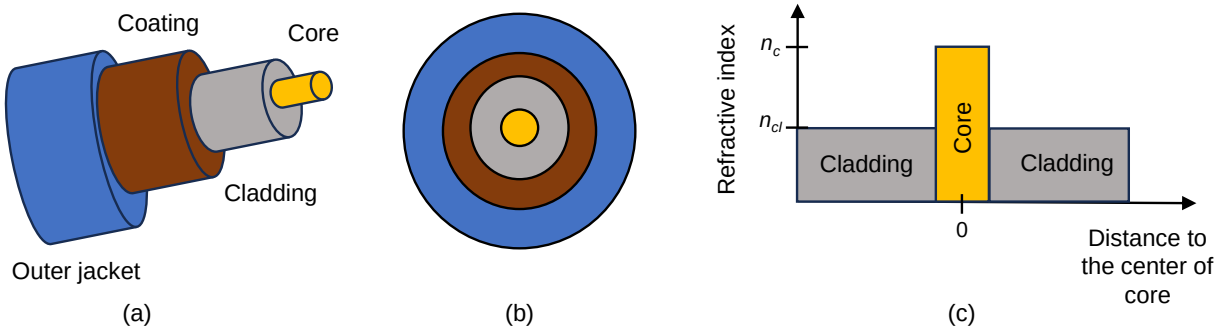


FIGURE 1.5 – Single-mode fiber (SMF) schematic representation and index distribution in the cladding/core

v Summary

In summary, many different multiplexing solutions can be implemented in optical communication systems to increase the overall capacity and meet the increasing data streaming demands. In our work, we mainly used WDM systems combined with PDM scheme. Solutions to optimize the performance of ultra wide band systems have also been investigated but SDM schemes are not considered in this thesis.

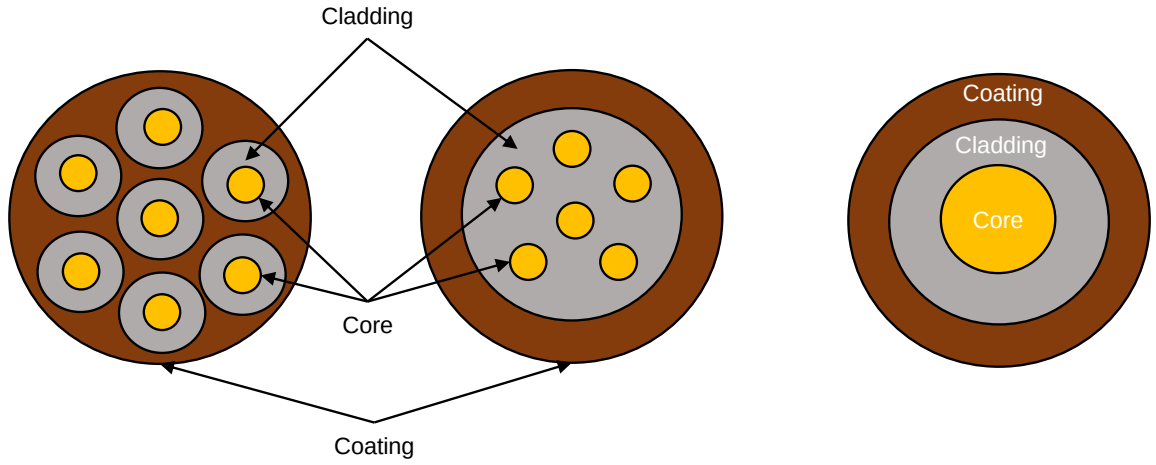


FIGURE 1.6 – SDM fiber section

1.1.3 Transmitter

The transmitter is used to convert source information into an analog signal which can propagate through the physical channel. This sub-section presents an overview of the modulation format, pulse-shaping DSP technique and the global structure of a coherent transmitter.

i Modulation format

In digital communications, bits of coded information are converted with a modulation mapper into a sequence of symbols before transmitting through the channel. We plotted the mapping between bits and symbols for different modulation format in Fig. 1.7.

The on-off keying (OOK) format was initially implemented, as direct-detection had been considered, before replace by more advanced modulation format. This two-level modulation format assigns the bit "0" to the lower level of the signal power and bit "1" to the highest level. Then coherent detection combined with DSP at receiver has considered since 2006. It allows sensitivities for both amplitude and phase. Thus, complex-valued constellations are commercially used for coherent systems by modulating the optical signal in both amplitude and phase domains over the I and Q quadratures. For instance, multi-level square 2^m quadrature QAM including phase shift keying (QPSK) are predominantly used.

As shown in Fig. 1.7, Gray mapping ensure that neighbouring symbols of a constellation only differs by one bit. This mapping method allows a more robust configuration for transmitting over a noisy channel. Indeed, DSP to recover the transmitted symbols have more difficulty to estimate correctly a constellation from the neighbouring constellation leading to transmission errors. Therefore Gray mapping helps reducing the bit-error rate (BER) which is the number of incorrect bits received over the total number of transmitted bits for a given time.

The multi-level QAM constellation can be shaped or un-shaped. For a Gaussian channel, a Gaussian-distributed

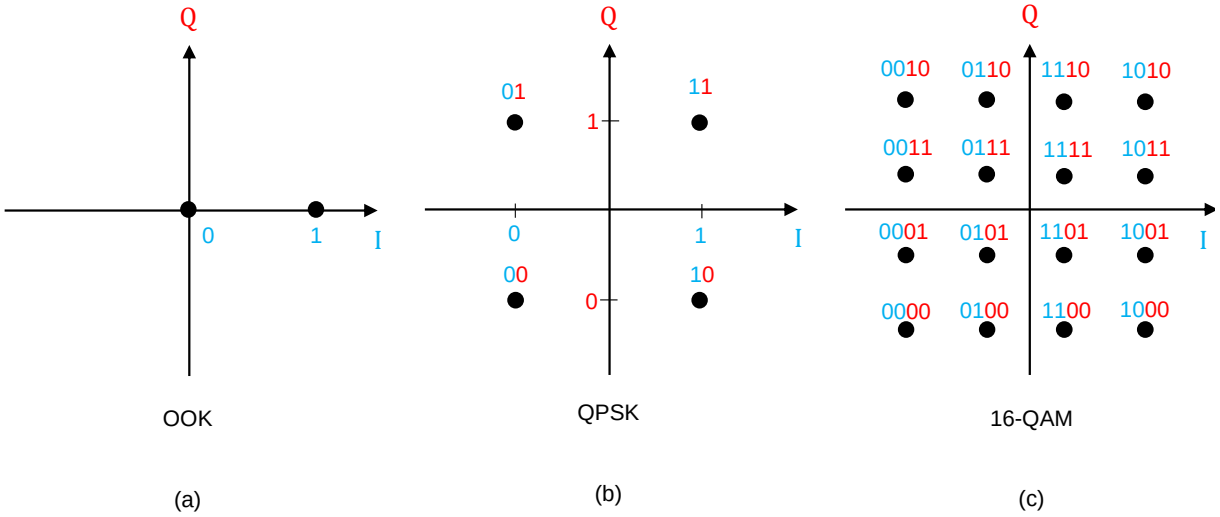


FIGURE 1.7 – Gray mapping between bits and symbols for three different modulation formats such as OOK, QPSK ($m = 2$) and 16-QAM ($m = 4$).

input maximize the mutual information thus probabilistic constellation shaping (PCS) was invented because it resembles more to a Gaussian-like distribution than the regular QAM constellation. For PCS-QAM, which is a popular modulation format, a distribution matcher is employed to generate the constellation points with nonuniform distribution $p_X(a)$ approximating the Gaussian distribution, according to a discrete Maxwell-Boltzmann distribution :

$$p_X(a) = \frac{\exp\left(-\nu|a|^2\right)}{\sum_{a\in\mathbf{A}} \exp\left(-\nu|a|^2\right)} \tag{1.1}$$

where $\mathbf{A}$ is the constellation alphabet, a a complex symbol $\in \mathbf{A}$ and $\nu \leq 0$ is a free parameter that we can tune to change the distribution.

ii Pulse shaping

After the creation of complex symbols $\mathbf{X} = \{x_1, x_2, ..., x_k, ..., x_n\}$ according to a desired modulation format, we have to convert the sequence into a signal waveform $s(t)$ suitable for transmission. The signal waveform can be written as :

$$s(t) = \sum_k x_k \psi_k(t) \tag{1.2}$$

where x_k are discrete symbols supposed to be transmitted and $\psi_k(t)$ are the basis waveforms.

To limit cross-talk in a WDM system, the Fourier transform of $\psi_k(t)$ in the frequency domain ($H_{ps}(f) = \mathbf{F}\{\psi_k(t)\}$) must have a limited bandwidth.

In addition to the cross-talk concern, inter-symbol interference (ISI) also need to be avoid. We can re-write the

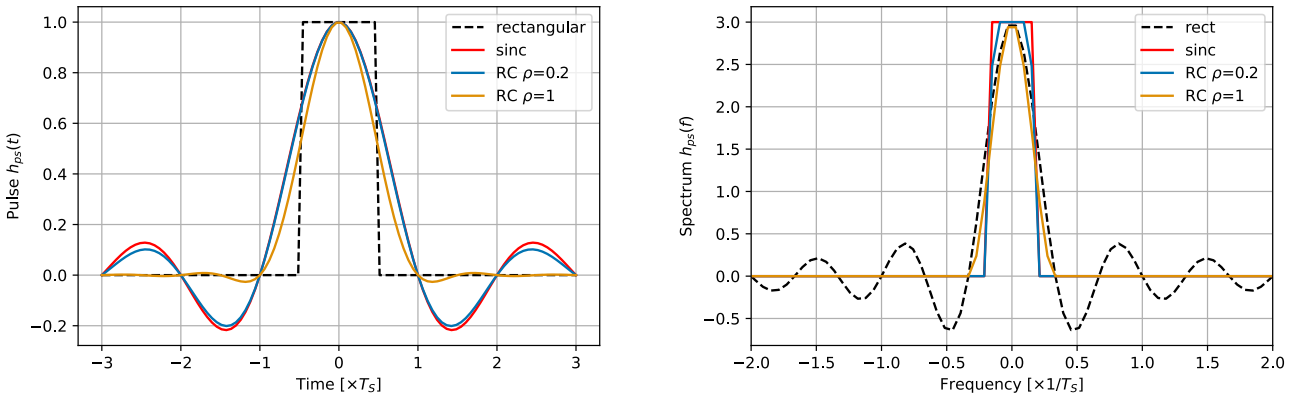


FIGURE 1.8 – Pulse shaping function in time and frequency domain including rectangular, sinc and RC function with $\rho = 1$ and $\rho = 0.2$

pulse-shaped signal waveform represented by the following expression :

$$s(t) = \sum_k x_k h_{ps}(t - kT_s) \tag{1.3}$$

with T_s the symbol duration time and the pulse shaping function.

To avoid ISI, the overall filter response $h_{filters}$ that includes the pulse-shaping function and the additional filter at the receiver should satisfy the following condition :

$$h_{filters}(t - kT_s) = \begin{cases} 0, & \text{if } t = k'T_s \text{ and } k' \neq k \\ 1 & \text{otherwise} \end{cases} \tag{1.4}$$

Under this condition, the pulse at the receiver after the filter is 0 during all sample times, except when sampling the k^{th} symbols.

Several candidate pulse shaping functions are shown in Fig. 1.8. Using a rect pulse is the simplest way to generate the waveform from symbols and it satisfies the first requirement (of Eq. (1.4)) limiting the ISI effect. Nevertheless, we can clearly see in Fig. 1.8 and in its expressions that rect pulse has an infinite bandwidth in the frequency domain. On the other hand, sinc pulse can achieve a limited bandwidth $W = 1/Ts$ avoiding cross-talk. But the disadvantage is that the pulse is infinite leading to important ISI over a larger number of symbols.

For those reasons, in practical, the root raised-cosine (RRC) pulse with a small roll-off factor ρ near 0 is predominantly used in optical communication. The RRC transfer function H_{rrc} is defined as the root-square of raised-cosine (RC) transfer function H_{rc} in the frequency domain and the RRC transfer function in the time-domain $h_{rrc}(t)$ is derived from $H_{rrc}(f)$.

The expression of the RC transfer function in the time domain h_{rc} and in the frequency domain H_{rc} are described

by :

$$h_{rc}(t) = \begin{cases} \frac{\pi}{4}\,\mathrm{sinc}\left(\frac{1}{2\rho}\right), & \text{if } t = \pm\frac{T_s}{2\rho} \\ \mathrm{sinc}\left(\frac{t}{T_s}\right)\frac{\cos\left(\pi\rho t/T_s\right)}{1-4\rho^2t^2/T_s^2}, & \text{otherwise} \end{cases} \tag{1.5}$$

$$H_{rc}(f) = \begin{cases} T_s, & \text{if } |f| \leq \frac{1-\rho}{2T_s} \\ \frac{T_s}{2}\left(1+\cos\left(\frac{\pi T_s}{\rho}\left(|f|-\frac{1-\rho}{2T_s}\right)\right)\right), & \text{if } \frac{1-\rho}{2T_s} < |f| \leq \frac{1+\rho}{2T_s} \\ 0, & \text{otherwise} \end{cases} \tag{1.6}$$

with h_{rc} satisfying the condition in Eq. (1.4) and the roll-off factor $\rho > 0$. First, this function decreases faster than sinc thus reduces the ISI. Second, it has a limited bandwidth $W = (1+\rho)/T_s$ that one can control by adjusting the roll-off factor ρ. In Fig. 1.8 (b), we draw the frequency response of RC pulse for two different roll-off factors.

At first glance, RRC pulses do not satisfy Eq. (1.4) so RRC will produce ISI. However, assuming the signal is transmitted through a channel with additive white Gaussian noise (AWGN), the matched filter is the optimal filter to obtain the highest SNR. The matched filter is the complex conjugate time reversal of the filter, so $h_{ps}^*(-t)$ for us. In this case, RRC filter is its own match filter because of its symmetry. So the configuration where we place a RRC filter at the transmitter and the same RRC filter at the receiver side, not only optimizes the SNR in presence of noise but also creates an overall RC waveform satisfying the two previous requirements : avoiding both ISI and cross-talk effects

iii Coherent transmitter

A representation of a coherent dual polarization transmitter, that generates two signals waveform $x^H(t)$ and $x^V(t)$ for two orthogonal polarizations (H for horizontal and V for vertical) from a sequence of complex symbols x_k^H and x_k^V is presented in Fig. 1.9. For each orthogonal polarization, the real and imaginary components (I and Q components) of their waveforms (after pulse-shaping filter) are extracted and converted to an electrical analog signal by using DACs. The four resulting electrical waveforms, denoted as $x^{HI}(t)$, $x^{HQ}(t)$, $x^{VI}(t)$, $x^{VQ}(t)$, are created as follow :

$$x^{HI}(t) = \sum_k Re(x_k^H)h_{rrc}(t-k/R_s) \tag{1.7}$$

$$x^{HQ}(t) = \sum_k Im(x_k^H)h_{rrc}(t-k/R_s) \tag{1.8}$$

$$x^{VI}(t) = \sum_k Re(x_k^V)h_{rrc}(t-k/R_s) \tag{1.9}$$

$$x^{VQ}(t) = \sum_k Im(x_k^V)h_{rrc}(t-k/R_s) \tag{1.10}$$

with the symbol rate $R_s = 1/T_s$

Most of the time, the four outgoing electrical analog signals need to be amplified by drivers before undergoing an electro-optical conversion with a modulator. Two in-phase and quadrature modulators (IQ-modulators) are used, one for each polarization, to modulate the real and imaginary parts of $x^H(t)$ and $x^V(t)$. The IQ-modulator is composed of two Mach-Zehnder modulators (MZM) to modulate a laser source carrier according to the electrical waveform signals. The IQ-modulator also exploits a $\pi/2$ hybrid to perform an appropriate phase shift between I and Q components before combining the two output components. Then, a polarization beam coupler (PBC) is placed at the output of the two IQ-modulators to combine the two polarizations. The same process is performed for all WDM channels then a WDM multiplexer combines all amplified channels (as described in subsection i) before sending the signal to the optical fiber.

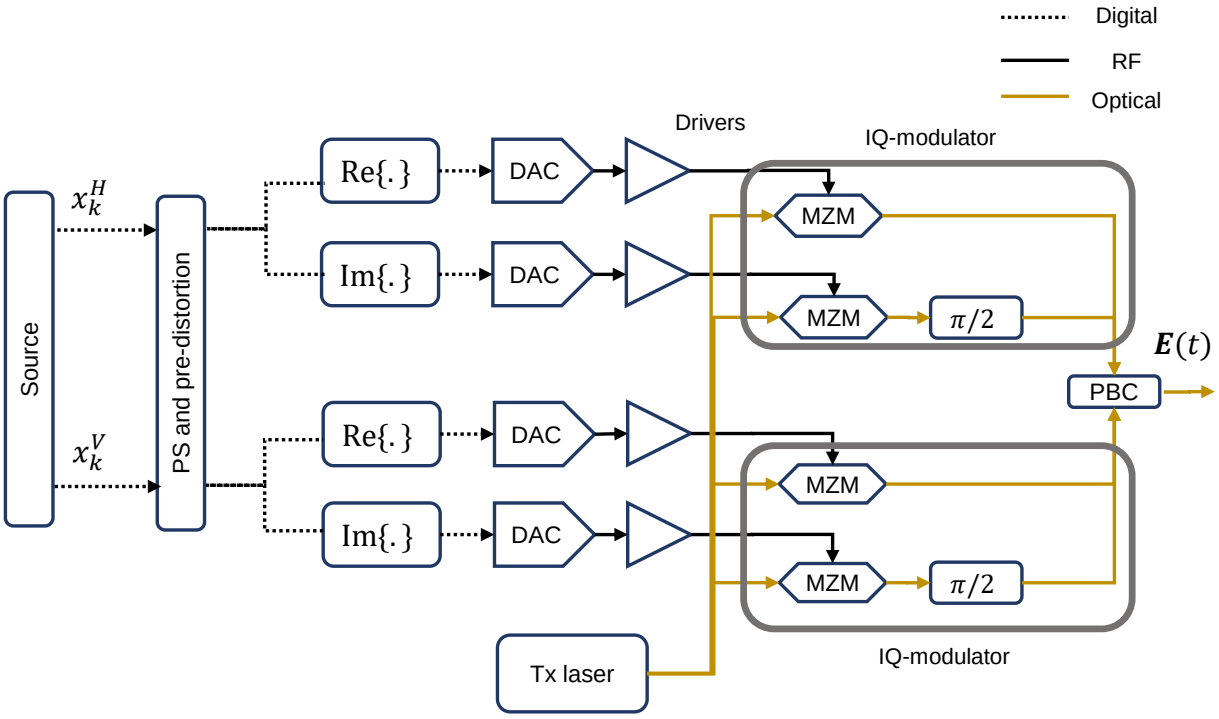


FIGURE 1.9 – Dual-polarization coherent transmitter. After the pulse-shaping, an additional pre-distortion filter can be applied to the transmitted waveform to compensate for the impairments due to the propagation.

1.1.4 Impairments during propagation

The refractive index is defined by $n = c/v_p$ with c the light speed in the vacuum and v_p the velocity in the medium. Fig. 1.5 illustrates the composition of a SMF fibre where the core and cladding parts are shown and their respective refractive indexes are plotted. This difference in refractive index with $n_c > n_{cl}$, combined with an appropriate choice of their diameter dimensions, makes the confinement of the optical field within the fiber core possible thanks to total internal reflection. The propagation of the optical signal inside the fiber core may encounter several effects. In this

sub-section, linear and nonlinear effects impacting the performance of optical transmission are presented. First, we explain the physical effect for a scalar propagation (without PDM), then the generalization to a dual-polarization propagation scheme is presented.

For a scalar optical field composed of WDM channels, its waveform can be written as :

$$E(z,t) = \sum_{k,s=0} x_{k,0}\psi_{k,0}(z,t)\exp\big(i\beta(\omega)z\big) + \sum_{k,s\neq 0} x_{k,s}\psi_{k,s}(z,t-t_s(z))\exp\big(-i\Omega_s t + i\Phi_s(0) + i\beta(\omega)z\big) \tag{1.11}$$

where the channel of interest (COI) corresponding to $s = 0$ and $s \neq 0$ corresponds to the other WDM channels' label. x_k is the kth symbol of the COI, $x_{k,s}$ is the kthsymbol of the sth adjacent channel, Ω_s is the center frequency detuning of the sth adjacent channel with respect to COI, $\beta(\omega)$ is the propagation constant, $t_s(z)$ is the time shift of the sth adjacent channel with respect to COI, $\Phi_s(0)$ is the initial ($z = 0$) relative phase offset of the sth adjacent channel with respect to COI, $\psi_{k,0}(z,t)$ is the pulse-shape of the COI, and $\psi_{k,s}(z,t)$ is the pulse-shape of the sth channel.

For the rest of the work, we suppose the pulse shaping is the same for all WDM channel where :

$$\psi_{0,s}(0,t) = \psi_0(0,t) \tag{1.12}$$

i Linear effects

i.A Fiber attenuation

Except in the vacuum, light propagating in any physical medium will be affected by the power attenuation. In this section, we delve into one major limitation for optical fiber : the fiber loss which is due to absorption and scattering effects.

For a optical field $E(z,t)$, its power P is defined by $P(z) = \langle |E(z,t)|^2 \rangle_t$. So the power decrease during the propagation is determined by :

$$P(z) = P_0 \exp(-\alpha z) \tag{1.13}$$

where P_0 is the input power at $z = 0$ and α the attenuation coefficient in [m^{-1}] and z the distance in [m].

In optical communications, the attenuation coefficient α_p is commonly expressed in [dB/km] :

$$\alpha_p[\text{dB/km}] = \frac{10}{z}\log_{10}\left(\frac{P_0}{P(z)}\right) \tag{1.14}$$

with z the distance in [km].

This attenuation coefficient depends on the fiber type and wavelength of the optical signal. The attenuation profile $\alpha_p(\lambda)$ for a standard SMF (SSMF) taken from [17] is shown in Fig. 1.10. The C-band (C for "conventional") ranged between 1530–1565 nm has the minimum attenuation value with α_p around 0.2 dB/km for SSMF. This is the reason

why it has become the preferred choice for long-distance transmissions.

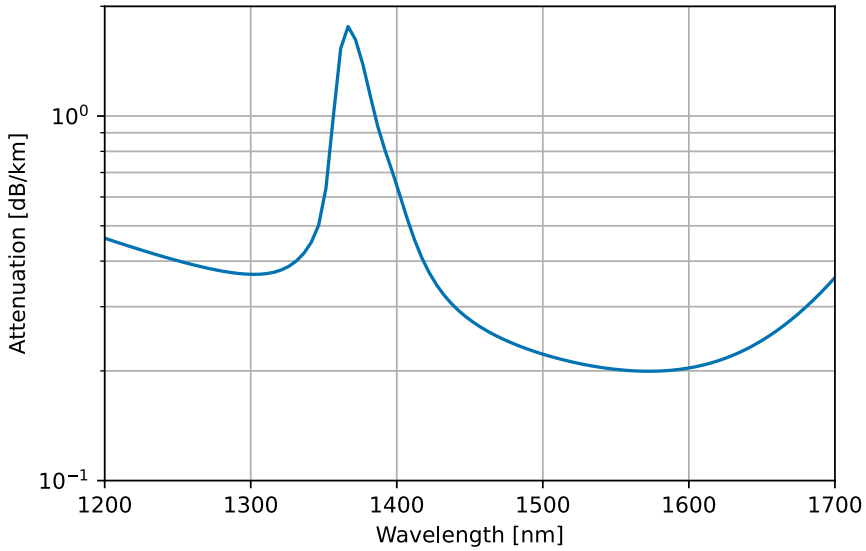


FIGURE 1.10 – Attenuation profile of a SSMF. Note that the absorption peak between 1350 and 1400 nm is due to OH^- ions. Modern fibers nearly eliminate the OH^- peak (for instance the G652-D fiber)

To compensate the fiber loss, long-haul optical systems can integrate periodic optical amplifiers for the transmission link (cf. subsection 1.1.6) .

i.B Chromatic dispersion

The speed of the light inside an optical fiber is not constant for all frequencies, so the different spectral components propagate at different velocities. The speed of the spectral component at angular frequency $\omega = 2\pi f$ is given by the phase velocity $v_p = c/n(\omega)$. This effect called chromatic dispersion (CD) leads to a pulse broadening problem during the propagation and causes ISI over long distances that affects the system's performance. If only CD effect is considered, the evolution of the scalar optical field E along the optical fiber can be described by :

$$\frac{\partial E}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 E}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 E}{\partial t^3} \tag{1.15}$$

where β_i is the i^{th} derivative of the propagation constant β with respect to ω. They are defined with the Taylor series development of $\beta(\omega) = \omega/v_p$ with respect to the COI's angular frequency ω_0 with the following expression ;

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{\beta_2}{2}(\omega - \omega_0)^2 + \frac{\beta_3}{6}(\omega - \omega_0)^3 + ... \tag{1.16}$$

β_0 in [km^{-1}] is the phase shift constant and the first order dispersion parameter β_1 [ps/km] is related to the

envelope propagation speed v_g (also called group velocity) as $v_g = 1/\beta_1$. The first two order parameters β_0 and β_1 do not cause any signal degradation during transmission. β_2 is called group velocity dispersion (GVD) corresponding to the acceleration of the spectral components of the pulse and is expressed in [ps^2/nm]. Finally, β_3 expressed in [ps^3/nm] is the GVD slope corresponding to the variation of GVD as a function of the angular frequency.

In optical communications, it is more common to use the dispersion factor, D, and its slope S_0, which are related to β_2 and β_3, to quantify the CD effects. D expressed in [ps.nm^{-1}.km^{-1}] quantifies the delay with respect to the COI and is defined as :

$$D(\lambda) = \frac{-2\pi c}{\lambda^2}\beta_2 \tag{1.17}$$

with c the light speed.

And the slope S_0 (expressed in [ps.(nm^{-2}·km^{-1})]) of the dispersion coefficient $D(\lambda)$ around reference wavelength λ_0 is defined with :

$$S_0 = \left(\frac{\partial D}{\partial \lambda}\right)_{\lambda_0} = \frac{4\pi c}{\lambda_0^3}\beta_2 + \frac{(2\pi c)^2}{\lambda_0^4}\beta_3 \tag{1.18}$$

For most of fibers deployed in optical transmission systems, D usually has a positive value (around 17 ps/nm/km for SSMF at 1550 nm). We plot the dispersion D with respect to the wavelength for a SMF (data from [18]) in Fig. 1.11. But fibers with negative dispersion constant also exist and they can be used in dispersion-managed (DM) systems to compensate the accumulated chromatic dispersion. For instance, a periodic configuration with several short spans of dispersion compensation fiber (DCF), which has the dispersion coefficient D around -50 to -80 ps/nm/km and up to around 120 ps/nm/km, can achieve a total accumulated CD around 0 ps/nm/km. However DM scheme lost its popularity because it introduces higher nonlinear effects and current deployed coherent systems prefer a dispersion un-managed (DU) scheme where CD is compensated by DSP thus offering more flexibility.

ii Nonlinear effects

In high power regime (high launched channel power), the performance of the system is predominantly limited by nonlinear effects. Indeed, the response of optical fiber for intense electromagnetic fields is nonlinear.

When nonlinear effects are also considered, the propagation is given by the the nonlinear Schrödinger equation (NLSE), which is derived from the Maxwell's equations [19] :

$$\frac{\partial E}{\partial z} = \underbrace{-\frac{\alpha}{2}E}_{\text{attenuation}} \underbrace{-j\frac{\beta_2}{2}\frac{\partial^2 E}{\partial t^2}}_{\text{dispersion}} \underbrace{+\frac{\beta_3}{6}\frac{\partial^3 E}{\partial t^3}}_{\text{dispersion slope}} \underbrace{+j\gamma|E|^2 E}_{\text{Kerr nonlinear effect}} \underbrace{-j\gamma T_R \frac{\partial |E|^2}{\partial t}E}_{\text{Raman nonlinear effect}} \tag{1.19}$$

where α is the attenuation coefficient, β_2 is the GVD, β_3 is the GVD slope, $\gamma = \frac{2\pi n_2}{\lambda_0 A_{eff}}$ is the nonlinear coefficient,λ_0 is the COI wavelength, A_{eff} is the effective area, n_2 is the nonlinear refractive index (see definition in [19]) and T_R is the Raman time constant. The last term in the equation can be neglected because the time-dependent effect of

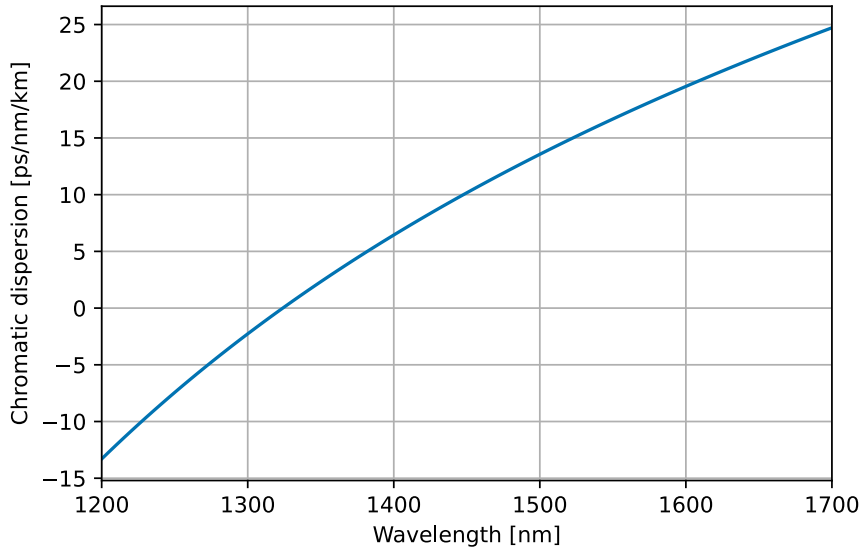


FIGURE 1.11 – CD profile of a SSMF

stimulated Raman scattering (SRS) crosstalk is negligible due to the longer impulse duration in optical communication.

In this subsection, a brief description of the two nonlinear effects is provided. A more detailed description about their respective models are presented afterwards (in section 3.3 for Kerr non-linearities and section 2.2 for Raman effects).

ii.A Kerr effect

Kerr effects cause the change of the refractive index of the physical medium depending on the optical power of input electromagnetic field $|E|^2$ governed by [19, 12] :

$$n\ (\omega, |E|^2) = n(\omega) + n_2|E|^2 \tag{1.20}$$

with $n(\omega)$ the index n evaluated at ω depending of the CD influence and n_2 the nonlinear refractive index.

The main Kerr nonlinear impairments can be categorized in two classes : intra-channel effect (interaction of COI with itself) and inter-channel effect (interaction between COI and neighbouring WDM channels). The effect that accounts for intra-channel non-linearities is called self-channel interference (SCI). On the other hand, we can distinguish two different types of cross-channel interference (XCI) accounting for inter-channel non-linearities : degenerate XCI (when the interaction happens with only one single WDM neighbouring channel) and non-denegerate XCI (when the COI distortions arise from 2 other channels)

To quantify the nonlinear noise due to Kerr effect and its impact on WDM transmission systems, the NLSE has to be solved through heavy numerical simulations. Thus, several analytical models have been proposed. In Chapter 3 subsection 3.3), we give an overview of the analytical models to estimate the Kerr nonlinear interference variance.

ii.B Stimulated Raman scattering

SRS is a non-elastic scattering effects originate from the interaction of the optical field with the silica molecules in the optical fiber where the photon transfers part of its energy to the silica molecules [19, 12].

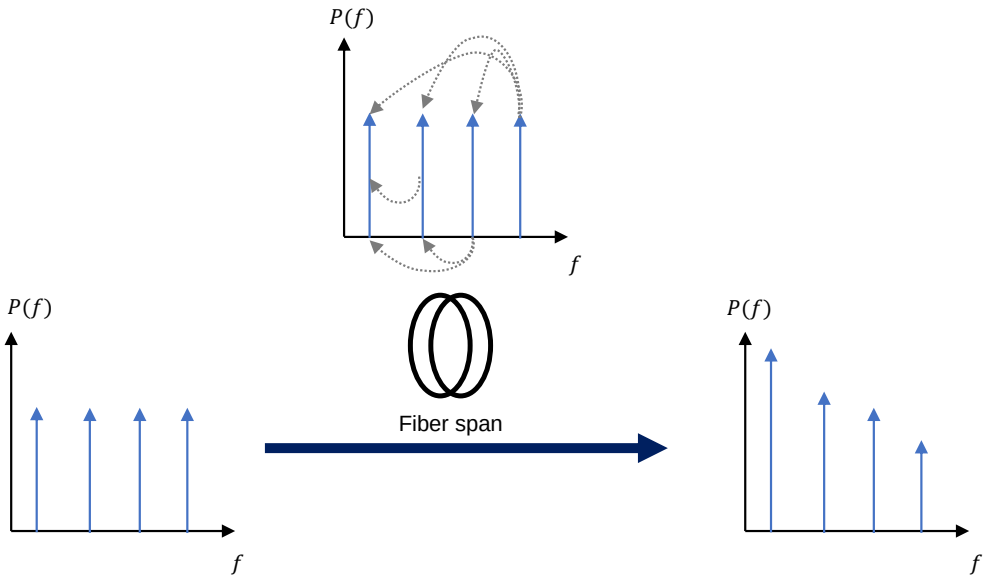


FIGURE 1.12 – Schematic representation of I-SRS in the fiber for WDM system resulting in a power tilt

The intra-channel SRS effect (I-SRS) effects occur in WDM systems and considers optical propagating waves at higher frequency as "pumps" and waves at lower frequency as "signals" thus a power tilt in the signal power profile happens, as shown in Fig. 1.12

This SRS nonlinear effect can also be exploited to build amplifier called Raman amplifier but using additional lasers at higher frequency to transfer their energy to the nonlinear medium that enables amplification of the optical signal at lower frequency. The Raman-induced stimulated emission only occurs when the optical field power is above the SRS threshold value which depends on the type of fiber. The maximum efficiency of SRS power transfer is found around 13.2 THz (100 nm) frequency shift from the original frequency.

Raman effect modelling and equations as well as Raman effects impact and application in fiber optic communication systems are are further discussed in the chapter 2 section 2.2

iii Polarization effects

We now consider the optical field in a vectorial form composed of two orthogonal polarisation states : $\mathbf{E}(z,t) = [E^H(z,t), E^V(z,t)]^T$

iii.A Polarization mode dispersion

Optical signals propagate at different velocities for different polarizations because the refractive index of the fiber core also depends on the polarization. This phenomenon is called birefringence. Indeed, real life fibers do not conserve a perfect cylindrical symmetry due to manufacturing imperfections, external stress on the fiber or even temperature. The delay between the modes propagating over the different axes called differential group delay (DGD) in [ps] is defined by :

$$\Delta t = L|\beta_\mathrm{f} - \beta_\mathrm{s}| \tag{1.21}$$

with β_f and β_s are propagation constants for the fast and slow axes and L is propagation distance.

For standard fibers, the fiber constraints vary randomly with time and distance due to the instability of the environment. So does the induced birefringence. So a real fiber as a concatenation of infinitesimal birefringent sections randomly coupled Consequently, DGD is a stochastic process and PMD is defined as its mean value. For current system, PMD can reach a small value around 0.05 ps/$\sqrt{\mathrm{km}}$.

Due to DGD, the two orthogonal polarization input are mixed together during propagation but DSP can be used to track and separate them.

iii.B Polarization dependent loss

Another effect affecting polarization-sensitive systems is the polarization-dependent loss (PDL). Indeed, optical components in the optical link (for instance amplifiers, couplers, etc.) may have a different behaviour at each polarisation. At the output of a PDL element, two effects may appear : on the one hand, the signal attenuation can be different between the two polarization, on the other hand the two polarization tributaries can be no longer orthogonal.

The impact of PDL is out of the scope of this thesis and is not taken into account in the propagation modeling.

iii.C Manakov equation

For dual-polarization case, Eq. (1.19) can be be extended to the well-know Manakov equation :

$$\frac{\partial \mathbf{E}}{\partial z} = -\frac{\alpha}{2}\mathbf{E} - j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 \mathbf{E}}{\partial t^3} + j\frac{8}{9}\gamma \mathbf{E}^\dagger \mathbf{E}\mathbf{E} \tag{1.22}$$

where the superscript $\dagger$ stands for Hermitian conjugation operation (complex conjugation followed by matrix transpose). It is worth to notice that the coefficient $8/9$ weighting the Kerr effect contribution comes from the averaging of the birefringence-induced polarization rotations assuming a random birefringence in the fiber.

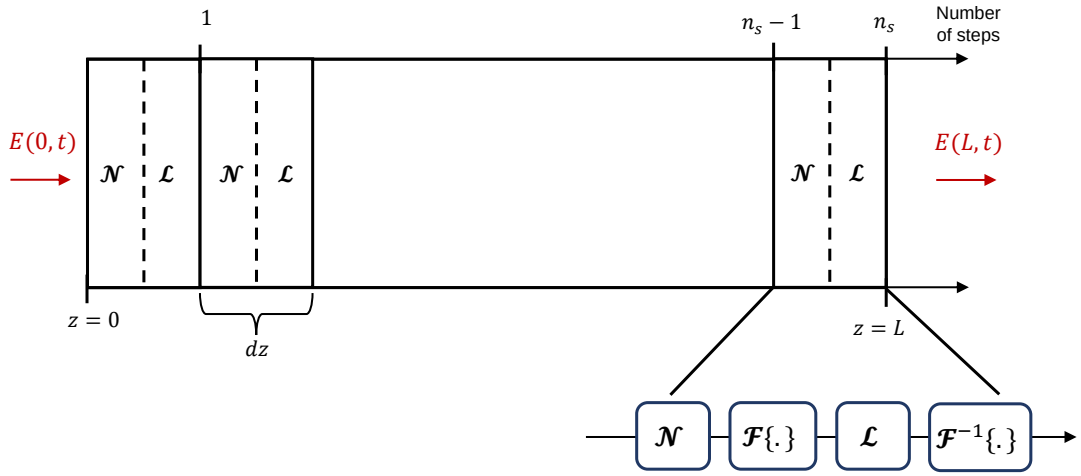


FIGURE 1.13 – SSFM representation

1.1.5 Split-step Fourier simulation method

Since the NLSE equation governing the optical signal propagation has no exact analytic solutions, the split-step Fourier method was proposed to simulate the propagation through an optical fiber [19]. This numerical method solve the NLSE equation with n_s steps : it divides the entire fiber into small steps with length dz. From the transmitter to the end of the link at $z = L$, for each step, a linear or a nonlinear operator will be performed. Fig. 1.13 illustrates a general representation of the SSFM.

We can re-write Eq. (1.19) in the form :

$$\frac{\partial E}{\partial z} = (\hat{L} + \hat{N})E \tag{1.23}$$

where $\hat{L}$ is the linear operator that accounts for dispersion and absorption and $\hat{N}$ is a nonlinear operator that accounts for Kerr nonlinear effect (SRS effect neglicted here). These operators are given by :

$$\hat{L} = -\frac{\alpha}{2} - j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3} \tag{1.24}$$

$$\hat{N} = j\gamma\left(|E|^2\right) \tag{1.25}$$

Finally we have :

$$E(z + dz, t) = exp(dz\hat{N})exp(dz\hat{L})E(z,t) \tag{1.26}$$

The linear operation can be evaluated in the Fourier domain (cf. Fig. 1.13). The step size dz has to be chosen carefully because the SSFM is more accurate when used with smaller step size but the number of computations inversely increase with the step size.

A modified version called symmetrical SSFM can improve the performance by inserting the nonlinear operator

in the middle of two linear operator :

$$E(z+dz,t) = exp\left(\frac{dz}{2}\hat{L}\right) exp\left(dz\hat{N}\right) exp(\frac{dz}{2}\hat{L})E(z,t) \tag{1.27}$$

The SSFM has been widely adopted to evaluate the accuracy of the analytical models to compute the NLI but its computational time can be very long : minutes (for a single channel) to several hours depending on the numbers of total spans and the number of co-propagating channels (and even days of computation for a long-haul system with fully-loaded UWB WDM scheme).

1.1.6 Loss compensation

To compensate the attenuation of the optical power, optical amplifiers can be used. Indeed, without amplifiers that can compensate the fiber loss, we need to regenerate the signal after a few tens of km of transmission. Current long-haul systems implement periodic optical amplification to amplify the whole WDM channels. So, the transmission link consists of consecutive fibre spans separated by optical amplifiers.

i Amplifier types

In this section, three different types of commonly used optical amplifiers are described.

i.A EDFA

The most widely used optical amplifier is the erbium-doped fiber amplifier (EDFA), proposed in late 1980 [20, 21]. Laser pumps are sent into a silica fiber that has been doped with erbium ions to amplify the attenuated low-power signal. Since the amplification medium is not the transmission fiber, EDFA amplifies the signal in a so-called lumped way. The excitation of the erbium ions leads to two physical phenomena : the stimulated emission of photons at same frequency as the signal (causing signal amplification) and the spontaneous emission of photons generating noise, also called amplified spontaneous noise (ASE).

EDFA amplification provides an optical amplification window of $35-40$ nm bandwidth in either C (Fig. 1.14) or L band, depending on the pump configuration and doped fiber length. Because of the EDFA gain spectrum shape, a gain flattering filter (GFF) is usually inserted compensate the non-flat spectrum gain.

i.B Raman amplifier

Raman amplifier is based on the SRS effect with external pumps with higher frequency to amplify the transmitted signal. During a SRS, a new photon is generated at the signal frequency, that is downshifted from the pumps frequencies. The frequency difference between the pump and the signal photon is called the Stokes shift, and

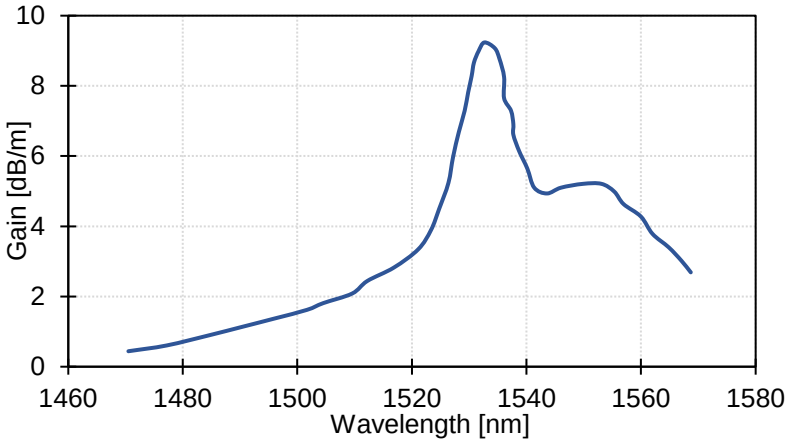


FIGURE 1.14 – Gain profile of an EDFA whose core was codoped with germania (taken from [18])

in standard transmission fibers, the peak of this frequency shift is about 13.2 THz. The gain spectrum is shown in Fig. 1.15 and multiple lasers at different frequencies can be used to achieve a broadband amplifier for WDM transmission. The pump lasers can be either injected in the same direction as the one of the transmitted signal (co-propagating) or in the opposite direction (contra-propagating) or in both direction (bi-directional). More details about SRS and Raman amplification are provided in Chapter 2 Section 2.2, in particular the ordinary differential equations (ODEs) governing the evolution of the signal and pump powers.

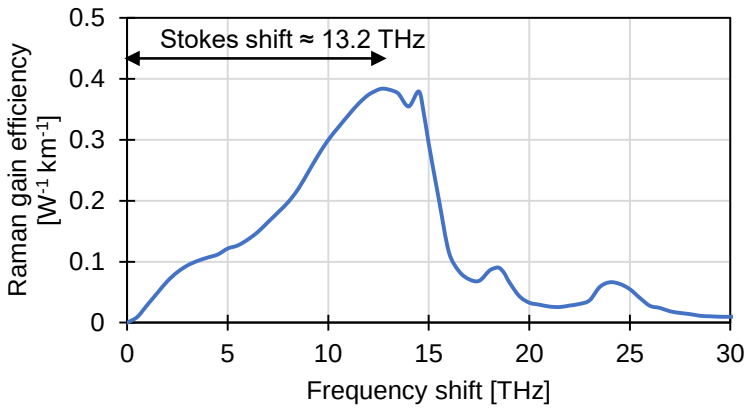


FIGURE 1.15 – Raman spectrum at 1550 nm for SSMF

i.C SOA

Semiconductor optical amplifiers (SOA) are composed of a Fabry-Perot cavity in semiconductor gain medium pumped with an electric current where the light is amplified through stimulated emission. Compared to EDFAs, the relaxation time of the excited state for SOAs is lower ($\simeq 10$ ps instead of > 1 µs for EDFAs) Consequently, SOA gain

varies with input power of the optical signal thus can lead to nonlinear distortions which is the main drawback of this amplifier. Despite this disadvantage, its compact size and ability to cover wide bandwidth make them very attractive for UWB communication. In [22], the authors experimentally demonstrated a data rate of 115.9 Tb/s over 100 km, based on a 100+nm-wide SOA.

ii Amplification noise

The noise figure of an amplifier F is defined by :

$$F = \frac{SNR_{in}}{SNR_{out}} \tag{1.28}$$

with SNR_{in} and SNR_{out} the amplifier input and output SNR and we can define it in dB which is more commonly used as :

$$NF \text{ in [dB]} = 10\log_{10}(F) \tag{1.29}$$

The noise figure can be evaluated as :

$$F = \frac{P_{ASE,B_{ref}}}{\hbar\omega_0 B_{ref} G} + \frac{1}{G} \tag{1.30}$$

with the $P_{ASE,B_{ref}}$ integrated ASE power in B_{ref}, the reference bandwidth, $\omega 0$ the reference angular frequency $\hbar$ is the reduced Planck constant and G the amplifier gain.

For EDFA, the amount of ASE added, integrated over the reference bandwidth B_{ref}, is :

$$P_{ASE,B_{ref}} = N_{ASE} B_{ref} \tag{1.31}$$

where N_{ASE} in [W/Hz] is the power spectral density per polarisation. Standard EDFAs' noise figure (NF) is usually between $4.5 - 5.5$ dB.

For dual-polarization, the power spectral density of a single EDFA is given by :

$$N_{ASE} = 2n_{sp}\hbar\omega_0(G-1) \tag{1.32}$$

$$F = 2n_{sp}\frac{(G-1)}{G} \tag{1.33}$$

The total noise factor of a chain of amplifiers with noise factor F_n and gain G_n can be expressed with the Friis formula :

$$NF_{tot} = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} \tag{1.34}$$

For a transmission link with N_{tot} spans, and assuming that $\lambda_0 \approx 1550$ nm and $B_{ref} = 12.5$ GHz (0.1 nm), we can

approximate the optical SNR (OSNR) after transmission as follows :

$$ONSR\ [\text{dB}] \approx 10\log_{10}\left(\frac{P_{in}}{(F-1)\hbar\omega_0 B_{ref} N_{tot}}\right) = 58[\text{dBm}] + P_{in}[\text{dBm}] - NF - 10\log_{10}(N_{tot}) \tag{1.35}$$

1.1.7 The coherent receiver and DSP

At the output of the optical link, channels are demultiplexd with a Demux then each channel pass through a coherent receiver to obtain tow electrical signals which are then converted in digital domain. Then DSP algorithms are performed to recover the corresponding symbols for each polarisation attribute.

i Opto-electric frontend

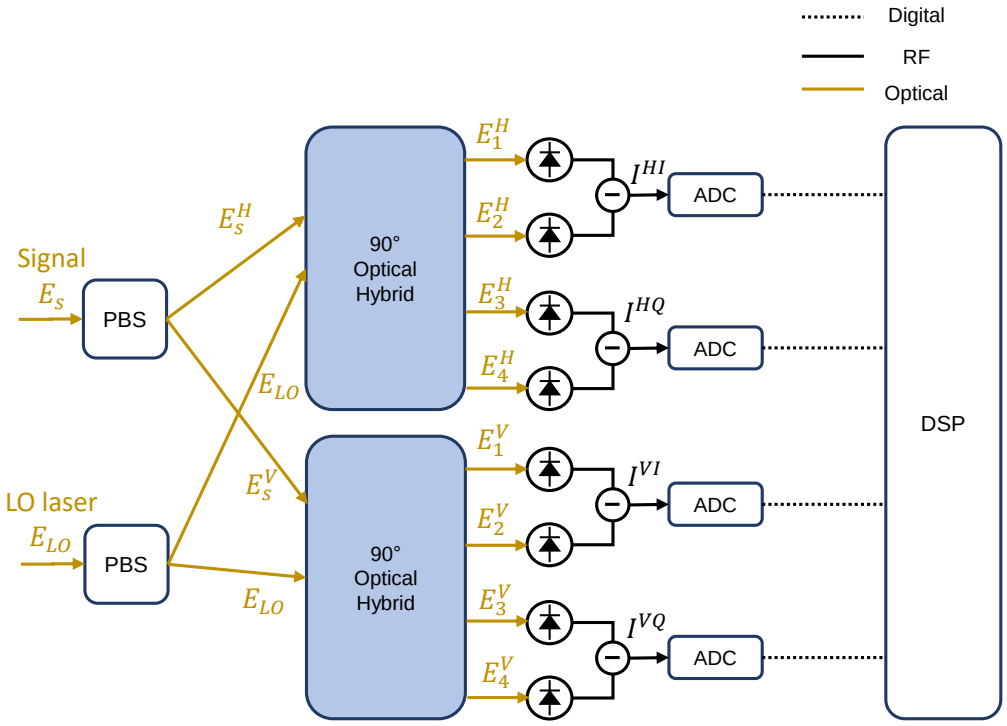


FIGURE 1.16 – Dual-polarization coherent receiver schematic.

The general representation of a coherent receiver used in a dual-polarization scheme is represented in 1.16. This receiver detect the I and Q components of two orthogonal but arbitrary polarizations E_S^V and E_S^H by the mean of a polarization beam splitter (PBS), $90°$-hybrid and a local oscillator (LO) laser with a frequency f_{LO} close to the signal's frequency f_{TX}. As a result, the eight optical fields at the output are expressed as follows :

$$\begin{bmatrix} E_1^H \\ E_2^H \\ E_3^H \\ E_4^H \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} +E_S^H + E_{LO} \\ -E_S^H + E_{LO} \\ +E_S^H + jE_{LO} \\ -E_S^H + jE_{LO} \end{bmatrix} \text{ and } \begin{bmatrix} E_1^V \\ E_2^V \\ E_3^V \\ E_4^V \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} +E_S^V + E_{LO} \\ -E_S^V + E_{LO} \\ +E_S^V + jE_{LO} \\ -E_S^V + jE_{LO} \end{bmatrix} \tag{1.36}$$

The four resulting photocurrents after detection with balanced photodiodes (i.e only the beating signal is preserved) are given by :

$$\begin{bmatrix} I^{HI} \\ I^{HQ} \\ I^{VI} \\ I^{VQ} \end{bmatrix} = 2\mathcal{R} \begin{bmatrix} \mathrm{Re}\left\{E_S^H E_{LO}^*\right\} \\ \mathrm{Im}\left\{E_S^H E_{LO}^*\right\} \\ \mathrm{Re}\left\{E_S^V E_{LO}^*\right\} \\ \mathrm{Im}\left\{E_S^V E_{LO}^*\right\} \end{bmatrix} \tag{1.37}$$

where $\mathcal{R}$ is the responsitivity of the photodiode, $\mathrm{Re}\{.\}$ and $\mathrm{Im}\{.\}$ give the real and imaginary part of a complex signal.

ii Digital signal processing

DSP's goal is to recover the symbol sequences sent over the two orthogonal polarizations from the four sampled photocurrents from the coherent receiver, after an analog-to digital conversion. Then the performance is be estimated with those recovered symbols y^H and y^V. The standard DSP, operating after the analog-to digital conversion, is illustrated in Fig. 1.17.

DSP is performed compensate for the accumulated CD with a filter, to mitigate PMD with an adaptive multiple-input multiple-output (MIMO) equalizer, compensate for the frequency and phase offsets between the transmitter (TX) and the LO lasers and equalize transmitter/receiver IQ imbalances and timing deskews with a adaptive post-equalizer.

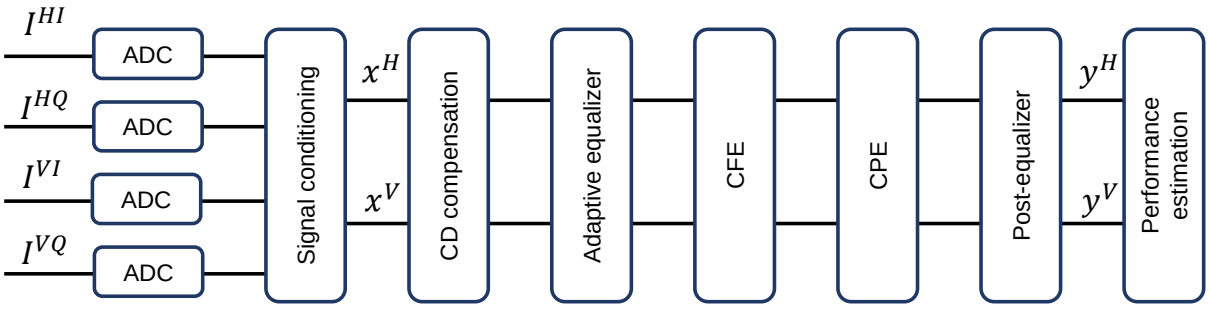


FIGURE 1.17 – Coherent DSP chain block to mitigate the distortion during the propagation. For the final step, a post-equalizer can be applied to compensate any phase or gain mismatch between I and Q components of each polarization

ii.A Signal conditioning

From the four photocurrents, four quantized are sampled at the sampling rate of the oscilloscope. To respect the Shannon-Nyquist theorem, first the received signal is re-sampled at a rate of 2 samples per symbol (SpS). Then we can use a brickwall bandpass filter to suppress out-of-band noise.

ii.B CD compensation

To mitigate the accumulated CD after propagating a distance L, a filter is applied with a transfer function given :

$$H^{CDC}(\omega) = e^{-j\omega^2 \frac{\beta_2}{2} L} \tag{1.38}$$

This frequency domain equalizer using overlap-and-save methods and fast Fourier transforms (FFT) [23] is able to minimize the computational complexity of the CD compensation.

ii.C Adaptive MIMO equalizer

Due to fiber birefringence, the two polarizations will be coupled during the propagation and the received two components will be orthogonal by with arbitrary polarization state. To compensate these impairments, an adaptive MIMO equalizer with a butterfly structure is applied where the output y_k^H and y_k^H is given by :

$$\begin{bmatrix} y_k^H \\ y_k^V \end{bmatrix} = \begin{bmatrix} (\mathbf{h}_k^{HH})^T & (\mathbf{h}_k^{VH})^T \\ (\mathbf{h}_k^{HV})^T & (\mathbf{h}_k^{VV})^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_k^H \\ \mathbf{x}_k^V \end{bmatrix} \tag{1.39}$$

with $\mathbf{x}_k^H = [x_{k-N_T}^H, ..., x_k^H, ..., x_{k+N_T}^H]^T$ a $(2N_T + 1)$ 1-D vector and the four vectors, $\mathbf{h}_k^{HH}, \mathbf{h}_k^{VH}, \mathbf{h}_k^{HV}, \mathbf{h}_k^{VV}$ each containing $(2N_T + 1)$ taps of the equalizer, can be estimated with blind approach such as constant modulus algorithm (CMA) [24]. This approach forces the output towards a constant amplitude on both polarization states by updating the equalizers' taps by minimizing the following error function ϵ_k on the two polarizations :

$$\begin{aligned} \epsilon_k^H &= 1 - \left|y_k^H\right|^2 \\ \epsilon_k^V &= 1 - \left|y_k^V\right|^2 \end{aligned} \tag{1.40}$$

Using the gradient descent method, the $2N_T + 1$ equalizer taps update can be written as :

$$\begin{aligned}
\mathbf{h}_{2k+2}^{HH} &:= \mathbf{h}_{2k}^{HH} + \mu \cdot \varepsilon_{2k}^{H} \cdot y_{2k}^{H} \cdot \left(\mathbf{x}_{2k}^{H}\right)^{*} \\
\mathbf{h}_{2k+2}^{VH} &:= \mathbf{h}_{2k}^{VH} + \mu \cdot \varepsilon_{2k}^{H} \cdot y_{2k}^{H} \cdot \left(\mathbf{x}_{2k}^{V}\right)^{*} \\
\mathbf{h}_{2k+2}^{HV} &:= \mathbf{h}_{2k}^{HV} + \mu \cdot \varepsilon_{2k}^{V} \cdot y_{2k}^{V} \cdot \left(\mathbf{x}_{2k}^{H}\right)^{*} \\
\mathbf{h}_{2k+2}^{VV} &:= \mathbf{h}_{2k}^{VV} + \mu \cdot \varepsilon_{2k}^{V} \cdot y_{2k}^{V} \cdot \left(\mathbf{x}_{2k}^{V}\right)^{*}
\end{aligned} \tag{1.41}$$

with μ as a constant parameter to control the speed of the convergence.

ii.D Frequency offset estimation

The LO laser used for the coherent receiver, has a frequency close the transmitter laser but slightly different. This leads to a small frequency offset and also phase offset because of the slowly time-varing phase noise (PN). Thus, the carrier recovery consist of two steps : first, an estimation of the frequency offset, followed by a phase estimation.

The problem of carrier estimation can be modeled as :

$$y_k = x_k e^{(j\phi_k k + j\theta_k)} + n_k \tag{1.42}$$

with y_k being previous DSP blocks output for one polarization(after CD compensation and polarization de-multiplexing), corrupted with respect to the true transmitted symbols x_k, $\phi_k = \delta\omega T_s$ the unknown phase due to the frequency offset $\delta\omega$ at symbol time T_s, θ_k the unknown discrete time-varing phase offset due to the PN of the laser and n_k an additive zero-mean circularly-symmetric complex Gaussian noise.

For 2^m QAM constellations, the estimation of ϕ_k can be performed by maximizating the periodogram as presented in [25] :

$$\hat{\phi}_k = \frac{1}{4} \underset{\phi}{\text{argmax}} \left| \frac{1}{2N+1} \sum_{m=-N}^{N} (y_{k+m})^4 e^{-j\phi m} \right| \tag{1.43}$$

with $2N + 1$ the number of samples used.

ii.E Phase estimation

After the frequency estimation, the signal for one single polarization can be modelled as :

$$y_k = x_k e^{j\theta_k} + n_k \tag{1.44}$$

with x_k the k^{th} sent symbols, n_k the additive Gaussian noise and θ_k the unknown phase noise we want to estimate.

Phase estimation can be achieved with the well-know blind phase search (BPS) algorithm. It consists in minimizing the following cost function [26] with $2N+1$ symbols to find $\hat{\theta}_k$:

$$\hat{\theta}_k = \underset{\theta}{\text{argmin}} \sum_{m=-N}^{N} \left| y_{k+m} - \hat{x}_{k+m} e^{-j\theta} \right|^2 \tag{1.45}$$

where x_k is the closest estimated symbol rotated by $e^{-j\theta}$. Square QAM modulation formats have a rotational symmetry of order 4 thus introducing an phase ambiguity. Consequently, the BPS algorithm tests phases between 0 and $\pi/2$ and then one can use the N_p know pilots symbols sent with the source symbol sequence to remove the cycle slip ambiguity.

1.1.8 Performance evaluation

To evaluate the performance, the SNR is one popular metric. If we assume that the received signal is impaired with an additive noise, the SNR can be estimated from the transmitted and received symbol sequences **x** and **y** as [27] :

$$\widehat{SNR} = \left[1 - \frac{\left| \mathbf{x}\mathbf{y}^{\dagger} \right|^2}{|\mathbf{x}|^2 |\mathbf{y}|^2} \right]^{-1} \tag{1.46}$$

with † the Hermitian operator representing the conjugate transpose.

Under the assumption of the additive white uncorrelated Gaussian noise model is used for all impairments, then the total noise variance is the sum of all different noises. Thus the resulting total SNR can be expressed as :

$$\frac{1}{SNR} = \frac{1}{SNR_{TRX}} + \frac{1}{SNR_{ASE}} + \frac{1}{SNR_{NLI}} \tag{1.47}$$

with SNR_{TRX} accounting for the transceiver impairments, SNR_{ASE} representing the ASE contribution and SNR_{NLI} for the nonlinear distortions. In this expression, we neglect other extra effects such as guided acoustic wave Brillouin scattering (GAWBS) (not studied in this thesis) or EEPN (see Chapter 3).

i Impairment sources

i.A Transceiver penalties

The first term in Eq. (1.47) accounts for TRX and DSP impairments and is independent of the input power, thus the maximum achievable SNR is :

$$SNR_{TRX} = \frac{1}{k_{TRX}} \tag{1.48}$$

with k_{TRX} the normalized variance and can be experimentally measured with a back-to-back (B2B) set-up on received constellation (i.e. at the DSP chain output).

i.B Amplified spontaneous emission noise

The second term in Eq. (1.47) represents the linear part of distortions due to ASE noise produced by amplifiers and is expressed as :

$$SNR_{ASE} = \frac{P_{ch}}{\sigma^2_{ASE}} = \frac{B_{ref}}{R_s} OSNR \tag{1.49}$$

where P_{ch} is the channel power, B_{ref} the reference bandwidth (usually 0.1 nm), R_s the symbol rate and OSNR is defined as :

$$OSNR = \frac{P_{ch}}{P_{ASE}} \tag{1.50}$$

and P_{ASE} is computed according to the used amplifiers type and characteristics. For EDFAs, its expression can be found in subsection ii.

i.C Nonlinear distortion

The last term in Eq. (1.47) accounts for the Kerr nonlinear effects. The variance of the NLI, σ^2_{NL}, can be expressed with the nonlinear coefficient a_{NLI} as :

$$\sigma^2_{NL} = a_{NLI} P^3_{ch} \tag{1.51}$$

with P_{ch} the channel-wise power. a_{NLI} depends on the fiber link characteristics, the WDM channels configuration and the COI characteristics such as modulation format, symbol rate, etc. .

Finally, the SNR accounting for the NLI is expressed as follows :

$$SNR_{NL} = \frac{P_{ch}}{\sigma^2_{NL}} = \frac{1}{a_{NLI} P^2_{ch}} \tag{1.52}$$

and the overall SNR of the transmission system accounting for TRX impairments, ASE noise and Kerr NLI can be written as :

$$SNR = \frac{P_{ch}}{k_{TRX} P_{ch} + \sigma^2_{ASE} + a_{NLI} P^3_{ch}} \tag{1.53}$$

k_{TRX} can be measured with a B2B configuration and σ^2_{ASE} can be either measured or computed with the noise figure of each amplifier. The computation of a_{NLI} can be achieved numerically with SSFM simulations (see subsection 1.1.5) or predicted with analytical models such as the frequency domain Gaussian noise model (GN) and enhanced GNM (EGN), and the time-domain perturbation approach (more details about the NLI modelling are described in section 3.3). The estimation of the NLI contribution is the main challenge for analytical models since the computation time can be very long. In this thesis, we described a method to accelerate the prediction of the NLI coefficient a_{NLI}

SNR can be maximized with a optimal power value called nonlinear threshold (NLT) is given by [28] :

$$P_{ch,NLT} = \left(\frac{\sigma^2_{ASE}}{2 a_{NLI}} \right)^{1/3} \tag{1.54}$$

and this maximum value of the SNR is thus :

$$SNR_{opt} = \left(k_{TRX} + 3 \left(\left(\frac{\sigma^2_{ASE}}{2} \right)^2 a_{NLI} \right)^{1/3} \right)^{-1} \tag{1.55}$$

for $P_{ch} < P_{ch,NLT}$, ASE noise is dominant and the system behave in a linear regime so the performance can be improved by increasing the launch power and the amplifier output. But after the NLT, Kerr effects become dominant and the performance is degraded with the increase of channel power. NLI is consequently a trade-off between those two limiting impairments and SNR_{NLT} is commonly used to predict the performance of transmission system.

ii Channel capacity

The concept of entropy was introduce by Shannon in 1948 and its expression for the random variable X defined as :

$$H(X) = -\sum_{x \in X} P(X = x) \log_2 \left(P(X = x) \right) \tag{1.56}$$

To quantify the average information shared between two random variables X and Y, we use the mutual information which can be expressed as an entropy difference [29] :

$$I(X;Y) = H(X) - H(X|Y) \tag{1.57}$$

and the capacity of the channel is defined as the maximum of the MI and for an AWGN channel with a complex Gaussian source, the capacity is given by the well-known Shannon capacity formula :

$$C = \max_{P_x} I(X;Y) = \log_2(1 + SNR) \tag{1.58}$$

where $SNR = P/N_0$ is the signal to noise ratio and C is measured in bits/s/Hz. The channel capacity represents the maximum achievable information rate for a given SNR value when an ideal modulation and coding scheme are employed.

1.2 Introduction to machine learning and artificial neural networks

ML have gained attention over the past years in the optical communication community [9] because of its powerful ability to learn from data using algorithms and computational statistics [30, 31, 32]. We may divide the ML problems into four categories : regression, classification, clustering and reinforcement learning. For regression, the ML outputs are real or continuous values whereas for a classification problem, the goal is to find the right class category to

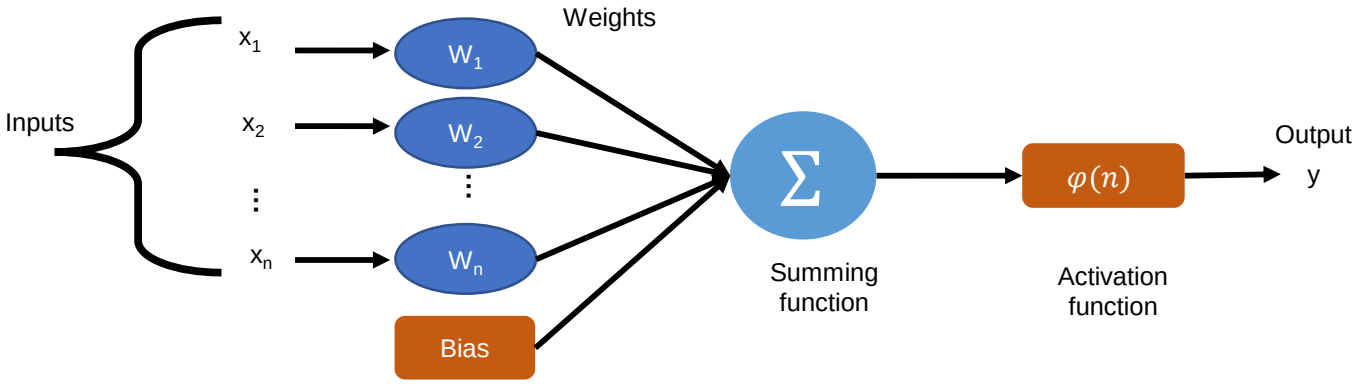


FIGURE 1.18 – Representation of the structure of one single artificial neuron

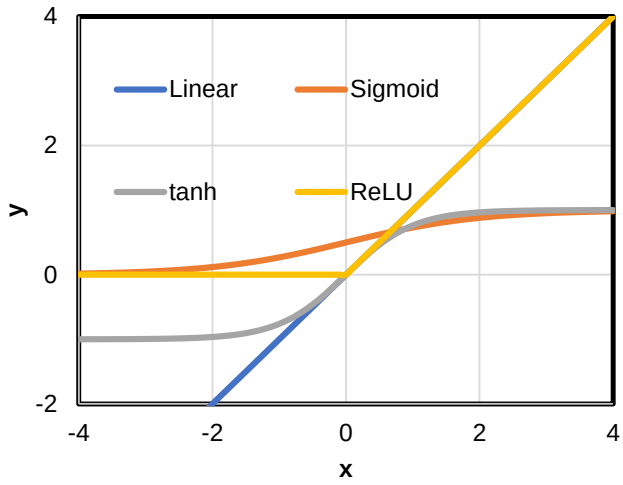


FIGURE 1.19 – Four popular activation functions for regression problems : linear, sigmoid, tanh and ReLU

which the data belongs. Clustering consists in dividing the data into several clusters based on their specificity. Consequently, regression and classification problems need a supervised learning algorithm, thus the data need to be labeled, however clustering problems rely on unsupervised algorithms and the data set does not need to be labeled, avoiding the time-consuming and expensive step of data labelling. Semi-supervised learning is also possible using only partially-labeled data. On the other hand, reinforcement learning consists in making decisions following a reward-based system where the algorithm learns the adequate behaviour from a defined environment. This method can be applied to games or to resolve optimization problems. In this thesis, the studied problems are all regression problems, since, for each of our applications, we want to predict an output vector for a given input vector.

1.2.1 Structure of Artificial Neural Networks

One popular ML algorithm for regression is the ANN which is inspired by the interconnected structure of biological neural networks. The most common feedforward ANN (FANN) is the multi-layer perceptron (MLP). The MLP is

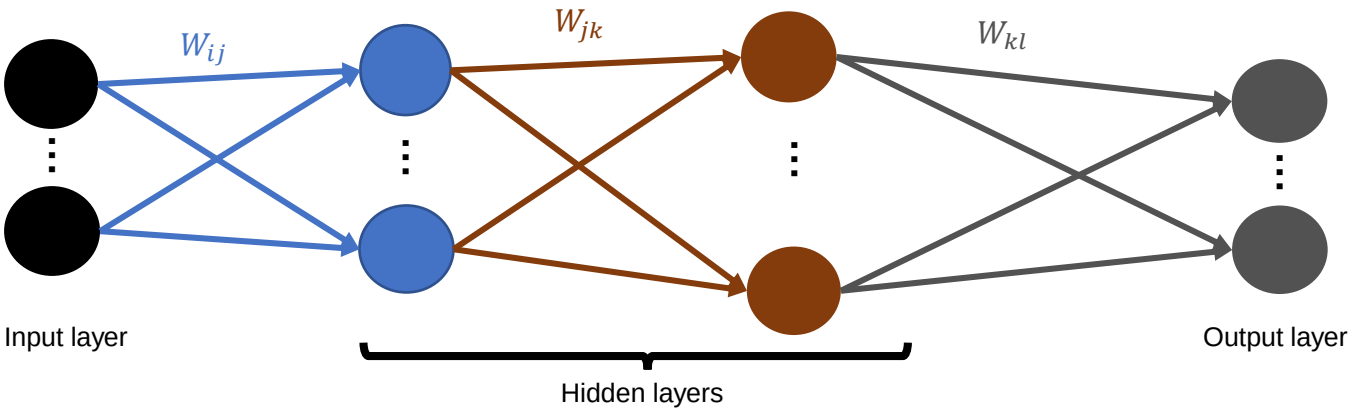


FIGURE 1.20 – Generic architecture of a classical MLP with two hidden layers

composed of several layers of neurons where each neuron in a layer is connected to every neuron in the subsequent layer through weighted connections. The general representation of a simple artificial neuron is depicted in Fig. 1.18. The neuron has multiple inputs and one output. The weights determine the strength of the influence that one neuron of layer number $l-1$ has on another neuron of layer number$l$. Those layers composed of neurons can be fully connected or partially connected (i.e. at least one neuron in layer number l is not connected to all neurons of the layer number $l-1$). Fig. 1.20 shows a fully connected MLP with one input layer, one output layer and two layers between them, called hidden layers. The output of the layer l, $\mathbf{a}^{(l)}$ is computed as follows :

$$\mathbf{a}^{(l)} = \varphi(\mathbf{z}^{(l)}) \tag{1.59}$$

with

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \tag{1.60}$$

$\mathbf{z}^{(l)}$ is the weighted sum of the inputs, $\mathbf{W}^{(l)}$ the weights, $\mathbf{b}^{(l)}$ the bias and $\varphi(.)$ the activation function of layer l. The first layer is the input layer so $\mathbf{a}^{(0)} = \mathbf{x}$ and the last layer, indexed as number N_l,l is the output layer so $\mathbf{a}^{(N_l)} = \mathbf{y}$.

Fig. 1.19 shows four different popular activation functions commonly used in regression problems : the linear function, the sigmoid, the hyperbolic tangent (tanh) and the rectified linear unit (ReLU). Their equations are represented by the following expressions.

$$\varphi_{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{1.61}$$

$$\varphi_{tanh}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{1.62}$$

$$\varphi_{ReLU}(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases} \tag{1.63}$$

1.2.2 Training step

After choosing the desired ML model and before training the model with data, we need to initialize its parameter set θ e.g. for the ANN, θ is the set of weights and bias values to be initialized. Usually, we use Xavier [33] or Kaiming [34] initialization methods for the weights. Then, the bias can either be set to zero or also initialized randomly.

On the other hand, we also have to divide the data into three independent classes : training, validation, and test data sets. The training data set is dedicated to train the model by updating the parameters. For ANNs, the weights and biases are updated through this step. The validation data set is used to evaluate the model or compare trained models between them. Finally, the test data set is used to test and assess the overall performance of the final model. The three data set have to be non-overlapping to get solid results and conclusions because it is evident that ML models always perform better on already seen data compared to new data. So if the validation/test data has already been fed to the model during training stage, the result would be misleadingly better.

The widely used back-propagation procedure to train a feed-forward ANN by updating the parameters was proposed in 1986 by [35]. The algorithm is composed of two steps : the forward pass and the backward pass. In the forward pass, each sample of the training data that constitutes the input vector or the matrix is fed to the input layer and goes through all the hidden layers to finally reach the output layer with a predicted output vector or output matrix. Then, a loss value is computed to evaluate the deviation of the prediction from the target output, called ground truth (GT) using a cost function. In regression problems, we usually choose to use the mean square error (MSE) or mean absolute error (MAE) defined in Eq. (1.64) and (1.65) respectively. Finally, according to the gradients computed through a chain of calculus operations, the backward pass starts from the output layer and propagates the errors backward to update the parameters of the network. By iteratively updating the parameters based on the gradients, the network gradually improves its performance.

$$J_{MSE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{1.64}$$

$$J_{MAE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{1.65}$$

where y_i is the ground truth target value, $\hat{y}_i$ the predicted output value and N the number of samples.

There are several optimization algorithms to update the parameters during the back-propagation step to minimize the cost function. The basic one is the gradient descent. It consists in calculating the gradients of the loss function with respect to the parameters and adjusting the parameters in the opposite direction of the gradients to minimize the loss. Eq. (1.66) to Eq. (1.71) show the fundamental steps of the update algorithm. $^{(l)}$ and $^{(L)}$ are respectively the error in the l^{th} layer and in the output layer. Using the chain rule, their expressions are given by :

$$\boldsymbol{\delta}^{(L)} = \nabla_{\mathbf{z}^{(L)}} J = \nabla_{\mathbf{a}^{(L)}} J \odot \varphi'(\mathbf{z}^{(L)}) \tag{1.66}$$

$$\delta^{(l)} = ((\mathbf{W}^{(l+1)})^T \delta^{(l+1)}) \odot \varphi'(\mathbf{z}^{(l)}) \tag{1.67}$$

with $\nabla_{\mathbf{z}^{(L)}} J$ the gradient of the loss J with respect to the weighted sum $\mathbf{z}^{(L)}$, φ' is the derivative of the activation function and $\odot$ represents element-wise multiplication. By combining Eq. (1.66) with Eq. (1.67), we can compute the error $\delta^{(l)}$ for any layer in the network. Starting from computing $\delta^{(L)}$, then $\delta^{(L-1)}$, then $\delta^{(L-2)}$ and so on, all the way back through the network, hence the term of back-propagation. Moreover, the gradients of the loss with respect to the weights and biases can be computed as follows :

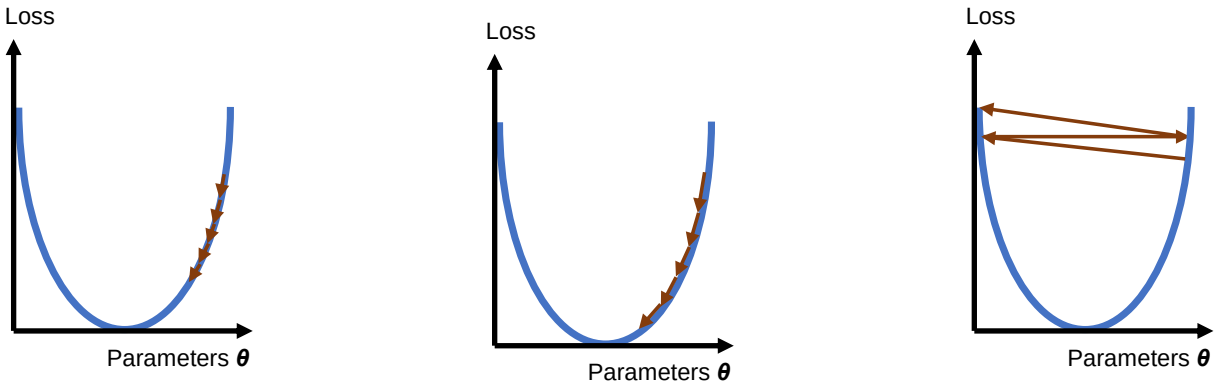
$$\nabla_{\mathbf{W}^{(l)}} J = \nabla_{\mathbf{z}^{(l)}} J \cdot \mathbf{a}^{(l-1)T} = \delta^{(l)} \cdot \mathbf{a}^{(l-1)T} \tag{1.68}$$

$$\nabla_{\mathbf{b}^{(l)}} J = \nabla_{\mathbf{z}^{(l)}} J = \delta^{(l)} \tag{1.69}$$

Finally, weights and biases are updated with one important hyper-parameter : the step size η, called the learning rate.

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \cdot \nabla_{\mathbf{W}^{(l)}} J \tag{1.70}$$

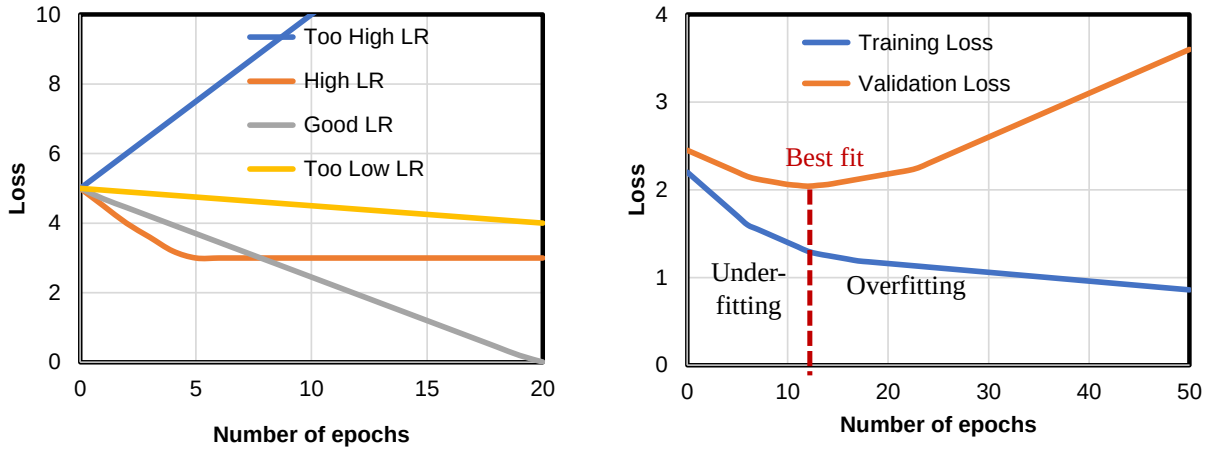
$$\mathbf{b}^{(l)} \leftarrow \mathbf{b}^{(l)} - \eta \cdot \nabla_{\mathbf{b}^{(l)}} J \tag{1.71}$$



(a) (b) (c)

FIGURE 1.21 – Impact of learning rate (LR) during the gradient descent steps. (a) LR is too low, so it requires more steps before reaching the minimum point ; (b) efficient learning with an optimal LR ; (c) large LR leads to oscillations

Fig. 1.21 illustrates the gradient descent step for different learning rates with the length of the arrow representing the learning rate value. A higher learning rate can lead to faster convergence, as the algorithm takes larger steps in the parameter space towards the minimum. However, using a learning rate that is too high can lead to instability because the algorithm may oscillate or even diverge entirely. Fig. 1.22 which represents the loss evolution during training for different learning rate values, also illustrates the impact of learning rate. If the chosen learning rate is too high, the loss can explode and the minimum will never be reached, but if it is too small, the algorithm can be stuck in a local minimum or may be too slow to converge.



(a) (b)

FIGURE 1.22 – Evolution of loss (a) during training stage using different learning rates ; (b) for the training set and validation set

Aside from the gradient descent algorithm, another very popular approach is the Adam (Adaptive Moment Estimation) [36], its popularity is mainly due to its high performance among several optimizers. Adam optimizer works similarly as the gradient descent algorithm but it adapts the learning rate for each parameter based on the historical gradient information. This allows larger updates in sparse gradients and smaller updates in dense gradients. It also incorporates momentum, which helps the optimizer move faster in the relevant directions and smoothes out the optimization process. It also helps to avoid sub-optimal solutions in which a local minimum instead of a global minimum is reached. Adam has demonstrated good performance and robustness across a wide range of machine learning tasks specially when the neural network is deep.

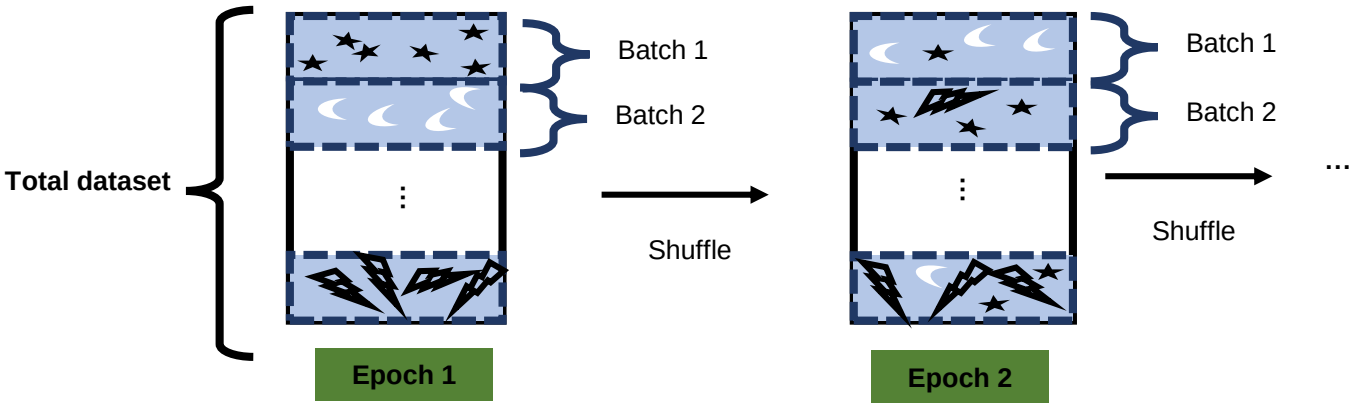


FIGURE 1.23 – Batch processing : dataset are divided into several mini-batches and a random shuffling is applied before each epoch to remix them

We can perform the update of parameters during this back-propagation step with all the data set or with batches of the data set. Batch is a subset of the data set that can be used in one iteration. Batch processing has several advantages : for a large data set, processing the entire data set can be be computationally expensive and slow

and requires a large amount of memory. Furthermore, mini-batches can fully take advantage of parallel computing, specially when using modern hardware such as GPUs and TPUs.

Consequently, to exploit the benefits of mini-batch gradient descent, the training data sets are divided into mini-batches then they are fed to the ANN through the forward pass, as depicted in Fig. 1.23. Hence, N in Eq. (1.64) and Eq. (1.65) can either be the number of total samples (for batch gradient descent when the parameter-update step only happens after all training samples have been evaluated) or the number of samples in a batch (for mini-batch gradient descent that computes the error for each mini-batch during the forward pass). Random shuffle before each training epoch reveals beneficial, so that the data set can be remixed and divided to reduce the bias and possible data imbalances in batches. The forward pass followed by the backward update step over the entire training data set is called an epoch.

Since the available data set is never of infinite size, we need to pay attention to two common problems, over-fitting and under-fitting. Over-fitting leads to a trained model which outperforms with training data but gives a very poor result with new data. This situation has to be avoided because a good model should have the ability to generalize. Fig. 1.24 illustrates three possible scenarios for training a model with the given data set. When the model under-fits, the performance will be poor on training data and unseen data but if the model over-fits, the performance will be very good over the training data set but will not give the best prediction for unseen data. To avoid over-fitting we can use the validation data set to stop the training right after reaching the sweet point (number of epochs that ensure the smallest errors on both validation and training dataset). This method is called early-stopping. As shown in Fig. 1.22(b), after a certain number of epochs, if we keep on training the model with the same data set, the loss will still decrease for the training set but it becomes flat or even begins to increase for the validation data set. Another way to prevent over-fitting is to use a regularisation method by adding a regularization term in the cost function or using other techniques like dropout [37] where neurons and connections of the network are randomly selected to be ignored with a given probability, during an epoch of the training stage. This helps in avoiding the co-adaption problem (i.e. when some layers correct mistakes from prior layers instead of contributing completely to the ANN), and makes the network more robust. On the other hand, under-fitting occurs when the model did not learn correctly the mapping between the input and the output. This situation leads to poor prediction performance on both training data and new data. This can happen, for instance, when the ML model does not fit to solve the problem (the model may be too simple model with respect to the complex problem or the hyper-parameters may not be optimized) or when the used data set is inconsistent.

1.2.3 Hyper-parameters

A hyper-parameter is a high-level parameter that concerns either the learning algorithm (for instance the learning rate, the cost function, the optimization algorithm, etc.) or the model structure (the number of neurons and layers, the

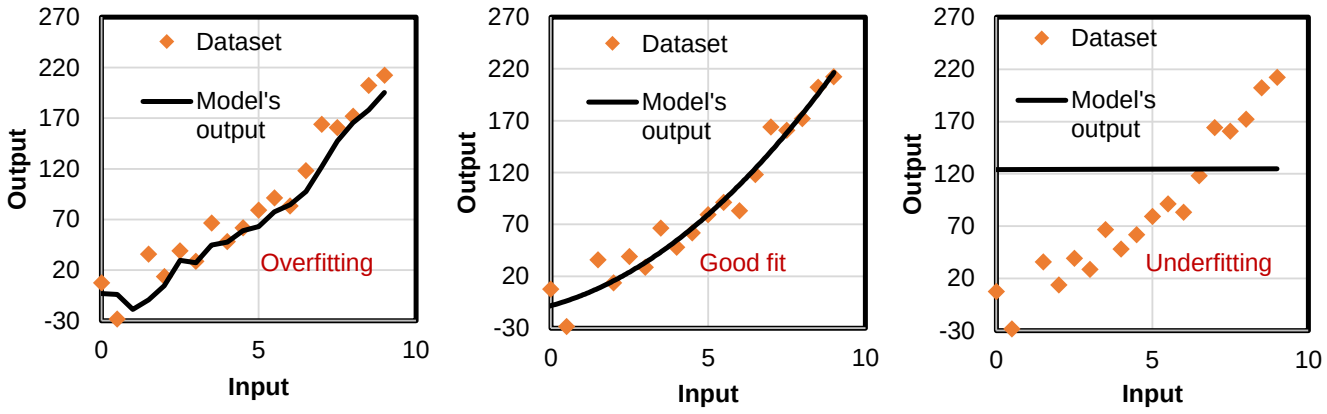


FIGURE 1.24 – The trained model (a) over-fits the data set ; (b) is a good fit ; (c) under-fits the data set.

used activation functions, etc.). They are called hyper-parameters to distinguish them from the model parameters (weights and biases for ANN) and so are not trainable parameters.

However, hyper-parameters are as important as the model selection itself, as they have a tremendous impact on the performance of the trained ML model. A tuning step is needed to choose carefully the best combination of all hyper-parameters. The validation data set can be useful for this process, as we can compare the performance of all trained models (using the same data set, under different hyper-parameter configurations) with the same unseen data set. This also explains why sometimes we need an independent and non-overlapping test data set, so that the best chosen model can be tested for performance evaluation. Several methods can be used to achieve a proper selection of the hyper-parameters, such as the grid search or random search. Grid search consists of defining a vector for each hyper-parameter in the search space, and performs an exhaustive search over all the possible combination. This method is basic but not efficient. [38] showed that random search leads to a better result. Random search consists in a random sampling of each hyper-parameter under a given statistical distribution (for instance the normal distribution) in the search space.

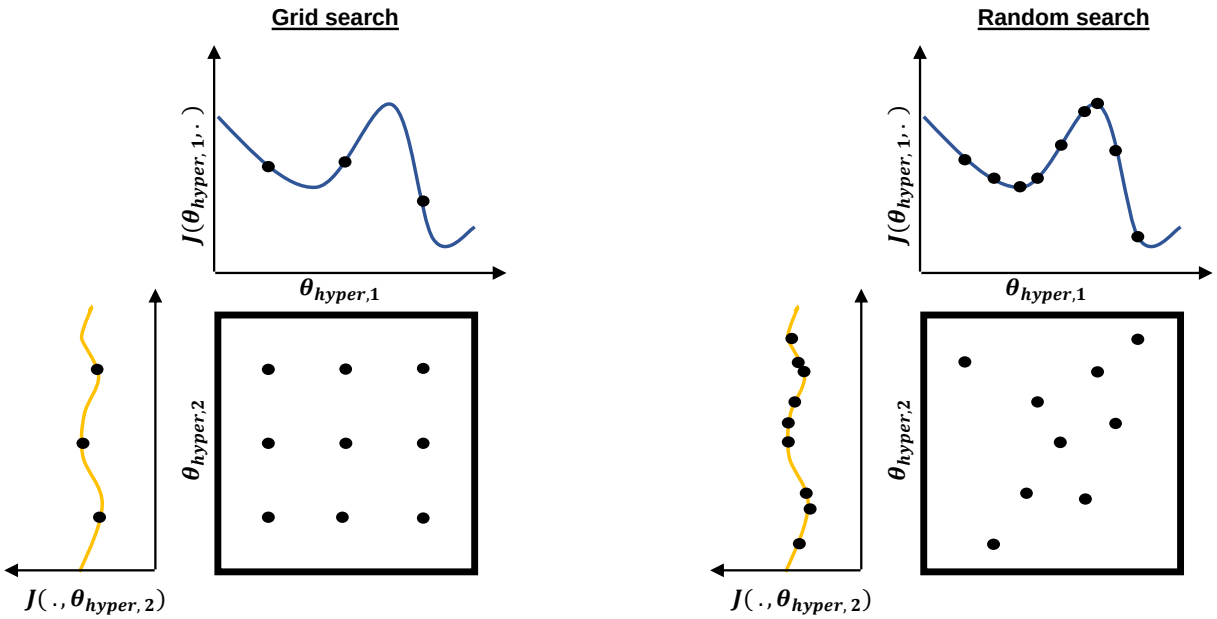
In Fig. 1.25, we show an illustration of grid and random search for 2 hyper-parameters $\theta_{hyper,1}$ and $\theta_{hyper,2}$ over the search space (black square). We also plotted the evolution of the cost function $J(\theta_{hyper,1}, \theta_{hyper,2})$ with respect to each hyper-parameter. We can assume that the total cost function is a combination of $J(\theta_{hyper,1}, .)$ and $J(., \theta_{hyper,2})$. From the curves in the figure, we can call $\theta_{hyper,1}$ the important hyper-parameter as the cost function J mainly depends on it and $\theta_{hyper,2}$ is a less important hyper-parameter. This behavior is also observed in the case of general problems with multiple hyper-parameters. With this conceptual illustration of the hyper-parameter tuning, we can notice that the grid search is inefficient compared to the random search. Indeed, only three values of the important hyper-parameter $\theta_{hyper,1}$ are sampled in nine trials whereas for the random research method, nine samples of $\theta_{hyper,1}$ can be generated, making the search for the best hyper-parameters configuration more effective. The global

minimum of the cost function J is closer to one of the nine sampled configuration using random search than using grid search.

In summary, a grid of points can provide an even coverage in the original multidimensional space, but projections onto the different hyper-parameters sub-spaces (here, the $\theta_{hyper,1}$ or $\theta_{hyper,2}$ sub-spaces) produce an inefficient coverage of the subspace itself. In contrast, random points are slightly less evenly distributed in the original multidimensional space, but far more evenly distributed in the hyper-parameter sub-space.

In addition, one can choose to stop the search based on different metrics : for instance, when the maximum number of tryouts is reached or when we reach an acceptable error. Another effective stopping way is letting the random search process run until the minimum error for the validation data set (over all trials) stops decreasing.

Bayesian optimization methods are also promising solutions to choose the hyper parameters [39] by using knowledge from the previous tryout to choose the candidate hyper-parameter values for the next try. This iteration process continues until the convergence to the minimum of J.



(a) (b)

FIGURE 1.25 – Grid search and random search of nine trials for optimizing a function $J(\theta_{hyper,1}, \theta_{hyper,2})$ in the search space represented by the black square.

1.2.4 Impact of data quality

In addition to the previously described techniques for training robust ML models, high-quality data is essential to yield accurate ML models. However, raw data is often noisy, incomplete, or contains outliers ; hence analyzing data and pre-processing it are critical steps before feeding it to the model.

Data cleaning ensures the reliability and accuracy of the training data, that directly impacts the model's performance. Data cleaning consists of pre-processing the data set to deal with missing values, outliers, and other data inconsistencies. Outliers can also be detected and either removed or treated separately.

Data imbalance occurs when the distribution of classes in the data set is significantly skewed, with one class being much more prevalent than others. This can lead to less accurate and biased models that favor the majority class. For instance, for a binary-classification problem with 2 classes A and B, if the proportion of labeled data from class A is highly disproportional compared to B. The trained model will also tend to label the predicted result as from the class A. To address data imbalance problem, various techniques can be employed. One of them is oversampling involving replicating examples from the minority class. Random under-sampling that removes examples from the majority class is also often useful. These techniques aim to improve the model's performance on underrepresented classes. Data augmentation, for the minority class data, can also achieve the goal by increasing the amount of data by generating new data points from existing data. For instance, for tasks involving image inputs, different image processing techniques can be used including gray-scaling, re-scaling, flipping, etc.

In regression tasks, the range of output values (and input values) has also a important impact, especially when the values show a significant difference in orders of magnitude. If the target data has a wide range with extremely small but also large values, then normalization (or standardization) method combined with a logarithmic function to pre-process the data is one effective way to help in compressing the data. Once the model has been trained, an inverse transformation is needed to convert and scale the predicted values into the original values.

1.3 Summary

In this chapter, first, we have introduced key elements of the optical coherent communication systems using WDM architecture. We have presented the models of the different sources of impairments during the propagation of the transmitted signal in a long-haul fiber-optic link as well as the DSP techniques that are used to recover the signal. The different metrics to quantify the final system performance were also recalled. Different optical amplification techniques were also presented with their benefits and disadvantages. Then, we have introduced the ML methods and particularly described the structure of ANNs and important elements of their training procedure leading to an accurate model that meets the needs of our applications that will presented in the next chapters.

Chapter 2

Raman amplifier design

This chapter focuses on the study of a machine learning based method for the design of an S+C+L UWB Raman amplifier in a hybrid Raman/SOA scheme. In the first part, we demonstrate experimentally the ability of an ANN to predict the overall loss profile taking into account several effects including fiber attenuation, inter-channel SRS and backward pumping, for any Raman pump current configuration. In the second part, we use another ANN to design the UWB Raman amplifier, by predicting the required pump currents given a target loss profile, and we experimentally validate our method in a 3×100 km SSMF configuration. Finally, we evaluate the performance of this inverse design ML method to realize an arbitrary pre-selected loss profile for a real-case multi-span transmission setup.

2.1 Introduction

Raman amplifiers were firstly proposed in 1972 [40], but their applications started after 2000s thanks to the development of high power and compact lasers. Instead of lumped amplification, one or multiple high power lasers are injected inside the in-line transmitted fiber (i.e they co/contra-propagate with the signal), called pump lasers so their power can be transferred to lower frequency transmitted signal hence amplifying the signal. Raman amplifiers present a significant advantage over EDFAs in terms of noise figure (NF) and their ability to provide arbitrary gain profiles over a broad bandwidth.

In subsection ii, we will see that the power evolution is governed by Eq. (2.5). To obtain the gain profile and NF of a Raman amplifier, the key parameters are : the number of pump lasers, their power and their frequency values, and the input powers of the WDM channels. This flexibility allows Raman amplifier to deliver arbitrary gain profiles for ultra-large band systems. Nevertheless, for each desired Raman amplifier configuration, heavy and time-consuming computations are still needed, if one uses a numerical method to solve simultaneously the set of coupled ODEs for all WDM channels. Besides, the SRS effect, which causes a power transfer from higher frequencies to

lower frequencies, also need to be considered in UWB systems. Fig. 2.1 illustrates different loss profiles, measured form a 100-km SSMF span output and input spectra, with different Raman amplifier power configurations in a contra-propagating scheme. The loss profiles account for backward Raman amplifier gains, fiber loss and impact of SRS. To account for SRS, given a WDM comb with N channels, an additional set of coupled ODEs, one for each transmitted channel, has to be solved likewise. Consequently, Raman amplifier gain, combined with fiber attenuation and inter-channel SRS effect, make the prediction of the spectral profile even more complex, because it requires a heavy computational effort.

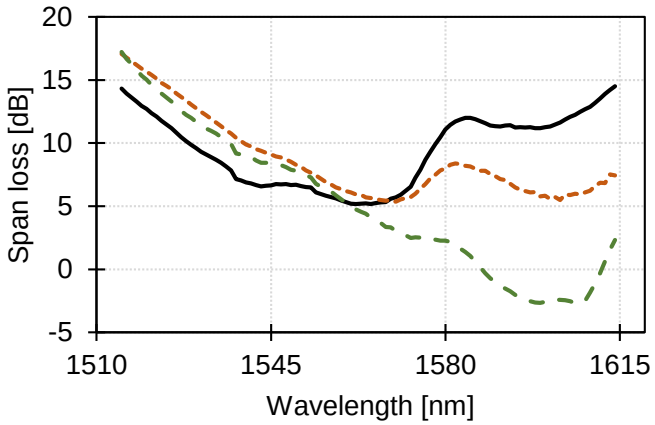


FIGURE 2.1 – Loss profiles for different backward Raman amplifier configurations

2.2 Raman effect in fiber optics

The spontaneous Raman scattering was named after Sir Chandrasekhara Venkata Raman, who discovered the effect in 1928. Photons of the incident optical field can be scattered by quantized vibrations called *optical phonon* generating a lower-frequency photon. Raman scattering can occur in all materials but in optical fibers, the dominant effect is caused by the Si-O-Si bound.

Raman scattering can also be stimulated by a pump leading to the nonlinear phenomenon of stimulated Raman scattering (SRS) that was briefly described in subsection ii of section 1.1.4. The initial signal photon is thus amplified by the pump photon. SRS can occur both in forward and backward directions.

2.2.1 Intra-channel SRS

For WDM signal transmission scheme, the SRS is an undesirable effect, specially for broadband transmission configuration. Indeed, an equalized WDM channels configuration is essential. However, this intra-channel SRS (I-SRS) effect induces a power tilt with an energy transfer of the optical propagating waves at higher frequency consi-

dered as "pumps" to waves at lower frequency considered as "signals", as shown in Chapter 1. The accumulated tilt effect leads to a degraded performance of some channels at higher frequencies in long distance transmissions.

The power evolution of the WDM signal propagating in the same positive direction (forward) is governed by a set of M coupled ordinary differential equations (ODE) with M being the total number of all frequency components [41, 6, 42]. The average power of each optical wave $P(\nu)$ at frequency ν obeys :

$$\begin{aligned}\frac{dP(z,\nu)}{dz} = &- \underbrace{\alpha(\nu)P(z,\nu)}_{\text{attenuation}} \\ &+ \int_{\mu>\nu} \left\{ \underbrace{C_r(|\nu-\mu|)P(z,\mu)P(z,\nu)}_{\text{Raman stimulated emission}} + \underbrace{2h\nu\, C_r(|\nu-\mu|)P(z,\mu)\left[1+n_{sp,r}(\mu-\nu)\right]}_{\text{Raman spontaneous emission}} \right\} d\mu \\ &- \int_{\mu<\nu} \left\{ \underbrace{\left(\frac{\mu}{\nu}\right) C_r(|\nu-\mu|)P(z,\mu)P(z,\nu)}_{\text{Raman stimulated absorption}} + \underbrace{4h\nu C_r(|\nu-\mu|)P(z,\mu)\left[1+n_{sp,r}(\nu-\mu)\right]}_{\text{Raman spontaneous absorption}} \right\} d\mu\end{aligned} \tag{2.1}$$

with $\alpha(\nu)$ the attenuation in [km^{-1}], P power in [W], $C_r(|\nu-\mu|)$ the Raman gain efficiency of the fiber between frequencies μ and ν, expressed in [W^{-1}·km^{-1}]. It is defined as follows :

$$C_r(|\nu-\mu|) = \frac{g_r(|\nu-\mu|)}{A_{eff}} \tag{2.2}$$

with g_r the measured Raman gain coefficient which is specific to the fiber and expressed in [m/W] and A_{eff} the effective core area. The Raman gain does not depend on the direction of the pumps and the signals, it mainly depends on the frequency difference between them. The normalized g_r of an SSMF is illustrated in Fig. 2.2. The gain bandwidth is over 40 THz wide, with the dominant peak near 13 THz frequency difference.

The temperature-dependant phonon occupancy factor $n_{sp,r}$ is defined as :

$$n_{sp,r}(\Delta\Omega) = \frac{1}{e^{\frac{h\Delta\Omega}{k_B T}} - 1} \tag{2.3}$$

with h the Plank's constant, k_B the Boltzmann constant,T the temperature of the fiber in kelvin and $\Delta\Omega$ the frequency difference.

This model includes physical effects such as attenuation, stimulated Raman scattering, spontaneous Raman scattering and its temperature dependence. The term $\left(\frac{\mu}{\nu}\right)$ accounts for the energy difference between photons at frequency μ and ν. The factor of 2 in the first term accounts for the two polarization modes of the fiber. An additional factor of 2 in the second term includes emission in both the forward and the backward directions, after the spontaneous absorption of photons at frequency ν that causes a depletion of the signal power $P(\nu)$.

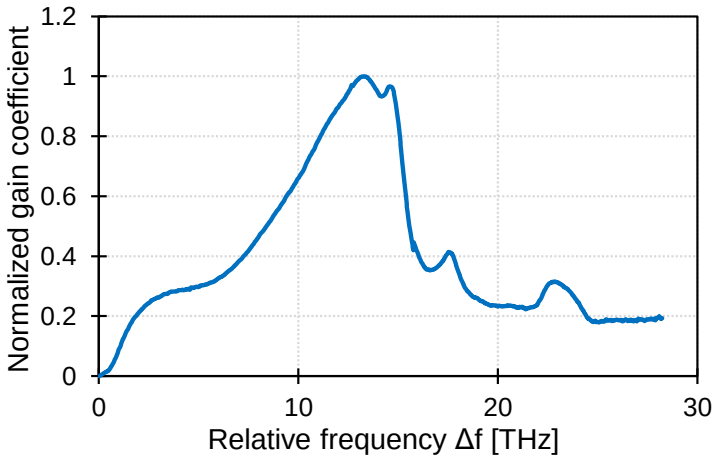


FIGURE 2.2 – Normalised Raman gain at 1550 nm for a SSMF adapted from [40]

2.2.2 Raman amplifier

Compared to the adverse effect of I-SRS, we could benefit from the SRS effect by using additional external high power pumps to amplify the propagating optical signal and thus build a so-called "Raman amplifier".

i Raman amplifier types

Raman amplifiers are mainly divided into two classes : distributed amplifiers and discrete amplifiers. For the discrete scheme, a lumped element is inserted into the transmission line as a gain medium. Shorter-length fiber span (a few kilometers) with a smaller effective area A_{eff}, hence larger Raman gain efficiency C_r such as dispersion compensation fibers (DCF) and highly nonlinear fibers, are commonly used. However, those fiber's CD and nonlinear coefficient γ can also significantly affect the generation and accumulation of nonlinear noise. On the other hand, the distributed amplification gained in popularity because it does not require any additional special fibers and can produce less nonlinear noise. In a distributed configuration, the amplifying medium is the transmission optical fiber itself. Fig. 2.3 shows a simple schematic of a distributed Raman amplifier in a a bi-directional configuration with one co-propagating pump laser and another one in a contra-propagating mode.

Since the Raman amplification does not depend on the propagation direction, the pumps and signals can propagate in different directions. We denote with a "+" sign the forward propagation and a "−" sign the backward one. For a signal propagating in the positive direction ($z > 0$), "+" denotes the co-propagating pumps (same direction than the signal) and "−" denotes the contra-propagation pumps (opposite direction of the signal). A bi-directional amplifier employs both co-propagating and contra-propagating pumps. The contra-propagating scheme is more common since it reduces the total optical power at the beginning of the fiber thus causing less nonlinear distortions because the pump power is not added to the signal power like when using the co-propagating scheme.

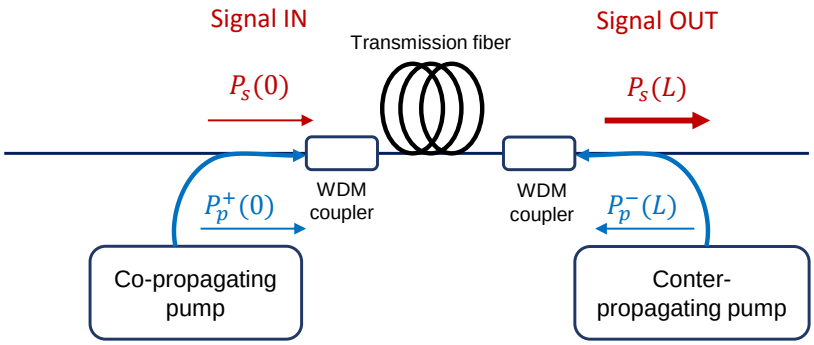


FIGURE 2.3 – Schematic illustration of distributed Raman amplifier with one single pump (forward or backward mode)

ii Raman amplification ODEs

Similarly to Eq. (2.1), the pump's power and signal's power evolutions are governed by a set of ODEs. Since the lasers used in the Raman amplifier operate at high power, the spontaneous Raman effect can be ignored because it would be weak compared to the other terms derived from the SRS effects.

Under this configuration with one pump and one signal, the set of two coupled ODEs that governs the evolution of pump and signal powers is expressed as follows :

$$\begin{cases} \frac{dP_s}{dz} = -\alpha(\nu_s)P_s + C_r(|\nu_s - \nu_p|)P_p^{\pm}P_s \\ \pm \quad \frac{dP_p^{\pm}}{dz} = -\alpha(\nu_p)P_s^{\pm} - \left(\frac{\nu_s}{\nu_p}\right) C_r(|\nu_s - \nu_p|)P_sP_p^{\pm} \end{cases} \tag{2.4}$$

with P_s the signal power at frequency ν_s and P_p the pump power at frequency ν_p. The sign $\pm$ corresponds to co-propagating pump P_p^+ and counter-propagating pump P_p^-, respectively, since the signal wave is propagating in the forward direction.

We can generalize Eq. (2.4) for a broadband Raman amplification. The optical power of each pump or signal component at frequency ν is governed by :

$$\begin{aligned} \pm\frac{dP^{\pm}(z,\nu)}{dz} =& -\alpha(\nu)P^{\pm}(z,\nu) \\ &+ \int_{\mu>\nu} \left\{C_r(|\nu-\mu|)P_{all}(z,\mu)P^{\pm}(z,\nu)\right\} d\mu \\ &- \int_{\mu<\nu} \left\{\left(\frac{\mu}{\nu}\right) C_r(|\nu-\mu|)P_{all}(z,\mu)P^{\pm}(z,\nu)\right\} d\mu \end{aligned} \tag{2.5}$$

where the sign $\pm$ corresponds to forward-propagating wave and backward-propagating wave respectively and the total pump power at frequency μ (accounting forward-propagating and backward-propagating waves) $P_{all}(z,\mu) = P(z,\mu)^+ + P(z,\mu)^-$.

Consequently, to obtain numerically the Raman amplifier gain profiles, one need to solve a set of $M+N$ coupled

ODEs with M frequency components of the signal and N laser pumps used in the Raman amplifier. Thus, the computation of the gain profiles are very time-consuming.

Like all optical amplifiers, Raman amplifiers also produce ASE noise generated by spontaneous Raman scattering and amplified by SRS. Similar to Eq. (2.1), the evolution of the ASE power $P_{ASE}(\nu)$ at frequency ν of a Raman amplifier is given by [6] as follows :

$$\begin{aligned}\pm\frac{dP^{\pm}_{ASE}(z,\nu)}{dz} = & \\ & -\alpha(\nu)P^{\pm}_{ASE}(z,\nu) \\ & +\int_{\mu>\nu}\left\{C_r(|\nu-\mu|)P_{all}(z,\mu)P^{\pm}_{ASE}(z,\nu)+2h\nu\, C_r(|\nu-\mu|)P_{all}(z,\mu)\left[1+n_{sp,r}(\mu-\nu)\right]\right\}d\mu \\ & -\int_{\mu<\nu}\left\{\left(\frac{\mu}{\nu}\right)C_r(|\nu-\mu|)P_{all}(z,\mu)P^{\pm}_{ASE}(z,\nu)+4h\nu\, C_r(|\nu-\mu|)P_{all}(z,\mu)\left[1+n_{sp,r}(\nu-\mu)\right]\right\}d\mu\end{aligned} \tag{2.6}$$

By ignoring the absorption effect of the ASE, the generation and amplification of total ASE is expressed as follows :

$$\begin{aligned}\pm\frac{dP^{\pm}_{ASE}}{dz} = & \underbrace{-\alpha(\nu)P^{\pm}_{ASE}}_{\text{attenuation}} \\ & +\int_{\mu>\nu}\left\{\underbrace{C_r(|\nu-\mu|)P_{all}(z,\mu)P^{\pm}_{ASE}}_{\text{amplification through SRS}}\underbrace{+2h\nu\, C_r(|\nu-\mu|)P_{all}(z,\mu)\left[1+n_{sp,r}(\mu-\nu)\right]}_{\text{generation through spontaneous emission}}\right\}d\mu\end{aligned} \tag{2.7}$$

with the boundary conditions that $P^{+}_{ASE}(0)=0$ and $P^{-}_{ASE}(L)=0$

Finally the NF is defined as [6] :

$$NF \approx \frac{2P^{+}_{ASE}(L)}{h\nu B_{ref}G_{net}}+\frac{1}{G_{net}} \tag{2.8}$$

with B_{ref} being the reference bandwidth and where the net gain is defined as follows :

$$G_{net}=\frac{P_s(L)}{P_s(0)} \tag{2.9}$$

For a transmission link composed of several fiber spans, we can use the chain rule in Eq. (1.34).

iii Advantages and challenges

Distributed Raman amplification exhibits several benefits compared to lumped amplification using EDFAs. First, it has been shown that the Raman amplifier has a lower total NF [6]. To understand this NF reduction, we can consider a link composed of one single fiber span and an amplifier as an example. For distributed Raman amplification, the signal and ASE are both attenuated (cf. Eq. (2.7)) in the fiber whereas for a lumped amplifier, only the signal will

be attenuated as the amplifier is placed at the end of the fiber span. Moreover, a multi-pump amplification scheme makes the design of broad-band amplifiers with a large gain bandwidth possible. This advantage makes it popular for ultra wide-band transmission experiments. Finally, another advantage is the flexibility of the gain spectrum design. A Raman amplifier achieving a desired gain spectrum can be tailored by adjusting the pump wavelengths and powers. However this advantage also leads to one of the main challenges in designing the adequate configuration since it involves solving a set of coupled ODE though time-consuming numerical solvers.

2.2.3 Hybrid Raman amplification

Conventional amplification methods such as Raman amplifiers, EDFAs and SOAs were used in deployed WDM networks. Each of these amplification techniques has its own drawbacks and benefits. The main issue of SOAs is their production of a large amount of ASE as well as the introduction of serious distortions to the optical signal. On the other hand, EDFAs produce less signal distortions and less ASE but their gain spectrum bandwidth is narrower, only about 40 nm. Besides, EDFA's gain profiles are not flat. For Raman amplifiers, they introduce even less noise and they allow for the achievement of flexible gain profiles as it can be adjusted by varying the number of pumps as well as the frequency and power of each pump. However, high power pumps are often needed to achieve high gains so its use as a unique amplification solution is not economically viable. In addition, high power pumps can also induce higher nonlinear distortions which impact the system performance.

Hybrid optical amplification techniques are a promising technology for future WDM multi-terabit systems providing good transmission performance and energy efficiency [2, 3, 4, 5]. Indeed, hybrid optical amplifiers are designed to achieve large gain bandwidth, to minimize the impairments, due to ASE or nonlinear distortions, and to maximize the transmission length. Hybrid Raman/EDFA [43] and hybrid Raman/SOA [44, 45] schemes showed promising results for UWB transmission and achieved world-record transmission capacities. For instance in [45], they achieved a 107-Tb/s transmission throughput over a continuous 103 nm optical bandwidth using this hybrid configuration.

2.2.4 Raman amplifier design challenges

One major advantage of a Raman amplifier is the ability to shape the spectral gain profile by changing the power and frequency of the pump lasers. However, it is problematic to inverse design the appropriate Raman amplifier, i.e. to obtain the pump lasers' configuration to obtain a desired gain profile. For instance, in a multi-span wide-band transmission system, flat-gain, tilted-gain or flat-power configurations are desired. A well-designed link should operate with almost flat power in every span which may be adjusted using Raman amplifiers. Arbitrary gain profiles are also needed in hybrid Raman amplification configurations (hybrid Raman-EDFA or hybrid Raman-SOA) and can be obtained thanks to the flexibility of Raman amplifiers.

Nevertheless, finding the right configuration of the pump lasers can be very time-consuming with numerical

methods, i.e. solving a system of nonlinear ODE, as discussed above. Moreover, the multi-dimensionality of the problem increases with the number of used pumps. Besides, the system of coupled ODE does not always guaranty a solution. So ML-based methods have been proposed to tackle this problem. Indeed, ML solutions become popular in the optical communications community and different applications have been explored in the last years [9]. From channel estimation and signal recovery to performance monitoring or prediction, ML showed its high potential [7, 9, 8]

Over the past few years, ML has also been applied to analyse and design broadband Raman amplifiers. Prior to this work, two different approaches have been proposed. First, [46] proposes to address this problem using iterative optimization algorithms such as genetic algorithms. However, one major drawback of this method is that every small change in the system configuration requires to restart the full time-consuming and high-latency optimization process. Besides, the initialization has a high impact on its performance and a solution is not always guaranteed. Other methods based on optimization loops [47] also encountered the same issue because they require multiple iterations to find an optimized solution and they have to be rerun for each desired Raman gain configuration.

Another approach is to directly use ML to learn the mapping between the Raman amplifier configuration parameters and the resulted gain profile. Once the ML model has been correctly trained, arbitrary gain profile can be used as the input of the model to make a prediction of the corresponding pump configuration. One major benefit of this method, compared to the optimization algorithm, is that the prediction process, called inference in ML language, is very fast. Another one is that it does not require re-training the ML model when the desired gain profile of the Raman amplifier changes, for fixed operation conditions (same transmission setup, same propagation orientations, etc.).

Prior to this work, the training dataset was essentially generated from pure simulations and ML-based models were mainly evaluated with a simulated setup [48, 49, 50]. In [51], the authors used experimental data to train and evaluate the ML framework for the Raman amplifier design in a hybrid Raman/EDFA scheme, but only in C+L band. In our study, we demonstrated an ML method for UWB Raman amplifier, in a hybrid Raman/SOA configuration, with experimental data. We show that the method is fast and accurate over a 100 nm S+C+L UWB spectrum with 5 Raman pumps. Not only the approach considers the Raman amplifier gain, but also takes into account the attenuation in the fiber span and the I-SRS effect.

2.3 Experimental setup

To generate the experimental data, we use the setup in Fig. 2.4. Three different ASE noise sources and an UWB WSS generate a 100 nm continuous spectrum, spreading from 1515 to 1615 nm. This spectrum is amplified by an UWB SOA before being sent to the fiber under test, which is a 100 km SSMF fiber span. We use backward Raman amplification with 5 pumps per polarization located at 1410, 1435, 1455, 1490, and 1510 nm. A $99\%/1\%$ coupler, an

optical switch and an optical spectrum analyzer (OSA) are used to obtain the UWB power spectrum at the span input and output respectively as shown in Fig. 2.4. Input and output calibrated spectra are respectively shown in black and green lines in dB scale in Fig. 2.4. We define the span loss profile as the difference between those power profiles, as represented in red in Fig. 2.4. This loss profile accounts for the backward Raman pumping contribution which depends on the values of the 5 pump currents, but also on the fiber attenuation and on the SRS occurring in high power UWB transmission. We set the same current value on both polarization states for each wavelength.

In this work, we operate the SOA to provide, at the fiber input, a 21 dBm optical spectrum with a 6 dB tilt over the whole bandwidth, to meet the operation conditions of a previous transmission experiment from our group [52]. We then randomly and independently choose the values of each of the pump currents from a uniform distribution between $[200, 1500]$ mA and measured 10000 random configurations. The data set has been arbitrarily split into 80% for training and 20% for validation respectively.

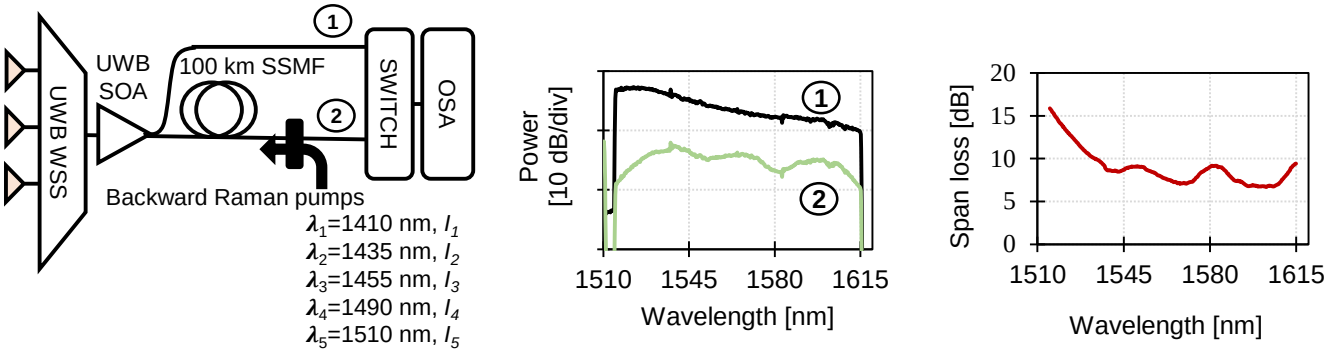


FIGURE 2.4 – Setup for a single span SSMF with 5 pumps for data collection, training and validation of the ML model.

2.4 Machine learning-aided model

ML can be a powerful tool in solving our problem. Indeed, if a small and simple ML model can be correctly trained, then the computation time for the predictions will be largely reduced without degrading the accuracy. So the most challenging part is the choice of the ML model (trade-off between accuracy and speed), and the way to train efficiently the chosen model. As presented in section 1.2.1 of Chapter 1, ANN is a popular ML model for regression problems thanks to its simple structure and very good results. In what follows, we will test and evaluate the performance of two ANN models for two use-cases.

2.4.1 Generative model

First, we aim to find a model that predicts the loss spectrum of the Raman amplifier (including Raman amplifier gain, SRS and attenuation in the span). We call this model "generative model" and we will use it for the design

of transmission lines, for instance to predict the behaviour of the Raman amplifier. Fig. 2.5 illustrates its principle and its ANN architecture. The ANN for the generative model is composed of an input layer with $n = 5$ neurons (corresponding to the number of Raman pump currents used in the experiment), an output layer with $m = 100$ neurons (corresponding to the spectral locations) to compute the loss profile and 2 hidden layers with $p = 150$ neurons. The activation function was selected to be ReLU for the hidden layers and linear for the output layer. Those hyper-parameters are chosen following a random search method. RMSE is the used cost function defined as :

$$J_{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2.10}$$

with y_i and $\hat{y}_i$ are the target and predicted loss values in dB.

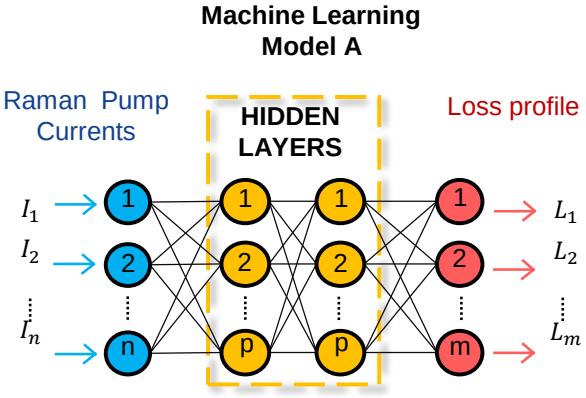


FIGURE 2.5 – ANN architecture for generative model A

After the training stage with 8000 configurations, we evaluate the performance of our ANNs over a validation set containing the remaining 2000 configurations. For each pump configuration of the validation set, we predict the loss profile with the ANN. Fig. 2.6 shows the prediction error distribution for 10 wavelengths in the spectrum : the ends of the whiskers indicate the 5^{th} and 95^{th} percentiles of the population, the box captures the half population between the 1^{st} quartile and 3^{rd} quartile, and the horizontal line inside the box indicates the population median. Outside cross markers correspond to predictions that are considered as outliers. We observe that for all wavelengths, the median error does not exceed ± 0.2 dB, and that 90% of the validation set show prediction errors less than ± 0.6 dB. To evaluate the performance of our method, we use the root mean square error (RMSE), defined in Eq. (2.10).

Fig. 2.7 shows the probability density function (PDF) and cumulative density function (CDF) of the RMSE between the initial measured profile and the predicted one over the whole validation set. The mean value of RMSE is 0.25 dB. Besides, the CDF indicates that 95% of the predictions give an RMSE less than 0.51 dB. Finally, we plot in Fig. 2.8 the true spectrum (solid line) and the prediction of our ANN (diamond markers), for the best fit with an RMSE of 0.099 dB and for the case corresponding to the 95th RMSE percentile with an RMSE of 0.51 dB. This figure clearly illustrates the high accuracy of the prediction for our 100 nm-wide optical signal.

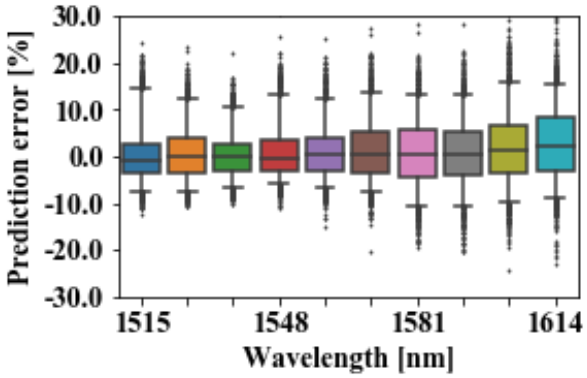


FIGURE 2.6 – Whisker plot of the loss profile error distribution vs. wavelength. The ends of the whiskers indicate the 5^{th} and 95^{th} percentiles of the population, the box captures the half population between the 1^{st} and 3^{rd} quartiles, and the horizontal line inside the box indicates the population median

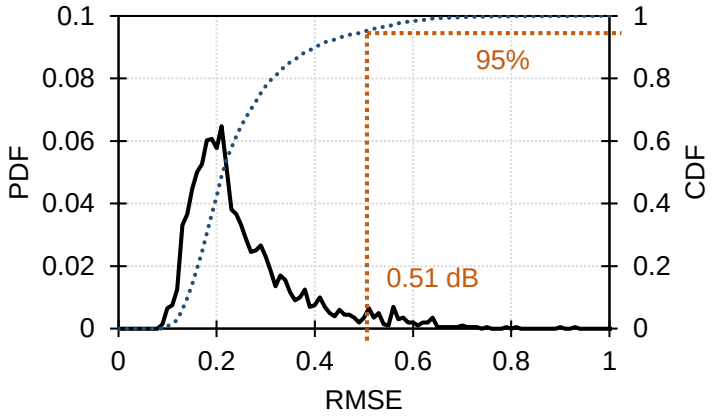


FIGURE 2.7 – RMSE of validation errors.

2.4.2 Inverse model

The role of the "inverse model" is to generate the configuration of the Raman pump according to a desired loss profile. In our case, the output of the inverse model is a vector of pump currents. Fig. 2.9 shows the chosen ANN architecture. The ANN architecture to model the inverse problem is composed of an input layer with $m = 100$ neurons, an output layer with $n = 5$ neurons and 2 hidden layers with $k = 300$ neurons. The selected activation functions are the same as the one in model A : ReLU for the hidden layers and linear for the output layer. Like the model A, those hyper-parameters are also chosen with a random search method. It is worth to notice that the number of neurons in the hidden layers is larger for the best inverse ANN model compared to the best generative ANN model ($k > p$). Indeed, the inverse design problem is more complex so it requires a larger ML model.

To evaluate the performance of our inverse model, we use the same validation set containing 2000 measurements. For each measurement in the validation data set, we feed the loss profile as the input of the ANN and we predict the pump current configuration. To assess the performance of our ANN, we first show in Fig. 2.10 (a), the

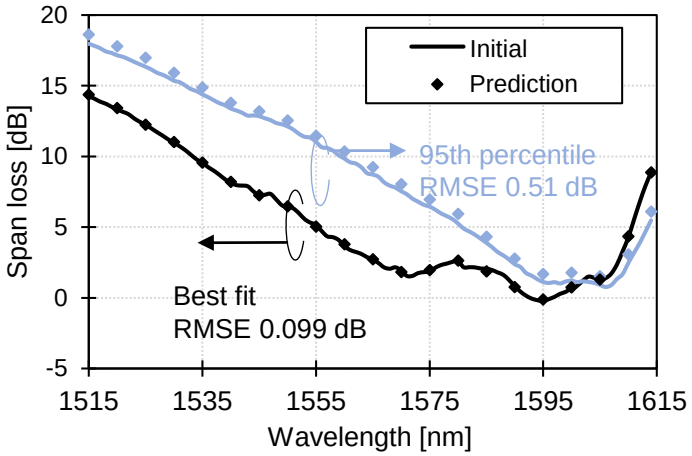


FIGURE 2.8 – Measured spectra and predicted spectra for the best prediction case and for the 95^{th}-RMSE-percentile respectively.

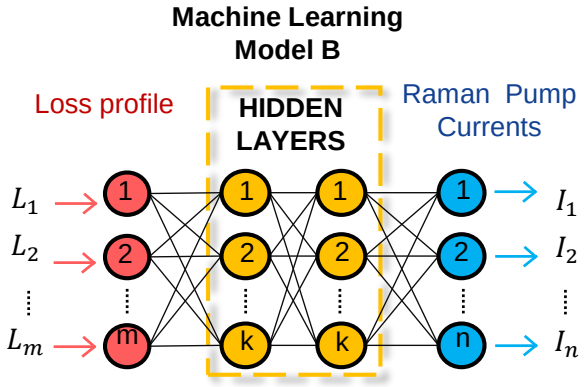


FIGURE 2.9 – ANN architecture for inverse model B.

relative current prediction errors for the whole validation set. The prediction error for the 5 currents $\{I_1, \ldots, I_5\}$ is less than $\pm 13\%$ for 90% of the cases. However, given that the solution for the current pump configuration is not unique, a high error in the pump current prediction does not always lead to a high error in the achieved loss profile. Hence, for the 2000 cases of the validation dataset, we re-measure the loss spectrum to observe the impact of current prediction error on the loss spectrum.

The proposed workflow show in Fig. 2.11 carries out the following steps in sequence :

— For the 2000 target loss profiles, we use our trained ANN to predict the currents ;

— We update the Raman amplifier's configuration with the obtained 2000 vectors of pump current vectors in our initial experimental ;

— We measure the obtained loss profile with those pump current configuration ;

— We then compute the error between the initial target loss profile and the re-measured loss profile.

The loss error distribution is given in Fig. 2.10(b) : 90% of the cases show a prediction error less than ± 0.9 dB, with 50% of the cases with ± 0.6 dB error. The PDF and CDF of the RMSE are shown in Fig. 2.12 showing good

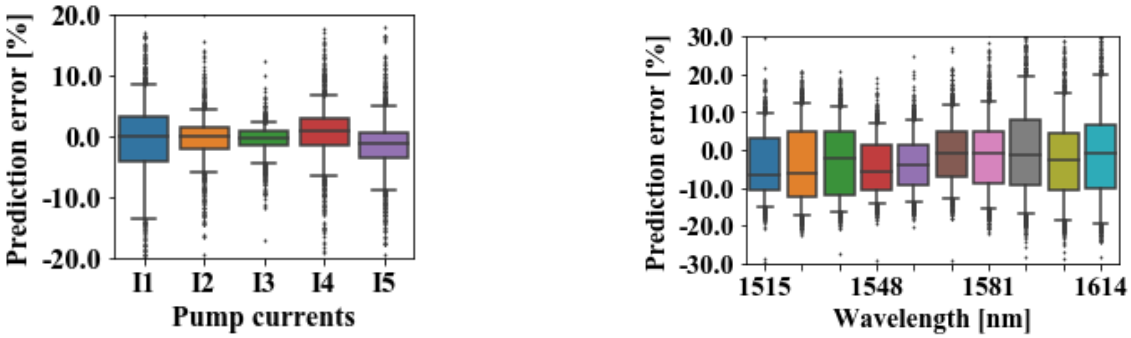


FIGURE 2.10 – Box-plot of the prediction errors : (a) pump current prediction error distribution ; (b) loss profile error distribution after re-measurement with predicted currents.

accuracy between target profile and re-measured profile after the prediction of current values, with a mean RMSE of 0.41 dB, and 95% of predictions resulting in loss RMSE less than 0.74 dB. We show in Fig. 2.13 the initial loss profiles as target profiles (solid line) and the re-measured loss profiles (diamond makers) using the predicted pump currents from the ANN model B for the best fit with an RMSE of 0.099 dB and for the case corresponding to the 95^{th} RMSE percentile, i.e. with RMSE of 0.74 dB.

For this Raman amplifier design method, we attribute the worse performance, compared to the one of the generative model, to the higher complexity of the prediction in the inverse problem and to the uncertainties of loss profile re-measurements after applying the predicted pump currents.

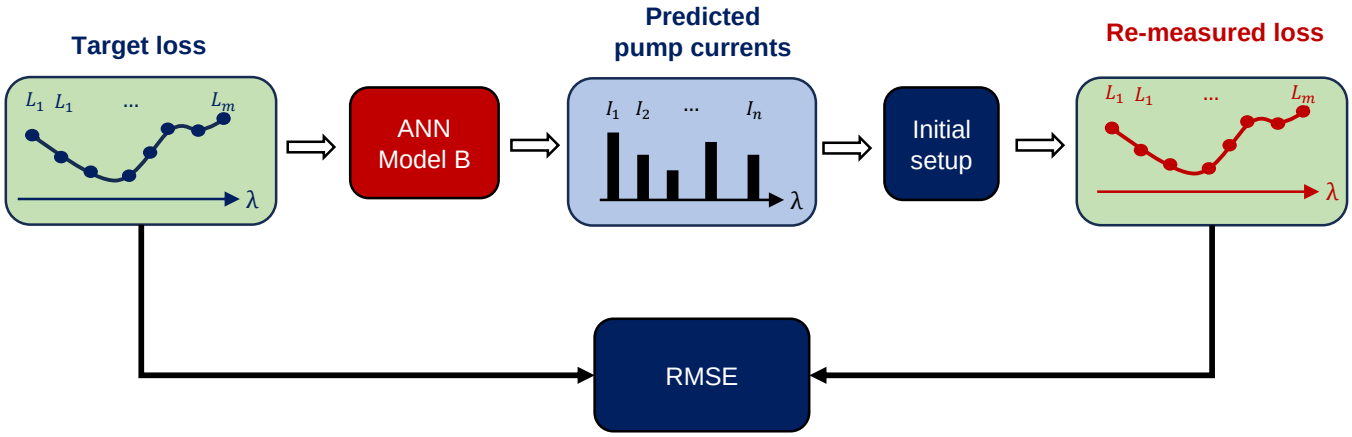


FIGURE 2.11 – The schematic flowchart to evaluate the performance of our ANN-based inverse model, the ML model B.

2.5 Application to a multi-span transmission

After analyzing the ANNs on a single-span system, we apply them in this section to optimize a multi-span transmission setup composed of a 3-spans homogeneous link so the three spans are identical. In a multi-span experiment, we usually aim at setting the transmission line such that the power spectrum is identical at each span

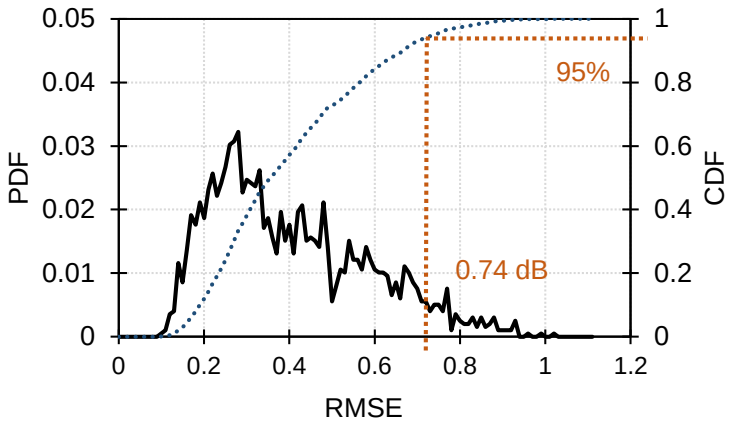


FIGURE 2.12 – RMSE of validation errors.

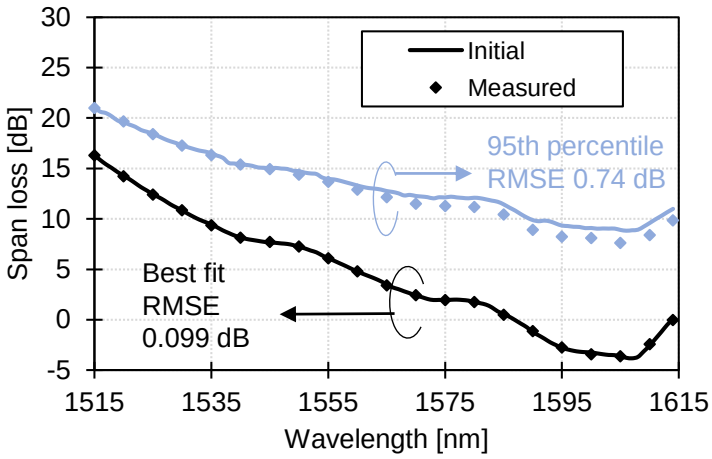


FIGURE 2.13 – Initially measured loss profile and re-measured loss profile using predicted pump currents, for the best prediction case and the one corresponding to the 95^{th} RMSE percentile.

input, as shown in the inset 1 of Fig. 2.14. With the power configuration described in a previous transmission work from our research group [52], we define our target loss profile as a linear profile with a tilt of 6 dB over 100 nm, yielding the power spectrum of the inset 2 of Fig. 2.14at the input of the next SOA.

We use the ANN of our design model to generate the corresponding pump current values, then measure the resulting loss profile for the first span and show the result in Fig. 2.15 (solid line in red). We also show on this figure the loss profile measured from an iterative optimization of the pump currents by hand-tuning without using the ML algorithm (solid line in blue). The loss profile resulting from the hand-tuned method oscillates around the target gain profile (dashed line) while the measured loss profile using ANN prediction of pump currents greatly limits the ripples. The magnitude of the oscillations, induced by the multi-pump design, is much smaller when using the trained ANN model B. The measured loss profile shows good agreement with the target loss profile, exhibiting a maximum absolute error of 1.52 dB and an RMSE of 0.76 dB. On the other hand, the hand-tuned solution shows a maximum absolute error of 3.9 dB and RMSE of 2.1 dB. We had similar results with the two other spans since the

three spans are exactly identical.

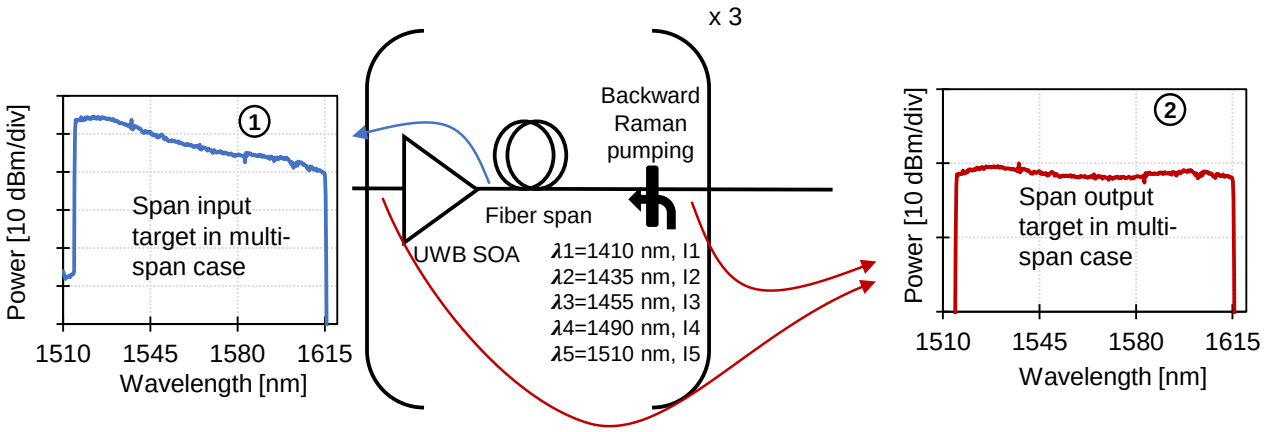


FIGURE 2.14 – Setup for multi-span usage

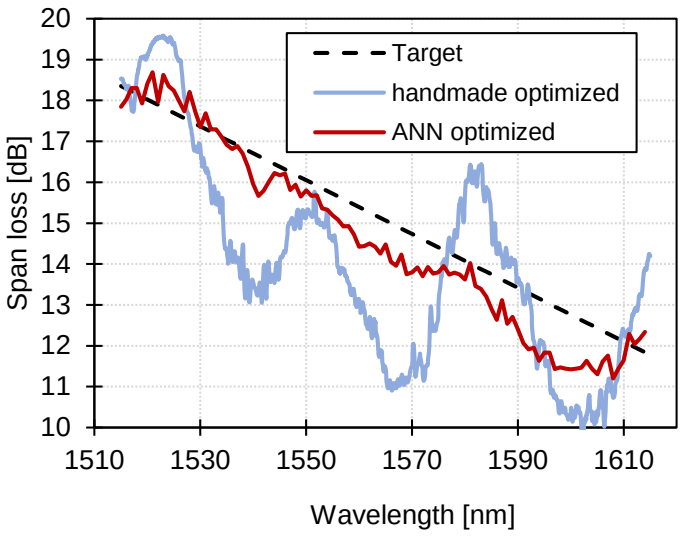


FIGURE 2.15 – Loss profile : target, measured with currents given by ANN B prediction, measured after hand tuning and iterating.

2.6 Conclusion

The aim of this chapter was to find an ML method to accurately and rapidly design Raman amplifiers in the context of UWB optical transmission system design. The key enabling technique is the ANN framework, in particular using three-layers ANN framework which presents a good compromise between complexity and precision.

In the first part of this chapter, we designed and experimentally tested the performance of our ANN model which predicts the 100 nm-wide loss spectrum given a set of pump currents. We achieved a mean RMSE of 0.25 dB with 95% of predictions with an RMSE less than 0.51 dB.

In the second part of this chapter, we described the inverse design ANN framework to predict the pump configuration according to a target loss profile and characterized its performance by fitting the predicted pump current vector to the initial setup and comparing the re-measured loss profile to the initial one. We achieved a mean RMSE of 0.41 dB with 95% of predictions with an RMSE less than 0.74 dB.

Finally, we applied our solution to a real-case transmission experiment and achieved outstanding performance. Following the requirements of previous experiments, we wanted the ANN to deliver a target loss profile for which we ignored whether a plausible pump-current configuration exists. We demonstrated the capacity of our solution to tackle this problem with a maximum absolute error of 1.52 dB and RMSE of 0.76 dB. Therefore, the proposed method can be employed in the multi-span long haul network to realize desired distributed Raman gain profiles and achieve an optimization of the overall system throughput for S+C+L UWB transmissions.

To conclude, we reported on the experimental demonstration of the use of ANNs to learn the mapping between Raman pump currents and UWB loss profile over a continuous 100 nm-wide optical spectrum in a 100 km-long SSMF span. This ML framework can be used for S+C+L UWB Raman amplifier design.

2.7 Recent developments

After this work was published, other research teams continued to study this field and proposed further ML methods to address the challenge of Raman amplifier design.

First, fine-optimization [53, 54, 55] methods and ensemble methods [56] were used to to further enhance the accuracy of the designed Raman amplifier through a gradient descent method by combining the predicted pump configuration with a previously trained ML model (which gives a prediction of the gain profile given a set of pump parameters).

More recently, a physics-aided ML framework has also be proposed [57, 58] to make the ODE fully differentiable and then a gradient-based optimization was used to achieve an optimal design. This approach avoids re-generation of the training dataset and re-training of the model, if the constraint of the amplifier differs from its initial conditions ; for instance, if the total number of pumps changes.

Besides, other configurations of the Raman amplifier have also been investigated mixing discrete and distributed amplification schemes using forward, backward or bi-directional amplifiers.

Chapter 3

KerrNet : a fast, accurate and general QoT tool using machine learning

In this chapter, we focus on the estimation of the Kerr nonlinear interference (NLI) variance (in absence of nonlinear compensation) in C-band using a ML-based framework to accelerate the numerical computation of the time-domain regular perturbation (TD-RP) analytical model. Our solution provides an reliable and ultra-fast quality of transmission (QoT) estimation tool for both existing and future optical networks.

First, we present a review of previous studies. Second, we review the basics of the TD-RP model derived from the first order perturbation method in the time domain. Then, we explain and validate our solution, called *KerrNet*, in a simplified case of a twenty-span homogeneous transmission link. Finally, we extend our solution to the general case with arbitrary WDM combs configuration with heterogeneous spans composed of different fiber spans (non-identical fiber types and lengths).

3.1 Introduction

Optical fiber communication systems have revolutionized the way we transmit data over long distances, forming the backbone of modern telecommunications networks. The success of these systems hinges on the ability to accurately estimate and manage various impairments that can degrade signal quality. Driven by a rapid increase of the throughput demand and loads, the control and optimization of the capacity of our current and future networks, is becoming a main concern. To enable dynamic traffic-driven service provisioning, a robust and real-time QoT estimation method is required.

Different sources of impairments that affect the signal quality were presented in Chapter 1 subsection i. While other impairments such as transceiver penalties and ASE noise are more readily estimated, the most challenging

but crucial part in QoT estimation, is the calculation of the Kerr NLI noise variance. Indeed, transceiver impairments are often well-characterized and can be quantified through back-to-back characterizations. These parameters can be directly measured or calculated based on transceivers' specifications. As for the ASE noise, it primarily depends on the inline optical amplifiers' characteristics and can be estimated through measurements or simple models based on amplifier NF.

On the other hand, the nonlinear distortions estimation, arising from the Kerr effect, is the most challenging step considering computational time. Thus, low margin optical network planning and optimization is becoming increasingly complex because of the large number of parameters and NLI computation time. Since the complexity rises with the number of channels, it is essential to have a low-complexity method that can give accurate predictions for wide-band WDM systems.

The accurate estimation of NLI based on split-step Fourier method (SSFM) simulations is prohibitively complex. Consequently, analytical models have been developed using the regular perturbation (RP) method in either frequency domain or time domain. However, the proposed analytical model either contain multi-dimensional integrals that make them unsuitable for real-time use, or otherwise assume an "ideal" system. The Gaussian noise (GN) model [28] is based on RP method in the frequency domain and many assumptions are made to simplify the model and reduce the computation time significantly, albeit at the cost of a reduced accuracy. In particular, the modulation format impact is ignored. An upgraded version of the GN model called enhanced GN (EGN) model [10] is modulation-aware by introducing a corrective term computed with multi-dimensional integrals at the expense of increased computational complexity. Consequently, the exact EGN model does not meet the need of ultra-fast QoT estimation and closed-form EGN models have been developed with an unavoidable sacrifice on the accuracy and generalization ability of the model.

The time-domain regular perturbation (TD-RP) model proposed by [12], exactly expresses nonlinear distortions up to the first-order of fiber nonlinear coefficient. This model also requires the computation of multi-dimensional integrals via Monte-Carlo (MC) integration. While accurate up to the first-order, the EGN and its time-domain equivalent TD-RP models' bottleneck is the numerical evaluation of various multi-dimensional integrals.

In the recent years, many research groups have published on applications of ML in estimating the QoT of optical networks, aiming to develop ultra-fast and reliable tools for network planning and control, and system design optimization [7]. The ML-based techniques can be classified into two categories : either they use binary classification to decide whether or not a specific light-path is feasible [59] ; or they use regression to predict crucial metrics, for instance, bit error rate, SNR , NLI variance [60, 61], etc. For the performance metric estimation, some authors have proposed to use ML to approximate the accurate EGN theory, or to improve the approximate closed form formulations by introducing and fine-tuning fiddling parameters [60, 61, 62] ; however, relying on simplifying hypotheses and introducing fitting parameters require that the training be done by sampling over an extremely large space of physical parameters (wavelength-dependent dispersion and attenuation coefficients, fiber lengths, etc.) and all sys-

tem configurations (number of channels, modulation formats, channel spacing, loading scheme, etc.) ; therefore the generalization and reliability of the resulting tool is a challenge. The same sampling and generalization issues are faced by ML-based estimation tools that purely relies on the outputs of SSFM to train the ML model [63].

In our study, we assume that all physical parameters of the link are known and can be used directly as input features for our QoT models. The study on parameter uncertainty, which affects the performance of the QoT estimation, has also been investigated in several papers [64, 65].

In this chapter, we present a bank of feed-forward artificial neural networks (FANN)s, that we call *KerrNet*, to exactly reproduce the semi-analytical TD-RP model in a general setting of optical transmission links. The FANNs reproduce the exact TD-RP theory and significantly reduce the computation time. Our ML-based KerrNet solution requires training only over the various values of physical parameters (fiber span parameters, symbol-rate) but not over the transmission system configurations (channel power, modulation format). Thus, this approach significantly simplifies the sampling requirements and leads to a more robust and general tool. Moreover, using the TD-RP model to train the FAANs can further accelerate the computation time of NLI variance σ_{NLI} when various transmission scenarios (channel power or the modulation format) need to be tested for the same optical link. Indeed, one major advantage of the chosen TD-RP theory is that the computation of σ_{NLI} relies on two distinct steps. The first step is time-consuming because of the computation of perturbation coefficients $\mathcal{X}$ (that will be defined later in this chapter) based on multi-dimensional integrals, which only depends on the physical parameters and are independent of modulation format, channel power and loading scheme (fully-loaded or sparsely occupied). Then the second step consists in computing σ_{NLI} with elementary algebraic operations with the obtained $\mathcal{X}$ coefficients and parameters such as the channel power, moment of the constellation, and channel occupation scheme. Consequently, it becomes unnecessary to redo from scratch the computation of σ_{NLI} if only the transmission scheme has been changed while the physical parameters remain unchanged (for instance when only the modulation format and/or channel power is different) while a re-computation is always needed for ML-models trained with data generated using other methods such as EGN or SSFM.

3.2 Nonlinear interference variance estimation

3.2.1 System under study

Let us consider a general transmission case where one need to predict the performance (we only consider the NLI variance σ^2_{NLI} in this chapter) of a transmitted signal in the WDM scheme. N_c WDM channels are transmitted from A to B through the designated target path inside a meshed network (cf. Fig. 3.1). The heterogeneous link is composed of N_s non-identical spans and nodes. A stretched version of the system under study is illustrated in Fig. 3.2.

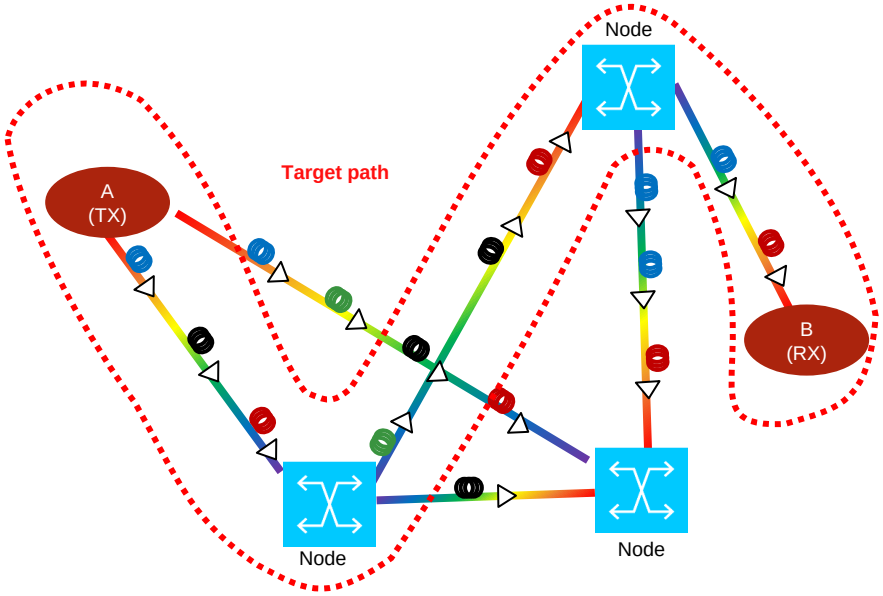


FIGURE 3.1 – Representation of the mesh network with 5 nodes

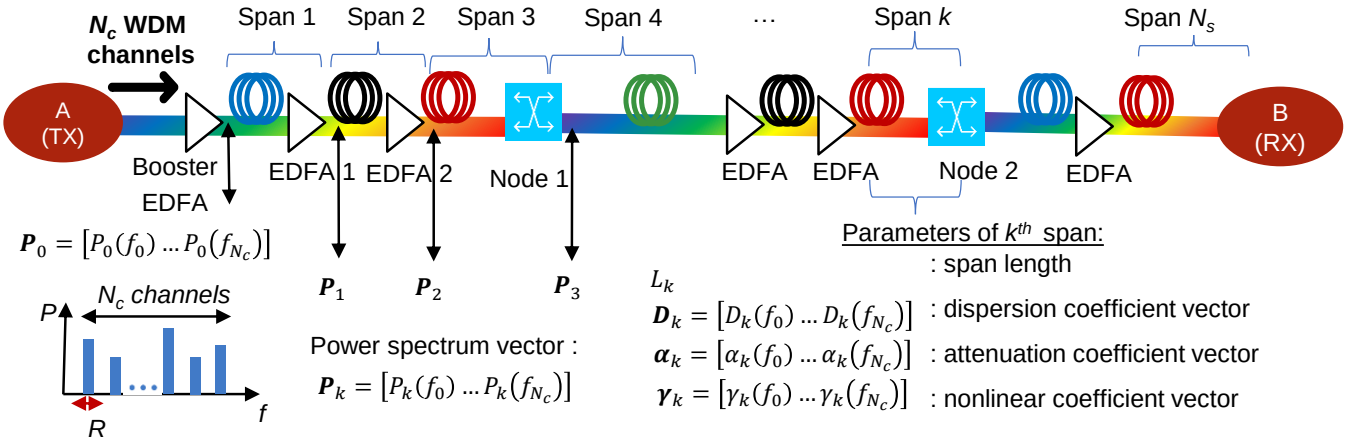


FIGURE 3.2 – Representation of the stretched optical transmission link

Each fiber span, indexed k is defined by its parameters : the dispersion coefficient profile $\mathbf{D_k}$, attenuation profile $\mathbf{a_k}$, nonlinear coefficient profile $_\mathbf{k}$, and the length L_k. The power spectrum $\mathbf{P}_k$ after the k^{th} span is non-flat due to residual gain ripples and/or tilts of EDFAs and/or non-flat response of the WSSs in the optical nodes.

Fig. 3.3 represents the configuration of our transmitted WDM channels : the N_c channels are transmitting in the C-band the same symbol rate R. The transmission configuration with variable symbol rate WDM channels is not studied in this thesis but the same principles can be extended to this application. We also suppose that all WDM channels are modulated with the same modulation format. Moreover, the rest of the study do not account for intra-channel SRS (ISRS) as it has been negligible over the C-band.

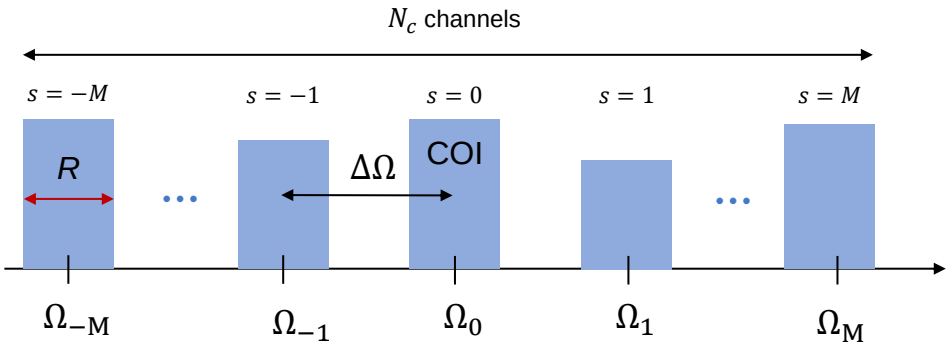


FIGURE 3.3 – The spectrum of $N_c = 2M + 1$ WDM channels. The channel of interest (COI) is indexed $s = 0$ and the center frequency of the s^{th} channel is $s\Delta\Omega$ with $\Delta\Omega$ the channel spacing. Each channel is transmitting at the same symbol-rate R

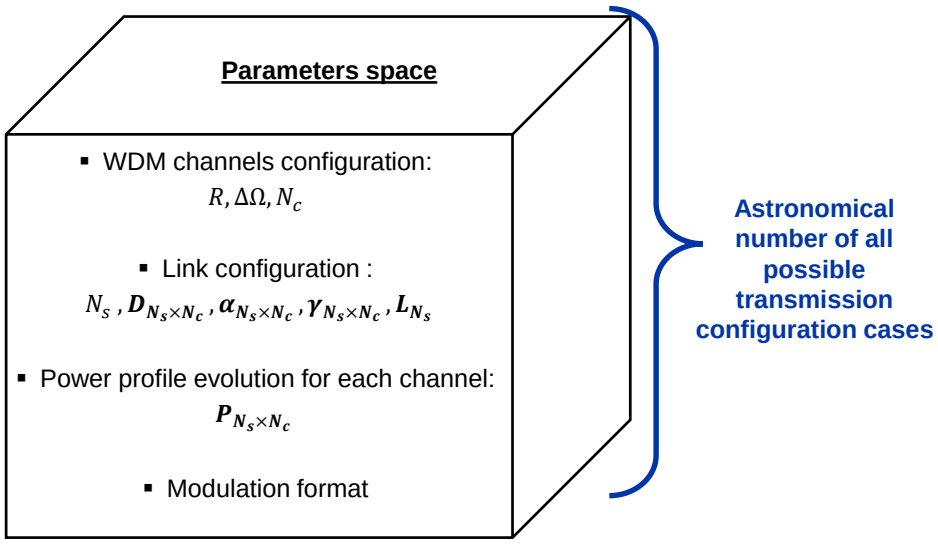


FIGURE 3.4 – Schematic representation of the parameters space influencing the NLI variance : the symbol rate R and channel spacing $\Delta\Omega$, and $\mathbf{D},$, are $N_s \times N_c$ matrix (with N_c the number of WDM channel and N_s the number of spans) containing the all spans' dispersion, attenuation, non-linear coefficients respectively for all WDM channels. $\mathbf{L}$ is a vector of size N_s containing the length of each fiber span of the link

3.2.2 Parameters space

We represent the parameters space in Fig. 3.4. It is worth to notice that the number of parameters that impact the variance of the NLI is extremely large. Thus, the expression of an analytical model, including the GN model, is complex. Consequently closed-form EGN and GN models have been proposed to simplify the expressions by making a trade-off between accuracy and computation speed. Thus, some authors proposes to leverage ML to approximate the accurate EGN theory, or to improve the approximate closed form formulations by introducing and fine-tuning fiddling parameters. However, it is obvious that this require sampling and training to be done over the entire huge space of all physical parameters and all system configurations. Therefore the generalization ability of the proposed methods is compromised. Furthermore, the sampling step is even more challenging since the number of parameters is not fixed but increases with the number of spans and the number of WDM channels. That is also

the reason why we do not rely on SSFM data to train an ML-based NLI estimation tool but use the TD-RP model instead.

In the next section, an overview of the above- mentioned analytical models are presented with an emphasis on the advantages of TD-RP and its suitability to train the ML-based NLI estimation tool.

3.3 Analytical model for nonlinear distortions

The NLI variance can be estimated by SSFM by solving numerically the NLSE. This method presented in subsection 1.1.5 is very time-consuming due to the usage of large amount of fast Fourier transforms and other mathematical operations (hours to days of computation time for a link). Thus, analytical model have been developed by different research groups using regular perturbation method to provide an analytical solution (to the first order) to NLSE for dual-polarization WDM transmissions over multi-span links. Indeed, nonlinear distortion is relatively small compared to the useful signal for relevant power levels in optical communications, thus, it can be treated as a small perturbation, hence the name "regular perturbation" method.

Fig. 3.5 show an overview of different models derived from RP methods to compute the NLI variance, using different approaches. To obtain the analytical solution, two approaches are possible. The first approach relies on an analysis in the spectral domain and leads to GN and EGN models. A different second approach is based on a time-domain analysis and the expressions derived from this model are the ones used in this thesis to compute the NLI variance.

All models presented here do not take into account the ISRS. For a matter of simplicity, we will present the analysis for a scalar field and then present the final solution generalized for a dual polarization case. For the same reason, only the second-order dispersion term is considered (β_3 and higher order terms are neglected but they can be included in the analysis without problem).

3.3.1 Frequency domain analysis

In this subsection, we review the essential parts of the frequency-domain analysis.

The NLSE in Eq. (1.19) can be re-written in frequency domain by applying the Fourier transform.

$$\frac{\partial}{\partial z}E(z,f) = -\frac{\alpha}{2}E(z,f) + j2\pi^2\beta_2 f^2 E(z,f) + j\gamma\left(E(z,f) * E^*(z,-f) * E(z,f)\right) \tag{3.1}$$

with $E(z,f) = \mathcal{F}\{E(z,t)\}$ where $\mathcal{F}\{.\}$ is the Fourier transform and $f * g$ denotes the convolution operation

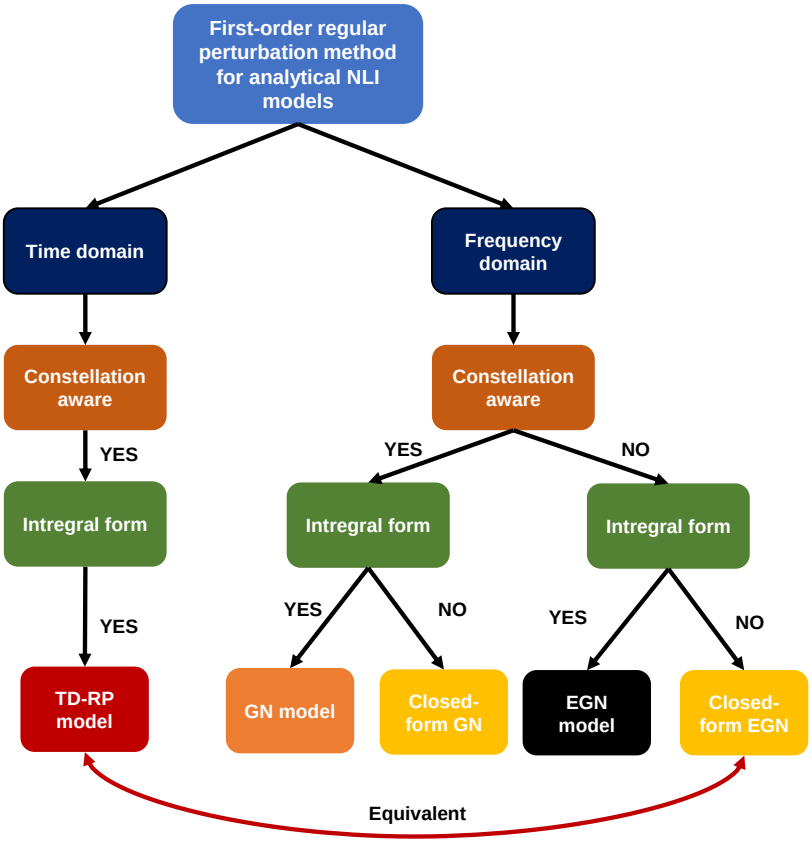


FIGURE 3.5 – NLI analytical model classes

i Regular first-order perturbation method

A regular first-order perturbation (RP1) approach was exploited where the optical field is written as a perturbation series with respect to the nonlinearity parameter γ as :

$$E(f) = E^{(0)}(f) + \gamma E^{(1)}(f) + \mathbb{O}(\gamma^2) \tag{3.2}$$

By substituting (3.2) into (3.1) and neglecting second-order terms $\mathbb{O}(\gamma^2)$, we obtain :

$$\frac{\partial}{\partial z}E^{(0)}(f) = (-\frac{\alpha}{2} + j2\pi^2\beta_2 f^2)E^{(0)}(f) \tag{3.3}$$

$$\frac{\partial}{\partial z}E^{(1)}(f) = [-\frac{\alpha}{2} + j2\pi^2\beta_2 f^2]E^{(1)}(f) + j[E^{(0)}(f) * E^{(0)*}(-f) * E^{(0)}(f)] \tag{3.4}$$

Thus, the zeroth-order solution is simply the linear part of the NLSE given by :

$$E^{(0)}(z,f) = E^{(0)}(0,f)\exp\left(-\frac{\alpha}{2}z + j2\pi^2\beta_2 f^2 z\right) \tag{3.5}$$

The initial condition for the first-order solution at $z = 0$ is :

$$E^{(1)}(z = 0, f) = 0 \tag{3.6}$$

So using Eq. (3.6) and Eq. (3.4), the first-order solution is given by :

$$E^{(1)}(z, f) = \exp\left((-\frac{\alpha}{2} + j2\pi^2\beta_2 f^2)z\right) \int_0^z \frac{E^{(0)}(z', f) * E^{(0)*}(z', -f) * E^{(0)}(z', f)}{\exp\left((-\frac{\alpha}{2} + j2\pi^2\beta_2 f^2)z'\right)} dz' \tag{3.7}$$

ii Gaussian noise model

To simplify the analytical model, GN model has been proposed by Poggiolini, Carena, *et al.* in [13], [66], and [28] with two key assumptions : the non-linear interference (NLI) is treated as an additive Gaussian noise. Moreover, the WDM signal is supposed to be composed of an infinite frequency comb that is shaped according to the power spectral density (PSD) of the signal. Each discrete spectral tone carries a zero-mean complex Gaussian distributed symbol with uncorrelated phase and quadrature components. Under those assumptions, the signal in frequency can be written as :

$$E(z = 0, f) = \sqrt{f_0 G_{WDM}} \sum_{n=-\infty}^{\infty} \xi_n \delta(f - nf_0) \tag{3.8}$$

where G_{WDM} is the PSD of the signal at the transmitter ($z = 0$), ξ_n are the Gaussian distributed random symbols for the spectral tones and $f_0 \to 0$ for a "continuous spectrum". Inserting Eq. (3.8) and Eq. (3.7) in the NLSE equation for dual polarisation case, the PSD of the NLI for a WDM signal at the end of the link $G_{NLI}^{GN}(f)$ is given in [28]. This expression is the famous so-called GN model reference formula (GNRF).

In a simplified system approximation, the expression can be simple and a fast computation of NLI variance is possible but in the general case, for heterogeneous links and general configuration of WDM channels, this expression is still complex.

Moreover, it has been shown that GN model overestimates the amount of NLI with respect to QAM formats due to the neglected modulation format dependence [67]. Indeed, due to the Gaussian assumption of the signal, GN model is unable to predict the modulation format dependence property of NLI.

iii Enhanced Gaussian noise model

In order to accurately estimate the NLI variance, it is required to have a modulation-format aware model. An upgraded version of the GN model, called enhanced GN-model (EGN), was presented and validated in [10], [68] and [69]. To take into account the transmitted constellation influence on NLI, we need to drop the assumption that the transmitted signal statistically behaves as Gaussian noise.

Finally, the PSD of the total NLI can be written as a superposition of a Gaussian constellation (with $G_{NLI}^{GN}(f)$

predicted by the GN model) and a modulation format dependent correction term $G_{NLI}^{corr}(f)$:

$$G_{NLI}^{EGN}(f) = G_{NLI}^{GN}(f) + G_{NLI}^{corr}(f) \tag{3.9}$$

More details about the expression of G_{NLI}^{corr} are provided in [10]. It is worth mentioning that three or more nested integrals are computed in any of the NLI contributions (for the corrective term) with Monte Carlo (MC) integration (see subsection 3.3.3 for more details). The accuracy of the EGN model comes at the cost of a higher computation time.

3.3.2 Time-domain perturbation analysis

In this subsection, we describe the RP1 analysis in the time-domain developed in [70] and [11] to analyze the NLI. For the rest of this thesis, we will refer to it as *TD-RP* model. This model will be used to compute the NLI variance and then serve as a ground truth (GT) for the ML-based model described in section 3.2.

i General formulation

We assume that all WDM channels have the same pulse shaping (Eq. (1.12)) so we define the energy of the pulse-shapes ε by :

$$\varepsilon = \int_{-\infty}^{\infty} |\Psi_0(0,t)|^2 \tag{3.10}$$

To simplify the developments, we define a normalized total optical field $U(z,t)$ as :

$$E(z,t) = \Psi(z)U(z,t) \tag{3.11}$$

where $\Psi(z)$ is the power enveloppe with the initial condition $\Psi(0) = \sqrt{\varepsilon}$ leading to :

$$E(z,t) = f(z)\sqrt{\varepsilon}U(z,t) \tag{3.12}$$

with $f(z)$ the normalized power profile function expressed as :

$$f(z) = \exp\left\{\int_0^z dz' \left[g(z') - \alpha(z')\right]\right\} \tag{3.13}$$

with $g(z)$ and $\alpha(z)$ the local power gain and attenuation coefficients. When we assume a lumped amplification with EDFAs operating at constant gain mode (each EDFA exactly compensates the attenuation in the fiber span), we have :

$$f(z) = \exp\left\{-\alpha(z - z_{n-1})\right\} \text{ for } z_{n-1} \leq z \leq z_0 \text{ and } n = 1, \cdots, N_s \tag{3.14}$$

Then, the expression of the normalized scalar NLSE can be derived from [70],[11] :

$$\frac{\partial U}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 U}{\partial t^2} + j\varepsilon\gamma f(z)|U|^2 U + \frac{n(z,t)}{\sqrt{\varepsilon f(z)}} \tag{3.15}$$

with $n(z,t)$ the ASE noise source.

For the dual polarization case, the normalized Manakov equation is expressed as follows :

$$\frac{\partial \mathbf{U}}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{U}}{\partial t^2} + j\frac{8}{9}\bar{\varepsilon}\gamma f(z)\mathbf{U}^{\dagger}\mathbf{U}\mathbf{U} + \frac{\mathbf{n(z,t)}}{\sqrt{\bar{\varepsilon} f(z)}} \tag{3.16}$$

with $\mathbf{U}(z,t) = [U^H(z,t), U^V(z,t)]^T$, $\mathbf{n}(z,t) = [n^H(z,t), n^V(z,t)]^T$ and $\bar{\varepsilon} = \varepsilon/2$.

In the following part, we will use the scalar NLSE for simplicity and then provide the analytical solution for a dual polarization configuration. Similar to the RP1 method used in subsection 3.3.1 and Eq. (3.2), we can approximate $U(z,t)$ as a perturbation series with respect to γ but here, in the time domain :

$$U(z,t) = u^{(0)}(z,t) + \gamma u^{(1)}(z,t) + \mathbb{O}(\gamma^2) \tag{3.17}$$

By substituting Eq. (3.17) in Eq. (3.15) we obtain :

$$\frac{\partial u^{(0)}(z,t)}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 u^{(0)}(z,t)}{\partial t^2} + \frac{n(z,t)}{\sqrt{\varepsilon f(z)}} \tag{3.18}$$

and

$$\frac{\partial u^{(1)}(z,t)}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 u^{(1)}(z,t)}{\partial t^2} + j\gamma\varepsilon f(z)|u^{(0)}(z,t)|^2 u^{(0)}(z,t) \tag{3.19}$$

The zeroth order solution is expressed as :

$$u^{(0)}(z,t) = \hat{D}_z\left[u^{(0)}(0,t) + u_{ASE}(z,t)\right] \tag{3.20}$$

with $\hat{D}_z$ the dispersion operator defined in time domaine by :

$$\hat{D}_z(x(z,t)) = \mathcal{F}^{-1}\left[\exp\left(j\frac{\omega^2}{2}\int_0^z \beta_2(z")dz"\right)\ \mathcal{F}[x(z,t)]\right] \tag{3.21}$$

where $\mathcal{F}\{.\}$ is the Fourier transform and $\mathcal{F}^{-1}\{.\}$ the inverse Fourier transform operation.

The first order solution is expressed as follows :

$$u^{(1)}(z,t) = j\gamma\hat{D}_z\left[\int_0^z f(z')dz'\right] \times \hat{D}_z^{\dagger}\left[\left|u^{(0)}(z,t)\right|^2 u^{(0)}(z,t)\right] \tag{3.22}$$

Finally, the nonlinear signal-signal distortion $\Delta x_{k,NL}$ on the k^{th} symbol of the COI can be computed [11]. According

to the central limit theorem, $\Delta x_{k,NL}$ are zero-mean Gaussian random variables : $\Delta x_{k,NL} \sim \mathcal{N}(0, \sigma^2_{SS})$. The variance of signal-signal interaction σ^2_{SS} is computed in [70],[71] and [32] as :

$$\begin{aligned}\sigma^2_{SS}(P) = \gamma^2 P^2 \Bigg\{ & 2\mathcal{X}_1 + \left(\frac{\mu_4}{\mu_2^2} - 2\right)[\mathcal{X}_2 + 4\mathcal{X}_3 + 4\mathcal{X}_4] \\ & + \left(\frac{\mu_6}{\mu_2^3} - 9\frac{\mu_4}{\mu_2^2} + 12\right)\mathcal{X}_5 + 4\sum_s \left[\mathcal{X}_{1,s} + \left(\frac{\mu_4}{\mu_2^2} - 2\right)\mathcal{X}_{3,s}\right] \\ & + \sum_s \sum_{s'} \mathcal{X}_{1,s,s'} \Bigg\}\end{aligned} \tag{3.23}$$

with $P = \varepsilon/T$ the same average power for all WDM channels, T the symbol duration, μ_n the n$^{\text{th}}$ moment of the constellation. In this study, we only consider the signal-signal interaction and noise-signal nonlinear interaction (NSNI) term is neglected (interaction between the signal and the noise produced by amplifiers). A general theory involving nonlinear signal-signal and signal-noise interaction theory is presented in [11]

Finally, the variance of NLI is given by :

$$\begin{aligned}\sigma^2_{NLI} = \sigma^2_{SS,DP} = \frac{16}{81}\gamma^2 P^2 \Bigg\{ & 3\mathcal{X}_1 + \left(\frac{\mu_4}{\mu_2^2} - 2\right)[\mathcal{X}_2 + 5\mathcal{X}_3 + 4\mathcal{X}_4] \\ & + \left(\frac{\mu_6}{\mu_2^3} - 9\frac{\mu_4}{\mu_2^2} + 12\right)\mathcal{X}_5 + 4\sum_s \left[6\mathcal{X}_{1,s} + 5\left(\frac{\mu_4}{\mu_2^2} - 2\right)\mathcal{X}_{3,s}\right] \\ & + 2\sum_s \sum_{s'} \mathcal{X}_{1,s,s'} \Bigg\}\end{aligned} \tag{3.24}$$

with $\sigma^2_{SS,DP}$ the total variance of the signal-signal nonlinear distortion for a dual polarization scheme. The coefficients $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_3$, $\mathcal{X}_4$, $\mathcal{X}_5$ correspond to intra-channel (SCI) effects. The $\mathcal{X}$-coefficients with sub-index s corresponds to degenerate inter-channel (XCI) and the sub-index s, s' for the non-degenerate inter-channel four-wave mixing (FWM) terms. The $\mathcal{X}_1$, $\mathcal{X}_{1,s}$ and $\mathcal{X}_{1,s,s}$ coefficients correspond to the modulation independent contribution to the total NLI variance whereas $\mathcal{X}_3$ and $\mathcal{X}_{3,s}$ coefficients correspond to the modulation dependent contribution.

The different $\mathcal{X}$-coefficients are computed using Monte Carlo (MC) integration over multidimensional integrals. Their expressions are presented in the Appendix. The MC integration method is presented briefly in subsection 3.3.3

One major advantage of using the TD-RP model to train the ML-based KerrNet QoT tool is that the model calculations do not need to be run at different launch powers and modulation formats. This highly reduces the parameters space for sampling and training. It also reduces the computation time of NLI variance since the $\mathcal{X}$ coefficients are unchanged and do not require to be re-computed with multi-dimensional integration. Once all $\mathcal{X}$ coefficients are computed, elementary algebraic operations are performed with the configuration parameters of the WDM channels (channel power, channel loading scheme) as described in Eq. (3.24).

For the rest of the study, the perturbative coefficients $\mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_5$ in Eq. (3.24) are neglected because, for scenarios

addressed in this study, their small values are negligible. Moreover, we do not take into account the non-degenerate FWM effects in this study so $\mathcal{X}_{s,s\prime}$ is also ignored.

ii Cross-correlation function

A different approach is to compute NLI using the cross correlation function between i^{th} span and j^{th} span. For a transmission link with a total length of $z = L_{tot}$ km composed of N_s fiber spans amplified by EDFAs at the end of each span. The i^{th} and j^{th} span end at $z = z_i$ and $z = z_j$ respectively.

We can re-write Eq. (3.24) for the four $\mathcal{X}$ coefficients as follows :

$$\begin{aligned}
\mathcal{X}_{1,00}(L_{tot}, f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \rho_1^{(sc)}(i,j) \\
\mathcal{X}_{3,00}(L_{tot}, f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \rho_3^{(sc)}(i,j) \\
\mathcal{X}_{1,0s}(L_{tot}, f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \rho_1^{(xc)}(i,j) \\
\mathcal{X}_{1,0s}(L_{tot}, f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \rho_3^{(xc)}(i,j)
\end{aligned} \tag{3.25}$$

with

$$\begin{aligned}
\rho_1^{(sc)}(i,j) &= \frac{1}{T}\int_{\mathbb{R}^3} \frac{d^3\omega}{(2\pi)^3} \int_{z_{i-1}}^{z_i} dz\, H_{\overrightarrow{\omega}}(z) \int_{z_{j-1}}^{z_j} dz'\, H^*_{\overrightarrow{\omega}}(z') \\
\rho_3^{(sc)}(i,j) &= \int_{\mathbb{R}^3} \frac{d^3\omega}{(2\pi)^4} \int_{z_{i-1}}^{z_i} dz\, H_{\overrightarrow{\omega}}(z) \int_{z_{j-1}}^{z_j} dz'\, H^*_{\overrightarrow{\omega}\prime\prime}(z') \\
\rho_1^{(xc)}(i,j) &= \frac{1}{T}\int_{\mathbb{R}^3} \frac{d^3\omega}{(2\pi)^3} \int_{z_{i-1}}^{z_i} dz\, H_{\overrightarrow{\omega},s}(z) \int_{z_{j-1}}^{z_j} dz'\, H^*_{\overrightarrow{\omega},s}(z') \\
\rho_3^{(xc)}(i,j) &= \int_{\mathbb{R}^3} \frac{d^3\omega}{(2\pi)^4} \int_{z_{i-1}}^{z_i} dz\, H_{\overrightarrow{\omega},s}(z) \int_{z_{j-1}}^{z_j} dz'\, H^*_{\overrightarrow{\omega}\prime\prime,s}(z')
\end{aligned} \tag{3.26}$$

where the expressions for the kernel function H and $\overrightarrow{\omega}$ $\overrightarrow{\omega}\prime\prime$ $\overrightarrow{\omega}, s$ and $\overrightarrow{\omega}\prime\prime, s$ are defined in the Appendix.

The function $\rho(i,j)$ in Eq. (3.26) corresponds to the universal cross-correlation function between the NLIs generated at i^{th} and j^{th} span (whose lengths, dispersion and loss coefficients are L_i, L_j, D_i, D_j, and α_i, α_j, respectively). The labels $'sc'$ and $'xc'$ stand for 'self-channel' and degenerate 'cross-channel' interference, respectively.

iii Comparison of the two versions of the TD-RP model

The first version of the TD-RP model ('TD-RP V1') has been commonly adopted for NLI variance computation thanks to its faster speed compared to the second version ('TD-RP V2', using cross-correlation function ρ) which is slower by a factor equal to the total number of span N_s.

On the other hand, the 'TD-RP V2' model has the advantage of being more general and reusable. If the NLI

Model	TD-RP V1 : product to sums	TD-RP V2 : sum of products
Speed	Faster	Slower by N_s
Compatible with ANN	×	✓
Re-usability and scalability	×	✓

TABLE 3.1 – Summary of the comparison between the two versions of the TD-RP model

variance needs to be computed for a new path (composed of the previous N_s-spans link plus one additional span), we can reuse the old data containing the N_s previously computed ρ and only compute the cross-correlation coefficients ρ between the $N_s + 1^{th}$ span and the N_s other spans. Consequently, the computation time is reduced by reusing the old data since we do not need to recompute from scratch the $(N_s + 1)^2$ integral based coefficients but only the additional $(2N_s - 1)$ coefficients.

Another advantage of using the 'TD-RP V2' model to build an ML-based framework is that this model has a fixed number of input parameters for arbitrary links. The size of the input features to compute $\mathcal{X}$ for 'TD-RP V1' model increases with the number of spans whereas the size of the input features to compute ρ with the 'TD-RP V2' model is fixed and do not increase with the total number of spans. This property of the 'TD-RP V2' model makes it perfectly suitable for an ANN structure with a fixed number of input neurons.

A summary of the advantages and the drawback of the two versions of the model is presented in Tab. 3.1

iv Summary

To conclude, this time-based regular perturbation model provides us an accurate (up to the first-order) computation of the NLI variance. Without the Gaussian-distributed symbols assumption, this model is modulation format sensitive. This TD-RP model is equivalent to the frequency domain EGN model. Compared to SSFM, its computational speed has been seriously reduced but still requires a large computation effort by computing the multi-dimensional integration for all spans and all channels of a WDM transmission.

The novelty of our work, presented in the next section of this chapter, lies in proposing a bank of shallow FANNs to efficiently interpolate the smooth function of either $\mathcal{X}$ or ρ as the key building block to compute the nonlinear interference variance. Our KerrNet method provides an acceleration of the exact models for NLI variance without compromising the accuracy of the model.

3.3.3 Monte-Carlo integration

Monte-Carlo (MC) integration has been used in the computation of NLI variance for integral-based models, such as in the work of Dar *et al.* [67], [72]. This integration method is also adopted for GN/EGN model or the TD-RP model that we use.

MC integration method proved its efficiency compared to a deterministic approach to perform numerical integration in the case of evaluating multi-dimensional integrals since the standard deviation of the error is independent of

the number of dimensions [73]. MC integration can be seen as the probabilistic interpretation of an integral. For a function f, an estimation of its integral can be expressed as :

$$\int_{\mathcal{V}} f(x)dx \approx \frac{V}{N_{MC}} \sum_{j=1}^{N_{MC}} f(\bar{x}_j) \tag{3.27}$$

where N_{MC} is the number of $\bar{x_j}$ samples generated following a uniform distribution in the subset $\mathcal{V}$ which has a volume V, and $f(\bar{x_j})$ is the integral evaluated for the sample $\bar{x_j}$.

The choice of N_{MC} number of integration points is important as it affects the performance since the standard deviation of the error scales with $1/\sqrt{N_{MC}}$. But the computation speed is also determined by N_{MC}.

3.4 KerrNet : an ANN-based framework to speed-up the TD-RP model

Our proposed solution to compute the NLI variance σ^2_{NLI} is based on the TD-RP theoretical semi-analytical model which is exact up to the first order of perturbation and a ML framework with a bank of small FANNs. Thus, the ground truth (GT) data is the output of the MC integration result of TP-RP model. We only use ANNs to predict the perturbative coefficients $\mathcal{X}$ or the cross-correlation coefficients ρ and then the computation of the overall σ^2_{NLI} can be obtained by simply applying elementary multiplications and summations involving power profiles, nonlinear coefficients, and modulation moments. Since $\mathcal{X}$ and ρ only depends on the physical parameters, this drastically reduces the parameters space for sampling and training, hence this approach results in a better generalization of the trained ML-framework. The longest step of the NLI variance computation is the calculation of $\mathcal{X}$ or ρ which will be only performed once even if one wants to test different power configurations or modulation formats or even removing several channels in the WDM combs. This is not the case if other methods such as EGN or SSMF are used where the computation of σ^2_{NLI} also needs to be run at different launch powers and modulation formats.

This perturbation analysis-based model has been extensively validated in [11] comparing the theoretical semi-analytical SNR values (i.e., using MC integration to compute theoretical X coefficients, and then SNR) vs. simulated SNR values (i.e., symbol-level MC and split-step Fourier method for transmission, followed by ideal coherent detection and SNR extraction) for various submarine and terrestrial links. Hence, in this work we only compare our ML model (indexed "ML") and the theoretical semi-analytical model (indexed "th").

3.5 Nonlinear interference estimation in homogeneous links

In this section, we first present our study for a homogeneous link i.e. all the fiber spans are identical : same type and same length. In this case, we can simplify the vector of the physical parameters $\mathbf{D}$, $\mathbf{L}$ to a single value. We also assume that all WDM channels have the same average launched power per channel, P, the same symbol rate R

and the same modulation format. Besides, we assume that all EDFAs, placed at the end of each span, operate in the constant output power mode and compensate exactly the fiber attenuation occurring in each span.

The NLI variance σ^2_{NLI} and the nonlinear coefficient a can be computed according to the following equation :

$$\sigma^2_{NLI} = P^2 a = \frac{16}{81}\gamma^2 P^2 \left(3\mathcal{X}_{1,00}\right) + \left(\frac{\mu_4}{\mu_2^2} - 2\right)\left(5\mathcal{X}_{3,00}\right) + 4\sum_s \left[6\mathcal{X}_{1,0s} + 5\left(\frac{\mu_4}{\mu_2^2} - 2\right)\mathcal{X}_{3,0s}\right] \tag{3.28}$$

where various perturbation coefficients $\mathcal{X}$ are calculated with multi-dimensional triple and quadruple frequency-domain integrals (computed with numerical MC integration as described in subsection 3.3.3) and the non-degenerate FWM effects are ignored in our study.

The variable s stands for the index of the adjacent WDM channels ; $s = 0$ for the COI. The $\mathcal{X}$ coefficients with 00 sub-index represent the nonlinear self-channel interference of the COI, and $\mathcal{X}$ coefficients with $0s$ sub-index represent the degenerate nonlinear cross-channel interference between the COI and the s^{th} adjacent channel. We propose to compute the four types of $\mathcal{X}$ coefficients, i.e., $\mathcal{X}_{1,00}$, $\mathcal{X}_{3,00}$, $\mathcal{X}_{1,0s}$ and $\mathcal{X}_{3,0s}$, each with a dedicated FANN to approximate the time-consuming MC integration.

3.5.1 FANN framework

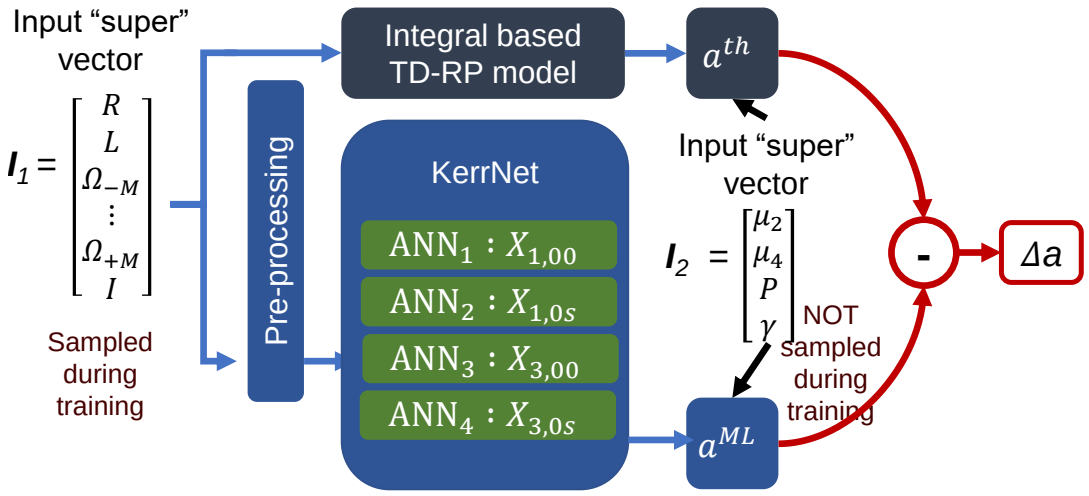


FIGURE 3.6 – Computation of the nonlinear coefficient a via the calculation of four X coefficients with a bank of four small ANNs

Fig. 3.6 illustrates the computation of the four perturbation coefficients $\mathcal{X}$ with our KerrNet tool, composed of four FAANs, with its input features defined by the super-vector $\mathbf{I}_1$ containing the physical parameters. Consequently, the training dataset is generated only by sampling the input super-vector $\mathbf{I}_1$. Once trained, KerrNet automatically works for all choices of the input super-vector $\mathbf{I}_2$ containing the link configuration parameters since for both TD-RP model and for KerrNet, the passage from X coefficients to the nonlinear coefficient a requires the same elementary multiplications and summations involving power profiles, nonlinear coefficients, and modulation moments defined in

Parameters	Range
L [km]	10 to 200, step = 1
I	1 to 20, step =1
R [GBd]	20 to 150, step = 1
s	0 to fully-loaded in C-band, step = 1
$\Delta\Omega$ [GHz]	12.5

TABLE 3.2 – Parameters space to generate data sample

Eq. (3.28). This significantly reduced the parameters space to generate the training dataset since $\mathbf{I}_2$ is not involved in the training step. Then, the ML-estimated nonlinear coefficient a^{ML} is compared with the theoretical value a^{th}, computed with the integral-based TD-RP model and the deviation Δa is calculated in [dB] to evaluate the performance of the KerrNet framework.

We use a dedicated FANN to predict each of the four different types of $\mathcal{X}$ coefficients because the ranges of the absolute values of these four types of $\mathcal{X}$ coefficients are different by at least one order of magnitude. Using this prior knowledge this "divide-and conquer" approach help in improving the accuracy of the prediction.

ANN_1 to ANN_4 are trained using Pytorch library. The ANNs have three layers where the activation function is the Relu function for the two first layers and the linear function for the output layer. The cost function is the MSE. Several learning rates have been tested. Hyper-parameters are chosen with independent hyper-parameters optimization according to the random search method presented in section 1.2). Moreover, hyper-parameters are deliberately geared towards the smallest size FANNs with acceptable accuracy, since reducing computational time is the main objective of this study. For this reason, we decided to train FANNs with only two hidden layers and each layers are composed of $50, 70, 90$ or 110 neurons. Of course, data generation and pre-processing as well as training these ANNs take time, but once done, the resulting tool will be faster in predicting NLI than the semi-analytical approach by several orders of magnitude in execution time.

In the first study presented in this section, we only considered identical spans of SSMF, so $X_{1,00}$ and $X_{3,00}$ depend on three parameters : symbol rate R, span length L, and the span index I of the 20-span link. $X_{1,0s}$ and $X_{3,0s}$ depend on 4 parameters : R, L, I and the frequency spacing between the COI and s^{th} adjacent channel $\Omega_s = s\Delta\Omega$ with $\Delta\Omega$ the frequency slot granularity. The parameters space of the dataset is summarized in Tab. 3.2.

The dataset was generated by calling the semi-analytical tool based on MC integration, with 10^7 integration points if $R > 80$ GBd or with 1.5×10^6 integration points otherwise (we need more integration points when the symbol rate increases). This choice was made after studying the MC statistical error to guarantee that the error in evaluating the nonlinear coefficient is less than ± 0.05 dB. The dataset is generated by sampling non-uniformly the input super-vector I_1 of all physical parameters (range of the parameters is summarized in Tab. 3.2).

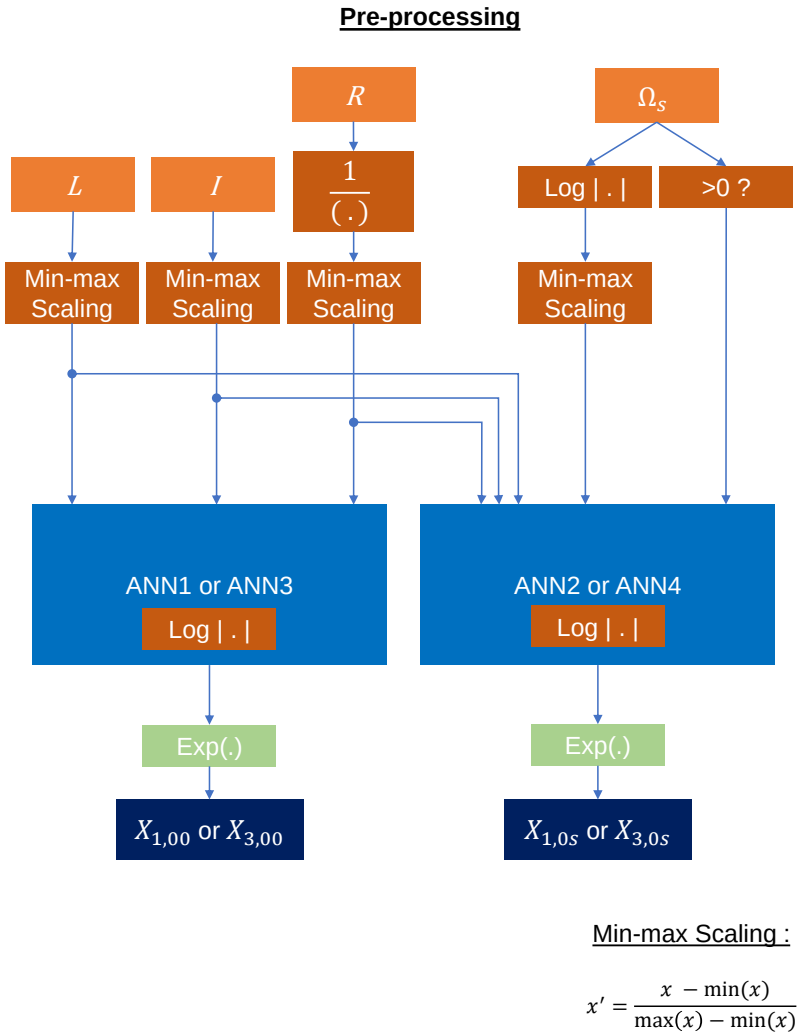


FIGURE 3.7 – Flowchart of the data pre-processing step while training the ML framework

3.5.2 Impact of data and pre-processing

As explained in section 1.2, the data quality and the pre-processing step to transform the dataset may have a major impact on the performance of an ML model. To evaluate the impact of data pre-processing and the distribution used to generate the input parameters on ther performance of the ML framework, we trained the FAANs under three configurations :

— Configuration (a) that serves as a reference where the entire dataset is generated using uniform distribution for all input parameters and without the data pre-processing step.

— Configuration (b) is trained with a dataset generated uniformly but this time, a data pre-processing step is performed.

— Configuration (c) where the dataset is generated following a non-uniform distribution and also includes the data pre-processing step.

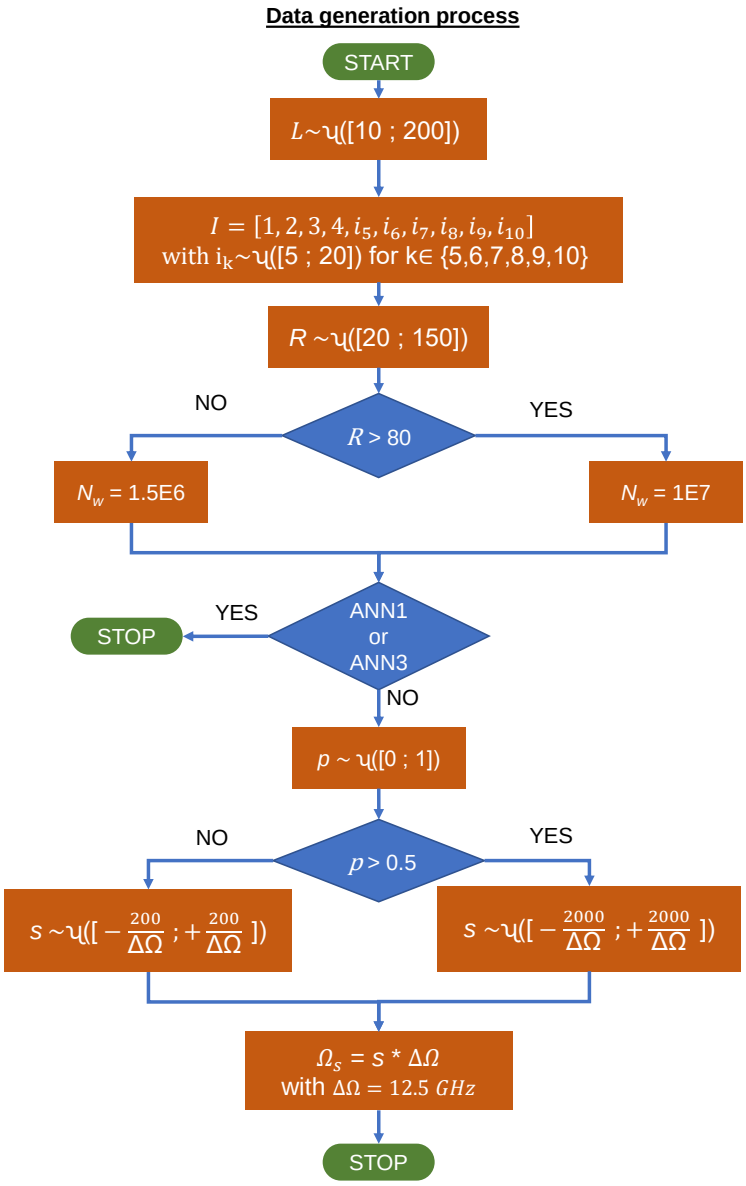


FIGURE 3.8 – Flowchart to generate the data

For a fair comparison, the three models are trained with the same quantity of data (4×10^5 data points for $\mathcal{X}_{1,00}$ or $\mathcal{X}_{3,00}$; and 8×10^5 points for $\mathcal{X}_{1,0s}$ or $\mathcal{X}_{3,0s}$).

The three models are also evaluated with the same test dataset. The test dataset is composed of 10^4 data points representing 500 different homogeneous link configurations. Similar to the training stage, our link configurations were made of twenty SSMF spans, and the span length L was generated randomly with the uniform distribution from typical fiber lengths ($L \in [20, 40, 60, 80, 100, 120]$ km). The symbol rate R was also generated uniformly : $R \sim \mathcal{U}[20, 150]$ GBd. All configurations are fully loaded in C-band (i.e., $\Omega_s = s \times \Delta\Omega \in [-2, 2]$ THz) and $\Delta\Omega$ depends on R : $\Delta\Omega = k \times 12.5$ GHz, with k the smallest integer that avoids channel overlapping ($12.5k \geq R$). The chosen modulation format is QPSK.

The pre-processing is described in Fig. 3.7. We used min-max scaling to normalize the inputs of ANNs and the

parameter Ω_s was separated into two auxiliary inputs $\log|\Omega_s|$ and $\text{sign}(\Omega_s)$, and we used $1/R$ instead of R at the ANN input. We also use $\log(\mathcal{X})$ as the output of the ANN instead of using $\mathcal{X}$ to "compress" the dataset because of its values' large range.

The non-uniform distribution to generate the dataset is illustrated by the flowchart in Fig. 3.8. After careful optimizations, we settled for generating $R \sim \mathcal{U}[20, 150]$ GBd and $L \sim \mathcal{U}[10, 200]$ km, where $\mathcal{U}[a, b]$ stands for uniform distribution between a and b, and Ω_s as following : $\Omega_s \sim \mathcal{U}[-200, 200]$ GHz with a probability $p = 0.5$ and $\Omega_s \sim \mathcal{U}[-2, 2]$ THz with a probability $q = (1 - p) = 0.5$. Besides, for each $\{R, L, \Omega_s\}$ configuration, the span index variable I runs through $[1, 2, 3, 4, i_5, \cdots, i_{10}]$ with $i_k \sim \mathcal{U}[5, 20]$ for $k \in \{5, 6, \cdots, 10\}$.

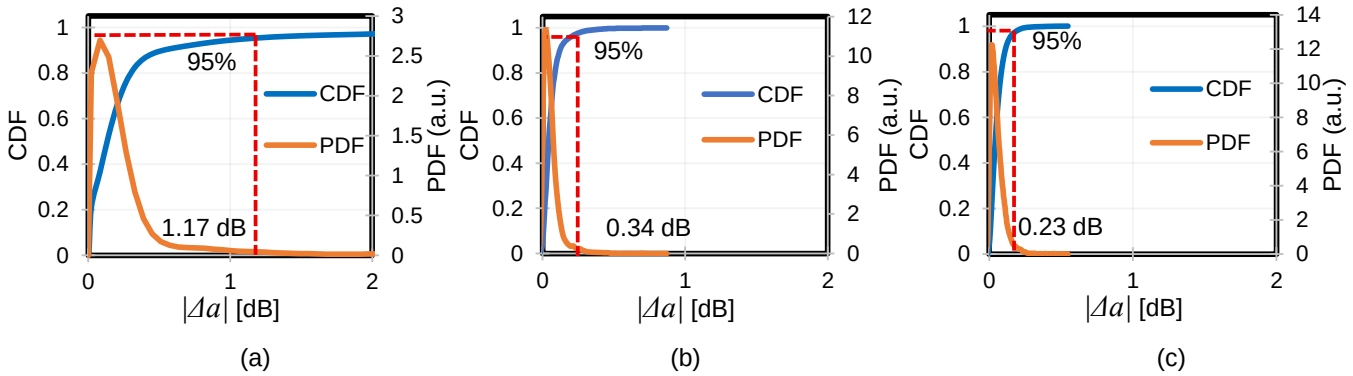


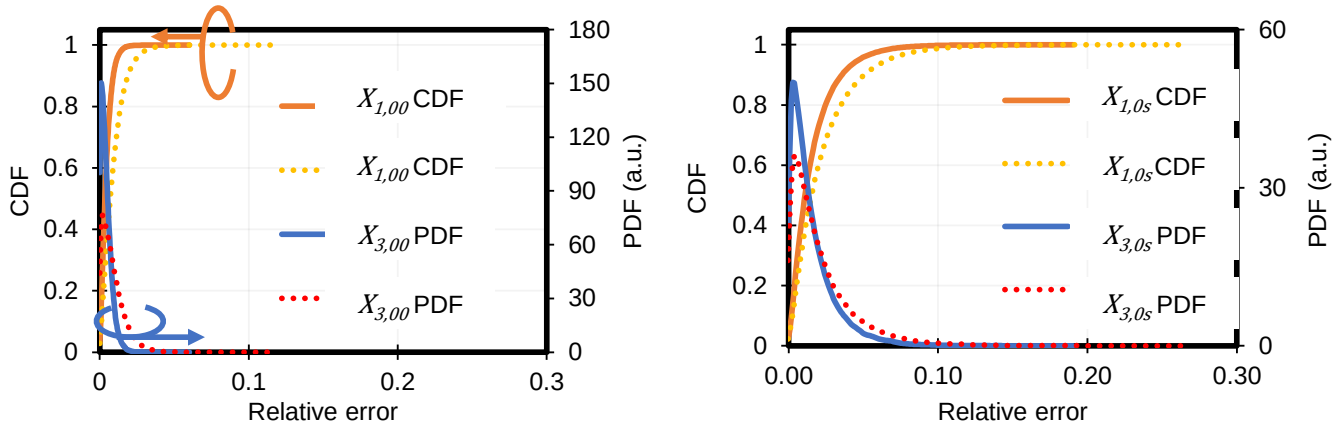
FIGURE 3.9 – PDF and CDF of errors evaluated for the non-linear coefficient $|\Delta a|$ in dB for three different training configurations : (a) with dataset generated uniformly and without data pre-processing, (b) with dataset generated uniformly and with data pre-processing, (c) with dataset generated non-uniformly and with data pre-processing

We evaluate the performance of the three trained ML models by computing the absolute error $|\Delta a| = a^{th} - a^{ML}$ [dB] of the nonlinear coefficient on the test dataset between the ML prediction and the theoretical result (computed with MC integration with the TD-RP model). The probability density function (PDF) and the cumulative distribution function (CDF) of the error $|\Delta a|$ are plotted in Fig. 3.9. The curves shows the benefit of the chosen pre-processing step. The case (a) shows the highest errors with a mean absolute error of 0.27 dB and max absolute error of 11.60 dB whereas the case (b) only have a mean absolute error of 0.06 dB and max absolute error of 0.84 dB. Moreover, generating data using the uniform distribution for every parameter gave poorer results after training. 95% of the predictions of the case (b) have an absolute error less than 0.34 dB but this value can be further decreased to 0.23 dB when we apply a non-uniform distribution to sample the dataset. The max absolute error is also reduced from 0.84 dB to 0.53 dB. Indeed, important link configurations for the training occur rarely, hence we need to use a non-uniform distribution of some input features to get a "more balanced" dataset. To conclude, we will use the configuration (c) for our KerrNet to achieve the best performance.

3.5.3 Final result

Finally, after finding the best training process (case (c)) as explained in the previous subsection, the training of the four FAANs was re-performed with ten times more data : 4 million ($\mathcal{X}_{1,00}$ or $\mathcal{X}_{3,00}$) or 8 million ($\mathcal{X}_{1,0s}$ or $\mathcal{X}_{3,0s}$) data points generated according to the chosen distribution and pre-processed similarly to the case (c). The dataset is split in 80% for training and 20% for validation. The same early-stopping method was used with the validation dataset to stop the training process when the precision stops increasing and the random search method was used to find the best hyper-parameters : ANN_1 to ANN_4 are composed of two hidden layers with 70, 110, 90, and 50 neurons respectively.

We evaluated our four trained ANNs separately with 10^5 ($\mathcal{X}_{1,00}$ or $\mathcal{X}_{3,00}$) data points or 2×10^5 ($\mathcal{X}_{1,0s}$ or $\mathcal{X}_{3,0s}$) data points using the relative error = $\left|\frac{Target-Predicted}{Target}\right|$. The independent test dataset was generated with the same parameters space listed in Tab. 3.2, but using the uniform distribution for all parameters R, L, I and Ω_s.



(a) (b)

FIGURE 3.10 – Error statistics. Probability Density Function (PDF) and Cumulative Density Function (CDF) of the relative error for (a) $\mathcal{X}_{1,00}$ and $\mathcal{X}_{3,00}$; (b) for $\mathcal{X}_{1,0s}$ and $\mathcal{X}_{3,0s}$

Fig. 3.10 illustrates the error statistics of KerrNet's prediction for $\mathcal{X}$ coefficients : 90% of samples show errors below 1% for $\mathcal{X}_{1,00}$, 2% for $\mathcal{X}_{3,00}$, 4% for $\mathcal{X}_{1,0s}$ and 5% for $\mathcal{X}_{3,0s}$. Errors are larger for $\mathcal{X}$ coefficients corresponding to the cross-channel nonlinear interference (the ones with sub-index s) compared to the ones related to self-channel nonlinear interference due to the higher sensitivity of the perturbation coefficients to the channel spacing.

To evaluate the accuracy of our KerrNet tool, the important quantity to compute is the overall error between the ground-truth value of the nonlinear interference coefficient a^{th} and the predicted a^{ML} in dB. For a^{th}, we only consider the TD-RP theory's result. Indeed, this perturbation analysis-based model has been extensively validated in [11] and re-generating stimulative result a^{sim} with SSFM is extremely long.

In order to demonstrate the efficiency of the proposed approach, we also calculated the absolute error $|\Delta a|$ [dB] of the nonlinear coefficient on the same test dataset than the one used to compare the impact of pre-processing and data generation methods (cf subsection 1.2 in Chapter 1) so on 10^4 data points representing 500 different

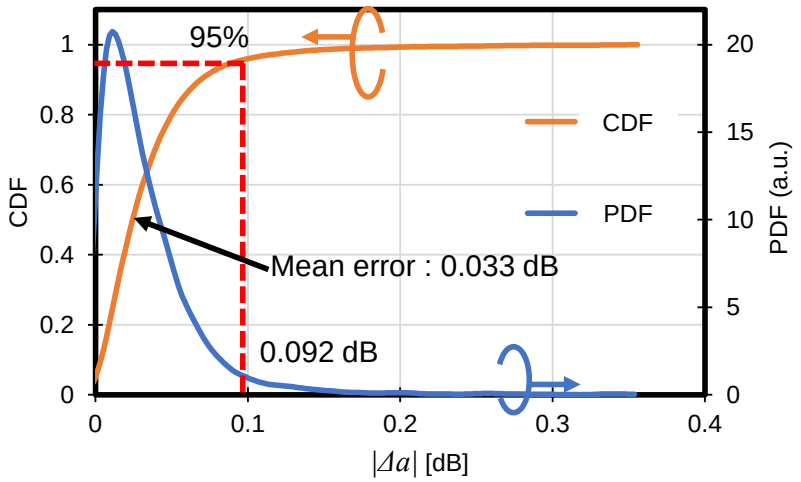


FIGURE 3.11 – PDF and CDF of the absolute error of the nonlinear coefficient $|\Delta a|$ in dB over test dataset

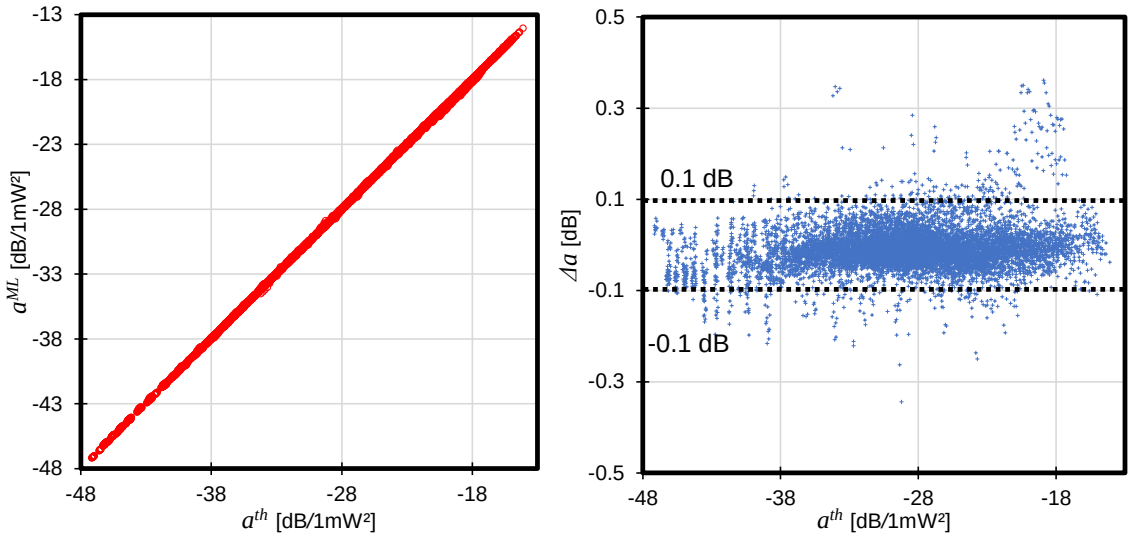


FIGURE 3.12 – (a) Scatter plot of the model predictions over the test dataset a^{ML} vs. theoretical estimation a^{th} and statistics prediction errors and (b) the absolute error $|\Delta a|$ vs. a^{th}

homogeneous link configurations composed of twenty spans of SSMF. The modulation format chosen is still QPSK.

For each link configuration, we computed the nonlinear coefficient $10\log_{10}(a/1\ \mathrm{mW}^2)$ using $\mathcal{X}$ coefficients calculated with our KerrNet, and with the MC integration for the theoretical results. Fig. 3.11 illustrates the statistics of the NLI error where the CDF (blue line) and PDF (orange line) of $|\Delta a|$ are plotted and show the high accuracy of our KerrNet solution. On 95% of the dataset, the absolute error between those two methods is less than 0.092 dB, and the mean absolute error is 0.033 dB.

Fig. 3.12 shows the scatter plot a^{ML} vs. a^{th} and the absolute error $|\Delta a|$ as a function of ground truth theoretical a^{th}. This figure shows the accuracy of the predicted nonlinear coefficient a^{ML} with most of the errors below 0.1 dB and also shows the KerrNet framework has a stable performance in different transmission schemes impacted by different strengths of the nonlinear effects.

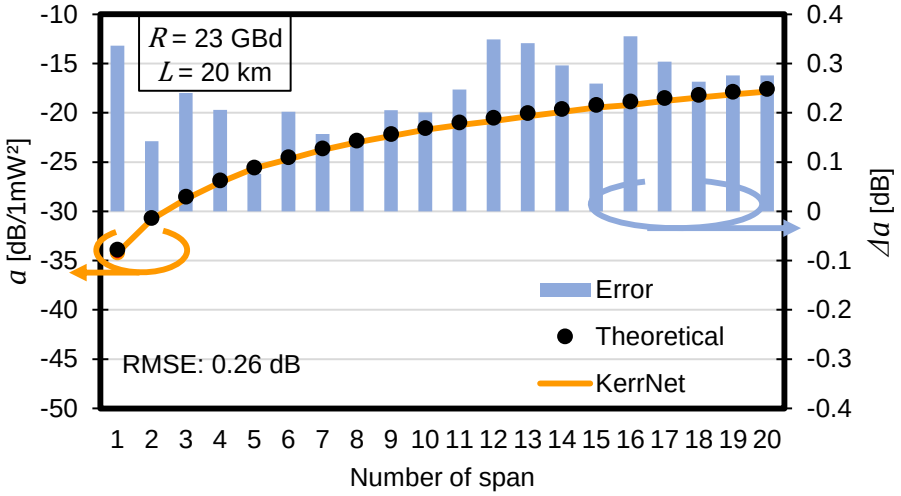


FIGURE 3.13 – Evolution of the nonlinear coefficient a with respect to the number of spans and their absolute errors Δa in dB, for the worst prediction case

We also evaluated the RMSE of each of the 500 links and plotted the evolution of a and the absolute errors $|\Delta a|$ with respect to the number of spans I for three selected cases : in Fig. 3.13 and Fig. 3.14, we plotted the worst and best prediction cases (according to the computed RMSE) respectively. For the worst case, the prediction gives an RMSE of 0.26 dB and 0.01 dB for the best prediction case. In Fig. 3.15, we have drawn the evolution of a for the link corresponding to 95^{th} percentile for evaluated RMSE. This figure also illustrates the high accuracy of our KerrNet with 95% of links prediction having most of errors less than 0.1 dB.

3.5.4 Speed comparison

As explained in the subsection 3.3.3, there is a trade-off between the integration points (higher number leads to , more precise but computational time higher) and the MC uncertainty. For instance, MC requires less integration points N_{mc} for a smaller symbol rate R compared to higher R. Consequently, we compare the computation speed for two extreme scenarios : on the one hand, WDM combs transmitting 20 GBd (case (A)) have the highest number of co-propagating channels, so a larger number of $\mathcal{X}$ coefficients need to be computed ; on the other hand, WDM combs transmitting 150 GBd (case (B)) have the smallest number of co-propagating channels but requires higher MC integration points N_{mc}, so each $\mathcal{X}$ coefficients computation time would be longer.

For the case (A), the evaluation of computation time is performed with a twenty-span link with 50km-long SSMF span, fully loaded with 161 channels, each transmitting at 20 GBd. For the case (B), the same link configuration is used but fully loaded with 27 channels transmitting at 150 GBd .

We timed the computation of the $\mathcal{X}$ coefficients with those two methods, on an "*Intel Xeon Processor E5-2640 v4*" CPU. The results are summarized in Tab. 3.3. For case (A), KerrNet computation time is 1.7×10^{-3} s vs. 3.9×10^{2} s

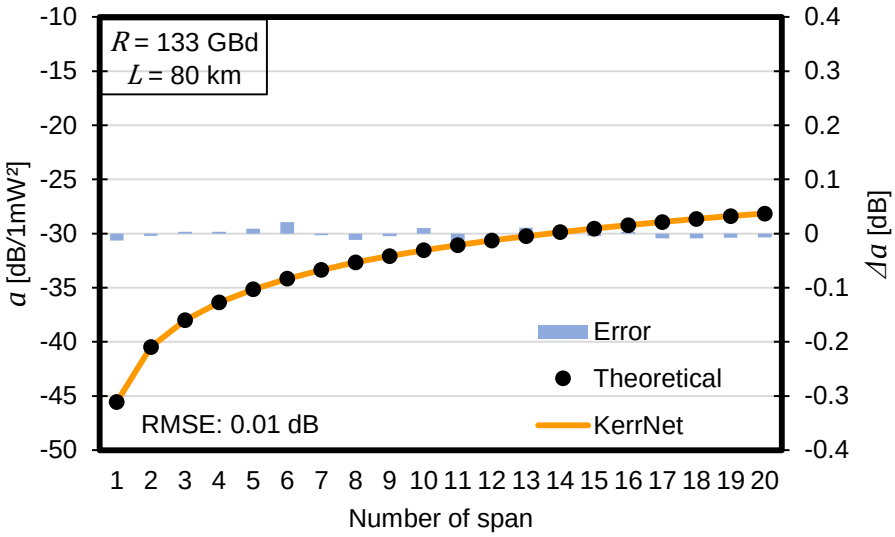


FIGURE 3.14 – Evolution of the nonlinear coefficient a with respect to the number of spans and their absolute errors Δa in dB,for the best prediction case

	case (A)	case (B)
WDM configuration	fully loaded (C band) with 161 channels	fully loaded with 27 channels
symbol rate R	$R = 20$ GBd	$R = 150$ GBd
Number of integration points N_w	1.5×10^6	1×10^7
Computation time with MC integration [seconds]	3.9×10^2	4.6×10^2
KerrNet computation time [seconds]	1.7×10^{-3}	3.4×10^{-4}
Speed-up factor	2.3×10^5	1.35×10^6

TABLE 3.3 – Computation time comparaison

for theoretical semi-analytical tool based on MC integration with 1.5×10^6 integration points, and for case (B), it is 3.4×10^{-4} s for KerrNet vs. 4.59×10^2 s for the semi-analytical tool with $1E7$ integration points. Hence, machine learning helped us to reach at least five orders of magnitude speeding using the same CPU device. We also clocked the computational time on a GPU, and the speed-up factor is even larger if we use large batches during the feed-forward inference with our KerrNet : we could obtain 2.9×10^{-4} s for case (A) and 5.4×10^{-5} s for case (B).

3.6 Nonlinear interference estimation in heterogeneous links

In this section, we present the generalization of the work presented in the previous section in which we only addressed homogeneous link cases. A representation of the general system under study was shown in Fig. 3.2, but we have made the assumption later that fiber parameters are constant for all channels (i.e. dispersion slope or channel dependent attenuation are ignored).

In this section, we deal with arbitrary loss/gain profiles (wavelength dependence) at each span. These gain/loss

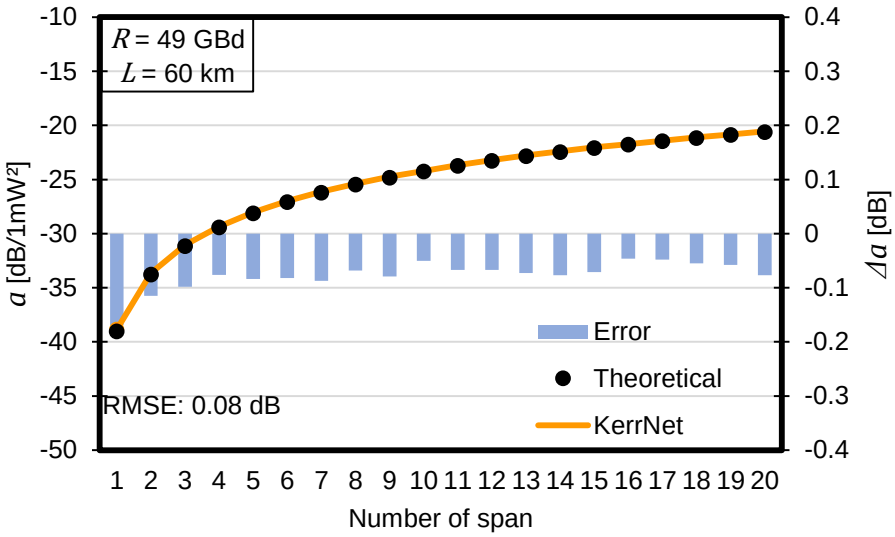


FIGURE 3.15 – Evolution of the nonlinear coefficient a with respect to the number of spans and their absolute errors in Δa dB, for the case corresponding to the 95^{th} percentile of RMSE

profiles can be due to residual gain ripples and/or tilts of EDFAs and/or non-flat response of the WSSs in optical nodes . With this approach, the channel power is no longer considered constant for all WDM channels and changes from span to span. Fig. 3.16 illustrates the transmission link under study with some examples of the arbitrary gain/loss wavelength dependent profiles $g_k(f)$ at the input of the k^{th} span.

For instance, at the input of the first span of the link, the power profile is given by $P_1(f) = P_c \times g_1(f)$ with P_c the initial average channel power at the output of the transmitter and g_1 is the residual gain/loss profile arising from the booster EDFA. Moreover, we consider that γ is constant for all channels (but γ can also be wavelength dependent since it does not appear in the computation of ρ thus it is not part of the input feature of the KerrNet framework either).

As described in subsection ii of 3.3.2, the computation of the nonlinear coefficient a (thus, also the NLI variance σ^2_{NLI}) can also be performed by the means of the cross-correlation function ρ to compute the perturbative coefficients $\mathcal{X}$ (cf Eq. (3.25)) and then applying Eq. (3.28).

Finally, to integrate the new features in this extended version of the KerrNet, all four $\mathcal{X}$ coefficients can be re-written using the cross-correlation function ρ leading to a new expression to compute σ^2_{NLI} and a as follows :

$$\sigma^2_{NLI} = P_c^2 a = \frac{16}{81} P_c^2 \left(3\mathcal{X}_{1,00}\right) + \left(\frac{\mu_4}{\mu_2^2} - 2\right)\left(5\mathcal{X}_{3,00}\right) + 4\sum_s \left[6\mathcal{X}_{1,0s} + 5\left(\frac{\mu_4}{\mu_2^2} - 2\right)\mathcal{X}_{3,0s}\right] \tag{3.29}$$

with P_c the mean channel power at the input of the link and an example of the expression for one of the four $\mathcal{X}$ coefficients is given by :

$$\mathcal{X}_{1,00}(L_{tot}, f_0) = \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \gamma_i \gamma_j g_i(f_0) g_j(f_0) \rho_1^{(sc)}(i,j) \tag{3.30}$$

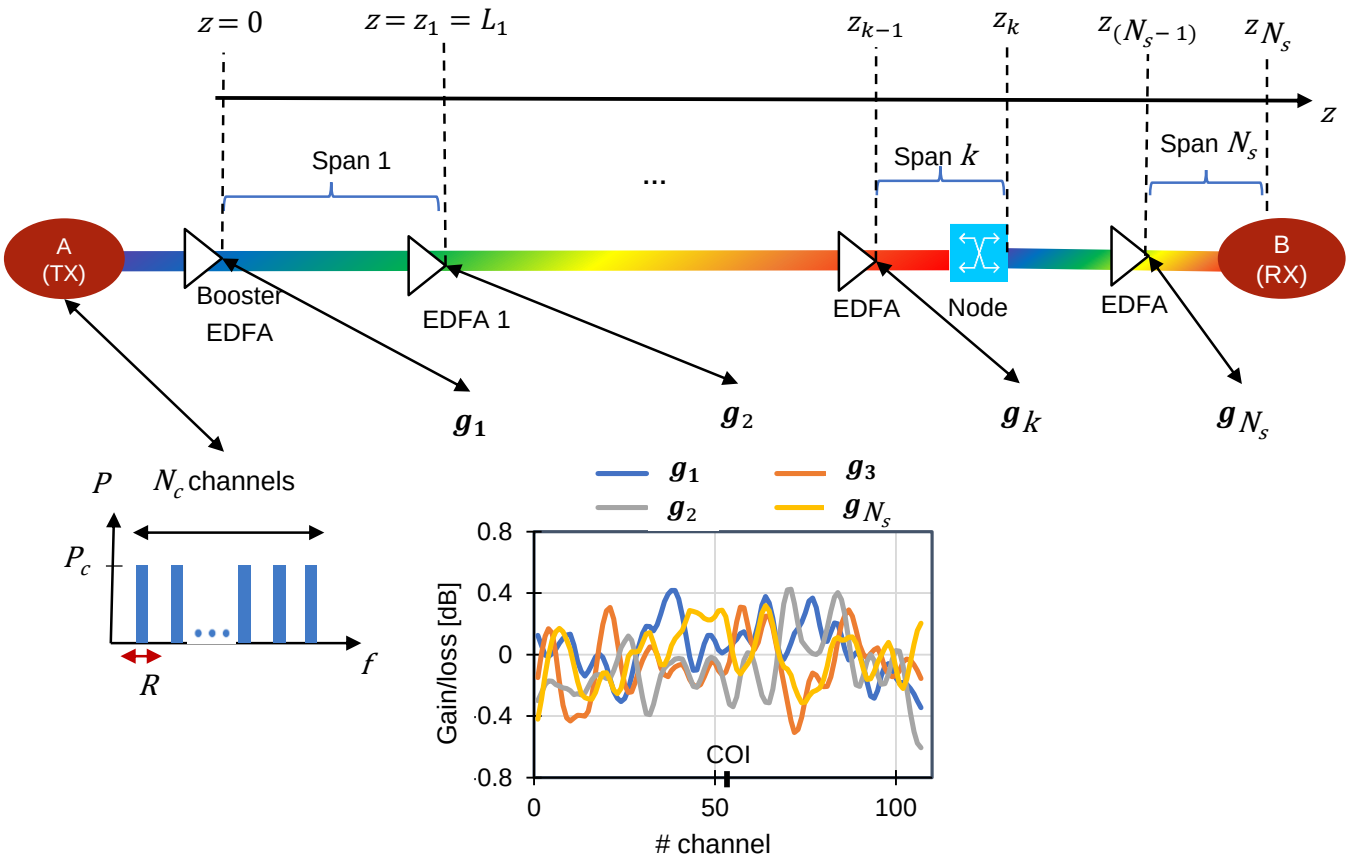


FIGURE 3.16 – Representation of the transmission link composed of arbitrary spans with wavelength-dependent local gain/loss affecting the channel power

where $g_i(f_0)$ and $g_j(f_0)$ is the gain/loss factor at the input of the i^{th} and j^{th} span respectively, evaluated for the COI (frequency f_0) and γ_i, γ_j are the nonlinear coefficients for those two spans. $\rho_1^{(sc)}(i,j)$ represents the cross-correlation interaction between i^{th} and j^{th} spans for self-channel interference which does not contribute to the modulation effect. The expressions of ρ are defined in Eq. 3.26.

For a matter of simplicity, here we only write the expression of $\mathcal{X}_{1,00}$ as an example (see the Appendix for the expression of $\mathcal{X}_{3,00}$, $\mathcal{X}_{1,0s}$ and $\mathcal{X}_{3,0s}$ using cross-correlation functions).

The fiber spans that compose our transmission link have different lengths, dispersion, loss and nonlinear coefficients so finally the cross-correlation function ρ depends on a reduced parameters space with only 12 parameters :

$$\rho(i,i) = \rho(bz_i, bz_j, D_i, D_j, \alpha_i, \alpha_j, L_i, L_j, f_0, f_s, R_0, R_s) \tag{3.31}$$

with bz, D, α, l corresponding to the accumulated dispersion, dispersion coefficient, fiber attenuation and span length respectively (for the i^{th} and j^{th} spans), and f_0, R_0, f_s, R_s correspond to the frequency and the symbol rate of the COI and s^{th} channels respectively. The goal of this work lies in proposing a bank of eight shallow FANNs to efficiently interpolate the eight 12-variable smooth functions of ρ in Eq. (3.31) as the key building block to compute the NLI

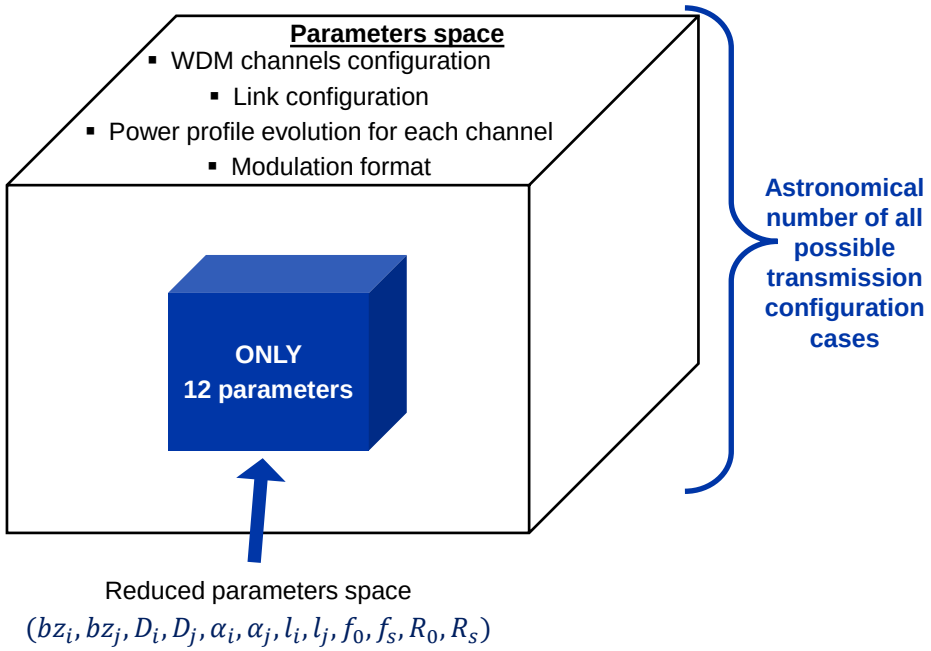


FIGURE 3.17 – Reduced parameters space with only 12 input features for arbitrary heterogeneous link performance estimation using the cross-correlation function ρ

variance for an arbitrary link.

In the following, we present the results only for the case $R_0 = R_s = R$ and for the COI at 1550 nm so f_0 is constant among our generated dataset, thus f_0 will not be included in the input parameters for the KerrNet and we used Ω_s (which is the center frequency de-tuning of the s'^{th} adjacent channel) instead of f_s. To summarize, the input features of our KerrNet version presented below only include 10 input parameters.

3.6.1 KerrNet configuration

We propose to compute eight types of ρ coefficients : $\rho_1^{(sc,ss)}$, $\rho_1^{(xc,ss)}$, $\rho_1^{(sc,ds)}$, $\rho_1^{(xc,ds)}$, $\rho_3^{(sc,ss)}$, $\rho_3^{(xc,ss)}$, $\rho_3^{(sc,ds)}$, $\rho_3^{(xc,ds)}$ with eight different and dedicated FANNs called ANN_1 to ANN_8. The labels sc and xc stand for SCI and degenerate XCI respectively, and the labels ss and ds stand for 'same-span' ($z_i = z_j$) and 'different-span' ($z_i \neq z_j$) NLI contributions, respectively. ρ coefficients with the sub-index 3 will contribute to the modulation-dependent terms in Eq. (3.29) and those with the sub-index 1 will not contribute to it.

We train eight FANNs (ANN_1 to ANN_8), each with only two hidden layers - to reduce the computational time - corresponding to the eight aforementioned types of cross-correlation functions ρ (see Fig. 3.18). The activation function is the ReLU for the two first layers and the linear function for the output layer. Hyper-parameters are determined with a random search. The dataset is generated by sampling the input super-vector $\mathbf{I}_1$ of all physical parameters (in this study $\mathbf{I}_1$ contains 10 parameters). Once trained, KerrNet automatically works for all choices of the input super-vector $\mathbf{I}_2$ containing the link configuration parameters since for both theory and for KerrNet, the passage from ρ coefficients

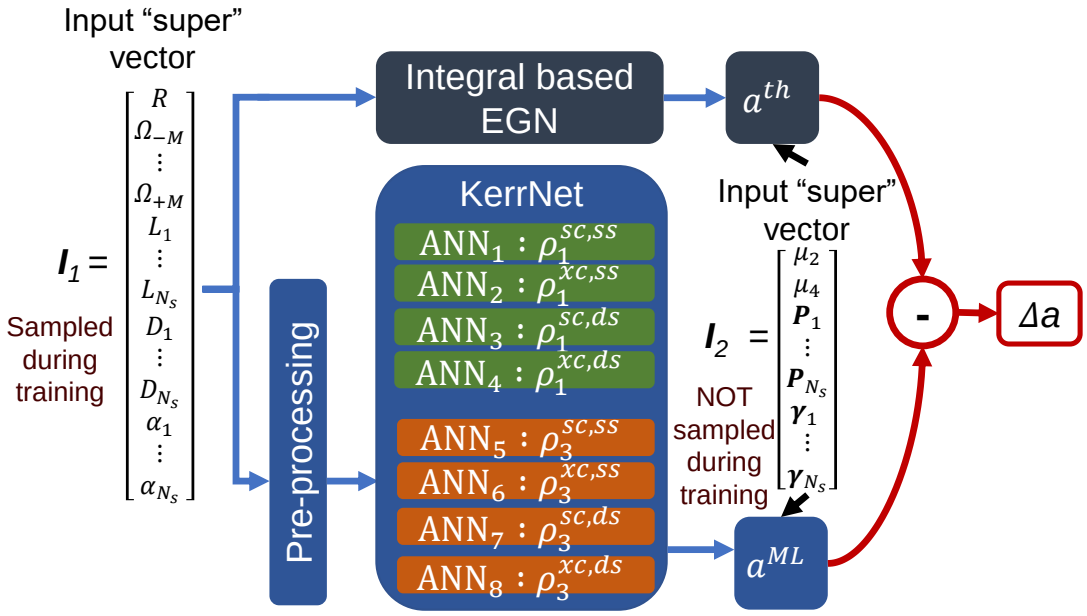


FIGURE 3.18 – Computation of a^{ML}, a^{th} and the absolute error $|\Delta a|$ between KerrNet estimation and the TD-RP theoretical result. KerrNet framework with 8 small ANNs with two hidden layers each. The number of total channel $N_c = 2M + 1$ and N_s is the total number of spans

Parameters	Range
D [ps/km/nm]	[2, 20], step = 0.01
α [dB/km]	[0.13, 0.30], step = 0.01
L [km]	[50, 100], step =1
L_{tot} [km]	[0, 10 000]
R [GBd]	[20, 150], step = 1
Ω_s [THz]	[-2,2] (C band), step = 1GHz
γ[1/W/km]	1.3174
Residual gain/loss [dB]	[-1,1]
Modulation format	QPSK, QAM, PCS (with various entropy values H)

TABLE 3.4 – Parameters range

to NLI requires the same elementary multiplications and summations involving power profiles, nonlinear coefficients, and modulation moments.

The range of the parameters is described in Tab 3.4. The upper six rows of the table summarize the physical parameters space for training and/or validating our FANNs (i.e., for $\mathbf{I}_1$) while the lower three rows include the system configuration parameters space only for validation (i.e., for $\mathbf{I}_2$).

3.6.2 Data generation and pre-processing

In this study with arbitrary heterogeneous links where the NLI is estimated by learning and interpolating the cross-correlation function ρ with FANNs, similar to the homogeneous case, we also apply this 'divide and conquer' method which has already proved its capabilities. Besides, the eight ρ coefficients ($\rho_1^{(sc,ss)}$, $\rho_1^{(xc,ss)}$, $\rho_1^{(sc,ds)}$, $\rho_1^{(xc,ds)}$, $\rho_3^{(sc,ss)}$,

ANN	Input parameters
ANN_1 for $\rho_1^{(sc,ss)}$	(D, α, l, R)
ANN_2 for $\rho_1^{(xc,ss)}$	$(D, \alpha, l, R, \Omega_s)$
ANN_3 for $\rho_1^{(sc,ds)}$	$(dbz, D_i, D_j, \alpha_i, \alpha_j, L_i, L_j, R)$
ANN_4 for $\rho_1^{(xc,ds)}$	$(dbz, D_i, D_j, \alpha_i, \alpha_j, L_i, L_j, R, \Omega_s)$
ANN_5 for $\rho_3^{(sc,ss)}$	(bz, D, α, l, R)
ANN_6 for $\rho_3^{(xc,ss)}$	$(bz, D, \alpha, l, R, \Omega_s)$
ANN_7 for $\rho_3^{(sc,ds)}$	$(dbz, bz_i, D_i, D_j, \alpha_i, \alpha_j, L_i, L_j, R)$
ANN_8 for $\rho_3^{(xc,ds)}$	$(dbz, bz_i, D_i, D_j, \alpha_i, \alpha_j, L_i, L_j, R, \Omega_s)$

TABLE 3.5 – Input parameters for each FANN

ANN	Input parameters
ANN_1 for $\rho_1^{(sc,ss)}$	$m = 0$
ANN_2 for $\rho_1^{(xc,ss)}$	$m = 0$
ANN_3 for $\rho_1^{(sc,ds)}$	$m = 1, \epsilon_{dbz} = 0$
ANN_4 for $\rho_1^{(xc,ds)}$	$m = 1, \epsilon_{dbz} = 0$
ANN_5 for $\rho_3^{(sc,ss)}$	$m = 1, \epsilon_{bz} = 5 \times 10^{-25}$
ANN_6 for $\rho_3^{(xc,ss)}$	$m = 1, \epsilon_{bz} = 1 \times 10^{-25}$
ANN_7 for $\rho_3^{(sc,ds)}$	$m = 1, \epsilon_{dbz} = 5 \times 10^{-25}, \epsilon_{bz} = 1 \times 10^{-25}$
ANN_8 for $\rho_3^{(xc,ds)}$	$m = 1, \epsilon_{dbz} = 1 \times 10^{-25}, \epsilon_{bz} = 5 \times 10^{-25}$

TABLE 3.6 – Pre-processing parameters for each FANN

$\rho_3^{(xc,ss)}$, $\rho_3^{(sc,ds)}$ and $\rho_3^{(xc,ds)}$) are different by at least one order of magnitude and depend on different parameters. We summarize in Tab. 3.5 the input feature for each of the eight FANNs (with their basic form, without mentioning the pre-processed form of parameters). For instance, the values of the accumulated dispersion for the i^{th} and j^{th} span bz_i and bz_j respectively, are less important to compute $\rho_1^{(sc,ds)}$ and $\rho_1^{(xc,ds)}$ but it is their difference $dbz = bz_i - bz_j$ which affects the computation result of these two ρ coefficients. At the contrary, the values of bz_i and dbz are both important to calculate $\rho_3^{(sc,ds)}$ and $\rho_3^{(xc,ds)}$.

With our previous study with the homogeneous link case, we noted that data generation and pre-processing have a huge impact on the performance of the FANNs. The pre-processing step employed for our KerrNet framework is illustrated in Fig. 3.19. We use the $\log$ function to compress input and output features which have a large range, to manage the data dynamic range. In Fig. 3.19, three pre-processing parameters (in yellow) appears : m, ϵ_{bz} and ϵ_{dbz}. Indeed, to avoid having very small numbers (near 0 or null) inside the $\log$ function, we also add the quantity ϵ before computing its $\log$ value. We called the quantity added to the parameter bz (accumulated dispersion), ϵ_{bz} and ϵ_{dbz} for dbz (difference of accumulated dispersion between two spans).

Since the number of integration points n_w is finite, there is always a tiny uncertainty on the results computed with MC integration. Thus, we always take the average value of ρ over five computations. Moreover, we noticed that for

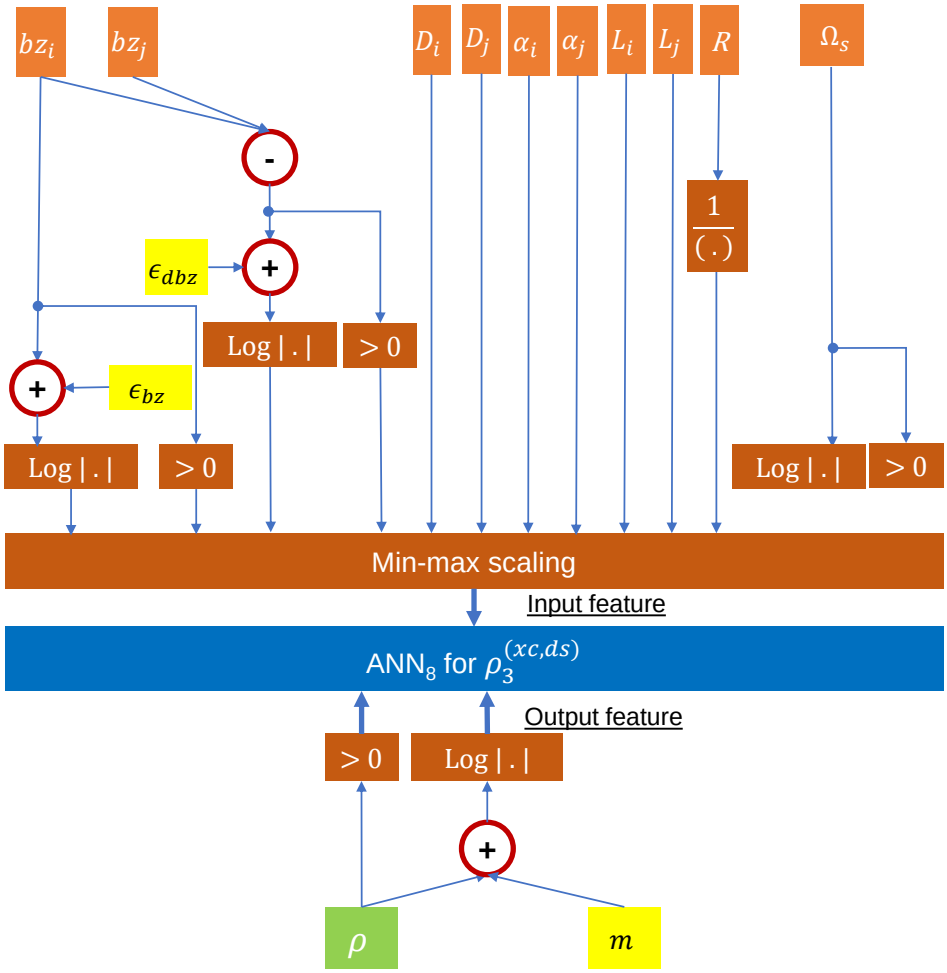


FIGURE 3.19 – Flowchart of the an example of the data pre-processing step while training the ANN$_8$ of the KerrNet framework for heterogeneous link

small values of ρ, the uncertainty of the MC integration result leads to a larger uncertainty when the $\log$ is applied leading to larger MSE during the training. To avoid this issue, we propose to smooth the oscillations by adding a constant m before computing the $\log$: the output of the FANN will be $\log|\rho + m|$ instead of unprocessed value of ρ. Finally, the min-max scaling is applied for all input features.

The training dataset quality is crucial in ML applications. Using a uniform distribution over the range of the parameters to generate the dataset is the simplest way and commonly adopted to train an ML framework but it is not optimal in our case. This would result in poor performance and large errors for the NLI estimation. Consequently, without a well-optimized distribution to generate the optimal training data - which is not a trivial task and requires a lot of effort - an ML-based solution with a high prediction accuracy is not feasible. Indeed, due to the large number of parameters, the important case which depends on the combination of several parameters, occurs more rarely, compared to noisy and small ρ coefficients.

To conclude, the distribution for the generation of each parameter has to be meticulously tailored. Second, the input and output vectors have to be pre-processed before feeding them to any of the FANNs. This also requires a fine

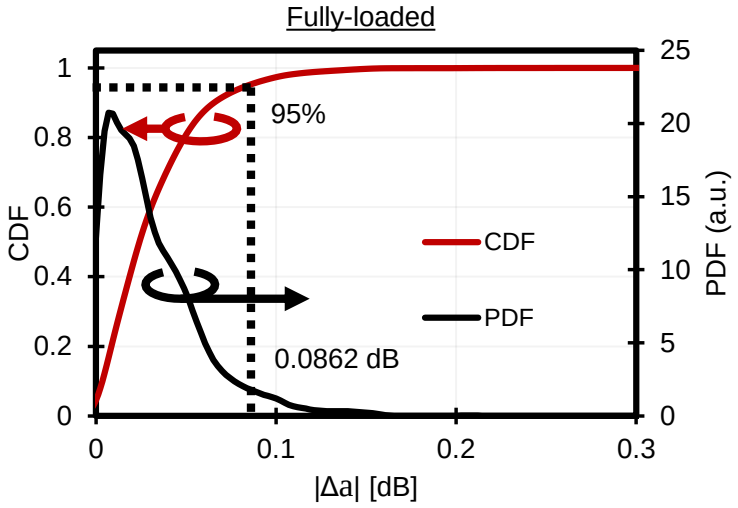


FIGURE 3.20 – PDF and CDF of the absolute error of the nonlinear coefficient $|\Delta a|$ in dB over test dataset without residual loss/gain and fully-loaded scheme.

investigation of the range of the input parameters and the nonlinear coefficient dependency on the input parameters.

3.6.3 Results without residual per-span gain/loss

After finding the adequate distribution to generate the dataset for each of the eight FAANs, we generated 10^5 data points for ANN$_1$ (for $\rho_1^{(sc,ss)}$) and ANN$_2$ (for $\rho_1^{(sc,ds)}$) and 2×10^5 data points for the six other FAANs. The generated dataset is split into training/validation dataset with an $80/20$ ratio. The validation dataset is used to avoid overfitting with the early-stop method as described in section 1.2 and for hyper-parameters searching.

Similar to the homogeneous link results, to evaluate the performance of KerrNet, we also only compare our ML prediction a^{ML} computed by KerrNet, with a^{th} computed with the theoretical integral-based TD-RP model. The performance evaluation was performed over a randomly generated test dataset : 7570 different and arbitrary transmission configurations are generated in total. The fiber characteristic vectors $\mathbf{D}$ and $\boldsymbol{\alpha}$, together with the fiber length vector $\mathbf{L}$, the symbol rate R, and the number of total spans N_s, are parts of the input "super vector" $\mathbf{I}_1$ generated with a uniform distribution over their range (cf Tab. 3.4) during the performance evaluation. The channel spacing is $12.5k$ GHz, where k is a random integer, and channel overlap is avoided. The random input super-vector $\mathbf{I}_2$ consisted of the vectors μ_2, μ_4 of the 2nd and 4th moments of the modulation formats ranging through unshaped and probabilistic-amplitude-shaped (PAS) QAMs with different entropy values. We set $\gamma = 1.3174$ W^{-1}km^{-1}. All the above mentioned generated parameters and the resulting computed ρ constitute the test dataset to evaluate the overall performance of the KerrNet framework compared the the theoretical results. In this subsection, we do not consider any extra gain/loss profiles in any spans.

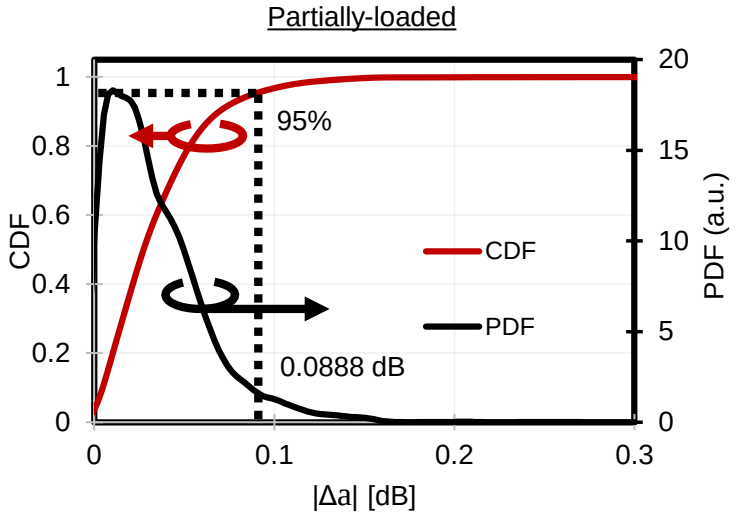


FIGURE 3.21 – PDF and CDF of the absolute error of the nonlinear coefficient $|\Delta a|$ in dB over test dataset without residual loss/gain and partially-loaded scheme

i Fully-loaded scheme

First, we present the result in a fully-loaded WDM scheme (in C-band). We plotted in Fig. 3.20 the PDF and the CDF of the absolute error in dB between the target theoretical estimation of the nonlinear coefficient a^{th} and the KerrNet framework's prediction a^{ML}. The mean value $|\Delta a|$ is 0.03 dB and 95% of the predictions give an absolute error less than 0.086 dB. This figure clearly illustrates the high accuracy of the KerrNet prediction for our arbitrary heterogeneous links with fully-loaded WDM channels.

ii Partially-loaded scheme

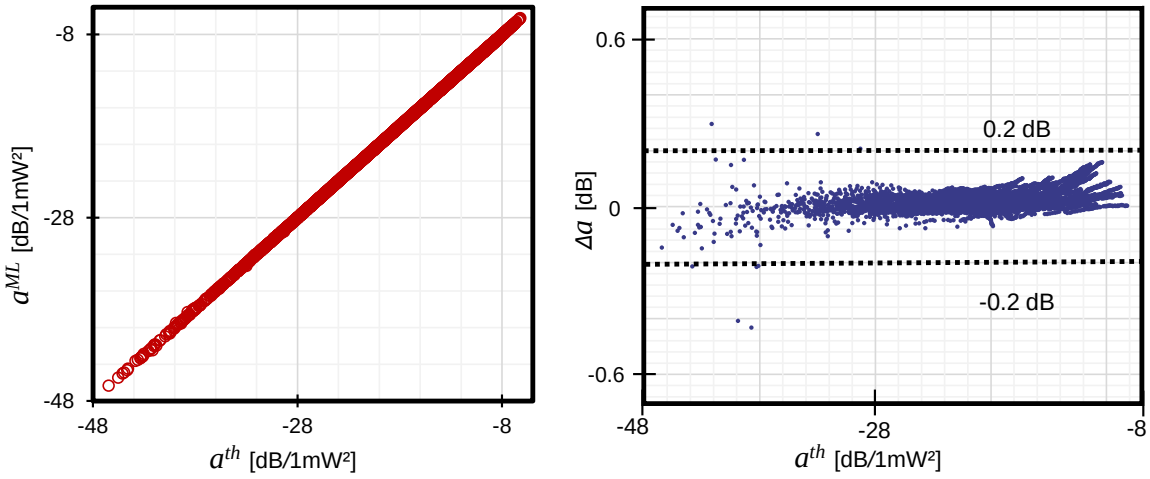


FIGURE 3.22 – (a) Scatter plot of the model predictions over the fully-loaded test dataset (with redisual gain/loss) a^{ML} vs. theoretical estimation a^{th} and statistics prediction errors and (b) the absolute error $|\Delta a|$ vs. a^{th}

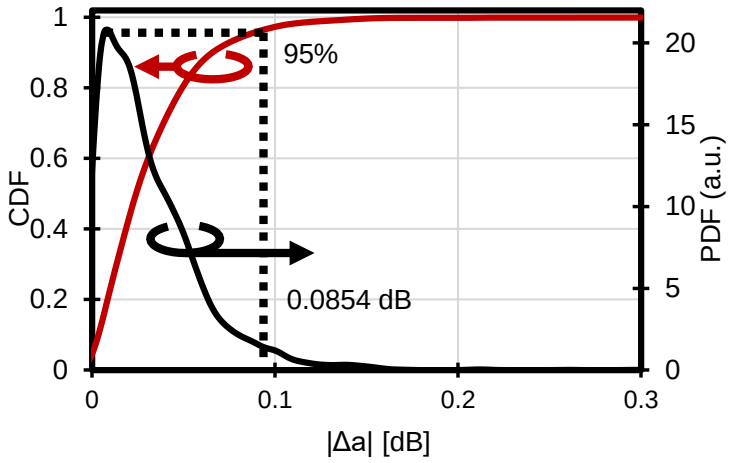


FIGURE 3.23 – PDF and CDF of the absolute error of the nonlinear coefficient $|\Delta a|$ in dB over the fully-loaded test dataset with residual loss/gain and fully-loaded scheme

Next, we present the result in a partially-loaded WDM scheme (in C-band). The WDM channels loading rate is between 38% and 100% generated uniformly. Fig. 3.21 illustrates the PDF and the CDF of $|\Delta a|$. The CDF indicates that 95% of the predictions give an absolute error less than 0.089 dB and the mean value $|\Delta a|$ is 0.034 dB. This figure shows that our KerrNet framework is a general QoT tool and it provides a highly accurate prediction of the nonlinear coefficient a for both fully-loaded and partially-loaded scheme. Besides, this also proves the high generalization capability of our solution since the retraining of the ML models is unnecessary to apply the KerrNet in arbitrary partially-loaded transmission scenarios. It is worth mentioning that if we use the SSFM or frequency-based EGN model as ground-truth data, an adequate sampling over all the WDM loading schemes will be indispensable and make the data generation and training much more complex.

3.6.4 Results with residual per-span gain/loss

In this subsection, we present the performance of our KerrNet framework in the general setting of a transmission link with residual gain/loss for each span and where the WDM channels are fully-loaded. The gain/loss profiles are generated randomly between $[-1, 1]$ dB and independently for each span. Then we use a window function to smooth the generated profiles.

Fig. 3.22 shows in (a) the scatter plot of the KerrNet's predicted a^{ML} vs. the theoretical estimation of a^{th} and in (b) the error Δa vs. a^{th} as a function of ground truth theoretical a^{th}. We observe that for all intensity of the nonlinear effect (with a ranging between -48 dB/mW^2 and -8 dB/mW^2), our KerrNet solution has a stable performance.

In Fig. 3.23, we report the statistics of the NLI error $|\Delta a|$ where $\Delta a = a^{th} - a^{ML}$ [dB] over the generated test dataset. The mean absolute error is 0.031 dB. Moreover, the CDF shows that the deviation from the theoretical value of a^{th} is less than ± 0.085 dB for 95% of the cases and less than ± 0.13 dB for 99% of the cases.

Fig. 3.24 shows the evolution of a^{th} and a^{ML} over distance for six distinct link configurations from our test dataset.

The evolution of the nonlinear coefficient a versus the total transmission distance L_{tot} for heterogeneous links is not trivial, but KerrNet outputs (markers) match the theoretical values (lines). This figure also illustrates the quality of our KerrNet tool : despite the form of the curves, our results match the theoretical one, which is not trivial.

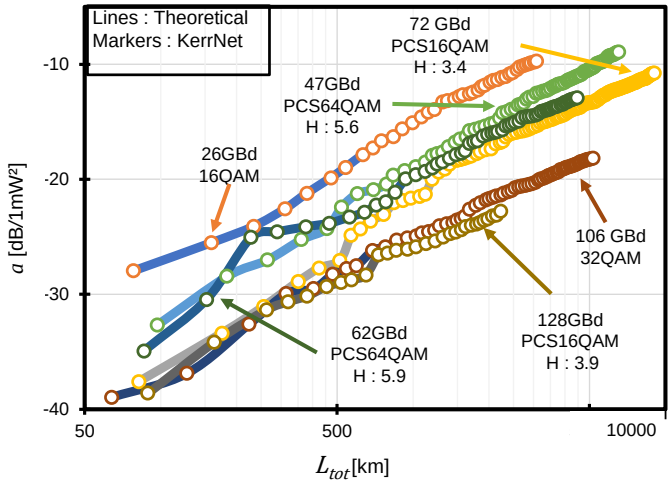


FIGURE 3.24 – NLI coefficient a vs. total transmission distance for six different links

3.6.5 Computation time

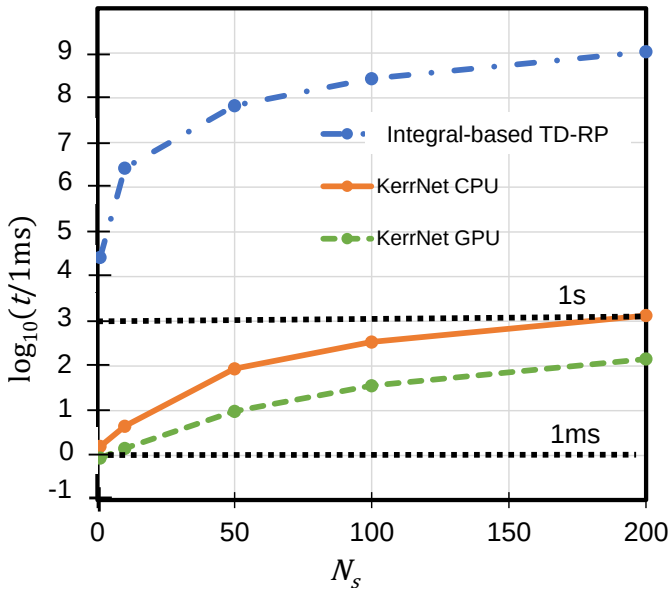


FIGURE 3.25 – Computation time with KerrNet and integral-based TD-RP theoretical model

In Fig. 3.25, we plot the theoretical and KerrNet computation time t vs. total number of spans N_s. The target configuration is a 200-span link fully loaded with 27 channels at 150 GBd. For the theoretical EGN computation, we used 3×10^6 integration points based on numerical error analysis of integral evaluations. We achieved a speed-

up factor up to 8×10^5 using a single CPU core and up to 7.5×10^6 using a single GPU. Moreover, thanks to the configuration that we have chosen (with a bank of eight small FANNs), our KerrNet is highly parallelable and the computation time can be further decreased using parallel processing enabling real-time applications.

3.7 Conclusion and perspectives

A new method for the QoT estimation in WDM system using machine learning is proposed and its accuracy and speed have been tested compared with an accurate semi-analytical model : the time-domain regular perturbation model. This tool, called KerrNet, speeds up the exact computation of the nonlinear interference variance up to six orders of magnitude without scarifying the accuracy. Moreover, we emphasised on and showed in this chapter the importance and non-trivial steps of the data generation and data pre-processing to successfully train the FANN-based ML framework.

For homogeneous links, the KerrNet is composed of four small FANNs and can predict the perturbative coefficients $\mathcal{X}$ of the semi-analytical TD-RP model with an exceptional level of accuracy (< 0.033 dB mean absolute error on estimation of a, the NLI coefficient with 95% of the prediction having an absolute error below 0.092 dB) and faster by up to five to six orders of magnitude. Our proposed method has been tested for 500 different links composed of twenty-spans of SSMF for a total transmission distance between 400 and 2400 km, and a C-band fully-loaded WDM scheme with symbol rate R between 20 and 150 GBd .

In the second part of this chapter, we reported an accurate ML-based framework to estimate the NLI for arbitrary heterogeneous links up to 10000 km, with WDM channels transmitting between 20 and 150 GBd and with arbitrary unshaped or shaped QAM formats (with various entropy values). This time, our KerrNet framework is composed of eight small FANNs to compute the perturbative coefficients $\mathcal{X}$ of the semi-analytical TD-RP model via the prediction of the cross-correlation function $\rho(i, j)$ between the i^{th} and j^{th} spans. We numerically evaluated the performance of our solution compared to the theoretical estimation at various configurations but without the need to perform the time-consuming re-training step for our trained FANNs. First, both fully-loaded and partially-loaded (loading rate between 38 and 100%) C-band WDM schemes are tested but without accounting the gain/loss profile at each span. The deviation from the theoretically estimated nonlinear coefficient Δa is less than ± 0.09 dB for boht fully-loaded and partially-loaded links for 95% of the cases from our tested dataset composed of 7570 distinct and arbitrary transmission heterogeneous links and different symbol rates and modulation formats. Finally, the evaluation was performed over the same test dataset but also integrating arbitrary residual gain/loss profiles at each span. The absolute error was below 0.085 dB for 95% of the cases.

Our implementation using a single GPU required only 36 ms for handling an example of full C-band transmission over 10000 km, but it can be parallelized in a straightforward way. With parallel processing and/or optimization of the neural network implementation, sub-millisecond execution times are feasible ; thus, enabling future real-time optical

network planning and control.

In this study, the dispersion and attenuation coefficients are considered constant for all WDM channels. An extension of this study for accounting for wavelength dependence of these effects can be achieved by re-training the ML model while increasing the number of its input features by adding six more input parameters. The six additional parameters include the fiber physical parameters value for the s^{th} neighbouring channel at the frequency f_s : i.e. $D_i(f_s), D_j(f_s), \alpha_i(f_s), \alpha_j(f_s), bz_i(f_s), bz_j(f_s)$ for the dispersion, attenuation and accumulated dispersion of the i^{th} and j^{th} span. However, there is no need to add the s^{th} channel's nonlinear coefficient γ value to the input features since γ does not account for the computation of ρ but only appears in the computation of the nonlinear coefficient a.

Chapter 4

Phenomenological estimation of the equalization enhanced phase noise

In this chapter, we focus on the characterization of the equalization enhanced phase noise (EEPN). In coherent dispersion-unmanaged (DU) transmission systems, the accumulated chromatic dispersion (CD) is compensated at the receiver side with DSP. The CD compensation is performed with an electronic equalizer as explained in 1.1.7 and is indispensable for high baud rate transmissions over long-distance fibers. However, this step generates an additional impairment from the interaction of the phase noise (PN) of the LO with the CD compensation equalizer called EEPN.

4.1 Introduction

The PN of a laser is the deviation of its electrical field from an ideal sinusoidal wave. It can also be described by a frequency-domain view of the frequency noise spectrum around the laser signal and the laser linewidth is related to this frequency noise PSD. Ideally, a laser has a Lorentzian spectral shape with a full width at half maximum (FWHM) defined as the laser linewidth.

In many applications, such as coherent optical fiber communication, this PN can impact the performance of the system. For coherent optical communications, two narrow-linewidth lasers are used : one at the transmitter side as a carrier, and one at the receiver side as a local oscillator (LO). As described in Chapter 1 in paragraph ii, the laser PN can be well mitigated by using the carrier phase estimation (CPE) in the DSP algorithm. But another impact of the phase noise called EEPN is also deteriorating the performance of DU transmission system. This effect emerges from the interaction between the widely used electronic CD compensation (CDC) and the PN of the LO laser. A first theoretical analysis of EEPN impairments was proposed by [14] using a time-domain impulse response approach

to compute the variance of EEPN.

It has been shown in [14] that if the CDC is performed at the receiver side, the transmitter PN do not induce EEPN because it goes through both the fiber channel (with CD effect) and the CDC equalizer, which perfectly cancels the effect. In [74], they also concluded that we can neglect the transmitter laser PN effect for EEPN with respect to the LO PN. Lau *et al.* have shown in [75] that designing a DSP algorithm to fully mitigate the EEPN, thus providing a significant performance improvement, is not easy. Moreover, the previous study has concluded that EEPN penalty increases with LO laser linewidth, accumulated dispersion and symbol rate. Thus, with recent transceivers operating near $100-150$ GBd, a reliable characterization method is needed specially for high speed and long haul transmission system.

However, the analytical model of EEPN assumes that the LO has a Lorentzian shape, and the EEPN is modeled as an additive, white and Gaussian noise. On the one hand, the laser linewidth measurement can be a tedious task requiring a dedicated setup with additional equipments. On the other hand, the ideal Lorentzian-shaped laser hypothesis does not hold for commercially available laser sources.

Besides, the work in [76] has shown that the DSP, in particular the CPE block, can mitigate partially the EEPN impairments. Thus, the estimation of the EEPN should be DSP-aware. Experimental characterization of the EEPN has been proposed in [77] but it requires using extremely a low PN laser as *"EEPN-free"* laser and back-to-back measurements are also indispensable.

In this chapter, we present a new experimental method to characterize the residual EEPN variance that truly affects the system's performance, without measuring the linewidth of the laser used as LO using the invalid notion of Lorentzian laser. One advantage of our method is that we do not rely on the use of extremely low PN lasers. Furthermore, this phenomenological method is also independent of the TX impairments, so it does not require a TX-impairments characterization.

4.2 Analytical estimation method

In this section, we describe the analytical estimation of EEPN firstly proposed by Shieh and Ho [14] and then discuss the limits of the model.

4.2.1 Equivalent model

To describe the EEPN effect, let us consider a multi-span coherent optical transmission system without inline dispersion compensation. The homogeneous link is composed of N_{span} identical fiber spans, each one amplified by an EDFA to compensate the power loss. Each fiber span's chromatic dispersion constant is D and the total length of the link is L. At the receiver side, a coherent receiver with an LO laser are used to convert the received optical field

into a digital signal. We only consider a simplified model with CD, ASE noise, and PN of the two lasers (Tx and LO). At the output of the coherent receiver, a DSP algorithm including CD compensation (considered ideal) and CPE are performed.

Fig. 4.1 shows the frequency-domain baseband-equivalent model that can be used to study the EEPN impairement. The first block of the simplified model accounts for both RRC pulse-shaping and CD accumulated during the propagation in the fiber with the transfer function $H(f)$. Then the ASE noise generated by all EDFAs n(t), considered AWGN is added to the model. Then, the signal is multiplied by the LO field for coherent detection. Finally, the DSP step begins with the application of an equalization filter $H^*(f)$ representing the RRC matched filter and sampling at rate $R_s = 1/T_s$ with T_s being the transmitted symbol duration, followed by the CPE block.

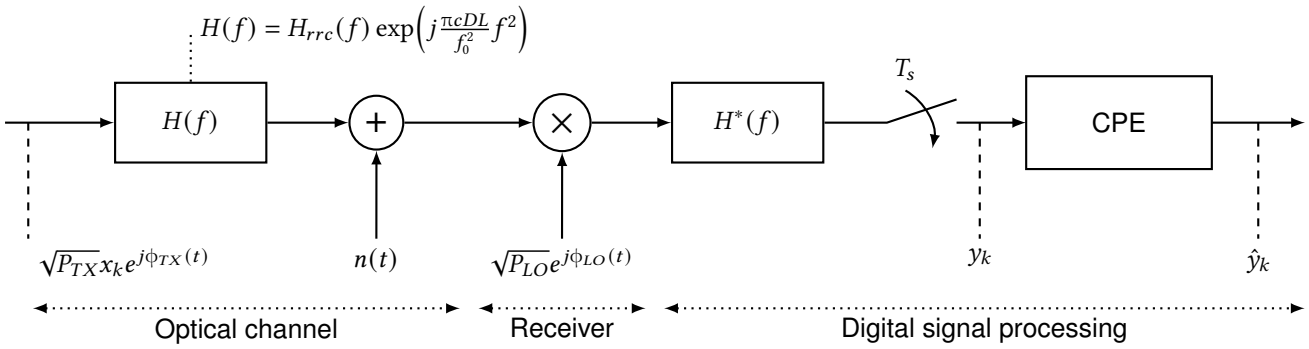


FIGURE 4.1 – Equivalent model schematic representation adapted from [76] with x_k transmitted symbols, $\Phi_{TX}(t)$: TX source laser PN process, $H(f)$: channel frequency response, $H_{rrc}(f)$: RRC function in frequency domain, C : the speed of light, D : fiber CD coefficient, L : total transmission length, f : frequency, f_0 : channel center frequency, T_s : symbol duration, $n(t)$: ASE noise, $\Phi_{LO}(t)$: LO PN process, P_{TX} : signal mean power at the transmitter, P_{LO} : LO laser power

4.2.2 Hypotheses of the ideal model and its limits

In the theoretical analysis, the transmitted symbols are assumed to be independent and identically distributed (iid) and filtered with an ideal Nyquist pulse shaping. We also suppose that the LO laser has a Lorentzian shape, so the PN can be approximated as a random walk following a Wiener process with the step size proportional to the laser linewidth $\Delta\nu$. EEPN is also assumed to be independent to LO field.

Finally, the EEPN arising from LO PN and the CD compensation filter is assumed to be a zero-mean additive, white and Gaussian noise having a variance defined as follows [14] :

$$\sigma^2_{EEPN} = \frac{\pi c(DL)R_s\Delta\nu}{2f_0^2} \tag{4.1}$$

with c the speed of light, D the dispersion coefficient, L the total length of transmission, $\Delta\nu$ the LO laser linewidth, and f_0 the LO frequency. We can notice that indeed, the EEPN variance increases proportionally with the accumulated

CD (D multiplied by distance L), the signal symbol rate R_s and the laser linewidth $\Delta\nu$.

However, the work in [76] studied the statistical properties of EEPN and concluded that EEPN noise is not a Gaussian-shaped noise, and highlighted the non-white nature of EEPN. They also demonstrated that it is strongly correlated to the LO field. Besides, the previous analyses of EEPN did not take into account the DSP impact, in particular the impact of CPE on the EPPN mitigation leading to an overestimation of its actual impact. Finally, lasers do not always show a Lorentzian shape (or equivalently a flat frequency noise spectrum, cf. Fig. 4.3) and the definition of their spectral shape and the value of their linewidth strongly depend on the observation window of these sources.

4.3 Phenomenological method

4.3.1 Characterization technique

Our characterization technique is based on the following well-known expression of the SNR of the signal at distance L, average power per channel P, and symbol rate R :

$$SNR = \frac{P}{N_0 LR + \eta(R)P + \sigma^2_{EEPN} PLR + aP^3} \tag{4.2}$$

where N_0 is the spectral level of the ASE, $\eta(R)P$ is the variance of the residual TRX noise in back-to-back, σ^2_{EEPN} is the EEPN noise variance coefficient that we aim to characterize in this work, and a is the Kerr nonlinear coefficient. To eliminate the dependence of the SNR on the variance of the residual TRX noise (which does not depend on the transmission distance), we choose to derive the Eq. (4.2) with respect to transmission distance L.

Let us define $s = 1/SNR$ and consider the following equation :

$$\begin{aligned} \frac{\partial s(P,L,R)}{\partial L} &= \frac{N_0 R}{P} + \sigma^2_{EEPN} R + \frac{a_{NLI}}{L} P^2 \\ &= \frac{\alpha}{P} + \beta + \gamma P^2 \end{aligned} \tag{4.3}$$

$\partial s/\partial L$ is called the slope in the following.

Starting from this equation to estimate the EEPN noise variance, we propose to first measure the quantity s for various values of transmission distance L, and per-channel power P, and compute the approximate slope of s with respect to L based on the experimental dataset. Finally, we propose to numerically fit the model of the slope vs. power as per Eq. (4.3)to the experimental slope vs. power curve and characterize the EEPN noise variance σ^2_{EEPN}. Moreover, we repeat the same procedure for various values of R for the sake of checking the consistency of the measured results.

4.3.2 Experimental setup

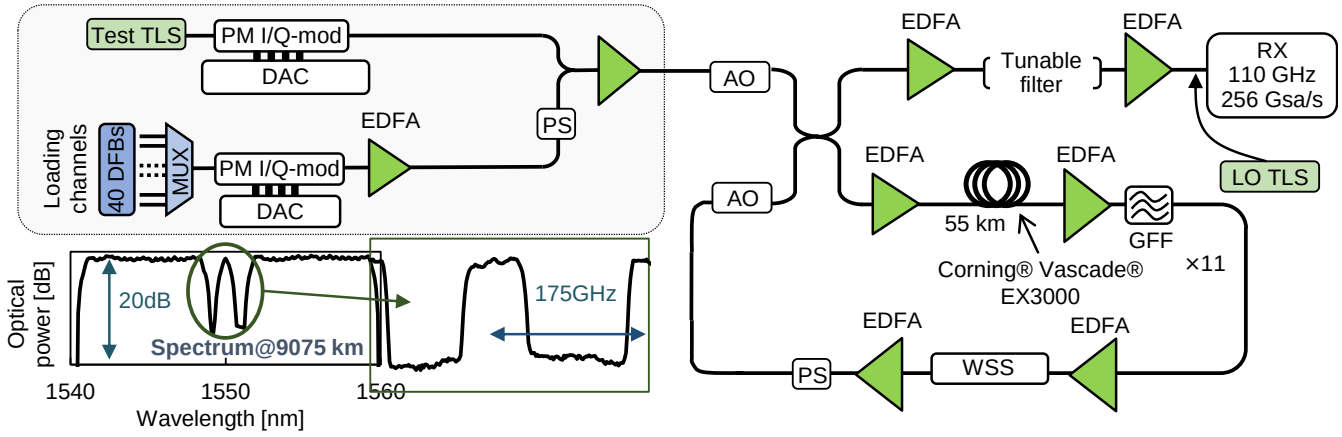


FIGURE 4.2 – Experimental setup

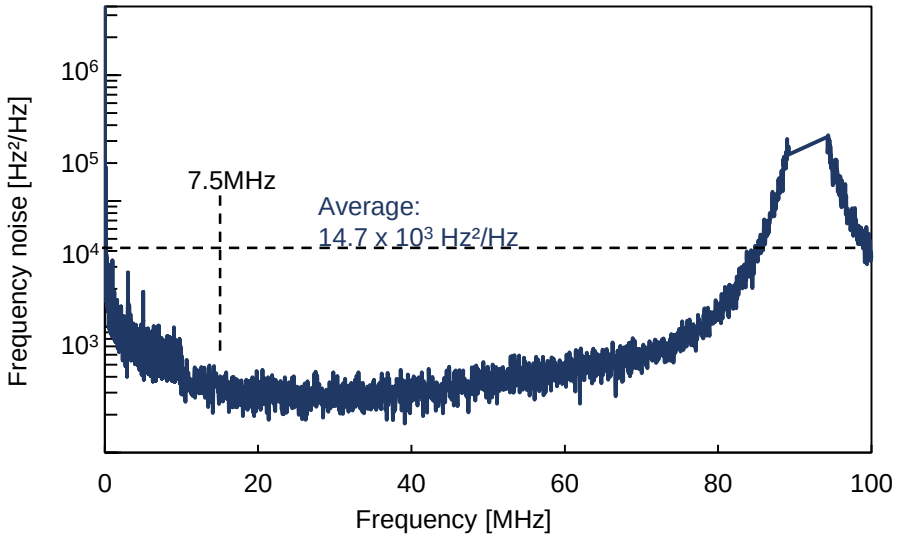


FIGURE 4.3 – The power spectral density (PSD) of the frequency noise of the TLS laser

The setup for our transmission experiment is shown in Fig. 2.4. The transmitted signal was synthesized using a WDM loading comb composed of 40 C-band DFB lasers spaced at 50 GHz, modulated with 49 GBd PDM QPSK signals. The test channel is made of a tunable laser source (TLS), separately modulated with PDM-QPSK signals at various symbol rates using a CMOS DAC operating at 120 GSamples/s. Digital pre-distortion is applied for the channel under test to provide a flat channel spectrum. The loading channels comb passed through a polarization scrambler (PS) before being multiplexed with the test channel. The test channel and the dummy WDM comb were multiplexed together and launched into the re-circulation loop. This loop consisted of 11 spans of 55 km Corning EX3000 fibers, with 0.157 dB/km loss coefficient, $D = 20.5$ ps/nm/km dispersion coefficient at 1550 nm, and 150 μm2 effective area. The span loss was compensated at the end of each span by a C-band EDFA followed by a GFF. A

50 GHz-grid-resolution WSS was used after the last span of the loop to equalize WDM channels across the whole C-band. Each loop thus emulated transmission over 605 km. We then performed transmission experiments at different total launched powers ranging from 11 to 17 dBm which were repeated for three different symbol rates of 80 GBd, 85 GBd and 90 GBd. For each configuration of launch power and symbol rates, we measured the channel under

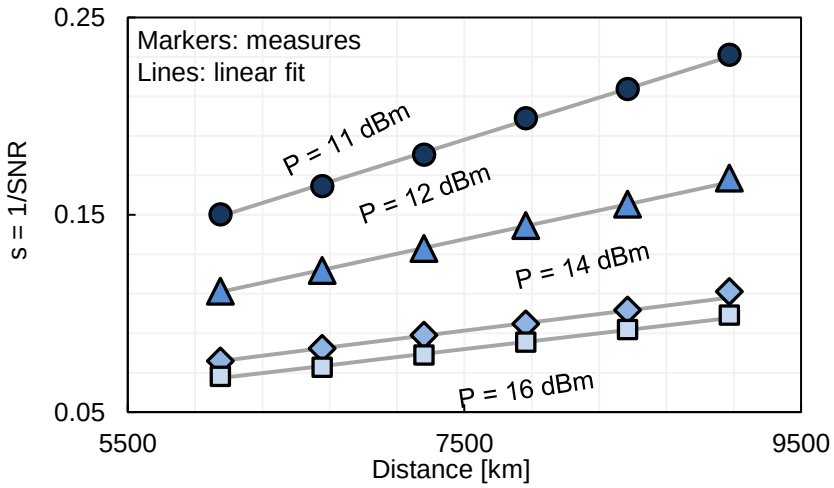


FIGURE 4.4 – The inverse of SNR, s, with respect to transmission distance for several configurations of P

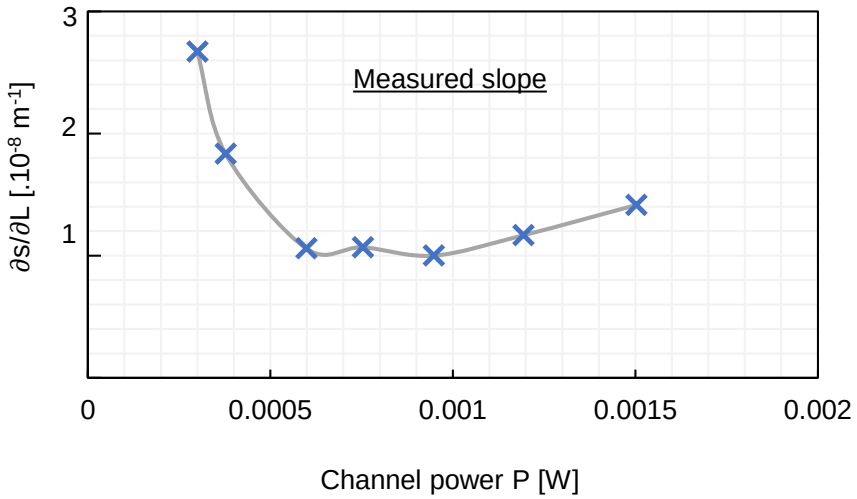


FIGURE 4.5 – The measured slope as a function of P

test after 10 to 15 loops by steps of one loop, so as to obtain distances ranging between 6050 and 9075 km by steps of 605 km. The signal at the loop output was received by a standard coherent receiver front-end with another TLS used as LO in this work. The frequency noise of the LO is illustrated in Fig. 4.3. The front-end signal is sampled at 256 Gsamples/s using a 110 GHz real-time sampling scope, and the standard DSP is applied off-line to the recorded sampled waveforms. The standard DSP suite consisted of chromatic dispersion compensation, complex MIMO 2×2 CMA, frequency offset compensation, and BPS carrier phase recovery with2% pilot overhead to remove cycle-slips,

followed by an LMS post-equalizer to mitigate transmitter imperfections.

4.3.3 Experimental results

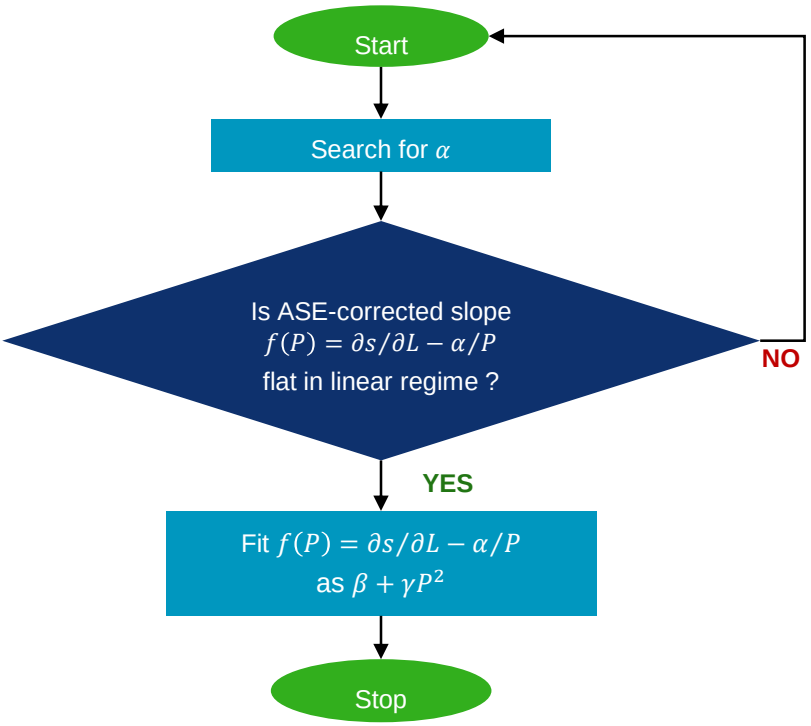


FIGURE 4.6 – The schematic flowchart to illustrate the extraction of the fitting parameters α, β and γ

The SNR of the received signals was calculated and the quantity $s = 1/SNR$ was obtained. Then, the approximate experimental slope of S vs. power was computed for each data set. We intend to numerically extract three parameters α, β, γ to fit the model of Eq. (4.3). i.e., $\partial s/\partial L = \alpha/P + \beta + \gamma P^2$ to the measured slope vs. power curves.

Fig. 4.4 illustrates experimental s vs. distance L curves for various powers, together with the linear fits in dashed lines using Huber cost function to suppress outliers' impact. The slope of these fitted lines are considered as the approximate experimental slope $\partial s/\partial L$. Fig. 4.5 now illustrates the measured experimental slope vs channel power P. We can see from this graph the evolution of the measured slope in the linear and slightly nonlinear regimes as a function of the channel power. To extract the EEPN noise variance as expected from Eq. (4.3), we proceeded in two steps. The flowchart illustrating this process is plotted in Fig. 4.6. First, we exhaustively searched for α, such that the ASE-corrected slope : $\partial s/\partial L - \alpha/P$ becomes as flat as possible for the powers corresponding to the linear regime below 14 dBm launch power (by minimizing the slope of the fitted line to the ASE-corrected slope).

Then, we plotted in Fig. 4.7 the ASE-corrected slopes for the measured channel powers at the baud rate of 85 GBd. As can be seen, the highest launched powers reveal the presence of nonlinearity as expected by the last term in Eq. (4.3). We finally interpolated the ASE-corrected slope according to the following form : $\beta + \gamma P^2$, i.e a constant term plus a parabolic law in order to estimate the $\gamma = \sigma^2_{EEPN} R$ and deduce the EEPN noise variance for

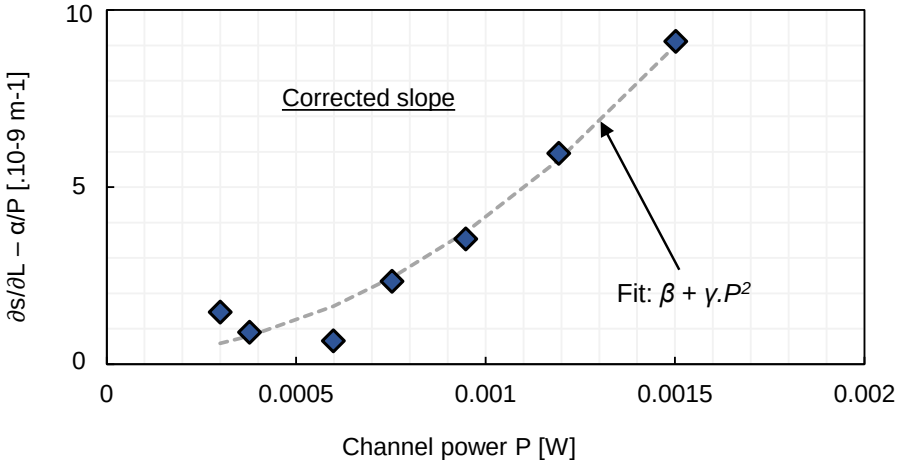


FIGURE 4.7 – The corrected slope as a function of P

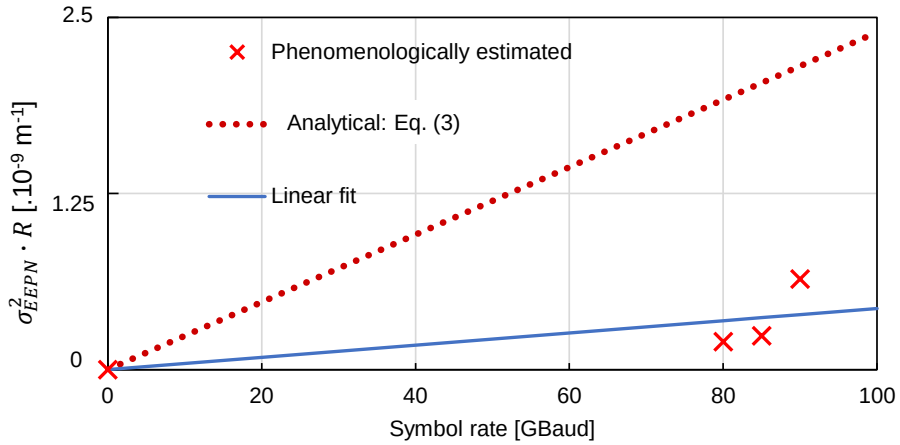


FIGURE 4.8 – The estimated EEPN noise $\times$ symbol-rate as a function of the nominal symbol-rate R

each symbol rate. Fig. 4.8 now shows the resulting EEPN noise variance estimated for our three measurement datasets at 80, 85 and 90 GBd.

Fig. 4.3 illustrates the PSD of the frequency noise of our LO TLS directly measured by a commercial optical phase noise test and measurement system. The LO TLS is clearly not a Lorentzian laser (otherwise the PSD of the frequency noise should be a flat line). Moreover, the PSD is not flat over the observation window pertinent to the EEPN characterization. This window is determined by the memory introduced by the CD compensation filter. For $R = 90$ GBd (i.e., $\Delta\lambda = 0.72$ nm), at $L = 9075$ km, the length of the impulse response of the CD filter is $DL\Delta\lambda \approx 133$ ns (~ 7.5 MHz^{-1}). This means that the PSD of the frequency noise should be integrated from 7.5 MHz onwards to obtain the variance of the phase noise of the LO under consideration here. This procedure allows us to estimate the actual frequency noise PSD by an equivalent constant level which results in the same area under the curve. The constant level is 14.7×10^3 Hz^2/Hz for our TLS LO. Then, following [78], we multiply this value by 2π and find the equivalent two-sided Lorentzian linewidth $\Delta\nu$ for our TLS LO. This value is then used to compute

the variance of EEPN given by the DSP-agnostic analytical expression in Eq. (4.1). The EEPN noise variance obtained from the DSP-agnostic analytical expression is also plotted in Fig. 4.8 for comparison. We can clearly see that the measured EEPN noise variances measured at the three symbol rates are consistent although well below the theoretical expectations taken from Eq. (4.1). This is consistent with previous published results [77] where experimental results showed that the analytical result overestimates the EEPN variance. Although there is a slight deviation among the estimated variances at different baud rates, we expect that the quality of estimation can be improved in future work if more data is collected to enhance the accuracy.

4.4 Conclusion

In this chapter, we detailed a new method to characterize the actual impact of EEPN in a real optical system. To estimate the EEPN variance, we conducted WDM transmission experiments in linear and slightly nonlinear regime without a priori knowledge of the laser phase noise characteristics. This phenomenological method, based on parameter extraction of a well-known analytical model for the evolution of the SNR along the link, is also independent from the transceiver imperfections.

Experimental validation of the technique has been performed over a submarine testbed with transmission distances ranging from 6050 to 9075 km. With the growing importance of EEPN effect in modern coherent systems over ultra-long haul distances, further improvement of the accuracy of the technique will be the subject of future works through the collection of more experimental data to enhance the accuracy of parameter extraction.

It is also worth to note that if the CDC is performed at the transmitter side, with a pre-dispersion compensation, the EEPN still exists but in this case it arises from the transmitter laser PN [74].

Conclusion

Over the past decades, optical network and submarine transmission systems have become the backbone of world transmission systems. The main goal of the thesis was to investigate several ML-based or data-driven methods to increase the transmission capacity or to analyze the performance of fiber-optic transmission systems for providing a reliable QoT estimation tool for low-margin design of the optical networks.

First, we can enhance the transmission capacity by increasing the bandwidth of the transmission system using ultra-wide band schemes. Since the bandwidth of EDFAs is not wide enough, hybrid Raman amplifier has gained popularity for its flexibility and performance but finding the optimal design of such amplifier is a complex task. Among possible solutions we considered, in this thesis, machine learning methods, and in particular ANNs, to design ultra-wideband Raman amplifier deployed in a hybrid amplification configuration with an SOA.

On the other hand, to meet the increasing internet traffic demands, we should fully exploit the optical network capacity with low-margin design. Consequently, a highly-accurate and ultra-fast QoT tool is indispensable including the estimation of each impairments impact. Among the major impairments, that occur during the propagation in the fiber, the estimation of the NLI and EEPN variances are the most challenging parts. To tackle this challenge, ANNs are also employed to estimate the most complex impairments contributions : the NLI variance estimation. Aside from ML-based solution, a new method was investigated to predict phenomenologically the impact of EEPN with experimental data.

Thesis summary and contributions

In Chapter 1, we first presented an overview of fundamental concepts of digital communication systems and the specificity of optical coherent transmissions. The state-of-the-art of the coherent systems and DSP algorithms are described, together with the presentation of different types of impairment and their contribution to the system's SNR.

In the same chapter, we also introduced a short description about basic knowledge of ML and ANN. In particular, their structure and the training process have been presented, as well as the influence of the important data processing step.

In Chapter 2, we illustrated the use of ANN to model the behavior of Raman amplifiers deployed in a hybrid

SOA/Raman configuration, to predict their gain profiles for a given pump currents configuration. Besides, we proposed another ANN to inverse design the ultra-wideband hybrid Raman amplifier for a given target gain profile. Moreover, we demonstrated that this solution is suitable for multi-span usage for arbitrary gain profile. We achieved a high accuracy ML-based framwork with 95% of predictions with an RMSE less than 0.51 dB for the generative model and 95% of predictions with an RMSE less than 0.74 dB for the inverse model.

In Chapter 3 we introduced ANN-based solution, called KerrNet, to provide a highly-accurate, ultra-fast and general QoT estimation tool for the Kerr NLI variance estimation for WDM transmission over the C-band. This framework is based on the theoretical time-domain regular perturbation model which is accurate up to the first-order of perturbation. The analysis has been conducted for both homogeneous and heterogeneous transmission link. First, we introduced a bank of four ANNs to estimate the perturbative coefficients $\mathcal{X}$ for homogeneous link with one-to-twenty spans of SSMF and fully-loaded WDM channels transmitting at R between 20 and 150 GBd in C-band. Then, the computation of the nonlinear coefficient a is straight-froward. We were able to achieve an absolute prediction error less than 0.092 dB for 95% of the cases between the KerrNet-predicted and TD-RP model-estimated nonlinear coefficient a while achieving a speed-up factor of five to six orders of magnitude of the TD-RP model computation time.

The second step consists in an extended version of our KerrNet framework to a more general configuration with heterogeneous link, up to $10,000$ km, composed of arbitrary fiber parameters (random dispersion and attenuation coefficients) and arbitrary fiber span length. Our proposed ML-based solution learns the cross-correlation function ρ between two spans instead of the perturbative coefficients $\mathcal{X}$ to guaranty a fixed number of parameters as input feature for the FANNs. Then the $\mathcal{X}$ is obtained by summing over the cross-correlation coefficients ρ for all spans and the nonlinear coefficient a is computed. The analysis has been conducted in both fully–loaded and partially–loaded scenarios and the general case, where the residual per span gain/loss profiles are included, have also be investigated. The prediction errors distributions were presented for all above-mentioned scenarios for arbitrary modulation format, symbol rate, WDM loading scheme and arbitrary wavelength-dependent gain/loss coefficients for each span. And the same level of low prediction errors over various scenarios highlighted the accuracy and the generalizability of our KerrNet solution. Meanwhile, we time the computation time for both KerrNet framework and integral-based TD-RP semi-analytical model (using MC integration). The obtained results are very promising : our proposed KerrNet framework can provide an speed-up up to six orders of magnitude while ensuring an average prediction error less than 0.13 dB for 99% of the tested cases composed of 7570 distinct and arbitrary heterogeneous transmission links, different symbol rate, modulation formats and per-span gain/loss profiles. Moreover, we investigated the importance of the data-generation and pre-processing step in the performance of our FANNs and highlighted the necessity of an optimized and tailored distribution for input parameters generation and a well-designed pre-processing algorithm.

Finally, in the last part of this work (Chapter 4), we investigated a new method to phenomenologically characterize the EEPN variance value in ultra-long haul WDM transmission experiments without a priori knowledge of the LO

laser PN characteristics. Based on parameter extraction of the well-known analytical model for the evolution of the SNR along the link, the technique is also independent from the transceiver imperfections and only measure the residual EEPN variance after CPE, that really affects the system's performance, contrary to other methods. We carried out experimental validation of the technique over a submarine test-bed with transmission distances ranging from 6050 to 9075 km for three symbol rate.

Perspectives

In this thesis, ML-based or data-driven methods to enhance or predict the optical transmission system's performance have been presented and these pave the way for future investigations.

Further extensions of the proposed ML-based solution presented in Chapter 2 would include a joint optimization for the hybrid Raman/SOA amplifier by also modelling the SOA with an ML framework.

The presented KerrNet framework (in Chapter 3) has showed promising results and there is still much room for improving this ML-based approach : future works of the proposed methodology could also include transmission link scenarios with negative dispersion fibers as well as transmission configurations composed of variable symbol rate WDM channels. Moreover, I-SRS can also be included in the framework for the NLI estimation of UWB systems. Besides, one challenging step in all ML framework is the data acquisition part, because of the high amount of data needed for training the ML framework. Several techniques, such as transfer learning, could also be implemented to accelerate the data collection stage.

Furthermore, we have seen in Chapter 4 that the phenomenological method to estimate the EEPN variance relies on the estimation of the parameters of a fitting function, thus further improvement of the accuracy of the technique would be possible through the collection of more experimental data.

Appendix A

Time domain-regular perturbation model for nonlinear interference variance computation

The Chapter 3, we presented an summary and some examples of the expressions used in TD-RP method to compute the NLI variance σ^2_{NLI} and the nonlinear coefficient a. This appendix aims to give more details about the computation of the perturbative coefficients $\mathcal{X}$ and the cross-correlation function ρ.

The $\mathcal{X}$-coefficients appearing in the variance of the signal-signal distortions inf Eq. (3.24) are defined in [11] as follows :

$$
\begin{aligned}
\mathcal{X}_{1,00} &= \frac{1}{T}\int_{\mathbb{R}^3}\frac{d^3\omega}{(2\pi)^3}\int_0^L dz H_{\vec{\omega}}(z)\int_0^L dz' H^*_{\vec{\omega}}(z') \\
\mathcal{X}_{2,00} &= \int_{\mathbb{R}^4}\frac{d^4\omega}{(2\pi)^4}\int_0^L dz H_{\vec{\omega}}(z)\int_0^L dz' H^*_{\vec{\omega}'}(z') \\
\mathcal{X}_{3,00} &= \int_{\mathbb{R}^4}\frac{d^4\omega}{(2\pi)^4}\int_0^L dz H_{\vec{\omega}}(z)\int_0^L dz' H^*_{\vec{\omega}}(z') \\
\mathcal{X}_{4,00} &= \int_{\mathbb{R}^3}\frac{d^3\omega}{(2\pi)^3}\int_0^L dz H^*_{\vec{\omega}''}(z) \\
\mathcal{X}_{5,00} &= T\int_{\mathbb{R}^5}\frac{d^5\omega}{(2\pi)^5}\int_0^L dz H_{\vec{\omega}}(z)\int_0^L dz' H^*_{\vec{\omega}'''}(z') \\
\mathcal{X}_{1,0s} &= \frac{1}{T}\int_{\mathbb{R}^3}\frac{d^3\omega}{(2\pi)^3}\int_0^L dz H_{\vec{\omega},s}(z)\int_0^L dz' H^*_{\vec{\omega},s}(z') \\
\mathcal{X}_{3,0s} &= \int_{\mathbb{R}^4}\frac{d^4\omega}{(2\pi)^4}\int_0^L dz H_{\vec{\omega},s}(z)\int_0^L dz' H^*_{\vec{\omega}'',s}(z') \\
\mathcal{X}_{1,ss'} &= \frac{1}{T}\int_{\mathbb{R}^3}\frac{d^3\omega}{(2\pi)^3}\int_0^L dz H_{\vec{\omega},s,s'}(z)\int_0^L dz' H^*_{\vec{\omega},s,s'}(z')
\end{aligned}
\tag{A.1}
$$

with kernel function H defined as :

$$H_{\vec{\omega},s,s'}(z) = \Pi_{\vec{\omega},s,s'} f(z) \times \exp\left\{ i\phi_{s,s'} + i\beta_2(z)\left[(\omega_2-\omega_3)(\omega_2-\omega_1) - ss'\Delta\Omega^2\right]\right\} \tag{A.2}$$

where $f(z)$ is the normalized power profile function and $\phi_{s,s'} = \phi_s(0) + \phi_{s'}(0) - \phi_{s+s'}(0)$ and $\phi_{s'}$ is the relative phase shift between COI and the $s'^{\,th}$ channel, $\Delta\Omega$ the channel spacing and $\beta(z)$ is the local GVD coefficient and $\Pi_{\vec{\omega},s,s'}$ is defined by :

$$\Pi_{\vec{\omega},s,s'} = \tilde{u}_0^{(0)}(0,\omega_1-\Omega_s)\,\tilde{u}_0^{(0)*}(0,\omega_2-\Omega_{s+s'}) \times \tilde{u}_0^{(0)}(0,\omega_3-\Omega_{s'})\,\tilde{u}_0^{(0)*}(0,\omega_1-\omega_2+\omega_3) \tag{A.3}$$

with $\Omega_s = s\Delta\Omega$ and $\tilde{u}_0^{(0)}(0,\omega)$ is the Fourier transform of the normalized base-band pulses assumed to be the Nyquist pulses defined as follows :

$$\tilde{u}_0^{(0)}(0,\omega) = \begin{cases} \sqrt{T}, & |\omega| < \frac{\pi}{T} \\ 0, & |\omega| \geq \frac{\pi}{T} \end{cases} \tag{A.4}$$

Note that the following notation simplifications are used : $H_{\vec{\omega},s} = H_{\vec{\omega},0,s}$, $\Pi_{\vec{\omega},s} = \Pi_{\vec{\omega},0,s'}$, $H_{\vec{\omega}} = H_{\vec{\omega},0,0}$ and $\Pi_{\vec{\omega}} = \Pi_{\vec{\omega},0,0}$ and we have also used the following shorthand notations

$$\begin{aligned} \vec{\omega}' &= (\omega_4, \omega_2, \omega_1+\omega_3-\omega_4) \\ \vec{\omega}'' &= (\omega_1, \omega_4, \omega_3-\omega_2+\omega_4) \\ \vec{\omega}''' &= (\omega_4, \omega_5, \omega_1-\omega_2+\omega_3-\omega_4+\omega_5) \end{aligned} \tag{A.5}$$

with ω_i , for $i = 1,\ldots,5$, are independent real dummy integration variables.

The four $\mathcal{X}$ coefficients can be re-written by the means of the cross-correlation function as follows :

$$\begin{aligned} \mathcal{X}_{1,00}(L_{tot},f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \gamma_i\gamma_j g_i(f_s)g_j(f_s)\rho_3^{(sc)}(i,j) \\ \mathcal{X}_{3,00}(L_{tot},f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \gamma_i\gamma_j g_i(f_s)g_j(f_s)\rho_1^{(xc)}(i,j) \\ \mathcal{X}_{1,0s}(L_{tot},f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \gamma_i\gamma_j g_i(f_s)g_j(f_s)\rho_3^{(xc)}(i,j) \\ \mathcal{X}_{3,0s}(L_{tot},f_s) &= \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \gamma_i\gamma_j g_i(f_s)g_j(f_s)\rho_1^{(sc)}(i,j) \end{aligned} \tag{A.6}$$

where $g_i(f_s)$ and $g_j(f_s)$ is the gain/loss factor at the input of the i^{th} and j^{th} span respectively, evaluated for the s^{th} channel (frequency f_s), γ_i and γ_j are the nonlinear coefficients for those two spans and $\rho(i,j)$ represents the cross-correlation interaction between i^{th} and j^{th} spans.

Bibliography

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

[2] Deepak Malik, Kuldip Pahwa, and Amit Wason. Performance optimization of soa, edfa, raman and hybrid optical amplifiers in wdm network with reduced channel spacing of 50ghz. *Optik*, 127(23) :11131–11137, 2016. ISSN 0030-4026. doi : https://doi.org/10.1016/j.ijleo.2016.09.047. URL https://www.sciencedirect.com/science/article/pii/S0030402616310555.

[3] Simranjit Singh and R.S. Kaler. Hybrid optical amplifiers for 64×10gbps dense wavelength division multiplexed system. *Optik*, 124(12) :1311–1313, 2013. ISSN 0030-4026. doi : https://doi.org/10.1016/j.ijleo.2012.03.073. URL https://www.sciencedirect.com/science/article/pii/S0030402612002756.

[4] Fukutaro Hamaoka, Masanori Nakamura, Seiji Okamoto, Kyo Minoguchi, Takeo Sasai, Asuka Matsushita, Etsushi Yamazaki, and Yoshiaki Kisaka. Ultra-wideband wdm transmission in s-, c-, and l-bands using signal power optimization scheme. *Journal of Lightwave Technology*, 37(8) :1764–1771, 2019. doi : 10.1109/JLT.2019.2894827.

[5] Jeremie Renaudier, Antonio Napoli, Maria Ionescu, Cosimo Calò, Gerrit Fiol, Vitaly Mikhailov, Wladek Forysiak, Nicolas Fontaine, Francesco Poletti, and Pierluigi Poggiolini. Devices and fibers for ultrawideband optical communications. *Proceedings of the IEEE*, 110(11) :1742–1759, 2022. doi : 10.1109/JPROC.2022.3203215.

[6] Jake Bromage. Raman amplification for fiber communications systems. *journal of lightwave technology*, 22(1) : 79, 2004.

[7] Yvan Pointurier. Machine learning techniques for quality of transmission estimation in optical networks. *Journal of Optical Communications and Networking*, 13(4) :B60–B71, 2021.

[8] Vladislav Neskorniuk, Andrea Carnio, Vinod Bajaj, Domenico Marsella, Sergei K. Turitsyn, Jaroslaw E. Prilepsky, and Vahid Aref. End-to-end deep learning of long-haul coherent optical fiber communications via regular perturbation model. In *2021 European Conference on Optical Communication (ECOC)*, pages 1–4, 2021. doi : 10.1109/ECOC52684.2021.9605928.

[9] Yapeng Xie, Yitong Wang, Sithamparanathan Kandeepan, and Ke Wang. Machine learning applications for short reach optical communication. *Photonics*, 9(1), 2022. ISSN 2304-6732. doi : 10.3390/photonics9010030. URL https://www.mdpi.com/2304-6732/9/1/30.

[10] Andrea Carena, Gabriella Bosco, Vittorio Curri, Yanchao Jiang, Pierluigi Poggiolini, and Fabrizio Forghieri. Egn model of non-linear fiber propagation. *Optics express*, 22(13) :16335–16362, 2014.

[11] Amirhossein Ghazisaeidi. A theory of nonlinear interactions between signal and amplified spontaneous emission noise in coherent wavelength division multiplexed systems. *Journal of Lightwave Technology*, 35(23) : 5150–5175, 2017.

[12] René-Jean Essiambre, Robert W Tkach, Roland Ryf, I Kaminow, T Li, and AE Willner. Fiber nonlinearity and capacity : Single-mode and multimode fibers. *Optical Fiber Telecommunications VI B*, pages 1–37, 2013.

[13] Pierluigi Poggiolini, Andrea Carena, Vittorio Curri, Gabriella Bosco, and Fabrizio Forghieri. Analytical modeling of nonlinear propagation in uncompensated optical transmission links. *IEEE Photonics technology letters*, 23 (11) :742–744, 2011.

[14] William Shieh and Keang-Po Ho. Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing. *Opt. Express*, 16(20) :15718–15727, Sep 2008. doi : 10.1364/OE.16.015718. URL https://opg.optica.org/oe/abstract.cfm?URI=oe-16-20-15718.

[15] Aymeric Arnould, Amirhossein Ghazisaeidi, Dylan Le Gac, Patrick Brindel, Mathilde Makhsiyan, Karim Mekhazni, Fabrice Blache, Nicolas Fontaine, David Neilson, Roland Ryf, Haoshuo Chen, Mohand Achouche, and Jérémie Renaudier. 103 nm ultra-wideband hybrid raman/soa transmission over 3 $\times$ 100 km ssmf. *J. Lightwave Technol.*, 38(2) :504–508, Jan 2020. URL https://opg.optica.org/jlt/abstract.cfm?URI=jlt-38-2-504.

[16] Amirhossein Ghazisaeidi, Aymeric Arnould, Maria Ionescu, Vahid Aref, Haik Mardoyan, Sophie Etienne, Mathieu Duval, Christian Bastide, Hans Bissessur, and Jeremie Renaudier. 99.35 tb/s ultra-wideband unrepeated transmission over 257 km using semiconductor optical amplifiers and distributed raman amplification. *Journal of Lightwave Technology*, 40(21) :7014–7019, 2022. doi : 10.1109/JLT.2022.3198518.

[17] E. Fred Schubert, Jaehee Cho, and Jong Kyu Kim. Light emitting diodes. 2003. URL https://api.semanticscholar.org/CorpusID:116310253.

[18] Govind P. Agrawal. Fiber-optic communication systems. 2021. URL https://api.semanticscholar.org/CorpusID:108425658.

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

[20] R. J. MEARS, L. REEKIE, I. M. JAUNCEY, and DAVID N. PAYNE. High-gain rare-earth-doped fiber amplifier at 1.54 μm. In *Optical Fiber Communication*, page WI2. Optica Publishing Group, 1987. doi : 10.1364/OFC.1987.WI2. URL https://opg.optica.org/abstract.cfm?URI=OFC-1987-WI2.

[21] C.R. Giles, E. Desurvire, J.R. Talman, J.R. Simpson, and P.C. Becker. 2-gbit/s signal amplification at lambda =1.53 mu m in an erbium-doped single-mode fiber amplifier. *Journal of Lightwave Technology*, 7(4) :651–656, 1989. doi : 10.1109/50.19091.

[22] J. Renaudier. Recent advances in ultra-wideband wdm transmission based on semiconductor optical amplifiers. In *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[23] T. Xu, G. Jacobsen, S. Popov, M. Forzati, J. Mårtensson, M. Mussolin, J. Li, K. Wang, Y. Zhang, and A. T. Friberg. Frequency-domain chromatic dispersion equalization using overlap-add methods in coherent optical system. *joc*, 32(2) :131–135, jun 2011. doi : 10.1515/joc.2011.022. URL https://doi.org/10.1515%2Fjoc.2011.022.

[24] Kazuro Kikuchi. Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers. *Opt. Express*, 19(10) :9868–9880, May 2011. doi : 10.1364/OE.19.009868.

[25] M. Selmi, C. Gosset, M. Noelle, P. Ciblat, and Y. Jaouën. Block-wise digital signal processing for polmux qam/psk optical coherent systems. *J. Lightwave Technol.*, 29(20) :3070–3082, Oct 2011.

[26] Timo Pfau, Sebastian Hoffmann, and Reinhold Noe. Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m -qam constellations. *Journal of Lightwave Technology*, 27(8) :989–999, 2009. doi : 10.1109/JLT.2008.2010511.

[27] Ivan Fernandez de Jauregui Ruiz, Amirhossein Ghazisaeidi, Omar Ait Sab, Philippe Plantady, Alain Calsat, Suwimol Dubost, Laurent Schmalen, Vincent Letellier, and Jeremie Renaudier. 25.4-tb/s transmission over transpacific distances using truncated probabilistically shaped pdm-64qam. *Journal of Lightwave Technology*, 36(6) :1354–1361, 2018. doi : 10.1109/JLT.2018.2791615.

[28] Pierluigi Poggiolini, G Bosco, A Carena, V Curri, Y Jiang, and F Forghieri. The gn-model of fiber non-linear propagation and its applications. *Journal of lightwave technology*, 32(4) :694–721, 2013.

[29] René-Jean Essiambre, Gerhard Kramer, Peter J. Winzer, Gerard J. Foschini, and Bernhard Goebel. Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4) :662–701, 2010. doi : 10.1109/JLT.2009.2039464.

[30] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997. ISBN 978-0-07-115467-3. URL https://books.google.fr/books?id=EoYBngEACAAJ.

[31] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning : From theory to algorithms*. Cambridge university press, 2014.

[32] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[33] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 9 :249–256, 2010.

[34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers : Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.

[35] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088) :533–536, 1986. ISSN 1476-4687. doi : 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

[36] Diederik Kingma and Jimmy Ba. Adam : A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

[37] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56) :1929–1958, 2014.

[38] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[39] Katharina Eggensperger, Matthias Feurer, Frank Hutter, James Bergstra, Jasper Snoek, Holger Hoos, Kevin Leyton-Brown, et al. Towards an empirical foundation for assessing bayesian optimization of hyperparameters. In *NIPS workshop on Bayesian Optimization in Theory and Practice*, volume 10, 2013.

[40] R. H. Stolen and E. P. Ippen. Raman gain in glass optical waveguides. *Applied Physics Letters*, 22(6) :276–278, March 1973. doi : 10.1063/1.1654637.

[41] Clifford Headley and Govind P Agrawal. *Raman amplification in fiber optical communication systems*. Academic press, 2005.

[42] Howard Kidorf, Karsten Rottwitt, Morten Nissov, Matthew Ma, and Eric Rabarijaona. Pump interactions in a 100-nm bandwidth raman amplifier. *IEEE Photonics Technology Letters*, 11(5) :530–532, 1999.

[43] M. Ionescu, D. Lavery, A. Edwards, E. Sillekens, L. Galdino, D. Semrau, R. I. Killey, W. Pelouch, S. Barnes, and P. Bayvel. 74.38 tb/s transmission over 6300 km single mode fiber with hybrid edfa/raman amplifiers. *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[44] M. Ionescu, A. Arnould, D. Le Gac, S. Etienne, A. Ghazisaeidi, M. Duval, C. Bastide, H. Bissessur, and J. Renaudier. 20.6 pb/s·km unrepeatered transmission without ropa : Uwb soa booster and backward raman amplification. *2020 European Conference on Optical Communications (ECOC)*, pages 1–4, 2020. doi : 10.1109/ECOC48923.2020.9333237.

[45] Aymeric Arnould, Amirhossein Ghazisaeidi, Dylan Le Gac, Patrick Brindel, Mathilde Makhsiyan, Karim Mekhazni, Fabrice Blache, Nicolas Fontaine, David Neilson, Roland Ryf, Haoshuo Chen, Mohand Achouche, and Jérémie Renaudier. 103 nm ultra-wideband hybrid raman/soa transmission over 3 $\times$ 100 km ssmf. *Journal of Lightwave Technology*, 38(2) :504–508, 2020. doi : 10.1109/JLT.2019.2946590.

[46] G C M Ferreira, S. Cani, Maria José Pontes, and Marcelo Eduardo Vieira Segatto. Optimization of distributed raman amplifiers using a hybrid genetic algorithm with geometric compensation technique. *IEEE Photonics Journal*, 3 :390–399, 2011.

[47] Hai ming Jiang, Kang Xie, and Ya fei Wang. Shooting algorithm and particle swarm optimization based raman fiber amplifiers gain spectra design. *Optics Communications*, 283(17) :3348–3352, 2010. ISSN 0030-4018. doi : https://doi.org/10.1016/j.optcom.2010.04.024. URL https://www.sciencedirect.com/science/article/pii/S0030401810003482.

[48] Junhe Zhou, Jianping Chen, Xinwan Li, Guiling Wu, Yiping Wang, and Wenning Jiang. Robust, compact, and flexible neural model for a fiber raman amplifier. *Journal of Lightwave Technology*, 24(6) :2362–2367, 2006. doi : 10.1109/JLT.2006.874602.

[49] Ann M. Rosa Brusin, Vittorio Curri, Darko Zibar, and Andrea Carena. An ultra-fast method for gain and noise prediction of raman amplifiers. In *45th European Conference on Optical Communication (ECOC 2019)*, pages 1–4, 2019. doi : 10.1049/cp.2019.0976.

[50] D. Zibar, A. Ferrari, V. Curri, and A. Carena. Machine learning-based raman amplifier design. In *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[51] Maria Ionescu. Machine learning for ultrawide bandwidth amplifier configuration. In *2019 21st International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, 2019. doi : 10.1109/ICTON.2019.8840453.

[52] J. Renaudier, A. Arnould, D. Le Gac, A. Ghazisaeidi, P. Brindel, M. Makhsiyan, A. Verdier, K. Mekhazni, F. Blache, H. Debregeas, A. Boutin, N. Fontaine, D. Neilson, R. Ryf, H. Chen, M. Achouche, and G. Char-

let. 107 tb/s transmission of 103-nm bandwidth over 3×100 km ssmf using ultra-wideband hybrid raman/soa repeaters. In *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[53] Darko Zibar, Ann Margareth Rosa Brusin, Uiara C. de Moura, Francesco Da Ros, Vittorio Curri, and Andrea Carena. Inverse system design using machine learning : The raman amplifier case. *Journal of Lightwave Technology*, 38(4) :736–753, 2020. doi : 10.1109/JLT.2019.2952179.

[54] Uiara Celine de Moura, Md Asif Iqbal, Morteza Kamalian, Lukasz Krzczanowicz, Francesco Da Ros, Ann Margareth Rosa Brusin, Andrea Carena, Wladek Forysiak, Sergei Turitsyn, and Darko Zibar. Multi–band programmable gain raman amplifier. *Journal of Lightwave Technology*, 39(2) :429–438, 2021. doi : 10.1109/JLT.2020.3033768.

[55] Uiara C. de Moura, Francesco Da Ros, A. Margareth Rosa Brusin, Andrea Carena, and Darko Zibar. Experimental characterization of raman amplifier optimization through inverse system design. *Journal of Lightwave Technology*, 39(4) :1162–1170, 2021. doi : 10.1109/JLT.2020.3036603.

[56] Anand Prakash, Jaisingh Thangaraj, Sharbani Roy, Shaury Srivastav, and Jitendra K. Mishra. Model-aware xgboost method towards optimum performance of flexible distributed raman amplifier. *IEEE Photonics Journal*, 15(4) :1–10, 2023. doi : 10.1109/JPHOT.2023.3286272.

[57] Metodi Plamenov Yankov, Francesco Da Ros, Uiara Celine de Moura, Andrea Carena, and Darko Zibar. Flexible raman amplifier optimization based on machine learning-aided physical stimulated raman scattering model. *Journal of Lightwave Technology*, 41(2) :508–514, 2023. doi : 10.1109/JLT.2022.3218137.

[58] Metodi P. Yankov, Darko Zibar, Andrea Carena, and Francesco Da Ros. Forward raman amplifier optimization using machine learning-aided physical modeling. In *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*, pages 1–3, 2022. doi : 10.23919/OECC/PSC53152.2022.9849857.

[59] Sandra Aladin, Anh Vu Stephan Tran, Stéphanie Allogba, and Christine Tremblay. Quality of transmission estimation and short-term performance forecast of lightpaths. *Journal of Lightwave Technology*, 38(10) :2807–2814, 2020.

[60] Jasper Müller, Sai Kireet Patri, Tobias Fehenberger, Helmut Griesser, Jörg-Peter Elbers, and Carmen Mas-Machuca. Qot estimation using egn-assisted machine learning for multi-period network planning. *J. Opt. Commun. Netw.*, 14(12) :1010–1019, Dec 2022. doi : 10.1364/JOCN.472632. URL https://opg.optica.org/jocn/abstract.cfm?URI=jocn-14-12-1010.

[61] Mahdi Ranjbar Zefreh, Fabrizio Forghieri, Stefano Piciaccia, and Pierluigi Poggiolini. Accurate closed-form real-time egn model formula leveraging machine-learning over 8500 thoroughly randomized full c-band systems. *Journal of Lightwave Technology*, 38(18) :4987–4999, 2020.

[62] F Arpanaei, B Shariati, P Safari, M Ranjbar Zefreh, JA Hernandez, A Carena, J Fischer, and D Larrabeiti. A novel approach for joint analytical and ml-assisted gsnr estimation in flexible optical network. In *European Conference and Exhibition on Optical Communication*, pages Tu5–60. Optica Publishing Group, 2022.

[63] Isaia Andrenacci, Matteo Lonardi, Petros Ramantanis, Élie Awwad, Ekhiñe Irurozki, and Stephan Clémençon. Fast and accurate nonlinear interference in-band spectrum prediction for sparse channel allocation. In *2023 International Conference on Optical Network Design and Modeling (ONDM)*, pages 1–5, 2023.

[64] Yan He, Zhiqun Zhai, Liang Dou, Lingling Wang, Yaxi Yan, Chongjin Xie, Chao Lu, and Alan Pak Tao Lau. Improved qot estimations through refined signal power measurements and data-driven parameter optimizations in a disaggregated and partially loaded live production network. *J. Opt. Commun. Netw.*, 15(9) :638–648, Sep 2023. doi : 10.1364/JOCN.496720. URL https://opg.optica.org/jocn/abstract.cfm?URI=jocn-15-9-638.

[65] Emmanuel Seve, J. Pesic, Camille Delezoide, S. Bigo, and Yvan Pointurier. Learning process for reducing uncertainties on network parameters and design margins. *Journal of Optical Communications and Networking*, 10 :A298, 02 2018. doi : 10.1364/JOCN.10.00A298.

[66] Andrea Carena, Vittorio Curri, Gabriella Bosco, Pierluigi Poggiolini, and F Forghieri. Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links. *Journal of Lightwave technology*, 30(10) :1524–1539, 2012.

[67] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Optics Express*, 21(22) :25685–25699, 2013.

[68] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Optics express*, 21 :25685–99, 11 2013. doi : 10.1364/OE.21.025685.

[69] Pierluigi Poggiolini and Yanchao Jiang. The egn model of nonlinear propagation in coherent optical transmission systems and its applications. 2014.

[70] Antonio Mecozzi and René-Jean Essiambre. Nonlinear shannon limit in pseudolinear coherent systems. *Journal of Lightwave Technology*, 30(12) :2011–2024, 2012.

[71] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Inter-channel nonlinear interference noise in wdm systems : Modeling and mitigation. *Journal of Lightwave Technology*, 33(5) :1044–1053, 2015. doi : 10.1109/JLT.2014.2384998.

[72] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Accumulation of nonlinear interference noise in fiber-optic systems. *Optics express*, 22(12) :14199–14211, 2014.

[73] Russel E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7 :1–49, 1998. doi : 10.1017/S0962492900002804.

[74] Gunnar Jacobsen, Marisol Lidn, Tianhua Xu, Sergei Popov, Ari T. Friberg, and Yimo Zhang. Influence of pre- and post-compensation of chromatic dispersion on equalization enhanced phase noise in coherent multilevel systems. 32(4) :257–261, 2011. doi : doi:10.1515/JOC.2011.053. URL https://doi.org/10.1515/JOC.2011.053.

[75] Alan Pak Tao Lau, William Shieh, and Keang-Po Ho. Equalization-enhanced phase noise for 100gb/s transmission and beyond with coherent detection. In *2010 IEEE International Conference on Communication Systems*, pages 456–460, 2010. doi : 10.1109/ICCS.2010.5686612.

[76] Aymeric Arnould and Amirhossein Ghazisaeidi. Equalization enhanced phase noise in coherent receivers : Dsp-aware analysis and shaped constellations. *Journal of Lightwave Technology*, 37(20) :5282–5290, 2019. doi : 10.1109/JLT.2019.2931841.

[77] Alexis Carbó Meseguer, Aymeric Arnould, Jean-Christophe Antona, Amirhossein Ghazisaeidi, Philippe Plantady, Suwimol Dubost, Alain Calsat, Elie Awwad, Jeremie Renaudier, and Vincent Letellier. Experimental characterization of equalization-enhanced phase noise in transoceanic transmission systems. In *45th European Conference on Optical Communication (ECOC 2019)*, pages 1–4, 2019. doi : 10.1049/cp.2019.0880.

[78] Gianni Di Domenico, Stéphane Schilt, and Pierre Thomann. Simple approach to the relation between laser frequency noise and laser line shape. *Appl. Opt.*, 49(25) :4801–4807, Sep 2010. doi : 10.1364/AO.49.004801. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-49-25-4801.

Publications

As first author

[P1] X. Ye and A. Ghazisaeidi, "KerrNet : Machine Learning to Speed up Exact Nonlinear Variance Computation of Arbitrary Links," in Optical Fiber Communication Conference (OFC) 2023.

[P2] X. Ye and A. Ghazisaeidi, "Phenomenological Characterization of the Electronically Enhanced Phase Noise in Transmission Experiments," in European Conference on Optical Communications (ECOC) 2022.

[P3] X. Ye and A. Ghazisaeidi, "Estimation de pertes de puissance pour les réseaux optiques," in Journées Nationales de l'Optique Guidée, Conférence OPTIQUE Nice 2022.

[P4] X. Ye and A. Ghazisaeidi, "KerrNet : Artificial Neural Networks to Speed up Perturbation Analysis-Based Models by Five Orders of Magnitude," in Optical Electronic Communication Conference (OECC) 2022.

[P5] X. Ye, A. Arnould, A. Ghazisaeidi, D. Le Gac and J. Renaudier, "Experimental Prediction and Design of Ultra-Wideband Raman Amplifiers Using Neural Networks", in Proc. OFC, paper W1K.3, 2020.

As co-author

[P6] S. Almonacil et al., "260-GBaud Single-Wavelength Coherent Transmission Over 100-km SSMF Based on Novel Arbitrary Waveform Generator and Thin-Film Niobate I/Q Modulator," in Journal of Lightwave Technology, 2023.

[P7] S. Almonacil et al., "Long-haul Transmission of 1 Tb/s Data Rate Channel with Inline Filtering based on 145 GBd Dual Polarization 16QAM," in Optical Fiber Communications Conference and Exhibition (OFC) 2023.

[P8] H. Mardoyan et al., "First 260-GBd Single-Carrier Coherent Transmission over 100 km Distance Based on Novel Arbitrary Waveform Generator and [...] ," in European Conference on Optical Communication (ECOC) 2022.

DOCTORAL SCHOOL

Titre : Intelligence Artificielle pour Optimiser la Capacité et Contrôler les Systèmes de Transmission sur Fibres Optiques

Mots clés : Intelligence Artificielle, Optimisation, Communications à Fibre Optique, Estimation de Performance

Résumé : La demande croissante de trafic internet a nécessité une augmentation continue de la capacité des systèmes de communication par fibre optique, fondement des réseaux de communication globaux. Cette thèse se propose des solutions innovantes pour relever les défis posés par l'amplification à bande ultra-large et l'estimation fine du bruit dans les systèmes de transmission optique. Les systèmes de communication par fibre optique ont connu une évolution significative pour répondre aux exigences croissantes en matière de capacité. Ils sont passés des amplificateurs optiques et de la détection cohérente aux formats de modulation et aux algorithmes de traitement numérique du signal avancés. Pour répondre à la demande croissante de trafic dans les réseaux optiques, l'intégration de schémas UWB et la mise en œuvre d'outil de conception de réseaux à faible marge sont devenues primordiales. Ce travail explore les aspects fondamentaux de l'amplification UWB. La prédiction précise des profils de gain Raman et la conception de configuration optimale sont primordiales, mais les méthodes conventionnelles s'avèrent très gourmandes en ressources de calcul. Dans ce contexte, l'apprentissage automatique apparaît comme un outil puissant, simplifiant la complexité et améliorant la précision dans ces scénarios. En outre, la thèse aborde le défi de la conception de systèmes à faible marge en développant un outil fiable de qualité de transmission. Les systèmes de transmission par fibre optique sont confrontés à diverses dégradations telles que l'atténuation de la fibre, le bruit des amplificateurs, le bruit de phase du laser, l'interférence non linéaire, etc. Alors que les détériorations linéaires peuvent être efficacement compensées et caractérisées, les méthodes traditionnelles peuvent manquer d'efficacité dans l'estimation de certaines détériorations non linéaires principales, car elles posent des problèmes en termes de précision et de complexité. Par conséquent, ce travail se penche sur les approches basées sur les données, y compris les modèles ML, afin de fournir une estimation efficace du bruit non linéaires Kerr et du bruit de phase renforcé électroniquement. En résumé, cette thèse s'appuie sur des méthodes d'apprentissage ou orientées données pour améliorer les performances des systèmes de transmission optique. Ces avancées sont prêtes à façonner l'avenir des systèmes de communication optique, en contribuant à des capacités plus élevées et à des transmissions plus fiables dans notre environnement numérique qui évolue rapidement.

Title : Applications of Artificial Intelligence to Control and Analyze the Performance of Fiber-Optic Transmission Systems

Keywords : Artificial Intelligence, Optimization, Fiber-Optic Communications, Quality of transmission

Abstract : The surging demands for internet traffic have necessitated continuous expansion in optical fiber communication systems capacity, cornerstone of global communication networks. This thesis delves into innovative solutions addressing the challenges posed by ultra-wideband (UWB) amplification and precise noise estimation in optical transmission systems. Optical fiber communication systems have undergone significant evolution to meet escalating capacity requirements. Progressing from optical amplifiers and coherent detection to advanced modulation format and digital signal processing (DSP) algorithms. To meet the need for higher traffic demands in optical networks, integrating UWB schemes and implementing low-margin network designs have become primordial. This work explores fundamental aspects of UWB amplification. Accurate prediction of Raman gain profiles and optimal pump configurations design is paramount, yet conventional methods prove computationally intensive. Here, Machine Learning (ML) emerges as a powerful tool, simplifying complexity and enhancing accuracy in these scenarios. Additionally, the thesis addresses the challenge of designing low-margin systems by developing a reliable Quality of Transmission (QoT) tool. Optical fiber transmission systems contend with diverse impairments such as fiber attenuation, ASE noise, laser phase noise, nonlinear interference (NLI), etc. While linear impairments can be effectively mitigated and characterized, traditional methods may falter in estimating some major nonlinear impairments, posing challenges in accuracy and complexity. Consequently, this work delves into data-driven approaches, including ML frameworks, to provide effective estimation of Kerr nonlinear impairments and electronically enhanced phase noise (EEPN) In summary, this thesis leverages ML and data-driven methods to enhance the performance of optical transmission systems. These advancements are poised to shape the future of optical communication systems, facilitating higher capacities and more reliable transmissions in our rapidly evolving digital environment.

Institut Polytechnique de Paris

91120 Palaiseau, France