Enhancing optical fiber transmission performance through advanced link and system design

Soltani, Mehran

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Soltani, M. (2023). *Enhancing optical fiber transmission performance through advanced link and system design*. Technical University of Denmark.

Technical University of Denmark

Ph.D. Thesis

Enhancing optical fiber transmission performance through advanced link and system design

Mehran Soltani

Supervisors:
Darko Zibar, Prof.
Francesco Da Ros, Assoc. Prof.
Andrea Carena, Prof.

June 2023

Ørsteds Plads, Byg. 343
DK-2800 Kongens Lyngby
Denmark

DTU Electro
Department of Electrical and Photonics Engineering

Preface

The work presented in this thesis was carried out as a part of my Ph.D. project from July 1st, 2020 to June 30th, 2023. The work took place at the Department of Electrical and Photonics Engineering of the Technical University of Denmark (DTU Electro), with an external stay of one month at Politecnico Di Torino (PoliTo), Turin, Italy as part of the Advanced Technology department. This Ph.D. project was financed by the European Research Council through the ERC-CoG FRECOM project under Grant Agreement 771878, and supervised by:

- Darko Zibar (main supervisor), Professor, Department of Electrical and Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

- Francesco Da Ros (co-supervisor), Senior Researcher, Department of Electrical and Photonics Engineering, Technical University of Denmark, 2860, Kgs. Lyngby, Denmark.

- Andrea Carena (co-supervisor), Professor, Department of Electronics and Telecommunications (DET), Polytecnico Di Torino, Turin, Italy.

Abstract

Distributed Raman amplification (DRA) is a key technology that can improve the performance of fiber optic communication systems. This amplification scheme provides several advantages over the Erbium-Doped Fiber Amplifiers (EDFAs), in terms of Noise Figure (NF), broadband gain, and flexibility in design by means of multi-pumping schemes. Due to its distributed amplification, DRA enables to control of the shape of signal power evolution in both frequency and fiber distance. This is crucial for attaining some of the long-term objectives in fiber optic communications, including optimization of Signal-to-Noise Ratio (SNR) and compensating for nonlinear impairments. However, the optimization of the pump power and wavelength values poses a challenge to DRA configurations.

In this thesis, we utilize Machine Learning (ML) and optimization techniques to design signal power evolution in two dimensions (2D), i.e. frequency and fiber distance, using Raman amplifiers. First, an inverse system model based on a Convolutional Neural Network (CNN) is used to map the 2D signal power profiles to their corresponding Raman pump power and wavelength values. The CNN model has shown a statistically low error in learning the inverse mapping. However, its performance is not accurate for designing 2D profiles of practical interest, such as a 2D flat or a 2D symmetric (with respect to the midpoint in the distance). To accurately design the practical 2D profiles, we use an online optimization framework based on Differential Evolution (DE). In this framework, the DE adjusts the pump power values online on the setup aiming to reduce the cost value between the desired and the designed 2D profiles. The DE framework is also combined with the CNN inverse model to achieve better accuracy, more reliable optimum values, and faster convergence.

Finally, we experimentally validate the performance of the CNN model, the DE, and the CNN-assisted DE framework using an amplifier setup employing four counter-propagating Raman pumps. Different target power profiles defined jointly in the entire C-band and in fiber distance are aimed to be designed. Moreover, the DE framework is tested and

showed promising performance in an experimental multi-objective design scenario to achieve 2D profiles with flat gain levels at the end of the span, jointly with minimum spectral excursion over the entire fiber length.

Resumé på Dansk

Distribueret Raman-forstærkning (DRA) er en nøgleteknologi, der kan forbedre ydeevnen af fiberoptiske kommunikationssystemer. Denne forstærkningsmetode giver flere fordele i forhold til erbium-dopede fiberforstærkere (EDFA'er), når det kommer til støjfaktor (NF), bredbåndsforstærkning og fleksibilitet i design ved hjælp af multipumpning. På grund af sin distribuerede forstærkning muliggør DRA styring af formen på signalstyrkens udvikling både i frekvens og fiberafstand. Dette er afgørende for at opnå nogle af de langsigtede mål inden for fiberoptisk kommunikation, herunder optimering af signal-to-noise ratio (SNR) og kompensation for ikke-lineære forstyrrelser. Dog udgør optimeringen af pumpekraft og bølgelængdeværdier en udfordring for DRA-konfigurationen.

I denne afhandling udnytter vi maskinlæring (ML) og optimeringsteknikker til at designe signalstyrkeudvikling i to dimensioner (2D), nemlig frekvens og fiberafstand, ved hjælp af Raman forstærkere. Først anvender vi en omvendt systemmodel baseret på et konvolutionsneuralt netværk (CNN) til at kortlægge 2D-signalstyrkeprofiler til deres tilsvarende Raman-pumpens effekt- og bølgelængdeværdier. CNN modellen har vist en statistisk lav fejl i indlæringen af den omvendte kortlægning. Dog er dens præstation ikke nøjagtig nok til at designe 2D-profiler af praktisk interesse, såsom en 2D-flad eller en 2D-symmetrisk profil (i forhold til midtpunktet i afstanden). For at kunne designe praktiske 2D-profiler præcist anvender vi en online optimeringsramme baseret på Differential Evolution (DE). I denne ramme justerer DE online pumpens effektværdier i opsætningen for at reducere omkostningsværdien mellem den ønskede 2D-profil og den designede 2D-profil. DE-rammen kombineres desuden med CNN-omvendtmodellen for at opnå bedre nøjagtighed, mere pålidelige optimale værdier og hurtigere konvergens i forhold til den selvstændige DE-ramme.

Endelig validerer vi eksperimentelt ydeevnen af CNN-modellen, DE og CNN-assisterede DE-rammen ved hjælp af en forstærkeropsætning

med fire modstrøms Raman-pumper. Forskellige målrettede effektprofiler, defineret over hele C-båndet og i fiberafstand, sigter mod at blive designet. Derudover testes DE-rammen og viser lovende præstation i et eksperimentelt multi-objektivt designscenario for at opnå 2D-profiler med flade forstærkningsniveauer ved enden af spændet sammen med minimal spektral afvigelse over hele fiberlængden.

Acknowledgement

The three years period of my PhD journey at DTU has come to an end, and I am deeply grateful for the support and contributions of numerous individuals, both on a scientific and personal level. Without their valuable assistance, none of my achievements would have been possible.

First, I would like to thank my main supervisor, Darko Zibar, for giving me this opportunity to be part of the PhD program at DTU, and for providing invaluable guidance throughout the project.

I am also thankful to my co-supervisors Francesco Da Ros and Andrea Carena for their invaluable advice, continuous support, and patience during my research. A special thanks again to Andrea for receiving me at Politecnico Di Torino during my external research stay.

I would like to express my gratitude to my colleagues at DTU, particularly the members of the Machine Learning in Photonic Systems (MLiPS) group. The friendly times we spent together during coffee breaks and lunch breaks, and the funny moments we shared during our fussball games have been memorable.

I would like to extend a special thanks once again to Francesco for being supportive, and intellectual and for teaching me how to make a good coffee in the morning. I am also thankful for the practical knowledge and skills I have acquired from him during our time in the lab.

I would like to express my heartfelt gratitude to my parents, Saleh and Ziba, and my sister Mozhgan, for their unconditional support throughout my life. Their love and belief in my ambitions have been a constant source of strength for me. None of my achievements would have been possible without their belief in me.

At last, I want to express my special thanks to my wife, Arezo. Thank you, Arezo, for coming to my life, and for all your encouragement. Thank you also for being there and for your care at crucial moments over the past two years.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASE	Amplified Spontaneous Emission
BVP	Boundary Value Problem
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
DE	Differential Evolution
DL	Deep Learning
DRA	Distributed Raman Amplifier
EDFA	Erbium-Doped Fiber Amplifier
GD	Gradient Descent
MAE	Maximum Absolute Error
MCF	Multi-Core Fibers
ML	Machine Learning
MMF	Multi-Mode Fibers
MPF	Multiple Parallel Fibers
MSE	Mean Square Error
NFT	Nonlinear Fourier Transform
NLSE	Nonlinear Schrödinger Equation
NN	Neural Network
ODE	Ordinary Differential Equation
OPC	Optical Phase Conjugation
OSNR	Optical Signal-to-Noise Ratio
OTDR	Optical Time-domain Reflectometer
PDF	Probability Distribution Function
QoS	Quality of Service

SDM	Space Division Multiplexing
SGD	Stochastic Gradient Descent
SMF	Single-Mode Fiber
SNR	Signal-to-Noise Ratio
SRS	Stimulation Raman Scattering
SSMF	Standard Single-Mode Fiber
WDM	Wavelength Division Multiplexing

Ph.D. Publications

The following publications have resulted from this Ph.D. project.

Articles in peer-reviewed journals

[J1] **M. Soltani**, F. Da Ros, A. Carena, and D. Zibar, "Experimental validation of machine-learning based spectral-spatial power evolution shaping using Raman amplifiers," *Opt. Express*, vol. 30, no. 25, pp. 45 958–45 969, Dec. 2022. DOI: `10.1364/OE.475873`.

[J2] **M. Soltani**, F. Da Ros, A. Carena, and D. Zibar, "Spectral and spatial power evolution design with machine learning-enabled Raman amplification," *Journal of Lightwave Technology*, vol. 40, no. 12, pp. 3546–3556, 2022. DOI: `10.1109/JLT.2022.3154471`.

[J3] **M. Soltani**, F. Da Ros, A. Carena, and D. Zibar, "Inverse design of a Raman amplifier in frequency and distance domains using convolutional neural networks," *Opt. Lett.*, vol. 46, no. 11, pp. 2650–2653, Jun. 2021. DOI: `10.1364/OL.422884`.

Contributions to peer-reviewed conferences

[C1] **M. Soltani**, F. Da Ros, A. Carena, and D. Zibar, "Experimental validation of spectral-spatial power evolution design using Raman amplifiers," in *European Conference on Optical Communication (ECOC) 2022*, Optica Publishing Group, 2022, We4A.2.

[C2] **M. Soltani**, F. Da Ros, A. Carena, and D. Zibar, "Distance and spectral power profile shaping using machine learning enabled Raman amplifiers," in *2021 IEEE Photonics Society Summer Topicals Meeting Series (SUM)*, 2021, pp. 1–2. DOI: `10.1109/SUM48717.2021.9505741`.

[C3] M. P. Yankov, **M. Soltani**, A. Carena, D. Zibar, and F. Da Ros, “Modeling of optical amplifiers: A comparison of black-, grey- and white-box modeling applied to the raman amplifier scenario,” in *European Conference on Optical Communication (ECOC) 2023 (Submitted)*, 2023.

[C4] F. Da Ros, M. P. Yankov, **M. Soltani**, A. Carena, and D. Zibar, “Modeling optical amplifiers: From inverse design to full system optimization,” in *2023 IEEE Photonics Society Summer Topicals Meeting Series (SUM)- (Submitted)*, 2023.

Contents

List of Figures

List of Tables

1

Introduction

Fiber optic communication systems play a pivotal role in our modern interconnected world as the majority of the Internet data traffic relies on these systems. Their importance lies in their ability to transmit vast amounts of data at high speeds over long distances [1]. These technologies have revolutionized communication by providing greater bandwidth, faster data transfer rates, and improved reliability, thanks to the introduction of new enabling technologies like wavelength division multiplexing (WDM), optical amplifiers, and coherent detection with advanced modulation formats [1]. Fiber optic communication systems are vital for supporting the ever-increasing demands of applications such as internet connectivity, cloud computing, video streaming, telecommunication networks, machine-to-machine communications, and more [2].

Long-term traffic growth is shown to be increasing at a pace of 60%/year, while the fiber capacities only scale at 20% per year [2]. With these enormous growth rates, bandwidth demands will soon reach the upper limit of the current optical fiber infrastructure, resulting in an increasingly critical disparity so-called *capacity crunch* [3]. Concerning this, there has been an extensive effort recently toward improving or upgrading the currently deployed optical networks. Space division multiplexing (SDM) is a widely investigate technology that is basically implemented using multiple parallel fibers (MPF), multi-core fibers (MCF), and/or multi-mode fibers (MMF), which can significantly increase the capacity of the current systems based on SSMF [4]. However, this approach requires deploying new types of fibers with high costs. Therefore, exploring technologies that enhance the capacity of the currently deployed fibers, such as increasing the available bandwidth, offers a more cost-efficient solution [5, 6].

Providing higher bandwidth in current optical communication systems can encounter several challenges. One of the main challenges lies in the selection of the appropriate optical amplifiers to compensate for signal degradation over the frequency and fiber distance. Distributed Raman amplifiers (DRAs) offer significant benefits in addressing these challenges. One major advantage of DRAs is their capability to provide distributed amplification along the fiber span, meaning that the signal is amplified continuously throughout its transmission [7]. By distributing the power along the fiber, DRAs improve the noise figure (NF) compared to Erbium-Doped Fiber Amplifiers (EDFAs), which require periodic amplification stations [8]. Moreover, DRAs have the flexibility to provide any desired gain or power profile in a broadband WDM system by properly adjusting the power and wavelength values of the multiple pump lasers operating in parallel [1]. Nonetheless, in order to achieve the desired gain or the signal power evolution in the context of a more flexible optical transmission, a flexible and reliable adjustment framework for Raman amplifiers will be required as the conventional approaches based on theoretical models and heuristic fine-tuning can be quite time-consuming and computationally expensive.

1.1 Motivation and the outline of the contributions

In recent years, machine learning (ML) models have gained more attention in optical communications due to the increasing desire to enhance the performance, efficiency, and reliability of these systems. By leveraging ML algorithms, it becomes possible to extract valuable insights from massive amounts of data generated by optical communication systems, enabling more accurate predictions and decision-making [9, 10]. DRAs can benefit from using ML models to overcome their existing limitations. In this work, ML models and optimization techniques are utilized to design the DRAs through numerical and experimental setups. The main contribution to this project can be summarized as follows:

- **DRA inverse system design**: DRAs' behavior in an optical communication system follows a theoretical model based on the Stimulated Raman Scattering (SRS) effect, which provides the relation between the input as the applied pumping configuration and the output as the resulting signal power evolution. On the other side, having an inverse model to predict the pumping configuration for a desired signal profile can be practically beneficial. Deriving the mathematical inverse model for SRS is not theoretically feasible. Concerning this, an inverse system based on ML models provides an approximation of DRA schemes, which accurately represent their behavior and can be reused in future optimizations.

- **Online DRA optimization framework**: DRAs can be used to achieve different signal power evolution shapes in the optical communication link. Moreover, different target profiles implicate satisfying various objectives such as spectral gain flatness, spatial power evolution flatness, or power symmetry along the distance. Approximate inverse models which are trained off-line, have certain limitations. Therefore, providing an online ML-based optimization framework to directly control the amplification setup is highly advantageous. The ML optimization framework provides system automation for future flexible systems, saves time demanded to optimize the DRAs, and can outperform human fine-tuning optimization.

1.2 Structure of the thesis

This thesis is organized as follows. Chapter 2 provides the theoretical fundamentals of DRAs schemes. In Chapter 3 , we briefly explain the concepts of the machine learning and optimization algorithms adopted in our investigations. In Chapter 4, we introduce our CNN-based model for inverse DRA design, followed by a numerical evaluation of this model for different amplification scenarios. In Chapter 5, we propose an online optimization framework that mainly consists of differential evolution (DE) as the online optimizer, assisted by the CNN inverse model for designing power profiles of practical interest. Chapter 6 includes the results of the experimental validation of the CNN model and the online optimization framework for designing various target power evolution profiles. Finally, in Chapter 7, we conclude the thesis and discuss the future work perspective.

2

Raman amplification in optical communication systems

As the demand for high-speed and high-capacity signal transmission continues to grow exponentially, the need for efficient and reliable amplification techniques becomes crucial to overcome the limitations of signal degradation over long distances. Distributed Raman amplifiers (DRAs) have emerged as a key solution to enhance signal quality and extend the reach of optical networks [7, 11].

DRAs are a class of fiber-based amplifiers that operate based on the Stimulated Raman Scattering (SRS) effect which involves the inelastic scattering of photons by molecular vibrations in a medium [12]. Due to SRS effect, energy is transferred from a higher-frequency pumping photon to a lower-frequency signal photon, while the excess energy is absorbed as optical phonons by the material, resulting in molecular vibrations [5].

The Erbium-Doped Fiber Amplifiers (EDFAs) have gained popularity in commercial communication systems due to their power efficiency and straightforward setup configuration [8, 13]. However, broadband amplification provided by DRAs, has made them an attractive alternative solution for wideband Wavelength Division Multiplexing (WDM) schemes [1]. Moreover, EDFAs are typically deployed at fixed intervals along the fiber, while DRAs offer a unique advantage by providing distributed amplification [1]. This means that the amplification process occurs continuously throughout the fiber, which can result in a uniform distribution of the signal power along the span. A uniform signal power distribution in the fiber, representing in a loss-less medium, minimizes the accumulated amplified spontaneous emission (ASE) noise at the receiver side and reduces the signal-to-noise ratio (SNR) degradation [14]. In addition, multiple Raman amplifiers operating at different wavelengths can be employed in parallel [1]. This attribute, along with the broad spectral gain and distributed amplification, enables Raman amplifiers to design any desired gain shape or signal power profile over a large bandwidth in WDM systems [7].

Considering the aforementioned features, DRAs improve the performance of optical communication systems by providing low noise figure (NF), high gain, and large bandwidth, which makes them ideal for increasing the capacity in long-haul and ultra-long-haul optical transmission [7]. By employing DRAs, the signal reach can be extended without the need for additional regenerators or repeaters, thereby reducing the complexity of the optical communication systems [15].

The SRS effect provides a particularly large gain spectrum, extending over 20–30 THz, with the peak at around 13 THz of frequency shift [1]. The Raman gain profile can change for different types of fibers with specific dependency on fiber effective area A_{eff} and fiber Raman gain coefficient g_R [16]. Particularly, g_R is measured in [m/W] and scales linearly with pump frequency [17].

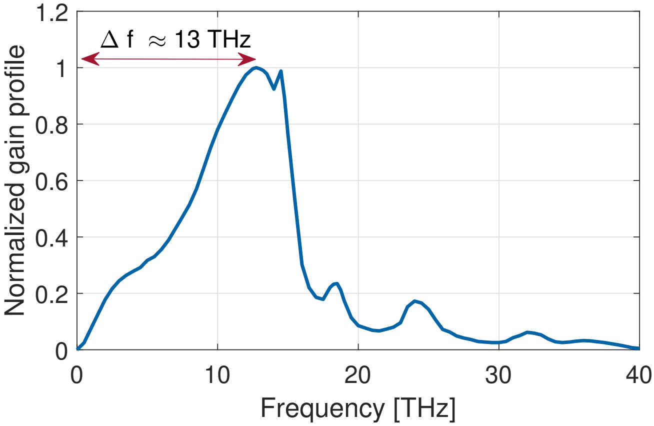
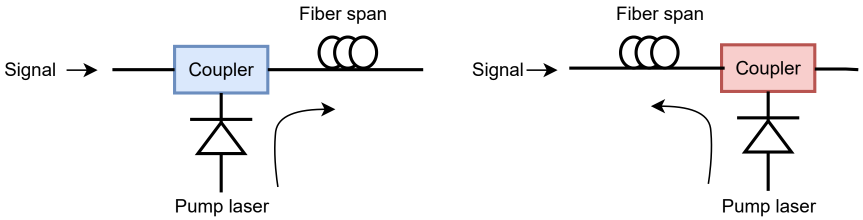


Fig. 2.1: Normalized Raman gain profile over the frequency for a fused silica fiber.

Fig. 2.1 illustrates the normalized Raman gain profile of fused silica fiber for co-polarized pump and signal. According to this plot, the maximum gain is provided by the pump when there is almost a 13 Thz space between the pump and the signal frequency, which is referred to as

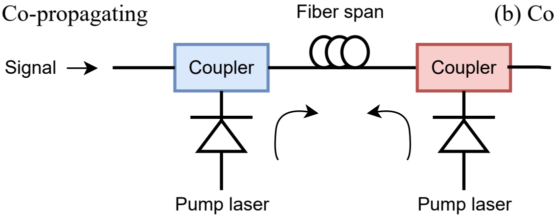
Stokes shift [1]. The Raman gain efficiency profile C_R in units [1/W/m] can be obtained by applying a proper scaling factor to the gain coefficient g_R [18] as follows:

$$C_R = g_R / A_{eff} \tag{2.1}$$



(a) Co-propagating

(b) Counter-propagating



(c) bidirectional propagating

Fig. 2.2: Different schematics for amplification based on pump direction.

Depending on the direction of the pump signal relative to the direction of the main signal in the fiber, three different amplification scenarios can be considered: 1) Co-propagating (forward), where the pump signals and the main signal propagate in the same direction, 2) Counter-propagating (backward), where pump and signal propagate in the opposite direction, and 3) Bidirectional propagating, where pump signals propagate in both directions. These schematics are shown in Fig. 2.2 (a)-(c), respectively.

In the presence of Raman amplification, the relation between the signal and pump power evolution over the fiber distance is described by a well-known set of non-linear ordinary differential equations [1]:

$$\frac{dP_{s,i}}{dz} = -\alpha_s P_{s,i} + C_R(f_{s,i}, f_{p,j})[P^{+}_{p,j} + P^{-}_{p,j}]P_{s,i} \tag{2.2}$$

$$\pm\frac{dP^{\pm}_{p,j}}{dz} = -\alpha_p P^{\pm}_{p,j} - (\frac{f_{p,j}}{f_{s,i}})C_R(f_{s,i}, f_{p,j})P_{s,i}P^{\pm}_{p,j} \tag{2.3}$$

where P_s is the signal power, $P^{+}_{p,j}$ and $P^{-}_{p,j}$ are the co- and counter-propagating pump powers, α_s and α_p represent the signal and the pumps attenuation coefficients measured in m^{-1}, and $f_{s,i}$ and $f_{p,j}$ are the signal and the pump frequencies, respectively. In these equations, i identifies the signal channel index for $i = \{1, ..., n_{ch}\}$, and j identifies the pump index for $j = \{1, ..., n_p\}$, with n_p number of pumps.

In Eq. 2.3, the $+$ and $-$ signs in $\pm$, and also used as the superscripts for $P^{+}_{p,j}$ and $P^{-}_{p,j}$, correspond to co-propagating and counter-propagating pumps, respectively. The first contribution in the power evolution in both Eq. 2.2 and Eq. 2.3 represents the power degradation due to the fiber loss. Instead, the second term in Eq. 2.2 determines the contribution of signal gain due to SRS, while the second term in Eq. 2.3 represents the pump depletion [1].

Considering the case of a single pump and signal with negligible depletion of the pump where the second term in Eq. 2.3 is omitted, the analytical formulation of the pump can be obtained as $P_p(z) = P_p(0)exp[-\alpha_p z]$ for a co-propagating pump, or as $P_p(z) = P_p(L)exp[-\alpha_p(L - z)]$ for a counter-propagating pump. For a single signal evolution and by neglecting the pump depletion, Eq. 2.2 can be solved as follows [19]:

$$P_s(z) = P_s(0)(-\alpha_s z + C_R P_p(0)(\frac{exp(-\alpha_p L)[exp(\alpha_p z) - 1]}{\alpha_p})) \tag{2.4}$$

$$P_s(z) = P_s(0)(-\alpha_s z + C_R P_p(0)(\frac{1 - exp(-\alpha_p z)}{\alpha_p})) \tag{2.5}$$

for counter- and co-propagating pumps, respectively.

According to Eqs. 2.4 and 2.5, the amount of pump power $P_p(0)$ required to fully compensate for the signal loss depends on the Raman gain efficiency C_R and the fiber attenuation at pump frequency α_p. To investigate this relation (assuming a fixed C_R profile), the signal power evolution for a 100 km fiber is compensated using a co- or a counter-propagating Raman pump. The resulting signal power profiles are shown in Fig.2.3 at different pump attenuation levels. Regarding the co-propagating pump cases, since the pump power is stronger at the

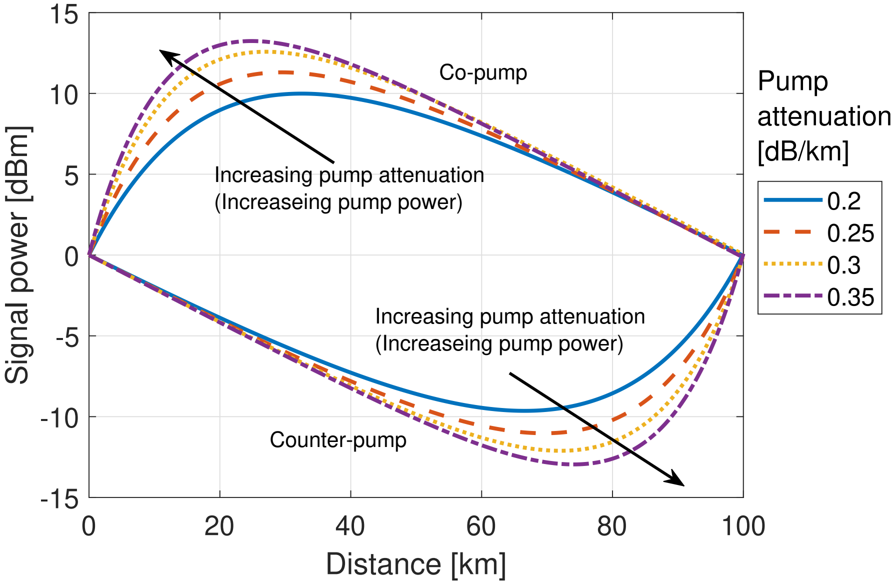


Fig. 2.3: Signal power evolution in a 100 km fiber, compensated using Raman amplification. The top curves show the co-pumping and the bottom ones show the counter-propagation results.

beginning of the fiber, the signal experiences more gain, and due to pump power reduction due to attenuation over higher distances, the signal power gradually gets reduced. Moreover, for higher fiber attenuation levels, the pump power reduction will be more severe and therefore, there will be more pump power needed to fully compensate for the signal loss. Considering the counter-propagating pump cases, the pump power gets attenuated until it gets to the beginning of the fiber, and therefore, the signal experiences less gain and linearly undergoes towards the end of the fiber. While the pump power is stronger at the end of the fiber and the signal experiences a stronger gain when it gets to the end of the fiber. As the attenuation of fiber at pump frequency increases, the gain becomes concentrated toward the fiber end, and eventually, there will be more pump power needed to fully compensate for the signal loss.

Focusing on the provided overall gain, the signal power at the end of the fiber of length L is formulated as the following [1]:

$$P_s(L) = P_s(0)exp[C_R P_0 L_{eff} - \alpha_S L] \tag{2.6}$$

where $P_0 = P_p(0)$ for co-propagating, or $P_0 = P_p(L)$ for counter-propagating pump, and L_{eff} is defined as:

$$L_{eff} = \frac{[1 - exp[-\alpha_p L]]}{\alpha_p} \tag{2.7}$$

According to these formulations, the Raman on-off gain in the small-signal region can be defined as the increase in signal power at the end of the fiber when the pump is turned on:

$$G_{on-off}[dB] = \frac{P_s(L)^{pump-on}}{P_s(L)^{pump-off}} = exp(C_R P_0 L_{eff}) \tag{2.8}$$

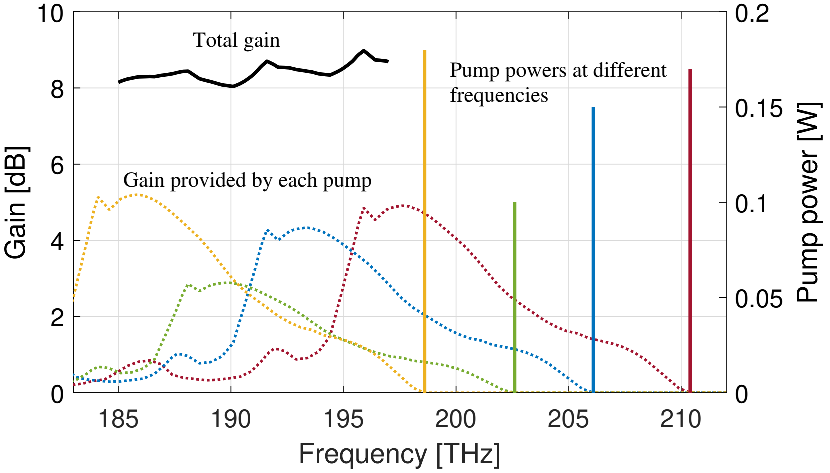


Fig. 2.4: Separate gain and the total gain provided by the Raman pumps operating at different powers and frequencies.

Gain flattening in wavelength-division WDM systems is a crucial technique employed to ensure uniform amplification of signals across multiple channels [20]. As pointed out earlier in this chapter, there is a potential to employ multiple Raman pumps at the same time and extend their resulting gain over a wide range of frequencies.

Fig.2.4 represents the gain profiles at the end of the fiber resulting from the utilization of pumps with different powers and operating at different frequencies for a 100-km long fiber. The total gain is obtained through the superposition of gains provided by different pumps across the frequency range, and it is depicted by the black solid curve. By appropriate adjustment of the number of pump lasers, their frequencies, and their powers, it is potentially feasible to achieve uniform gain profiles across any desired bandwidth [21]. This is because the Raman gain spectrum covers a broad range of frequencies, spanning 20-30 THz, with a relatively even distribution, allowing for significant bandwidth coverage. In this particular case, a relatively flat gain at 8.5 dB with less than 0.5 dB deviation is achieved between 185 THz and 197 THz by tuning the power values of four Raman pumps.

It is worth noting that the total gain profile in Fig.2.4 is depicted just for providing a visual intuition without considering the full SRS effect including signal-signal and pump-pump interactions. However, when designing broadband amplification schemes, it is crucial to consider these interactions, which will be discussed in the following section.

2.1 Full Raman amplification model

In order to design broadband Raman amplification schemes, besides considering the pump-signal interaction, it is also crucial to take into account other existing interactions such as the signal-signal and pump-pump. In a general amplification scenario with N_s signals and N_p pumps, Eqs. 2.2 and 2.3 can be modified as follows [19]:

$$\frac{1}{P_{si}}\frac{dP_{si}}{dz} = -\alpha_{si} + \sum_{j=1}^{N_p} C_{ps,ji}P_{pj}^{\pm} + \sum_{j=1}^{i-1} C_{ss,ji}P_{sj} - \sum_{j=i+1}^{N_s} \left(\frac{\upsilon_{si}}{\upsilon_{sj}}\right) C_{ss,ij}P_{sj} \tag{2.9}$$

$$\pm\frac{1}{P_{pi}^{\pm}}\frac{dP_{pi}}{dz} = -\alpha_{pi} - \sum_{j=1}^{N_s}\left(\frac{\upsilon_{pi}}{\upsilon_{sj}}\right)C_{ps,ij}P_{sj} + \sum_{j=1}^{i-1} C_{pp,ij}P_{pj}^{\pm} - \sum_{j=i+1}^{N_p}\left(\frac{\upsilon_{pi}}{\upsilon_{pj}}\right)C_{pp,ji}P_{pj}^{\pm} \tag{2.10}$$

where P_{si} and P_{pi} are the instantaneous power of the ith signal at frequency υ_s and ith pump at frequency υ_p, respectively. Moreover, α_s and α_p are the fiber attenuation values for the signal and the pump. C_{ps}, C_{ss}, C_{pp} are the Raman gain efficiency between the pumps and signals, signals and signals, and pumps and pumps, respectively. The arrangement of pumps and signals assumes that they are ordered in increasing wavelength (decreasing in frequency). In Eq. 2.9, the first term corresponds to the linear loss of the fiber, while the second term represents the power obtained from the pump wavelengths. The third and fourth terms account for the power gained from shorter wavelength signals and the power lost to longer wavelength signals, respectively. Similarly, in Eq. 2.10, the first term corresponds to the linear loss of the fiber, and the second term denotes the power lost to the signals, leading to amplifier saturation. The third and fourth terms account for the Raman interactions between the pump wavelengths, with the third term representing the power gained from shorter wavelengths and the fourth term indicating the power lost to longer wavelength pumps.

It is worth noting that in these equations, the ASE noise term and the Rayleigh Back-scattering effect are omitted. These effects can be investigated in a separate set of ODEs.

2.2 Multiple pump interactions

Considering the Eqs. 2.9 and 2.10, the pump-pump, signal-signal and signal-pump interactions can produce tilt in the signal power evolution and also in the gain profile. Moreover, the pumps at shorter wavelengths transfer energy to the pumps at longer wavelengths. In order to have a flat gain spectrum, most of the gain needs to be provided by the pump with the longest wavelength [19]. Therefore, the longest pump will need to be launched with high power. Nonetheless, by having a few shorter wavelength pumps, due to pump interactions, the longer wavelength pumps obtain power from multiple short wavelength pumps, and therefore, launching it with high power will not be required.

The power evolution over the distance is shown in Fig. 2.5 for a set of four signals with wavelengths 1530, 1550, 1580, and 1600 nm, and also for four counter-propagating pumps operating at wavelengths 1423, 1443, 1464 and 1495 nm, which are taking place almost at the peak gain wavelengths of the pumps, respectively. The shortest wavelength pump 1423 nm is launched with slightly higher power compared to the other pumps, while it is attenuated more dramatically by the other pumps and has the least power after almost 5 km propagation in the fiber span. On the other side, the pump operating at 1495 obtains gain in the first 10 km and loses its power slower, using the power provided by the pumps with shorter wavelengths. Regarding the signals, they initially

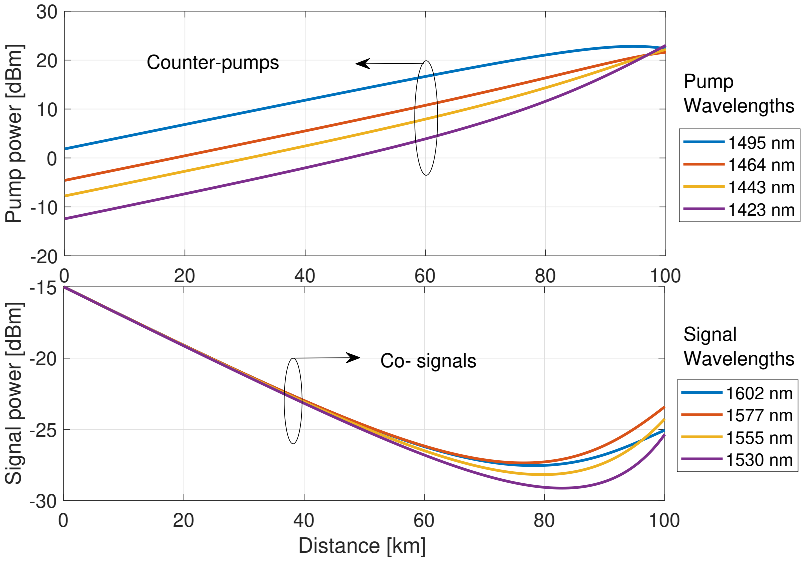


Fig. 2.5: The power evolution over the distance for a set of four signals with wavelengths 1530, 1550, 1580, and 1600 nm, and also for four counter-propagating pumps operating at wavelengths 1423, 1443, 1464, and 1495 nm.

experience linear loss in almost the first 60 km and then start interacting and eventually getting gain from the pumps. The signal at 1602 nm experiences most of its gain roughly at this point as it's getting mostly amplified by the pump at 1495 nm with the highest power level. On the other side, the signal at 1530 receives its highest gain as it gets closer to the end of the fiber.

2.3 Higher Order Pumping

In the Raman amplifier configurations discussed so far, the Raman pumps are one stokes shift away from the signal. Regardless of the relative direction of the pumps and signals, these configurations are mainly called first-order pumping schemes. An alternative configuration that can be investigated further is to use a second-order pump which is utilized to amplify the first-order pump. As a practical example, in [22] a signal at 1550 nm is amplified by a first-order pump operating at 1455 nm which is also combined with a second-order pump at 1366 nm. This combination has more advantages over an amplification setup that utilizes only first-order pumps especially when used in a bidirectional scheme. One main advantage is that the second-order pumps help with uniformly distribut-

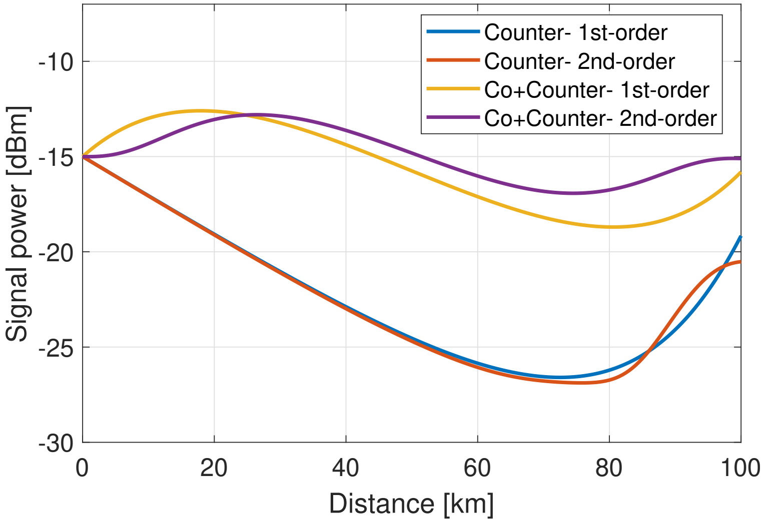


Fig. 2.6: Signal power evolution over the fiber distance for the different first-order and second-order Raman amplification schemes.

ing the signal power along the fiber span [23]. Having a transparent signal power distribution in the fiber reduces the accumulated ASE noise at the receiver side, which is one of the main goals in systems where the ASE is the dominant noise source [24]. Moreover, a co-propagating second-order Raman pump has fewer induced impairments affecting the signal compared to the first-order co-propagating pump.

The effect of using first-order and second-order pumps on the evolution of a single-channel signal operating at 1550 nm over a 100km fiber is shown in Fig.2.6. In this figure, both counter- or bidirectional propagating schemes are used as an example. The selected pump wavelengths are chosen as reported in [23]. It is evident that in the case of using bidirectional second-order pumping, the signal will experience less excursion compared to a first-order amplification. Moreover, according to [23], the second-order amplification provides a considerable improvement in NF over a first-order pumping scheme.

2.4 Summary

In this chapter, we briefly discussed the advantages and the theory behind the application of using DRAs to compensate for signal loss in transmission systems. We pointed out that DRAs are a key technology that are vital to improving the performance of fiber optic communication systems by providing several interesting characteristics over the EDFAs. In the next chapter, a brief overview of Machine Learning (ML) models, especially Deep Learning algorithms, and the concept of gradient-free optimization techniques will be discussed. The main focus of the next chapter will be to explain the basics of the ML models used in this project for designing DRAs.

3

Machine Learning and Optimization algorithms

3.1 Machine Learning and Artificial Neural Networks

Machine learning (ML) and artificial neural networks (ANN) have emerged as powerful tools in the field of computer science, revolutionizing various industries and shaping the way we interact with technology. In an era characterized by the abundance of data and the growing complexity of problems, ML provides a framework for enabling computers to learn and make predictions or decisions without being explicitly programmed.

At its core, ML is a branch of artificial intelligence (AI) that focuses on the development of algorithms and models capable of learning from and making predictions or decisions based on data. It encompasses a diverse set of techniques and methodologies that enable computers to automatically identify patterns, extract meaningful insights, and make accurate predictions or decisions [25].

One of the fundamental building blocks of ML is the ANN. Inspired by the structure and functioning of the human brain, an ANN is a computational model composed of interconnected nodes, called neurons, which work collaboratively to process and analyze data. Each neuron receives input signals, applies certain operations or transformations to these inputs, and produces an output signal that is passed on to other neurons [26].

ANNs are particularly adept at learning complex patterns and relationships in data, making them well-suited for a wide range of tasks such as image and speech recognition, natural language processing, and predictive analytics. The learning process in neural networks involves adjusting the strength or weights of connections between neurons based on the available data, enabling the network to capture and represent the underlying patterns and relationships in the data.

Over the years, advancements in computing power, the availability of large-scale datasets, and the development of more sophisticated algorithms have propelled machine learning and artificial neural networks to new heights. Deep learning, a sub-field of ML that focuses on neural networks with multiple layers, has emerged as a dominant approach, achieving remarkable breakthroughs in areas such as computer vision, speech synthesis, and autonomous systems [27].

3.2 Convolutional Neural Networks

One prominent type of neural network is the Convolutional Neural Network (CNN). CNNs are specifically designed to excel at processing and analyzing visual data, such as images and videos. They leverage the concept of convolution, a mathematical operation that involves sliding a filter or kernel across an input image to extract meaningful features [28].

The distinctive characteristic of CNNs is their ability to automatically learn hierarchical representations of visual features. The initial layers of a CNN learn simple features such as edges and corners, while subsequent layers learn more complex features like shapes and textures. This hierarchical learning enables CNNs to capture the intricate details and patterns within visual data, leading to superior performance in tasks such as object recognition, image classification, and image segmentation [29].

CNNs have significantly advanced the field of computer vision, achieving remarkable accuracy and efficiency in a variety of applications. They have been instrumental in autonomous vehicles, facial recognition systems, medical image analysis, and many other domains where visual data plays a crucial role [30]. The success of CNNs can be attributed to their ability to exploit spatial relationships and local correlations within images, effectively capturing and representing the relevant information for accurate decision-making.

Convolutional Neural Networks (CNNs) are composed of different layers that perform specific operations, enabling the network to effectively extract and process features from input data. Each layer plays a crucial role in the overall architecture of a CNN. Here, we'll discuss some of the key layers commonly found in CNNs [29]:

Convolutional Layer: The convolutional layer is the core component of a CNN. It performs the convolution operation, where a set of learnable filters (also called kernels) are applied to the input data. Each filter extracts different features from the input by sliding across the spatial dimensions and computing element-wise multiplications and summations. This layer helps capture local patterns and spatial dependencies in the input data.

Activation Layer: The activation layer introduces non-linearities into the network. It applies a non-linear activation function element-wise to the output of the previous layer. Common activation functions include ReLU (Rectified Linear Unit), sigmoid, and hyperbolic tangent. The activation layer enables the network to model complex relationships and introduces non-linearity into the feature maps.

Pooling Layer: Pooling layers reduce the spatial dimensions of the feature maps while retaining important information. Max pooling is a widely used pooling technique that selects the maximum value within a defined pool size, effectively downsampling the feature maps. Average pooling takes the average of values within the pool size. Pooling layers help reduce computational complexity, extract dominant features, and provide translation invariance.

Batch Normalization Layer: The batch normalization layer normalizes the activations of a previous layer, bringing them to zero mean and unit variance. It helps with stabilizing and accelerating training by reducing internal covariate shift, ensuring that each layer receives inputs in a more consistent distribution. This layer improves the generalization of the network and reduces the sensitivity to the initialization and learning rate.

Dropout Layer: The dropout layer is a regularization technique used during training to prevent overfitting. It randomly sets a fraction of the input units to zero at each training iteration. By forcing the network to rely on different subsets of units, dropout reduces the co-adaptation of neurons and promotes better generalization. Dropout is typically applied after fully connected layers.

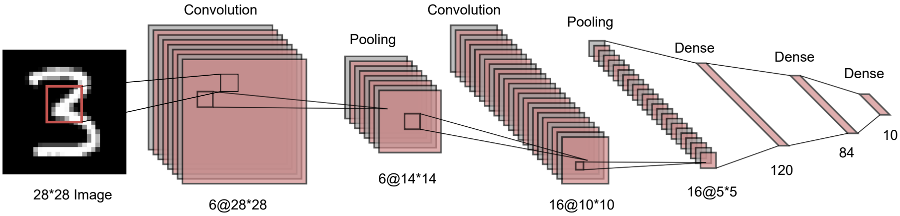


Fig. 3.1: Data flow in LeNet [31]. The input is a handwritten digit, and the output is a probability over 10 possible outcomes.

Fully Connected Layer: Fully connected layers, also known as dense layers, are traditional artificial neural network layers. They connect every neuron in the current layer to every neuron in the subsequent layer, allowing for complex relationships and high-level abstractions. Fully connected layers are commonly found at the end of CNNs to map the extracted features to specific output classes or regression values.

These layers can be stacked together to form the overall architecture of a CNN. The arrangement and number of layers depend on the specific task and complexity of the problem. The deepening of CNNs through the stacking of these layers enables the network to learn hierarchical representations of the input data, capturing both low-level and high-level features, and achieving remarkable performance in various computer vision tasks. For example, Fig.3.1 represents a CNN architecture called LeNet [31] which was designed to classify handwritten digits from the MNIST [32] dataset, consisting of 28x28 grayscale images of handwritten digits ranging from 0 to 9. This model consists of two parts: (a) a convolutional encoder consisting of two convolutional layers; and (b) a dense block consisting of three fully connected layers. More specifically, Convolutional layers allow LeNet to learn and extract spatial features from input images. These layers employ a set of filters or kernels that scan the input image, capturing important patterns and features at different spatial scales. The extracted features become increasingly complex as they pass through multiple convolutional layers.

After the convolutional layers, LeNet utilizes fully connected layers to map the extracted spatial features to specific output classes. These layers establish connections between all the neurons in the previous layer and the neurons in the subsequent layer, enabling the network to learn complex relationships and make accurate class predictions.

By combining the power of convolutional layers to extract spatial features and fully connected layers to map those features to classes, LeNet demonstrates the ability to recognize and classify handwritten digits in the MNIST dataset with impressive accuracy. LeNet's architecture and its success have had a profound impact on the development of CNNs and their widespread application in various computer vision tasks. It showcased the importance of spatial feature extraction and the effectiveness of deep learning approaches in solving complex problems related to data with image-based topology.

3.3 Gradient-free optimization

Gradient-based optimization methods, such as stochastic gradient descent (SGD) and its variants, are widely used in machine learning to update model parameters based on the gradients of the loss function with respect to those parameters [33]. However, these methods rely on the availability of differentiable functions and gradients, which may not be feasible in certain scenarios where the objective function is non-differentiable, noisy, or lacks explicit mathematical formulation. In such cases, gradient-free optimization techniques provide an alternative approach to searching for optimal solutions [34].

Gradient-free optimization methods aim to find the global or local optima of a function without using its gradients. Instead, they explore the function's landscape by sampling and evaluating different points iteratively. These techniques leverage various strategies, including random search, evolutionary algorithms, surrogate models, and heuristics, to guide the search toward better solutions. One such popular gradient-free optimization algorithm is Differential Evolution (DE) [35].

Differential Evolution (DE) is a population-based optimization algorithm that mimics the process of natural evolution. It was proposed by Rainer Storn and Kenneth Price in the late 1990s [36]. DE operates by maintaining a population of candidate solutions, called individuals, and iteratively evolves them towards better solutions by combining their information.

The basic steps of the DE algorithm are as follows [35]:

- *Initialization*: Generate an initial population of individuals, where each individual represents a potential solution to the optimization problem.

- *Mutation*: For each individual in the population, create a mutant vector by combining information from multiple individuals. This is achieved by adding a scaled difference vector to the individual's position.
- *Recombination*: Create a trial vector by recombining the mutant vector with the original individual, guided by a crossover probability. The trial vector represents a potential new solution.
- *Selection*: Compare the trial vector with the original individual. If the trial vector outperforms the original individual, replace it in the population; otherwise, keep the original individual.
- *Termination*: Repeat the mutation, recombination, and selection steps for a predefined number of iterations or until a termination criterion is met (e.g., reaching a maximum number of evaluations or achieving satisfactory performance).

DE utilizes the population dynamics and the stochastic search process to explore and exploit the solution space efficiently. It has been successfully applied to a wide range of optimization problems, including parameter tuning in machine learning algorithms, feature selection, function optimization, and design optimization.

The advantages of DE and other gradient-free optimization techniques include their ability to handle noisy or non-differentiable objective functions, avoid getting trapped in local optima, and provide robust and global search capabilities.

3.4 Summary

In this chapter, a brief introduction to the concept of ML algorithms was provided followed by an explanation of the CNN models used widely for pattern recognition in data with image topology. Afterward, a brief summary of the purpose of using gradient-free optimization techniques is presented. This was followed by providing a brief diagram of DE, as one of the most popular gradient-free techniques used for optimization.

4

Inverse design of Raman amplifiers using Convolutional Neural Networks

4.1 Introduction

The signal power evolution in an optical communication system is directly affected by the behavior of the optical amplifiers. Therefore, modeling and optimizing the amplifier configuration is highly desirable. Distributed Raman Amplifiers (DRAs) have been extensively researched recently as they provide several interesting advantages over the Erbium-Doped Fiber Amplifiers (EDFAs) [1, 11]. As pointed out in Chapter 2, due to the multi-pumping scheme, DRAs are a practical solution to amplify a broad range of wavelengths beyond the C-band which eventually results in increasing the available transmission capacity[37].

One of the main challenges in designing DRAs is to set up the appropriate configuration, i.e. to select the pump powers and their frequency values to achieve desired signal power evolution in the fiber span. Most of the research conducted in this area has approached designing desired gain spectra at the receiver side using machine learning (ML) and optimization techniques [37–41]. In [38] a neural network (NN) is combined with a genetic algorithm to design flat gain profiles. This algorithm is computationally expensive as it requires performing an iterative process with multiple integrations of the propagation equations, which must be performed for every target gain. However, [39] proposed an NN model to solve this issue by learning the inverse relationship between the pump powers and wavelength values and the corresponding gain profiles.

This approach was mainly conducted by using a synthetic dataset of gain profiles generated with random pump parameters. The trained NN model was then used to derive the pump parameter values for a desired gain profile. Moreover, the authors of [39] added two additional NN architectures so-called *fine-tuning* and *model-averaging* to their previous NN model to refine the predicted pump parameter values to increase the accuracy. The authors of [39] produced several other contributions involving the use of NNs for Raman amplification, including experimental validation of the proposed NN models, noise profile prediction, and optimization of a hybrid Raman/EDFA scheme [21, 37, 42]. Regarding the application of NN for gain spectral shaping, a different approach in [40] is taken, where a differentiable Raman amplification model is presented and used in the training procedure of an NN model to predict the pump power values to design groups of flat or tilted spectral gains.

Alternative to gain spectral shaping, utilizing DRAs in designing the signal power evolution jointly in frequency and along the fiber distance is also a beneficial approach to satisfy some of the long-time goals in optical communication systems. It has been demonstrated that controlling the signal power evolution jointly in frequency and fiber distance can pave the way to optimize the signal-to-noise ratio (SNR) and mitigate nonlinearity impairments [43]. More particularly, it is shown that an even distribution of the power along the fiber distance, which represents a lossless link, minimizes the accumulated amplified spontaneous emission (ASE) noise at the receiver [14, 15, 23]. It is proven that in a system where ASE is the dominant source of noise, the flat signal power in distance can effectively improve the optical signal-to-noise ratio (OSNR) for a fixed nonlinear weight, i.e. the path-averaged signal power [44]. Moreover, a medium with effective zero attenuation is considered a prerequisite for some applications such as transmission based on Nonlinear Fourier Transform (NFT) which relies on analytically solving a lossless Nonlinear Schrödinger equation (NLSE) [45, 46].

A second practically interesting signal power evolution is a symmetric profile with respect to the middle point in the distance [47–49]. This profile optimizes the performance of optical phase conjugation (OPC) [13, 50] which is mainly used for mitigating the nonlinear impairments in the system [49, 51].

A significant research effort has been conducted both numerically and experimentally to design practically interesting two-dimensional (2D) signal power profiles such as 2D flat and 2D symmetric power profiles. These profiles are mostly addressed by heuristically tuning the pump power values and simplifying the optimization process based on the physics of the setup under test [15, 24, 52, 53]. In these papers, the authors present amplification setups including second-order pumps combined with the FBG reflectors to turn the span into an ultra-long laser. This is mainly conducted to achieve a lossless transmission in frequency and distance. However, these systems relied on either optimizing the pumps heuristically [15, 24], optimizing forward pumps and then adjusting the backward pumps to match the desired gain for the central channel [52], or assuming equal power for backward and forward pumping, thus simplifying the optimization problem [53].

Finding the optimal pump power through a heuristic search requires precisely simplifying the optimization problem. This approach can be challenging, especially when the number of parameters to optimize is high. Moreover, this optimization process will be difficult and time-consuming to solve if there are several different target profiles, with different objectives and constraints, aimed to be designed using a specific amplifier setup under test. Furthermore, the strategy in heuristic search will need to be re-defined or adapted in case the system configuration including the number of pumps, their available power, and frequency, or the fiber parameters change. Considering these drawbacks mentioned regarding heuristic search, a general ML or optimization-based framework can be beneficial in finding the optimal values, as they have been very successful in a wide variety of applications.

In this Chapter, we propose an inverse model based on a supervised Deep Learning (DL) architecture to design Raman amplifiers in spectral and spatial (fiber distance) domains. The main purpose of the DL model is to find the mapping between the 2D spectral-spatial signal power profiles and their corresponding Raman pump parameters, i.e. pump power and wavelength values. First, the architecture of the proposed DL model which is based on a Convolutional Neural Network (CNN) architecture, will be explained in detail. Afterward, the amplifier setup for evaluating the CNN's performance will be presented followed by the synthetic dataset generation for model training. Finally, the numerical results will be demonstrated for both counter- and bidirectional propagating DRA schemes.

4.2 CNN-based model architecture for inverse DRA design

Generally, in a system with an input-output function denoted as $Y = f(X)$, finding models for both direct and inverse mappings can be of practical interest. This approach can be more crucial when it is challenging to derive or solve the mathematical or numerical formulation of the direct or inverse models. The direct mapping model aims at finding the output of the system given the input value. The inverse mapping instead aims to find the input in order to receive a desired output.

Considering designing the DRAs, first, we call the direct function $f(.)$ as the set of nonlinear differential equations presented for DRAs. These equations were provided with details in Chapter 2 through Eqs. 2.9 and 2.10. The direct model follows the Stimulated Raman Scattering (SRS) effect which is a physics-based phenomenon that maps the amplifier pumping characteristics such as the pump power and wavelength values to the spectral gain or spectral-spatial signal power evolution. To solve this forward mapping, both analytical and black-box modeling investigations can be performed. In the analytical approach, the direct model is treated as a boundary value problem with ordinary differential equations (ODEs), which can be solved with iterative models such as Runge-Kutta [54, 55]. When it comes to black-box modeling, a commonly employed strategy involves the utilization of ML models, as explained in the previous section. The inverse model instead aims at finding the amplifier configuration to achieve a desired spectral or joint spectral-spatial response. Nonetheless, due to the lack of an analytical closed-form formulation for SRS equations, a potentially reliable approach is to take a black-box modeling approach. In this chapter, our main focus is to provide an ML-based inverse system model to map the 2D power evolution profile defined in frequency-distance to the corresponding Raman pump power and wavelength values.

For the inverse system design scenario, the forward mapping can be described as $\mathbf{P}_s(f, z) = f([\mathbf{p}_p; \boldsymbol{\lambda}_p])$ where $\mathbf{P}_s(f, z) = [p_{ij}]_{N_{ch} \times N_z}$ is the 2D signal grid defined in spectral (f) and spatial (z) domains. More specifically, p_{ij} is signal power at i-th spectral channel and j-th distance index in a WDM system with N_{ch} number of channels and N_z distance points. Furthermore, $\mathbf{p}_p = [P_1, \cdots, P_{N_p}]^T$ is the pump power vector

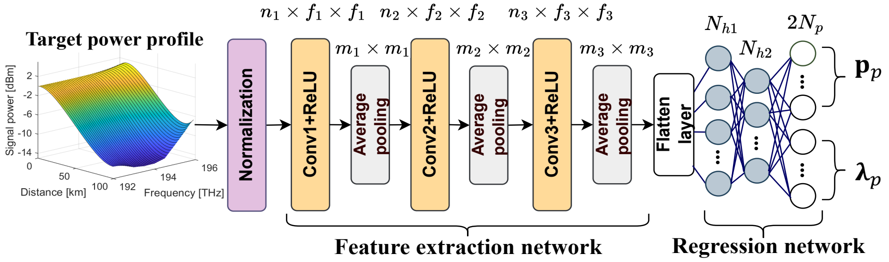


Fig. 4.1: Diagram of the CNN architecture for the inverse DRA design. The model input is the 2D target power profile and its output is the set of Raman pump power and wavelength values.

with and $\boldsymbol{\lambda}_p = [\lambda_1, \cdots, \lambda_{N_p}]^T$ is the pump wavelength vector, where T indicates the transpose operator.

The proposed inverse mapping learns the inverse function $f^{-1}(.)$ such that it can simultaneously predict the pump powers and wavelength values $[\mathbf{p}_p; \boldsymbol{\lambda}_p]$ for a given target power evolution profile $\mathbf{P}_s^t(f, z)$. The suggested inverse model should be able to perform a mapping between the 2D signal power profile defined in a $\mathbb{R}^{N_{ch} \times N_z}$ space and the Raman pump settings defined in one dimension $\mathbb{R}^{2N_p \times 1}$ space. One potential way to perform this mapping could be to apply NN-based frameworks proposed by [39, 40]. In this approach, the input signal power profile should be flattened to become a one-dimensional (1D) vector of length $N_{ch} \times N_z$, and the model will perform a 1D-to-1D mapping. This approach is not quite practical for this design case scenario as the proposed NN will have a large architecture with too many parameters to be trained. For example, considering a system with $N_{ch} = 50$ channels and a 100 km span length with a distance resolution of 1 km, $N_z = 100$, the number of the nodes of the input layer will be $N_{ch} \times N_z = 5000$. If the amplifier setup has four operating pumps $N_p = 4$, therefore the number of the NN output nodes will be $2N_p = 8$ (to predict both pump power and wavelength values). Hence, the mapping between a long input with 5000 nodes and the pumping configuration with 8 nodes requires a network with a high number of trainable parameters. Training this network will not only require a significant amount of time and training data but also expose it to potential issues such as overfitting and being trapped in local minima.

Another drawback of converting the 2D power evolution grid into a 1D vector is that it will remove the existing inherent spatial correlation in the input data. Based on a physical perspective, each point of the signal power in the frequency-distance grid has a high correlation with its neighboring point. Therefore, it is highly probable that they convey very similar information. However, fully-connected NNs are not capable of capturing spatial information of grid-shape data and reducing their redundancies. Concerning these two main issues with fully-connected NNs, we found that CNNs can be suitable for the DRA inverse system design since they are sort of DL models designed to process data with grid-like topologies, such as images. In regard to this, the CNNs can successfully capture the spectral and spatial dependencies in data with 2D form through the application of relevant filters [56] and weight sharing.

The diagram of the proposed CNN-based model for inverse DRA design is illustrated in Fig.4.1. In this diagram, the main goal is to predict the Raman pump powers and wavelength values for a given target 2D signal power profile $\mathbf{P}_s^t(f, z)$. The proposed diagram is made up of two separate networks trained end-to-end. The first network consists of three CNN layers each one followed by an average pooling layer. This network will be referred to as *feature extraction* network $R(.;\theta_R)$, defined as:

$$R(.;\theta_R) : \mathbb{R}^{N_{ch} \times N_z} \rightarrow \mathbb{R}^{q \times r \times n_3}, q \times r \times n_3 < N_{ch} \times N_z \tag{4.1}$$

with θ_R defined as the training parameters. The definition and the values of parameters r, q, and n_3 will be discussed below, after clarifying the data flow process in the network.

First, as a pre-processing step and before using the feature extraction network, the 2D power profile is normalized according to the minimum and maximum possible power values. Afterward, the normalized profile passes through the three CNN layers with n_1, n_2 and n_3 number of filters each with the size $f_1 \times f_1$, $f_2 \times f_2$ and $f_3 \times f_3$, respectively. In each CNN layer, it is necessary to choose a nonlinear activation function. There are several commonly used activation functions in the literature for different DL architectures [27]. Among these options, we select a rectified linear unit ($ReLU(x) = max(0, x)$) for each CNN layer due to its simplicity in gradient calculation, which can contribute to a faster training process [27].

As illustrated in Fig. 4.1, each CNN layer is followed by an average spatial pooling with the window size of $m_1 \times m_1$, $m_2 \times m_2$, and $m_3 \times m_3$, respectively. The purpose to use pooling layers is to reduce the spatial size of the input, making it easier to process and requiring less memory. Pooling layers also help to reduce the number of parameters and make the training process faster. Furthermore, another advantage of using pooling layers is to remove local redundancies in the input data. It is worth mentioning that each pooling layer reduces the length of the input with a scale equal to the pooling window size. Therefore, the CNN and pooling layers generate an informative and lower-dimension representation of the input 2D power profile. The output of the last pooling layer is a three-dimensional representation which can be referred to as the extracted features. These features consist of n_3 different two-dimensional 2D-grids generated by the different filters of the last layer each with the spatial sizes of $q = N_{ch}/(m_1 \times m_2 \times m_3)$ and $r = N_z/(m_1 \times m_2 \times m_3)$. The next step in inverse system learning is to use a model to map the resulting low-dimensional features to the pump powers and wavelength values. To this end, a feed-forward NN can be utilized as it is typically used towards the end of a CNN when the goal is to take the features learned by the previous layers and use them to make predictions. Regarding this, a flattened layer is used to take the extracted three-dimensional features into a one-dimensional array. The flattening can be performed here as the resulting 3D representation has much less spatial redundancy compared to the input, which makes it suitable to be used as an input to a fully-connected NN. Once the resulting representation is flattened, it passes through the fully-connected NN which is referred to as a *regression network*, defined as $F(.;\theta_F)$ with the following mapping criteria:

$$F(.;\theta_F) : \mathbb{R}^{(q \times r \times n_3) \times 1} \rightarrow \mathbb{R}^{2N_p \times 1} \tag{4.2}$$

with θ_F defined as its trainable parameters. The proposed regression network has two hidden fully-connected layers of size N_{h1} and N_{h2}. Since the network is supposed to predict both the pump powers and wavelengths, the last layer of the regression network has $2N_p$ neurons (double the number of Raman pumps). The values of N_{h1} and N_{h2} are optimized with grid search depending on the proposed pump configuration.

4.2.1 Dataset generation and training process

The proposed inverse system model in Fig. 4.1 is categorized as a supervised learning algorithm. This model learns the approximate mapping between the 2D power profiles and their corresponding pump power values. The models learns this mapping using a training dataset $D = \{\mathbf{X}_k, \mathbf{y}_k | k = 1, \cdots, k\}$ where k is the number of samples, the output $\mathbf{y}_k = [\mathbf{p}_{p_k}; \boldsymbol{\lambda}_{p_k}]$ and $\mathbf{X}_k = \mathbf{P}_k$ are the pumping configuration vector and the corresponding 2D signal profile of the k-th sample, respectively.

The dataset for training the inverse model can be generated by solving the analytical direct SRS model presented in Eqs. 2.9 and 2.10. Moreover, each sample in the dataset is generated by a random selection of the pump power and wavelength values (within the pre-defined constraints), which is denoted as $\mathbf{y}_k$, and the resulting 2D profile $\mathbf{X}_k$ is generated by solving Eqs. 2.9 and 2.10. More specifically, each value of the pump parameters denoted as the m-th value of the vector $\mathbf{y}$ is selected based on a uniform distribution U:

$$y_m \sim U[y_m^{min}, y_m^{max}] \tag{4.3}$$

in which y_m^{min} and y_m^{max} are the minimum and maximum values allowed to be taken by the m-th value of $\mathbf{y}$, respectively.

Once the dataset with k number of samples is generated, we follow the same procedure performed for supervised learning approaches, which is dividing the dataset into separate, non-overlapping sets called: training, testing, and validation. The overall model of the inverse design network can be described as:

$$\mathbf{y} = R(F(\mathbf{X}; \theta_F); \theta_R) \tag{4.4}$$

where R and F are jointly trained using the training set and their optimum corresponding trainable parameters θ_R^* and θ_F^* are found by solving the following optimization described as follows:

$$\theta_R^*, \theta_F^* = \underset{\theta_R, \theta_F}{\operatorname{argmin}} \frac{1}{L} \sum_{l=1}^{L} C(\hat{\mathbf{y}}_l, \mathbf{y}_l) \tag{4.5}$$

in which L is the number of training samples, $\hat{\mathbf{y}}_l = R(F(\mathbf{X}_l; \theta_F); \theta_R)$ is the network output for a given input $\mathbf{X}_l$, $\mathbf{y}_l$ is the true output value, and $C(\hat{\mathbf{y}}_l, \mathbf{y}_l)$ is the mean square error (MSE) value between $\mathbf{y}_l$ and $\hat{\mathbf{y}}_l$, formulized as the following:

$$C(\hat{\mathbf{y}}_l, \mathbf{y}_l) = \frac{1}{2N_p} \sum_{i=1}^{2N_p} (\mathbf{y}_l^i - \hat{\mathbf{y}_l}^i)^2 \tag{4.6}$$

The parameters θ_R^* and θ_F^* are initialized randomly, and updated afterward in an iterative process with back-propagation, utilizing the advanced optimization algorithm RMSprop [57]. RMSprop is an extension of the Stochastic Gradient Descent (SGD) algorithm, momentum method, and the foundation of the Adam algorithm, providing a fast and robust convergence for each parameter. Once the network is trained, the trained parameters $\theta = \{\theta_R, \theta_F\}$ will be fixed, and the network's performance in predicting the pump powers and wavelengths will be evaluated.

4.2.2 Numerical amplifier setup

In this section, we numerically generate a dataset with different pumping configurations to train the CNN network and evaluate its accuracy. To numerically solve the SRS equations, we have utilized the *Raman Solver* function provided by GNPy [58], which is an open-source library developed in Python programming language for analyzing optical networks.

One single span of a standard single-mode fiber (SSMF) is considered in our amplifier setup and the signal power evolution is jointly investigated over the distance and the entire C-band (between 192 and 196 THz). Three separate amplification cases are deployed for training and evaluation: two counter-propagating cases with 2 and 3 pumps and a bidirectional propagating case with 4 pumps (2co+2counter).

The ranges for pump powers and wavelengths are specified in Table 4.1. The superscripts (-) or (+) on the power ranges specify the counter or co-propagation of the corresponding pump, respectively.

The C-band (192 and 196 THz) is divided into 40 channels with 100 GHz spacing. The per-channel input signal power is set to 0 dBm, resulting in a total WDM signal power of 16 dBm. Additionally, a standard silica fiber with the following parameters is assumed: span length L_{span} =100 km, signal data attenuation $\alpha_s = 0.2\ dB/km$, pump power attenuation $\alpha_p = 0.25\ dB/km$, effective area $A_{eff} = 80\ \mu m^2$, non-linear coefficient $\gamma = 1.26\ 1/W/km$.

Tab. 4.1: Power and wavelength ranges for each DRA case

Parameter	2 pumps	3 pumps	4 pumps
$p_1[mW]$	$[40-400]^-$	$[30-300]^-$	$[30-300]^+$
$p_2[mW]$	$[40-400]^-$	$[30-300]^-$	$[30-300]^+$
$p_3[mW]$	-	$[30-300]^-$	$[30-300]^-$
$p_4[mW]$	-	-	$[30-300]^-$
$\lambda_1[nm]$	$[1414-1449]$	$[1414-1437.3]$	$[1414-1449]$
$\lambda_2[nm]$	$[1449-1484]$	$[1437.3-1460.3]$	$[1449-1484]$
$\lambda_3[nm]$	-	$[1460.3-1484]$	$[1414-1449]$
$\lambda_4[nm]$	-	-	$[1449-1484]$

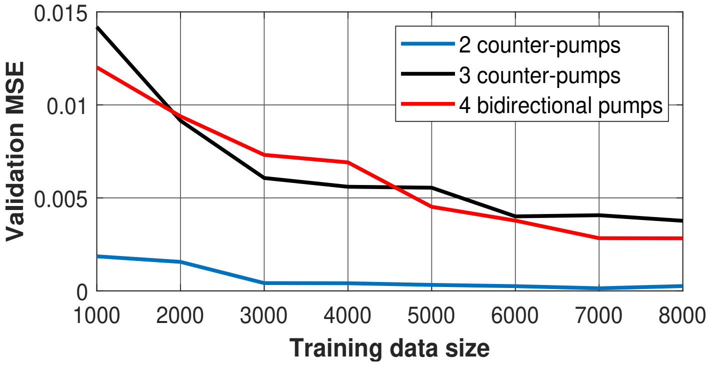


Fig. 4.2: MSE of the validation data-set for different pumping schemes as a function of the training data-set size

4.2.3 Numerical results

Once the amplification configuration for different pumping schemes is set, a dataset consisting of 11000 samples is generated and divided into 8000, 1000, and 2000 samples for training, validation, and testing, respectively. In order to determine the best training data size for each pumping case, we have generated training subsets of sizes from 1000 to 8000. Each subset is generated by random selection of the samples from the primary 8000-size dataset aimed for the training size. For each training subset, the CNN model is trained and its performance is evaluated using the resulting MSE value from the validation set.

Fig. 4.2 shows the validation MSE as a function of the size of the training dataset. Based on the validation MSE and also the training time, we realized that for 2 and 3 counter and 4 bidirectional cases, the best training data sizes are 5000, 6000, and 7000 samples, respectively. According to Fig. 4.2, increasing the training size will not result in a remarkable improvement. The model with the best training size for each pumping scenario is investigated further with different hyper-parameters values such as the number of CNN filters and their size, the size of the average pooling layers between each consecutive CCN layer, and the number of neurons in the hidden layers in the regression network. More specifically, the parameters of the *feature extraction* network, including the number of filters (n_1, n_2, n_3), filter sizes (f_1, f_2, f_3), and the average-pooling layer window sizes (m_1, m_2, m_3) have been set and evaluated based on the most common values in the literature. Regarding the number of filters for each layer, we tested 32 and 64 numbers and observed that 64 filters increase the training time extremely with no improvement in performance. Moreover, for each CNN layer, a filter size of 3×3 has less number of training parameters with slightly better validation MSE over a bigger 5×5 filter. For the average-pooling layers, a commonly used 2×2 window has a better MSE over a window of size 3×3.

In the regression network, we evaluated the validation MSE by setting the N_{h1} and N_{h2} values according to the set of $\{20, 40, 80, 100\}$ values. For 2 pumping case, $N_{h1} = 40$ and $N_{h2} = 40$ with ReLU activation, and contrarily, for both 3 and 4 pumping cases, $N_{h1} = 100$ and $N_{h2} = 40$ with ReLU activation function have a lower validation loss.

For all pumping schemes, the training batch size has been set to the common value of 128, and the learning rate of the RMSprop is set to 0.001. Furthermore, the best distance resolution for 2 and 3 pumps is 2 km, and for 4 pumps is 1 km. The higher resolution of the 4 bidirectional pump case is mainly because of its more signal variation along the distance compared to the other two counter-propagating schemes. The network output is the min-max normalized value of the pump setup, and its exact value can be obtained after the prediction, using the parameters' corresponding minimum and maximum values.

Once the network is trained with the above-mentioned parameters, it is utilized to predict the pump power and wavelength values of the 2D profiles in the test set. The test set has been generated with no overlap with the training and validation sets, to make sure that the model is not over-fitting and the test data have not been previously observed by the network. The most straightforward way to analyze the trained CNN model performance is to predict the pump power and wavelength values and compare them with their corresponding true

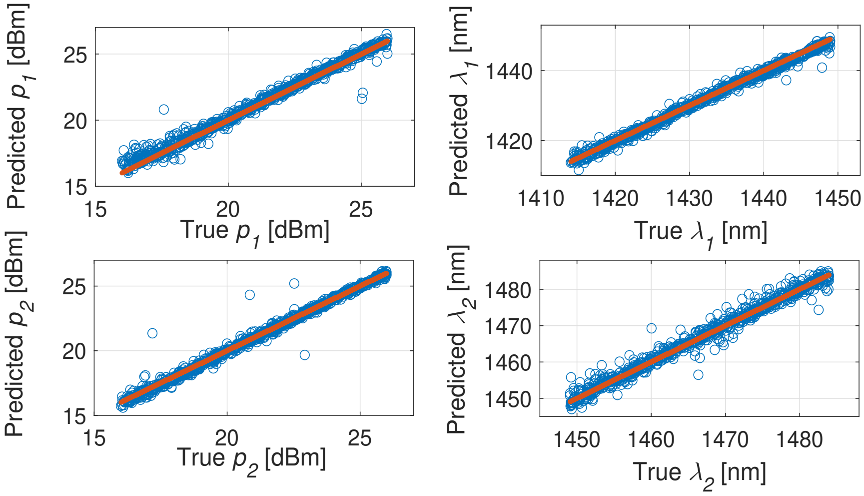


Fig. 4.3: True-predicted pump power and wavelength values for two backward propagating pump case.

values. The scatter-plot of the true versus predicted values for pump power and wavelength values for the amplification scenario with two counter-propagating pumps is shown in Fig.4.3. The solid red line in all sub-plots represents the ideal prediction where the predicted and the true parameters match perfectly. Instead of showing the plot for the 3 counter-propagating pumps case (due to its similar scatter-plot to Fig.4.3), we have shown the results for 4 bidirectional propagating pump in Fig.4.4. It is visually obvious that the model shows more deviation around the ideal prediction for the wavelength values, especially for λ_2 and λ_4, which correspond to the wavelength in the range between 1449 nm and 1484 nm for both co- and counter-propagating pumps. To have a quantitative analysis of the CNN model performance, the mean (μ) and the standard deviation (σ) of the MSE error for pump power and wavelength values are calculated and reported for 2, 3, and 4 pump scenarios in tables 4.2, 4.3 and 4.4, respectively. Regarding the reported MSE mean and standard deviation values, in case the number of pumps increases, the accuracy of the model decreases accordingly since the pumping scheme becomes more complex to solve and moreover, the signal power evolution will have more excursion.

One drawback of comparing the true versus predicted parameter values is that it is difficult to practically quantify the goodness of designing the profile using the proposed inverse CNN model. In practical scenarios,

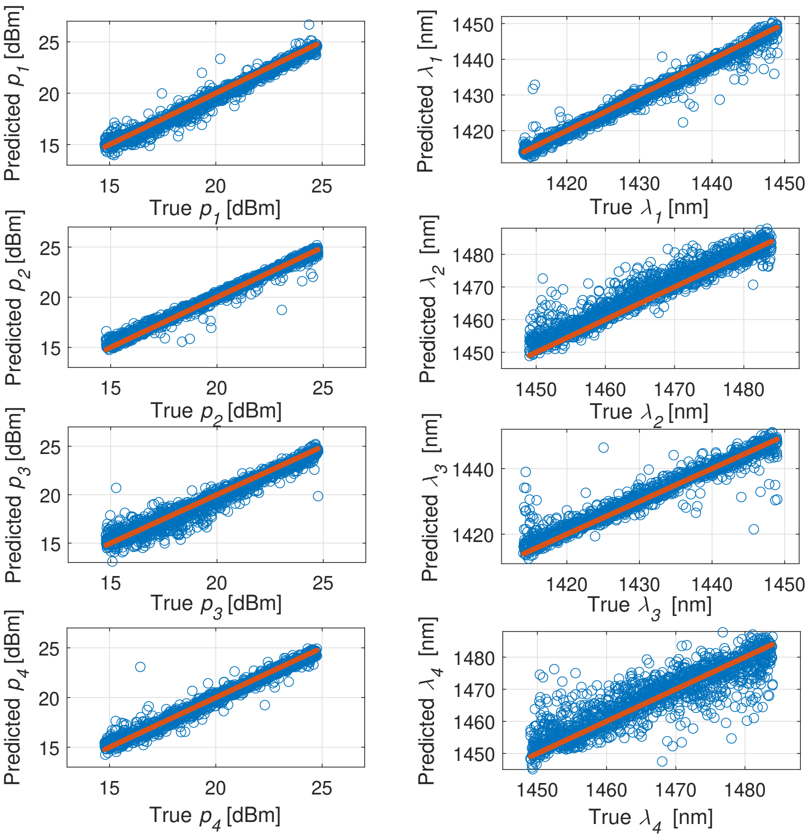


Fig. 4.4: True-predicted pump power and wavelength values for four bidirectional propagating pump case.

Tab. 4.2: Mean (μ) and standard deviation (σ) of MSE for 2 counter-propagating pumps case.

$Parameter$	p_1	p_2
$\mu[dB]$	0.29	0.22
$\sigma[dB]$	0.35	0.3

$Parameter$	λ_1	λ_2
$\mu[nm]$	1.37	1.25
$\sigma[nm]$	1.66	1.13

Tab. 4.3: Mean (μ) and standard deviation (σ) of MSE for 3 counter-propagating pumps case.

$Parameter$	p_1	p_2	p_3
$\mu[dB]$	0.38	0.64	0.57
$\sigma[dB]$	0.4	0.83	0.73

$Parameter$	λ_1	λ_2	λ_3
$\mu[nm]$	1.08	1.44	1.67
$\sigma[nm]$	1.12	1.66	1.64

Tab. 4.4: Mean (μ) and standard deviation (σ) of MSE for 4 bidirectional-propagating pumps case

$Parameter$	p_1	p_2	p_3	p_4
$\mu[dB]$	0.38	0.25	0.49	0.30
$\sigma[dB]$	0.33	0.28	0.49	0.33

$Parameter$	λ_1	λ_2	λ_3	λ_4
$\mu[nm]$	1.15	2.46	1.49	3.11
$\sigma[nm]$	1.33	2.32	2.1	3.36

the primary objective is to create a specific target 2D profile. Therefore, solely predicting the parameters may not be sufficient to evaluate the goodness of the design. We will likely need to apply the predicted values back to the system to determine if they produce the same target profile. Hence, we consider an alternative way to evaluate the model performance in designing a target 2D power profile $\mathbf{X}$, by applying the corresponding predicted pump power and wavelength values to the amplifier setup, and generating the resulting 2D power profile $\hat{\mathbf{X}}$. Next, the performance of the CNN is evaluated by quantifying the error, which can be measured as the maximum absolute error (MAE) between $\mathbf{X}$ and $\hat{\mathbf{X}}$. This can be formulated as follows:

$$\text{MAE [dB]} = \max_{f,z} |\mathbf{X} - \hat{\mathbf{X}}| \tag{4.7}$$

The schematic of the proposed evaluation approach is presented in Fig. 4.5. For all three pumping schemes, the MAE is calculated for the 2D profiles in the test set. Fig. 4.6 indicates the probability density function (PDF) for the maximum absolute error (MAE) of the reconstructed power profile. The μ and σ values of MAE for all three pumping cases in reported in table 4.5, asserting the overall good performance of the CNN

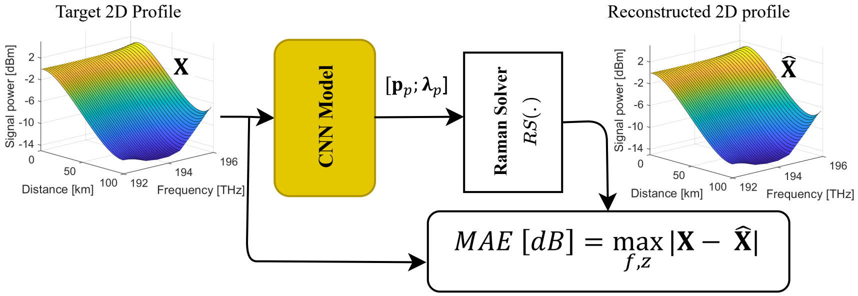


Fig. 4.5: Evaluation diagram of the CNN inverse model.

model in terms of the reconstruction error for 2D profiles defined in the whole C-band and along the fiber distance.

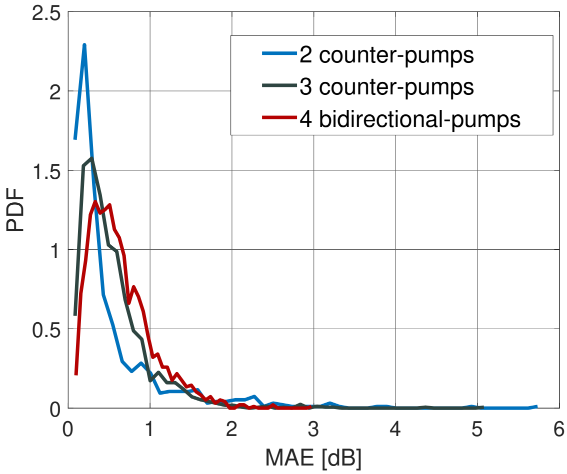


Fig. 4.6: Probability density function of the MAE for three pumping cases.

4.3 Fiber length-aware CNN model for inverse DRA design

In the previous section, all simulation results were performed assuming that the fiber-related parameters, especially the fiber length, are fixed. Considering a fixed type of fiber (SSMF) under test, one variable which can normally vary and have perturbations is the fiber length. If the fiber length in the numerical amplification setup changes, the process of

Tab. 4.5: Mean (μ) and standard deviation (σ) of reconstruction MAE for all three pumping scenarios.

Pumping case	2-counter	3-counter	4-bidirectional
$\mu[dB]$	0.51	0.54	0.64
$\sigma[dB]$	0.62	0.43	0.38

dataset generation and moreover, the training of the CNN model needs to be performed from the beginning.

The CNN architecture provided in Fig.4.1 is not capable of considering the fiber length as an input and performing the inverse system design for the proposed target fiber length. Additionally, generating thousands of samples for every possible fiber length and training a separate CNN model for each fiber length is not a practical approach to dealing with fiber length variability. Considering this, we aim to modify the CNN model such that it can accept both the target 2D power profile and the length of the fiber as the input, and predict the pump power values for that specific fiber length.

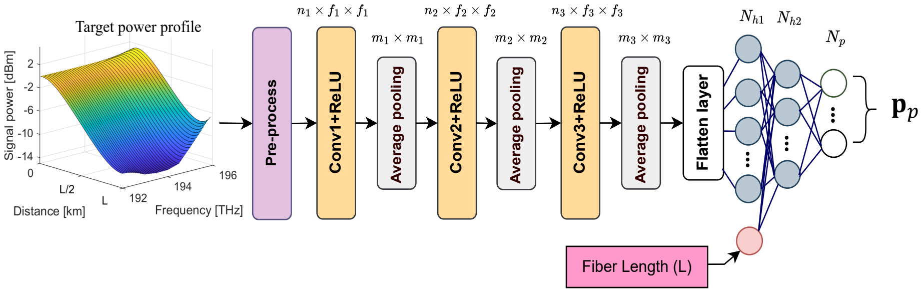


Fig. 4.7: Diagram of the proposed length-aware CNN architecture for the inverse DRA design based on target power profile in frequency and distance, considering fiber length as a variable.

The architecture of the proposed length-aware CNN model is shown in Fig.4.7. In this architecture, the target 2D profile is pre-processed and then used as the input to the feature extraction network (composed of the CNN and pooling layers). In the previous CNN architecture, the pre-processing was only including the min-max normalization of the power profile. However, here, since the target profile size changes for various fiber lengths, we also need to perform one further pre-processing step which is the adjustment of the input dimensions such that all of them have a unique input power evolution size. In relation to this, we take the maximum possible length of fibers under test as the reference and by performing interpolation, increase the length dimension of the profiles to match the reference size. The proposed 2D power profile adjustment makes all profiles with different lengths have the same dimensions. In order to differentiate these profiles and make the CNN model aware of the length of each input target 2D profile, we have added the length as the input to the regression model. In fact, the length parameter is concatenated to the low-dimensional feature representation of the target 2D power profile which is used further to predict the pump configuration using the regression network.

4.3.1 Data set generation and training process

Regarding the dataset, we consider generating 2D power profiles with lengths in the range between 70 km and 100 km with a discrete 5 km step size. For every discrete length label, 3000 2D profiles are generated which are divided into 1400 training, 800 validation, and 800 test 2D profiles. The whole training data, consisting of the 2D profiles and their corresponding length labels and the pump configuration values, are fed to train the length-aware CNN model in Fig.4.7.

Regarding the amplifier setup, we use a standard second-order system. more specifically, we use eight pumps including four co- and four counter-propagating ones. For each propagation direction, one second-order pump operating at 1366 nm and three first-order ones operating at 1425 nm, 1455 nm, and 1475 nm are considered. The available power range for each pump is reported in table 4.6. Unlike the setup introduced in the last section, and approaching a more practical scenario where the pump wavelengths are challenging to tune, we assume that the pump power values in this analysis are the only parameters to be predicted and the pump wavelength values are fixed. Considering this, there is no need to have the wavelength values to be predicted in the last layer of the network and therefore, the last layer of the network is fully dedicated to predicting the pump power values $\mathbf{p}$.

Once the length-aware CNN model is trained on the training data set, its performance is evaluated by measuring the MAE between each test 2D profile and the generated 2D profile after applying the predicted pump power values to the amplifier setup. After measuring the MAE for all test 2D profiles, its PDF for each corresponding length label is generated and shown in Fig.4.7. Addtionally, the mean μ and the standard deviation

Tab. 4.6: Wavelength values and power ranges of the Raman pumps.

Co-pumps	p_1	p_2	p_3	p_4
λ [nm]	1366	1425	1455	1475
$[\mathbf{p}_{LB}, \mathbf{p}_{UB}]$ [mW]	[200,1200]	[5,150]	[5,150]	[5,150]
Counter-pumps	p_5	p_6	p_7	p_8
λ [nm]	1366	1425	1455	1475
$[\mathbf{p}_{LB}, \mathbf{p}_{UB}]$ [mW]	[200,1200]	[5,150]	[5,150]	[5,150]

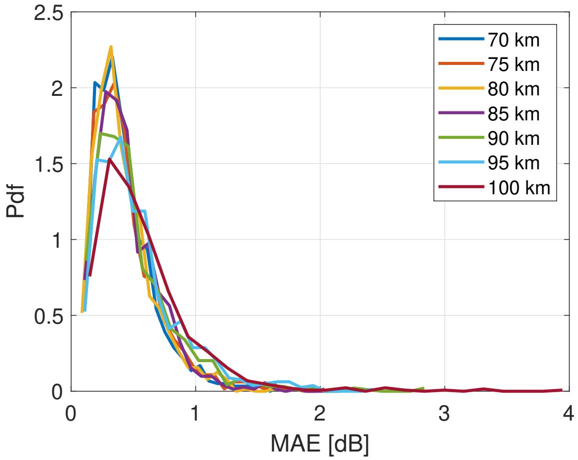


Fig. 4.8: Probability density function of the resulting value for the test dataset of different length values from 70 km to 100 km.

σ of the MAE for each length is reported in table 4.7. According to this table, the model performance is statistically good for predicting the pump power values for all test data with different length labels, showing less than 0.6 dB average error with less than 0.4 MAE standard deviation.

In the above-mentioned simulations, we have evaluated the performance of the model in predicting the pump power values for 2D profiles whose length labels already exist in the training set (70 km to 100 km with 5 km step size for both training and test data sets are considered).

Tab. 4.7: Mean (μ) and standard deviation (σ) of the CNN model MAE for test data with equal length labels as training data.

$L[km]$	70	75	80	85	90	95	100
$\mu[dB]$	0.43	0.44	0.42	0.45	0.49	0.53	0.57
$\sigma[dB]$	0.25	0.27	0.25	0.26	0.34	0.34	0.40

As the NN models are known for their good extrapolation performance [26], here we aim to evaluate the trained model's accuracy in designing 2D profiles whose length labels do not exist among the labels of the training data set. Approaching this, we have generated another set of test data 2D profiles with the length labels taken from the following list: $L[km] = [73, 78, 83, 88, 93, 98]$. For each length label in this list, 800 2D profiles are generated and the 2D profiles with their corresponding length labels are fed into the pre-trained length-aware CNN model. Once the pump power values are predicted, they are applied back to the amplifier setup, and their resulting 2D profile is generated. Similar to the analysis for the previous test data, the PDF of the MAE for each length label is calculated and shown in Fig.4.9.

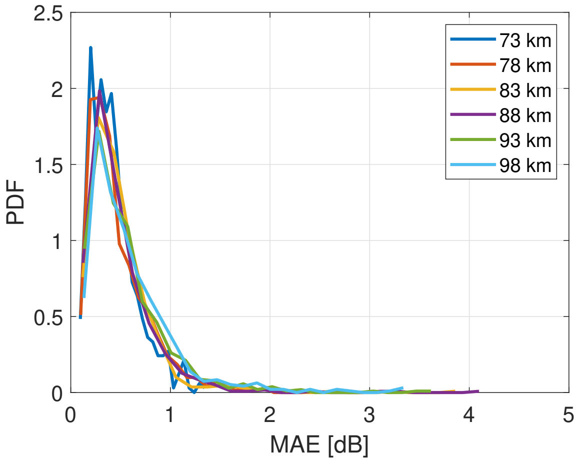


Fig. 4.9: Probability density function of MAE for 2D profiles whose length labels are not among the length labels used for training the CNN model.

Tab. 4.8: Mean (μ) and standard deviation (σ) of the CNN model MAE for test data with non-equal length labels as training data.

$L[km]$	73	78	83	88	93	98
$\mu[dB]$	0.42	0.47	0.49	0.47	0.54	0.58
$\sigma[dB]$	0.23	0.31	0.35	0.35	0.40	0.44

Additionally, the MAE mean μ and the standard deviation σ values are calculated and reported in table 4.8. The resulting average MAE is less than 0.6 dB for all length labels and its standard deviation is less than 0.45 dB, showing statistically good performance for pump power prediction for the new test data set. According to these results, we can confirm that the proposed length-aware CNN model has a good extrapolation performance as it is accurate in designing test 2D profiles when their length does not exactly match with any of the length labels existing in the training set. It is worth noting that the test length labels should lie within the minimum and maximum lengths in the training set with enough surrounding label values for better extrapolation results.

4.4 Summary

In this chapter, we presented a CNN-based model for inverse DRA design to shape the signal power evolution in 2D, i.e. in frequency and fiber distance domains. The proposed inverse model consists of two networks trained end-to-end: 1) A *feature extraction* network with 3 CNN layers utilized to extract an informative and low-dimensional set of features, representing the input 2D target power profile, and 2) A *regression* model built based on two fully-connected NN layers, aiming to predict the pump powers and wavelengths values based on the extracted features. Numerical simulations show that the proposed framework provides high accuracy in terms of predicting the pump parameters for both counter and bidirectional propagating pumps for the signal propagating in the whole C-band. Moreover, we proposed a length-aware CNN model in which besides the 2D power profile, the fiber length is also provided as an input to the network. Considering this, the network has the capability to perform inverse DRA design for different fiber length values.

5

ML-based online optimization framework for signal power evolution shaping using DRAs

5.1 Introduction

In Chapter 4, we proposed and numerically validated a Convolutional Neural Network (CNN) based inverse system model to map the 2D signal power profiles to the pump power and wavelength values in a distributed Raman amplifier (DRA) setup. The proposed CNN model was trained offline using a dataset, which was generated by applying a set of randomly selected pump power and wavelength values to the DRA setup and measuring their corresponding 2D power profiles. This model was tested on a numerical amplifier setup with first-order pumps with counter- and bidirectional propagating schemes and obtained low average and low standard deviation means absolute error (MAE) values for the test data. Nonetheless, most of the focus in the literature has been on designing 2D profiles that are practically interesting, and moreover challenging, such as 2D flat or 2D symmetric profiles [14, 15, 47, 48]. Regarding this, both numerical and experimental investigations in the literature have proven that practically interesting profiles such as 2D flat or 2D symmetric can be achieved using high-order Raman pumps in a bidirectional propagating scheme [47, 59]. Most common Raman amplification setups for designing the power evolution in fiber distance are made by combining second-order and first-order pumps in a bidirectional propagating scheme [59], or by using Raman pumps combined with Fiber Bragg Gratings (FBGs) as reflectors [14]. In [14, 15], an amplification structure based on second-order Raman pumps with FBGs has been presented to provide a quasi-lossless transmission medium by implementing an ultra-long Raman fiber laser. Moreover, in [47, 48], numerical optimization of narrow-band signal power asymmetry for OPC has been addressed by considering different amplification setups using first-order and second-order pumps. As also introduced in Chapter 4, many of the proposed approaches in the literature for designing 2D power profiles rely on either heuristically optimizing the pump parameters or simplifying the optimization process based on the specific target profile and the setup under test [14, 15, 24, 52, 53]. However, it would be highly beneficial to provide a framework that can automatically optimize the parameters and moreover, is flexible enough to be applicable to different power profiles and amplifier setups.

Practically desired 2D profiles such as 2D flat or 2D symmetric profiles have their own specific objective or constraints which need to be fulfilled. The objective functions for these profiles (power excursion or power symmetry) are non-differentiable with respect to the optimization parameters which makes it challenging to optimize using gradient-based learning methods. Moreover, these profiles are still challenging to be ideally obtained with the current amplification setups due to their physical limitations. Therefore, the performance of the inverse Machine Learning (ML) methods such as CNN, which have been trained on the data extracted from the setup, might not be very accurate in designing the aforementioned practical profiles. Considering these two challenges, a highly advantageous approach to design these profiles is to use an online gradient-free optimization framework to tune the parameters.

In this chapter, we propose an online framework based on a gradient-free optimization technique whose main goal is to minimize the cost functions in an iterative process directly on the amplifier setup. More specifically, we will deploy Differential Evolution (DE), known as an evolutionary-based optimization, to adjust the pump power values in an amplifier setup. DE is a derivative-free technique with the robustness and flexibility to capture solutions to complex optimization problems [60, 61].

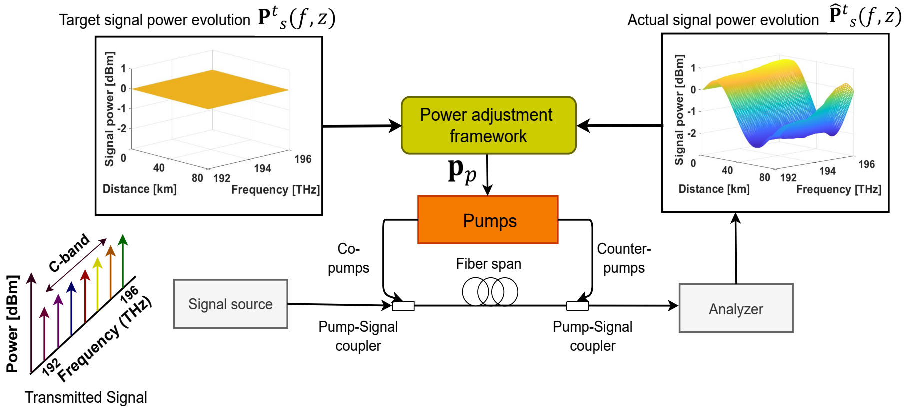


Fig. 5.1: An online optimization framework utilized to adjust the Raman pump power values in an amplification setup to design a target 2D signal power profile.

5.2 Online framework for 2D power evolution design using DRAs

Fig. 5.1 illustrates an amplification scheme in which an adjustment framework is utilized to optimize the pump power values to achieve a target 2D signal power evolution profile $\mathbf{P}^t(f,z)$. In this scheme, $\mathbf{P}^t(f,z)$ is used as the input to the framework which predicts the Raman pump power values $\mathbf{p}_p$, including both co- and counter-propagating pumps. The pump power values are then applied to the fiber span ends in which a wavelength division multiplexing (WDM) frequency comb signal with flat spectra in the entire C-band is propagating through. The signal power evolution shape is tailored in both frequency and fiber distance domains due to stimulated Raman scattering (SRS) phenomena and an analyzer measures the actual resulting 2D power profile $\hat{\mathbf{P}}^t(f,z)$. The resulting profile is used as the second input to the pump power adjustment framework where its cost value with the main target $\mathbf{P}^t_s(f,z)$ is calculated. Once the resulting cost value is evaluated, the framework updates the pump power values and applies them to the setup to be evaluated in the next iteration. This mechanism creates an *apply-feedback*, mainly aiming at fine-tuning the optimum parameters by reducing the cost values in an iterative process. This process continues until a certain convergence criterion such as a maximum number of iterations or a certain threshold for cost value is fulfilled.

5.3 Differential Evolution algorithm for 2D power evolution design

Evolutionary optimization techniques like DE, start with a population of individuals which are the primary possible solutions to the optimization problem. Once the population is initialized, the algorithm follows a set of rules trying to improve each individual in an iterative process in order to make it closer to the optimum point [35]. In case there is no prior information on how to initialize the individuals, the first option will be to randomly select them in the parameter space.

We consider DE as the main building block of the proposed online framework for optimizing the pump power values to design 2D signal power profiles. Considering this, we can define the individuals by a random selection of values within a lower-bound $\mathbf{p}_{LB}$ and an upper-bound $\mathbf{p}_{UB}$ value. The simplest way of specifying $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$, is to assign them the minimum and maximum possible values of pump powers, respectively. Fig. 5.2 shows the detailed numerical block diagram of the proposed DE framework used to optimize the set of pump power values $\mathbf{p}_p = [p_1, ..., p_{N_p}]$ to design a target 2D profile $\mathbf{P}^t(f, z)$. The DE framework is combined with the numerical direct model of the Raman amplification scheme referred to as Raman Solver $\mathcal{R}(\cdot)$ which calculates the signal power evolution in frequency and fiber distance by solving the SRS Eqs. 2.9 and 2.10, given the pump power values:

$$\mathbf{P}(f, z|\mathbf{p}_p) = \mathcal{R}(\mathbf{p}_p) \tag{5.1}$$

In the proposed framework in Fig. 5.2, once the $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$ values are defined, the population consisting of n individuals is initialized. Each individual in the population is a candidate for the optimization solution and it is represented by a vector of size N_p, named $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^{N_p}), i = 1, ..., n$, as a set of pump power values. Each individual passes through the Raman Solver $\mathcal{R}(\cdot)$ and its corresponding 2D power evolution profile is calculated. Assuming that we already have a pre-defined cost function $J(\cdot)$, it calculates the cost (error) between the resulting 2D profiles of all individuals and the main target 2D profile. After calculating the cost for each individual, an iterative process including mutation and crossover is performed and in each iteration, the individuals are adjusted and evaluated in a loop using Raman Solver to decrease the resulting error, aiming to converge to the optimum point with minimum cost value [35].

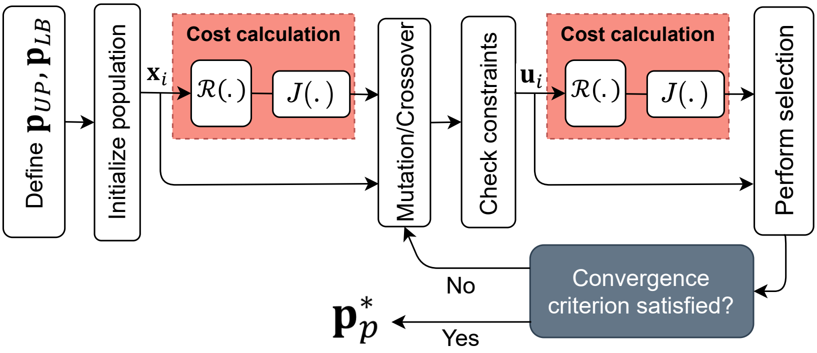


Fig. 5.2: DE framework, applied to adjust the pump power values in an online numerical amplification setup.

Providing further details, within the online DE loop, three individuals, namely $\mathbf{x}_{r_1}$, $\mathbf{x}_{r_2}$, and $\mathbf{x}_{r_3}$, are randomly selected from the population, with the condition that $r_1 \neq r_2 \neq r_3 \neq i$. Subsequently, a mutation process is carried out to generate a donor vector, denoted as $\mathbf{v}_i$, which has the same size as $\mathbf{x}_i$. The purpose of generating the donor vector $\mathbf{v}_i$ is to enhance the diversity of the population and prevent convergence to a local minimum. The donor vector can be defined as follows:

$$\mathbf{v}_i = \mathbf{x}_{r_1} + F \cdot (\mathbf{x}_{r_2} - \mathbf{x}_{r_3}) \tag{5.2}$$

where $F \in [0, 1]$ is the mutation factor, which controls the diversity of the population.

Following the mutation, the crossover process combines the elements of the donor vector $\mathbf{v}_i$ and the current target vector $\mathbf{x}_i$ to create the trial vector $\mathbf{u}_i$ in the following manner:

$$\mathbf{u}_i^j = \begin{cases} v_i^j, & \text{if } r_i^j \leq CR \, or \; j = j_{rand} \\ x_i^j, & \text{otherwise} \end{cases} \tag{5.3}$$

where u_i^j, v_i^j and x_i^j represent the jth element of $\mathbf{u}_i$, $\mathbf{v}_i$ and $\mathbf{x}_i$, respectively. CR is the crossover probability, $r_i^j \sim U(0, 1)$ is a uniform distribution generated for each j, and $j_{rand} \in 1, 2, ..., n$ is a random integer used to ensure that $\mathbf{x}_i \neq \mathbf{u}_i$. After performing the mutation and the crossover, the trial vector is checked if it follows the constraints $\mathbf{p}_{LB}$

Algorithm 1 Online pump power adjustment framework.

```
Input: Target power profile $\mathbf{P}^t(f,z)$, DE parameters: $\{n$: Population size, $CR$: Crossover probability, $F$: Mutation factor, $MaxEv$: Maximum number of cost function evaluations$\}$
Output: Pump power values $\mathbf{p}_p^*$
Optimization procedure:
Generate population: $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^n), i = 1, ..., N_p$, $\mathbf{p}_{LB} < \mathbf{x}_i < \mathbf{p}_{UB}$
Direct model: Calculate $\mathbf{P}(f, z|\mathbf{x}_i) = \mathcal{R}(\mathbf{x}_i)$
Calculate individual cost: $J(\mathbf{x}_i)$
While $Ev < MaxEv$:
    for $i = 1 : N_p$ do
        Select three random individuals $x_{r1}, x_{r2}, x_{r3}$ from the population where $i \neq r_1 \neq r_2 \neq r_3$
        Mutation : Form the donor vector as in Eq. 5.2
        Crossover: Form the trial vector as in Eq. 5.3
        Check constraints: Check if $\mathbf{p}_{LB} < \mathbf{u}_i < \mathbf{p}_{UB}$, otherwise, go to the next individual.
        Direct model: $\mathbf{P}(f, z|\mathbf{u}_i) = \mathcal{R}(\mathbf{u}_i)$
        Evaluate: if $J(\mathbf{u}_i) \leq J(\mathbf{x}_i)$, replace $\mathbf{x}_i$ with $\mathbf{u}_i$
        $Ev \leftarrow Ev + 1$
    End for
End while
```

and $\mathbf{p}_{UB}$. Afterward, $\mathbf{u}_i$ is passed through the Raman solver function to generate the corresponding power evolution profile $\mathbf{P}(f, z|\mathbf{u}_i) = \mathcal{R}(\mathbf{u}_i)$ and calculate the cost value J. The cost values of the trial vector $J(\mathbf{u}_i)$ and the target vector $J(\mathbf{x}_i)$ are compared with each other and the one with a lower cost value keeps staying in the population and the other one will be removed. The selection process can be formulated to form the new particle $\mathbf{x}_{i_{new}}$ as follows:

$$\mathbf{x}_{i_{new}} = \begin{cases} \mathbf{u}_i, & \text{if } J(\mathbf{u}_i) < J(\mathbf{x}_i) \\ \mathbf{x}_i, & \text{otherwise} \end{cases} \tag{5.4}$$

The DE loop continues further until it reaches a maximum number of iterations $MaxEv$ defined as the convergence criteria. The implementation procedure of the proposed CNN-assisted DE optimization scheme is summarized in Algorithm 1.

5.3.1 Numerical amplifier setup

In this section, we use the proposed online DE framework to perform pump power prediction for designing practically desired 2D flat and 2D symmetric signal power profiles. In order to enable the framework to approach these practical profiles, we use a standard second-order bidirectional numerical amplifier setup which consists of eight pumps amplifying the signal in the entire C-band in a single span of a standard single-mode fiber (SSMF). For each propagating direction, four pumps consisting of a second-order pump with the wavelength 1366 [nm] and three first-order pumps with the wavelengths 1425 [nm], 1455 [nm], and 1475 [nm], are considered. The pump wavelengths 1366 [nm] and 1455 [nm] are selected mainly based on the previously presented papers in which second-order and first-order pumps are utilized to achieve flat and symmetric power profiles [45–47]. The primary objective in selecting the 1425 [nm] and 1475 [nm] wavelengths is to ensure an adequate power supply for all channels across the entire C-band, as discussed in previous works [53, 62].

In the following analysis, we assume to have fixed pump wavelength values, approaching a more practical scenario where pump power values are mostly considered as the free parameters to be tuned. However, it is worth noting that the proposed framework is also capable of optimizing the wavelength values, as well. Ranges of the pump powers considered for $[\mathbf{p}_{LB}, \mathbf{p}_{UB}]$ are reported in table 5.1.

The desired 2D signal power profiles are designed in the entire C-band (ranging from 192 THz to 196 THz), divided into 40 channels, each having a bandwidth of 100 GHz. The input signal power per sub-channel is fixed at 0 dBm, resulting in a total signal power of 16 dBm. Moreover, the utilized SMF has the following specifications: length of the span $L_{span} = 80$ km, signal data attenuation of $\alpha_s = 0.2$ dB/km, second-order pumps $[p_1, p_5]$ attenuation $\alpha_P^{2nd} = 0.32$ dB/km, and first-order pumps $[p_2, p_3, p_4, p_6, p_7, p_8]$ attenuation $\alpha_P^{1st} = 0.25$ dB/km, effective area $A_{eff} = 80$ μm^2, non-linear coefficient $\gamma = 1.26$ 1/W/km, and Raman coefficient peak $g_R = 0.39$ 1/W/km. A standard silica Raman efficiency profile is assumed, as shown in Fig 2.1 [1].

To deploy the numerical Raman Solver model, we have used the GnPy library [58] which utilizes a Boundary Value Problem (BVP) solver based on residual error control to solve the set of Raman differential equations.

We set the distance resolution to $z_{res} = 500$ m which satisfies the residual error threshold $r_{th} = 10^{-6}$.

Tab. 5.1: Raman pump wavelengths and power ranges.

Co-pumps	p_1	p_2	p_3	p_4
λ [nm]	1366	1425	1455	1475
$[\mathbf{p}_{LB}, \mathbf{p}_{UB}]$	[200,1200]	[5,150]	[5,150]	[5,150]
Counter-pumps	p_5	p_6	p_7	p_8
λ [nm]	1366	1425	1455	1475
$[\mathbf{p}_{LB}, \mathbf{p}_{UB}]$	[200,1200]	[5,150]	[5,150]	[5,150]

5.3.2 Simulation results: 2D flat power profile

In this part, we focus on optimizing the pump power values to provide a 2D flat profile, meaning that it has the least signal power excursion jointly in frequency and fiber distance. Considering the cost function for power excursion minimization, we extend the formulation presented in the literature and define it to properly fit our wide-band with multi-channel (entire C-band) analysis [15, 63]. More specifically, We formulate power excursion minimization as an optimization problem aiming to find the optimum pump power values $\mathbf{p}_p^*$ that minimize the maximum power excursion among all distance and frequency signal power points in the fiber. In more precise terms, we establish three distinct cost functions to address the design of a 2D flat profile. Each cost function incorporates a particular constraint to shape the 2D profile. The three proposed cost functions are sequentially defined as follows.

First, the main objective is to minimize the maximum power excursion in frequency and along the fiber span J_0 [dB]. Maximum power excursion (which we will refer to as power excursion) is defined as the difference between the overall maximum and the overall minimum of the signal power level in frequency and fiber distance domains:

$$J_0(\mathbf{p}_p) = \max_{f,z}(\mathbf{P}(f, z|\mathbf{p}_p)) - \min_{f,z}(\mathbf{P}(f, z|\mathbf{p}_p)) \tag{5.5}$$

In a wide-band amplification system, another desirable requirement is to minimize the spectrum excursion at the receiver side in order to achieve optimal signal power distribution [37, 39, 40]. Considering this objective, we define a generalized cost function that aims to minimize the maximum spectrum excursion across all points in the fiber. This cost function is referred to as spectrum excursion J_1 [dB], and it is defined as follows:

$$J_1(\mathbf{p}_p) = \max_z[\max_f(\mathbf{P}(f, z|\mathbf{p}_p)) - \min_f(\mathbf{P}(f, z|\mathbf{p}_p))] \tag{5.6}$$

Furthermore, in the experimental configurations outlined in the existing literature concerning flat power evolution design [15, 63], the pump power values are adjusted through empirical methods to fully compensate for the signal loss and achieve a 0-dB gain between the transmitter and the receiver. In this context, the 0-dB gain J_2 between the transmitter and the receiver is regarded as the third cost value, and it is formulated as follows:

$$J_2(\mathbf{p}_p) = \max_f |\mathbf{P}(f, z = L|\mathbf{p}_p) - \mathbf{P}(f, z = 0|\mathbf{p}_p)| \tag{5.7}$$

where L is the fiber length and $|.|$ is the absolute value operator.

Considering these three cost functions, we define a multi-objective optimization problem for a 2D flat target profile which aims to find the optimum $\mathbf{p}_p$ such that all three cost functions are minimum. Given the complexity of each individual cost function, determining whether their minimum values coincide poses a challenge. Therefore, to simplify the optimization process and also to be able to interpret the impact of each cost on the overall optimization process, we make an approximation by converting it into a classical weighted-sum optimization. In this approach, one weight value is assigned to each cost function, defined as a hyper-parameter, and it is added to the other weighted costs as the following:

$$\mathbf{p}_p^* = \arg\min_{\mathbf{p}_p} \sum_{i=0}^{2} m_i J_i(\mathbf{p}_p) = \mathbf{m} \cdot \mathbf{J}^T(\mathbf{p}_p), \tag{5.8}$$

$$\text{such that } \mathbf{p}_{LB} \leq \mathbf{p}_p \leq \mathbf{p}_{UB}, \; m_i > 0, \; \sum_{i=0}^{2} m_i = 1$$

where $\mathbf{m} = [m_0, m_1, m_2]$ is a hyper-parameter vector of weights, the objective vector is $\mathbf{J}(\mathbf{p}_p) = [J_0(\mathbf{p}_p), J_1(\mathbf{p}_p), J_2(\mathbf{p}_p)]$, and T is the transpose operator. Considering the optimization problem defined in Eq. 5.8, the impact of each cost function $J_i, i = 0, 1, 2$ on the final optimum pump power values can be tuned by their corresponding weight value defined as m_i. This approach provides the flexibility to tune weights arbitrarily and depending on our target power evolution shape, designate more importance to our desired cost function in the optimization process. An advantage of the DE as a gradient-free optimization technique is that we can target problems with multiple objectives which we cannot solve easily with gradient-based ML inverse models like CNNs and fully-connect Neural Networks (FCNNs).

We utilize the DE framework presented in Fig. 5.2 to perform the optimization. According to Algorithm 1, DE has a set of hyper-parameters to be assigned with proper values. We also need to specify cost weight values $\mathbf{m} = [m_0, m_1, m_2]$. An approach for determining the optimal set of hyper-parameters is to conduct a grid search, exploring various combinations of all possible values and evaluating the performance of the DE algorithm accordingly. Nevertheless, employing a grid search for identifying the best hyper-parameter set is impractical due to its complexity and time-consuming nature, making the optimization process inefficient. In order to reduce the complexity, we set some of the DE hyper-parameters values such as the crossover probability and the mutation factor, as the standard values $CR = 0.5$ and $F = 0.8$, normally used for other applications [64].

During our simulation procedures, we assign a higher weight value to the power excursion cost function $J_0(\mathbf{p})$ weight, which is m_0, compared to the other two cost function weights. Nevertheless, it is worth noting that various weight configurations can be contemplated in this optimization process. We explore three distinct scenarios with different values of $\mathbf{m} = [m_0, m_1, m_2]$ to examine their respective impacts.

In the first case scenario, we set $\mathbf{m}^{(1)} = [1, 0, 0]$ where the objective is only to minimize the power excursion J_0. In the second scenario, we set $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$ where the objective is to minimize both the J_0 and the spectrum excursion J_1 having J_0 twice the impact compared to the impact of J_1. In the final scenario, we set $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$ where the objective is to minimize J_0 , J_1 and 0-dB gain variation J_2 giving equal impact to J_1 and J_2, each one with quarter impact of J_0.

Fig. 5.3 shows the results for the first case scenario where $\mathbf{m}^{(1)} = [1, 0, 0]$. Fig. 5.3 (a) represents the resulting power evolution of all channels over the fiber distance after applying the optimal pump power values found by the DE. Fig. 5.3 (b) shows the optimal pump power values of co- and counter-propagating pumps operating at different frequencies. Fig. 5.3 (c) represents the evolution of the cost value J_0 over the number of the DE iterations. The final power excursion value for the resulting 2D profile is $J_0 = 2.83$ dB.

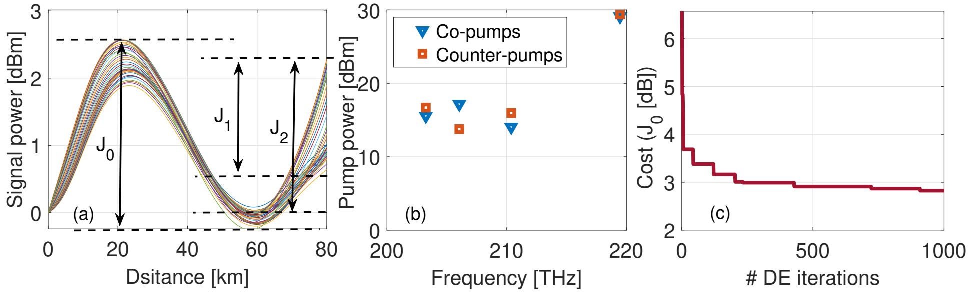


Fig. 5.3: Results for 2D flat power profile design using DE framework with $\mathbf{m}^{(1)} = [1, 0, 0]$. (a) Resulting 2D power profile. (b) The optimized pump power values for different pump frequencies. (c) The evolution of the cost value over the number of DE iterations.

In the previous case scenario, we observe that the resulting 2D power evolution, despite its low power excursion, has a high spectrum excursion of 1.64 dB, occurring at the end of the fiber. Regarding this, in the second design scenario, we have applied $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$, where the focus is on minimizing both power excursion J_0 and spectrum excursion J_1, still considering higher weight value for J_0. Similar to the previous case, have shown the results in Fig.5.4. Fig.5.4 (a) illustrates the 2D power evolution profile over the fiber distance, in which the pump power values are optimized to minimize both power excursion and spectrum excursion. In the resulting profile, the power excursion and spectrum excursion are $J_0 = 3.04$ dB and $J_1 = 0.82$ dB, respectively, asserting that the optimization framework has significantly improved the spectrum excursion (decreased from 1.64 dB to 0.82 dB) with a slight increase in power excursion (from 2.83 dB to 3.04 dB). Fig.5.4 (b) shows the resulting pump power values for this case. The evolution of cost over the number of DE iterations is shown in Fig.5.4 (c), asserting no considerable improvement in cost value after almost 400 DE iterations.

In the previous two case scenarios, we set $m_2 = 0$, meaning that we have not considered the 0-dB gain cost function $J_2(\mathbf{p})$ in the last

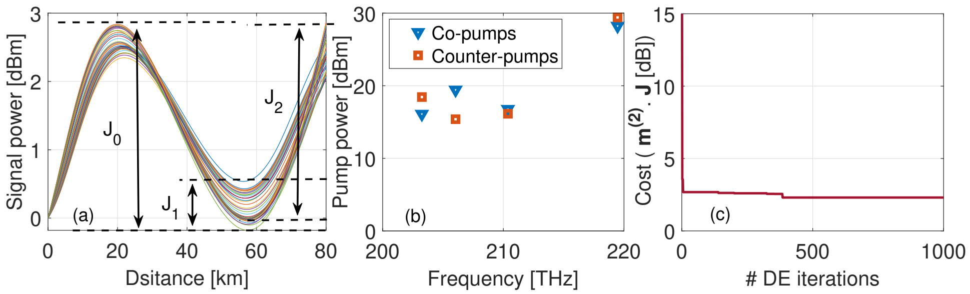


Fig. 5.4: Results for 2D flat power profile design using DE framework with $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$. (a) Resulting 2D power profile. (b) The optimized pump power values for different pump frequencies. (c) The evolution of the cost value over the number of DE iterations.

optimization processes. Considering the first and second scenario results in Fig.5.3 and Fig.5.4. It can be observed that for the first and second cases, the 0-dB gain cost value is $J_2 = 2.28$ dB and $J_2 = 2.86$ dB, which can be relatively high values for different applications. Therefore, in our third design case, we also take into account the minimization of the last cost function $J_2(\mathbf{p})$ by setting $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$, considering higher weight value for J_0, and lower but both equal weights for J_1 and J_2. Fig.5.5 (a) shows the 2D power evolution profile by targeting all three cost functions simultaneously resulting in cost values $J_0 = 3.11$ dB, $J_1 = 0.96$ dB and $J_2 = 1.18$. These results show significant improvement in J_2 with a slight increase in J_0 and J_1 values compared to the last two case scenarios. Moreover, Fig.5.5 (b) shows the resulting pump power values and Fig.5.5 (c) illustrates the evolution of the error over the number of DE iterations, asserting no further improvement after 300 iterations.

The cost values for all three cases are summarized in Table 5.2.

Tab. 5.2: Cost values achieved by DE for a 2D flat target profile.

Design case	DE with $\mathbf{m}^{(1)}$	DE with $\mathbf{m}^{(2)}$	DE with $\mathbf{m}^{(3)}$
J_0	**2.82** dB	3.04 dB	3.11 dB
J_1	1.63 dB	**0.82** dB	0.96 dB
J_2	2.28 dB	2.86 dB	**1.18** dB

It is worth noting that the DE framework initializes the population with a random sampling of the power values within the pre-specified

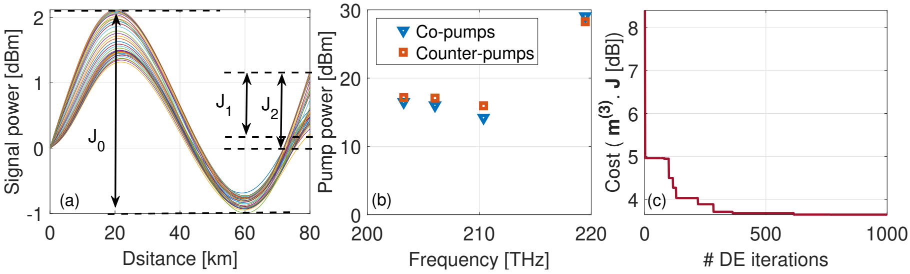


Fig. 5.5: Results for 2D flat power profile design using DE framework with $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$. (a) Resulting 2D power profile. (b) The optimized pump power values for different pump frequencies. (c) The evolution of the cost value over the number of DE iterations.

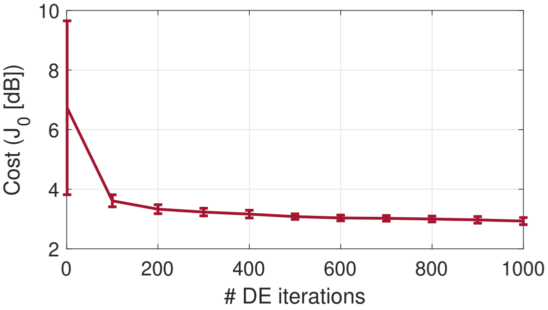


Fig. 5.6: Average cost (plot line) and standard deviation (error bars) for the DE framework with $\mathbf{m}^{(1)} = [1, 0, 0]$ for a 2D flat profile design.

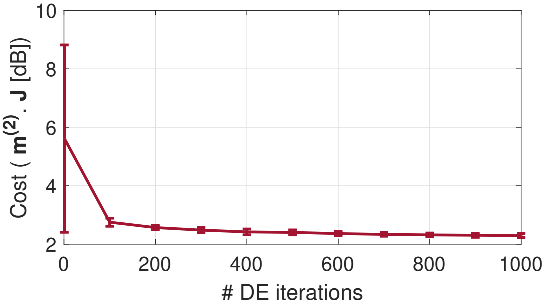


Fig. 5.7: Average cost (plot line) and standard deviation (error bars) for the DE framework with $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$ for a 2D flat profile design.

ranges. Therefore, the quality of the final point will be very sensitive to initialization. This means that the closer the sampled individuals are to

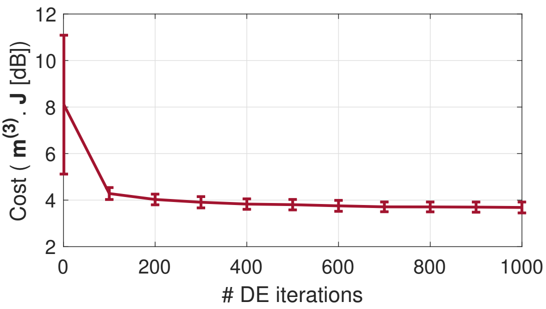


Fig. 5.8: Average cost (plot line) and standard deviation (error bars) for the DE framework with $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$ for a 2D flat profile design.

the optimal point, the faster the convergence will be, and furthermore, it will be less likely to get trapped in a local minimum. To investigate this, we conducted ten trials of the DE framework for each case scenario, using different initialization seeds. Then, we plotted the average error with error bars over the number of DE iterations, as shown in Fig. 5.6, 5.7 and 5.8, respectively. Comparing these figures, the third scenario with $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$ exhibits wider error bars, particularly at higher levels of DE iterations. This indicates the more uncertainty of its results and also its more likelihood to get trapped in a local minimum.

5.3.3 Simulation results: 2D symmetric power profile

The literature has primarily focused so far on addressing symmetric power evolution in distance within single-channel narrow-band amplification scenarios [47, 48]. The primary objective of designing symmetric power evolution is to minimize the asymmetry factor $A(f)$ at a specified frequency f_0, either through numerical or experimental methods, as defined below [47]:

$$A(f_0) = \frac{\int_0^{L/2} |\mathbf{P}(f_0, z) - \mathbf{P}(f_0, L - z)| dz}{\int_0^{L/2} \mathbf{P}(f_0, z) dz} \tag{5.9}$$

where the power profile $\mathbf{P}(f, z)$ is defined in linear scale [mW]. Extending the minimization model to a wide-band scenario, we define the proposed cost value J as the highest level of asymmetry among all the frequency channels within the bandwidth as follows:

$$J = \max_{f}(A(f)) \tag{5.10}$$

in which $A(f)$ and consequently J value, can be formulated as a function of the pump power values as the following:

$$J(\mathbf{p}_p) = \max_{f}\left[\frac{\int_0^{L/2} |\mathbf{P}(f,z|\mathbf{p}_p) - \mathbf{P}(f,L-z|\mathbf{p}_p)|dz}{\int_0^{L/2} \mathbf{P}(f,z|\mathbf{p}_p)dz}\right] \tag{5.11}$$

and the optimization problem is formulated as:

$$\mathbf{p}_p^* = \underset{\mathbf{p}_{pump}}{\arg\min}\ J(\mathbf{p}_p) \tag{5.12}$$

$$\text{such that} \quad \mathbf{p}_{LB} \leq \mathbf{p}_p \leq \mathbf{p}_{UB}$$

After formulating the optimization problem, we utilize the DE framework to determine the optimal values of pump powers that lead to a symmetric power evolution profile along the fiber distance. The resulting power evolution profile is shown in Fig.5.9 (a) with the maximum asymmetry value of $J = 20.05$ % over all frequency channels. Fig.5.9 (b) shows the allocation of the pump power values for different co- and counter-propagating pumps. Moreover, Fig.5.9 (c) shows the evolution of the cost over the number of the DE iterations.

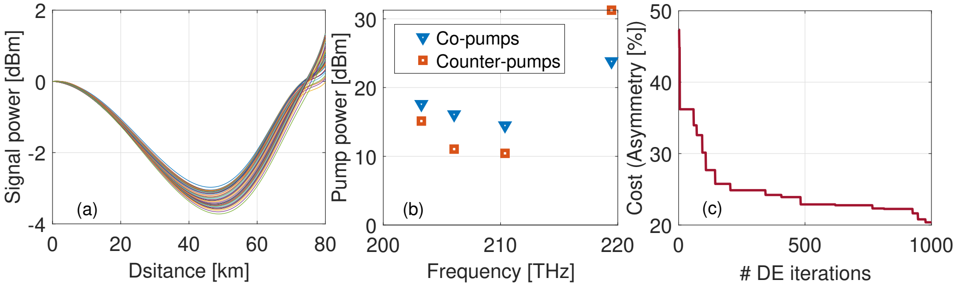


Fig. 5.9: Results of using DE framework for designing a 2D symmetric power profile. (a) Resulting power profile, (b) Resulting pump power values, (c) Evolution of the cost value over the number of DE iterations.

As discussed earlier, random initialization of the DE can affect its performance in terms of the quality of the results and the convergence speed. Similar to the 2D flat power evolution case, we have performed ten different trials of using the DE framework with different initialization seeds to optimize the pump powers to generate a symmetric power evolution. These trials are mainly performed to investigate how much it is probable to converge to different minimum values with different initial individuals. We have observed that some of the trials converge to a local minimum with very high cost (asymmetry) values. For example, Fig. 5.10 (a) shows a power evolution profile that has resulted from convergence to a local minimum. The maximum asymmetry value for this profile is $J = 32$ %. Fig. 5.10 (b) also asserts that there is no improvement in cost value after almost 100 iterations, which means that none of the individuals is initialized in the vicinity of the optimum value.

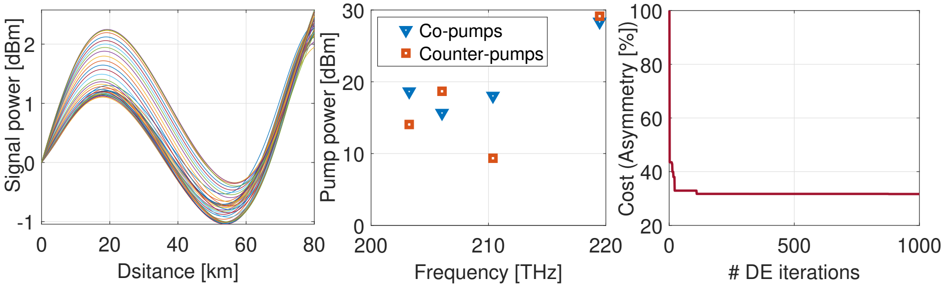


Fig. 5.10: Results of using DE framework for designing a 2D symmetric power profile trapped in local minima. (a) Resulting power profile, (b) Resulting pump power values, (c) Evolution of the cost value over the number of DE iterations.

Fig. 5.11 shows the average maximum asymmetry value plot with the standard deviation, represented as error bars, over the number of the DE iterations. This plot shows that the size of the error bars grows with increasing the DE iterations, asserting a higher probability of convergence into local minimum points every time we run the DE framework. Considering this, to have more reliable results, especially for the 2D symmetric power profile design, we will need to run the DE with different initialization and choose the trial with the lowest error as our final optimum point. This approach, besides providing a better solution, can be time-consuming and computationally expensive.

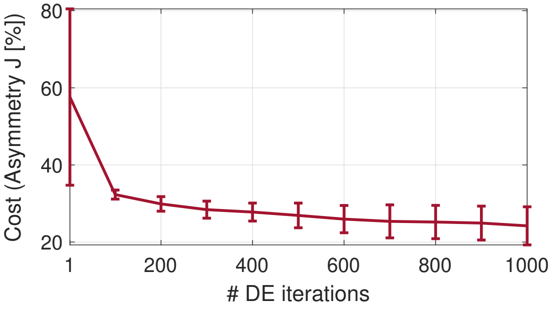


Fig. 5.11: Average cost (plot line) and standard deviation (error bars) of using the DE framework for a 2D symmetric profile.

5.4 CNN-assisted DE framework for 2D power evolution design

In the previous section we illustrated that the DE framework starts with a random initialization of the pump power values within the whole range of available power for designing a target 2D profile, This framework is proposed generally when there is no prior information provided to narrow the search region for the optimum set of pump power values. In this blind search approach, the framework would be very sensitive to the initial distribution of the candidates and liable to slow convergence and additionally, plunging into a local minimum. This uncertainty in results was shown in the last section, especially for a 2D symmetric profile in Fig. 5.10 and 5.11. The issues such as slow convergence and local minimum are more dominant when the cost function is more complex, and the number of dimensions in the search space, i.e. the number of pumps for our case, increases. In the case of higher dimensionality, the uncertainty in convergence to the optimum value decrease since the probability of the individuals taken in the vicinity of the optimum point will decrease.

In this section, we improve our chance to approach a more reliable and accurate pump power optimization for the proposed 2D power profile design problem by using the inverse CNN model as an initializer for the DE framework. As discussed in Chapter 4, the CNN model is a neural network trained offline, which learns the inverse mapping based on a data-set generated by random selection of the pump power values in the proposed amplification setup. This inverse model has shown statistically promising results in effectively mapping the achievable 2D power profiles to their corresponding pump power values while having not accurate results for practically challenging profiles. However, this model approximately predicts the pump power values, providing initial information for the solution, which can be fine-tuned using the DE framework to find the optimum pump power values.

The CNN prediction gives prior information about a space region of pump power values where a better solution to the optimization falls inside. By initializing the DE population based on the CNN results, we improve the convergence speed and also the quality of the final solution. Considering this, the fine-tuning process is performed with parameter constraints on a relatively narrower space region neighboring the set of values predicted by the CNN model.

The schematic of the proposed DE framework initialized with the CNN prediction which we refer to as *CNN-assisted DE* framework, is shown in Fig. 5.12. In this scheme, first, the 2D target power profile $\mathbf{P}^t(f, z)$ is used as the input to the pre-trained CNN model to predict the corresponding Raman pump power values as the initial solution $\mathbf{p}_p^{*'} = [p_1, ..., p_n]$. Afterward, the constraints $\mathbf{p}_{LB} = \mathbf{p}_p^{*'} - \boldsymbol{\Delta}_p \cdot \mathbf{p}_p^{*'}$ and $\mathbf{p}_{UB} = \mathbf{p}_p^{*'} + \boldsymbol{\Delta}_p \cdot \mathbf{p}_p^{*'}$ are defined, where $\boldsymbol{\Delta}_p = [\Delta_1, \Delta_2, ..., \Delta_n]$ is a hyper-parameter vector consisting of n number of unit-less scalar values. Each element of $\boldsymbol{\Delta}_p$ is multiplied by its corresponding initial pump power value in $\mathbf{p}_p^{*'}$. The resulting number for each pump power is considered as its amount of deviation and the constraints $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$, accordingly. After defining the $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$ values using the CNN prediction, the DE framework will perform the optimization, same as what was proposed in the last section (Fig.5.2).

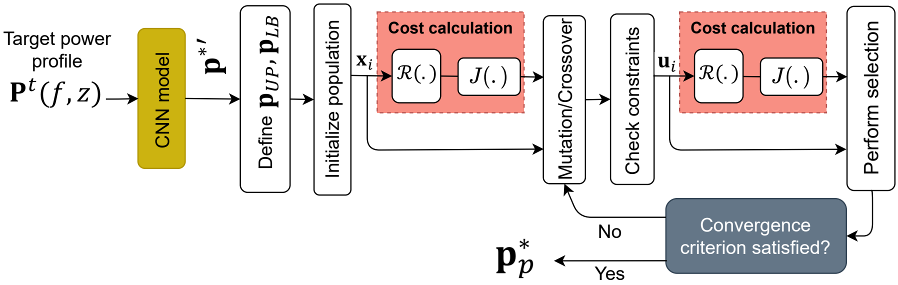


Fig. 5.12: Block diagram of the proposed CNN-assisted DE framework used for pump power optimization based on a specific 2D target power profile.

5.4.1 Simulation results: CNN model evaluation

In the proposed CNN-assisted DE framework, first, we need to train and evaluate the CNN model. To specify the proper size of training data, training sets with different sizes from 1000 to 5000 samples are used, and a validation set with 800 samples is used to assess the achieved accuracy across these training sizes. Validation MSE for datasets with different training size is shown in Fig. 5.13. Since there is not a considerable improvement in validation MSE error for training datasets with more than 3500 samples, we choose this value as the training data size for all further analyses.

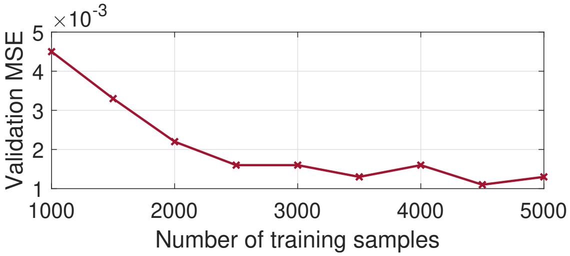


Fig. 5.13: CNN validation error the different training dataset sizes.

To measure the CNN accuracy, we use R-Squared (R^2) [65], which serves as our measurement tool to assess how well the trained network fits the data. R-Squared takes on values between 0 and 1, where a value of 1 signifies that the regression predictions perfectly explain the variations in the true values. Conversely, lower values of R^2 indicate that the regression model's outputs do not accurately vary in accordance with the true values. The R^2 values for the predicted pump powers in the test data set are presented in table 5.3. Notably, all pumps, except p_8, exhibit R^2 values higher than 0.90, indicating the good performance of the CNN in capturing the relationship between the 2D profiles and their corresponding pump power values in the proposed second-order amplification scheme.

Tab. 5.3: R^2 values of the pump power set for test data

Pump	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
R^2	0.98	0.93	0.99	0.97	0.95	0.96	0.95	0.86

As suggested in Chapter 4, an alternative approach to evaluating the accuracy of the trained CNN model involves measuring the MAE between the target and resulting 2D profiles. To calculate the MAE, we utilize the predicted pump power values for the ith test profile, apply them to the amplifier setup, and generate the corresponding 2D power profile $\hat{\mathbf{P}}^i(f,z)$. The MAE for the ith profile is computed as follows:

$$MAE^i[dB] = \max_{z,f} |\mathbf{P}^i(f,z) - \hat{\mathbf{P}}^i(f,z)| \tag{5.13}$$

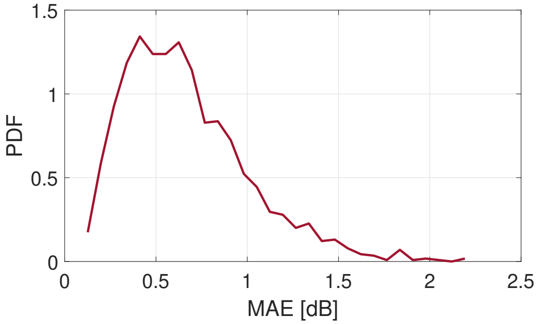


Fig. 5.14: PDF of the MAE values resulted from CNN inverse model prediction.

Here we perform the same process and calculate the MAE for all test 2D profiles. The probability density function (PDF) of the MAE for the test dataset is shown in Fig. 5.14. The low mean $\mu = 0.62$ dB and the standard deviation $\sigma = 0.33$ dB values for MAE also assert the good performance of the proposed CNN model for the prediction of the pump power values.

5.4.2 Simulation results: 2D flat profile

For the flat 2D power profile design, first we evaluate the CNN model performance by providing it with a 0-dBm flat 2D power profile input. More particularly, we use $\mathbf{P}^t(f,z) = 0$ dBm as the input to the CNN model to predict the pump power values and apply the predicted pump power values to the amplifier setup to measure the resulting 2D power profile. Fig. 5.15 (a) shows the 2D power evolution profile resulting from the CNN model prediction for a 0-dBm flat input with the predicted pump power values shown in 5.15 (b). All three cost values Eq. 5.5, 5.6 and

5.7 are specified on the resulted power profile with corresponding values of $J_0 = 3.58$ dB, $J_1 = 1.48$ dB and $J_2 = 0.97$ dB, respectively.

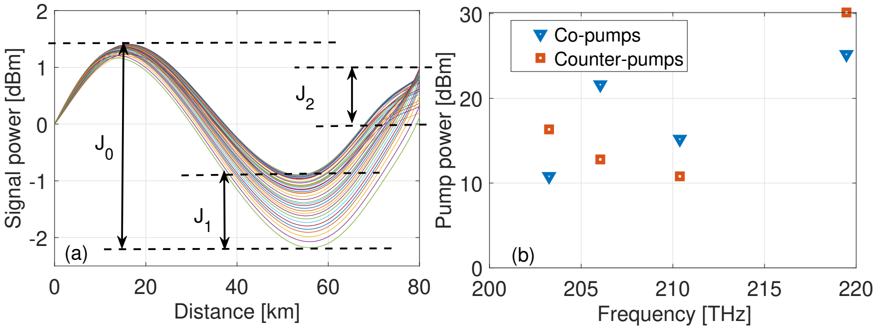


Fig. 5.15: CNN inverse model results for a flat 2D power profile. (a) Resulting 2D power profile. (b) Resulting pump power values.

The CNN model yields an approximate flat 2D profile, but there is potential for further improvement by fine-tuning the resulting pump power values using the online DE framework. Based on the CNN-assisted DE scheme proposed in Fig.5.12, the pump power values predicted using the CNN model $\mathbf{p}_p^{*'}$ are used to initialize the DE population. The population of the DE is initialized within the range between lower-bound $\mathbf{p}_{LB} = \mathbf{p}_p^{*'} - \boldsymbol{\Delta}_p \cdot \mathbf{p}_p^{*'T}$ and lower-bound $\mathbf{p}_{UB} = \mathbf{p}_p^{*'} + \boldsymbol{\Delta}_p \cdot \mathbf{p}_p^{*'T}$, in which $\boldsymbol{\Delta}_p = [\Delta_1, \Delta_2, ..., \Delta_{N_p}]$ represents the deviation of the pump power values around the CNN prediction $\mathbf{p}_p^{*'}$. Each element of $\boldsymbol{\Delta}_p$ is a scalar value corresponding to a pump power deviation. For each value of $\boldsymbol{\Delta}_p$, we use a consistent and fixed scaling factor for the first-order pumps, as well as the same value for the second-order pumps. Due to the amplification from the second-order pumps on the first-order ones, we assign a lower scaling factor value to the second-order pumps compared to the first-order ones. Since the pump p_5 has higher prediction value compared to the pump power value predicted for p_1, we consider 1.4 [W] as the maximum possible value and consequently $\Delta_1 = \Delta_5 = 0.35$. Addressing the first-order pumps, based on the highest value which is predicted for p_3, we consider more deviation to first-order pumps by setting their corresponding scaling factor value to 0.5. With this investigations, we finally set $\boldsymbol{\Delta}_p = [0.35, 0.5, 0.5, 0.5, 0.35, 0.5, 0.5, 0.5]$ (0.35 scaling value for the second-order pumps and 0.5 for the first-order pumps).

Once the lower-bound and upper-bound values are set according to the scaling factor $\boldsymbol{\Delta}_p$, the DE framework is applied to minimize the

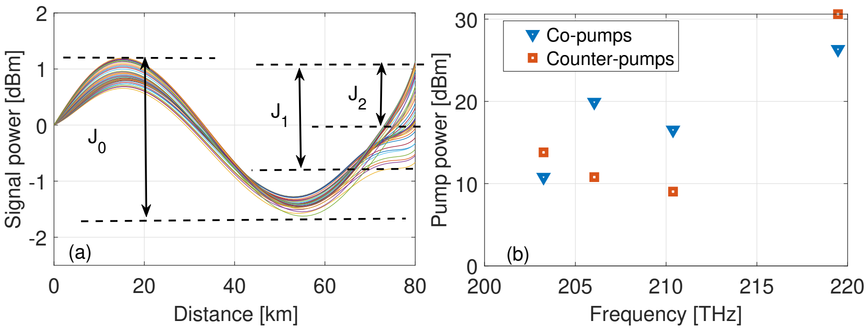


Fig. 5.16: CNN-assisted DE result for a flat 2D power profile, considering $\mathbf{m}^{(1)} = [1, 0, 0]$. (a) Resulting 2D power profile. (b) Resulting pump power values.

weighted sum of the cost values J_0, J_1 and J_2 according to solve the optimization scheme in Eq. 5.8. Similar to the last section's results, we have investigated three different optimization scenarios where three sets of weights $\mathbf{m} = [m_0, m_1, m_2]$ are applied and the results are investigated and compared. In the first scenario, we set $\mathbf{m}^{(1)} = [1, 0, 0]$ in which the cost value J_0 is aimed to be minimized, and the other cost values J_1 and J_2 are neglected. Fig.5.16 (a) shows the resulting power evolution profile for $\mathbf{m}^{(1)} = [1, 0, 0]$ with the corresponding cost values $J_0 = 2.81$ dB, $J_1 = 1.80$ dB and $J_2 = 1.14$ dB. The predicted pump power values for this case are shown in Fig.5.16 (b).

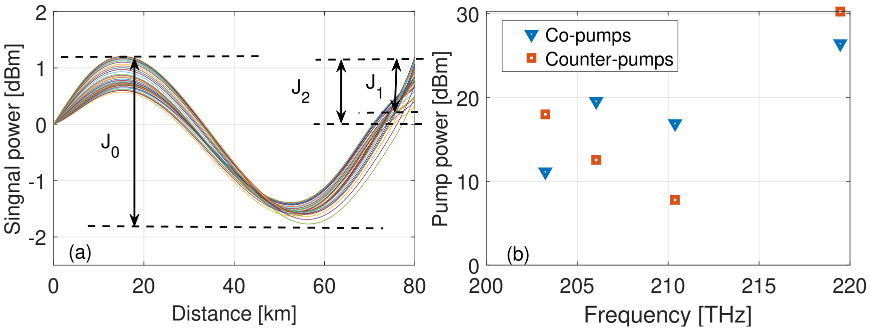


Fig. 5.17: CNN-assisted DE result for a flat 2D power profile, considering $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$. (a) Resulting 2D power profile. (b) Resulting pump power values.

In the second design scenario, we set $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$, meaning that both cost values J_0 and J_1 are minimized in the optimization process without involving the third cost value J_2. Fig. 5.17 (a) indicates the resulting power evolution profile and Fig.5.17 (b) shows the optimum pump power values for this design scenario. The cost values for this case are $J_0 = 2.97$ dB, $J_1 = 0.88$ dB, and $J_2 = 1.2$ dB, asserting the proposed framework's ability in minimizing the targeted cost value J_1 with a slight increase in power excursion cost J_0.

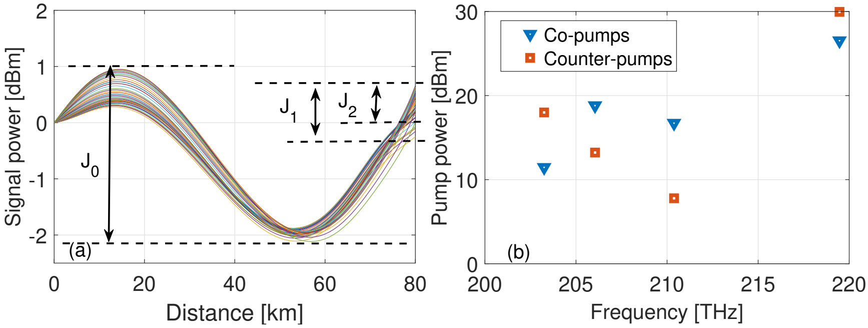


Fig. 5.18: CNN-assisted DE result for a flat 2D power profile, considering $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$. (a) Resulting 2D power profile. (b) Resulting pump power values..

In the third scenario, all cost values are attempted to be minimized with the corresponding weights $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$. The resulting 2D power profile for this case is shown in Fig.5.18 (a) and its resulting optimum pump power values are shown in Fig.5.18 (b). The cost values for this profile are $J_0 = 3.06$ dB, $J_1 = 0.9$ dB, and $J_2 = 0.65$ dB. These results assert a considerable improvement in the cost value J_2 with a slight compromise in the cost values J_0 and J_1. In order to provide a better comparison, the cost values for different case scenarios and their corresponding pump power values (in linear scale) are reported in Table 5.4 and Table 5.5, respectively.

Tab. 5.4: Cost function values achieved by CNN and CNN-assisted DE framework for a 2D flat input profile

Design case	CNN only	CNN+DE with $\mathbf{m}^{(1)}$	CNN+DE with $\mathbf{m}^{(2)}$	CNN+DE with $\mathbf{m}^{(3)}$
J_0	3.58 dB	**2.81** dB	2.97 dB	3.06 dB
J_1	1.48 dB	1.80 dB	**0.88** dB	0.90 dB
J_2	0.97 dB	1.14 dB	1.2 dB	**0.65 dB**

Tab. 5.5: Predicted pump power values by the CNN and CNN-assisted DE with different weights for 2D flat input power profile

model	CNN only	CNN+DE with $\mathbf{m}^{(1)}$	CNN+DE with $\mathbf{m}^{(2)}$	CNN+DE with $\mathbf{m}^{(3)}$
$p_1[mW]$	330	430	440	450
$p_2[mW]$	33	45	49	47
$p_3[mW]$	145	98	90	76
$p_4[mW]$	12	12	13	14
$p_5[mW]$	1030	1150	1060	990
$p_6[mW]$	12	8	6	6
$p_7[mW]$	19	12	18	21
$p_8[mW]$	43	24	63	63

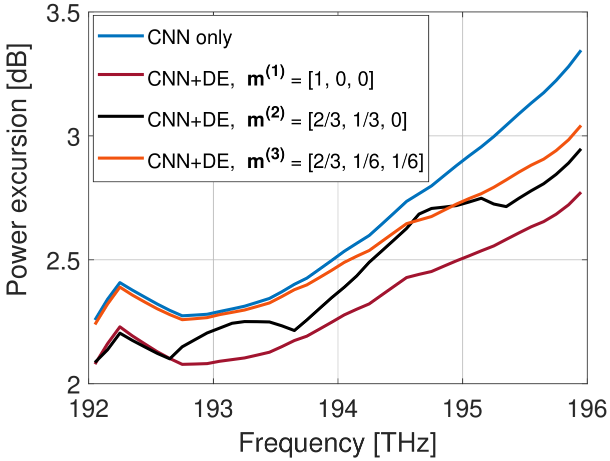


Fig. 5.19: Power excursion value computed as a function of frequency, shown for different pump power adjustment scenarios.

Besides the analysis provided so far for power excursion in the entire C-band and along the fiber distance, we will now examine the power excursion as a function of the signal frequency. This analysis aims to determine the frequencies that contribute the most to the power excursion. For each frequency channel in the previously proposed 2D profile design scenarios, the difference between the maximum and the minimum value of signal power level over the distance is calculated and shown in Fig. 5.19. This figure illustrates that the power excursion is generally increasing over the frequency for all scenarios. For all resulting profiles, the minimum power excursion which is less than 2.4 dB, is achieved almost at 192.7 THz. Additionally, The maximum value for all scenarios is measured at 196 THz. It is worth mentioning that the frequency-based power excursion value is less than or equal to the power excursion cost value J_0, calculated over the whole frequency-distance plane.

As pointed out in the last section (DE results), the DE framework without CNN-based initialization, has more potential to converge to a local minimum for different trials. To confirm this statement, we performed ten trials with random initialization of the DE and showed the evolution of the error with error bars over the number of DE iterations. To provide a visual comparison, we have also performed ten different trials for the CNN-assisted DE and showed the resulting error bars over the number of DE iterations together with the stand-alone DE error bars in Fig. 5.20. When comparing the results of the differential evolution (DE) with the CNN-assisted DE (referred to as CNN+DE), we observe that the CNN-assisted DE converges to the optimum point in approximately 100 iterations. This indicates a relatively faster convergence speed compared to the DE alone. Furthermore, it is worth noting that the error bars of the CNN-assisted DE are narrower compared to those of the DE alone. This implies greater reliability and a reduced likelihood of converging to a local minimum when CNN is employed for the DE initialization.

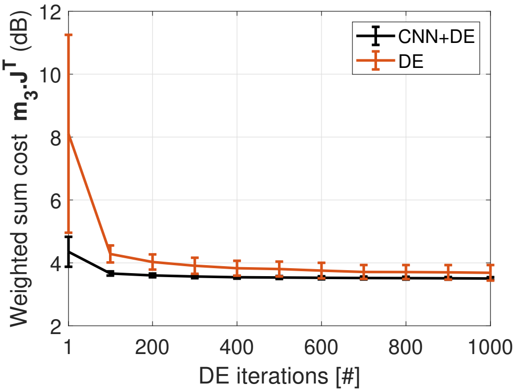


Fig. 5.20: Average cost (plot line) and standard deviation (error bars) for the CNN-assisted DE and the DE frameworks with random initialization for 2D flat profile design.

5.4.3 Simulation results: 2D symmetric profile

In the previous section, we utilized the DE framework also to design a 2D symmetric power evolution profile as a practically interesting profile used to optimize the OPC performance. As discussed, the symmetric 2D power profile was more prone to a local minimum when the pump power values are randomly selected between the minimum and maximum available power values. Here, we provide the simulation results on using the CNN-assisted DE for designing this practically desired power profile. Afterward, the CNN-assisted DE performance will be compared with the DE results presented in the previous section.

Firstly, we assess the CNN model's prediction for a symmetric power profile input. While the 2D flat profile is unique, there are multiple 2D profiles that can exhibit symmetry with respect to the midpoint in distance. In this case, we consider the second half-period of a sinusoidal signal as the target symmetric power profile for all channels. This signal is defined in logarithmic [dBm] scale with the following formulation:

$$\mathbf{P}^t(f, z) = 4sin(\pi z/L + \pi), \; \forall f, \; 0 < z < L \qquad (5.14)$$

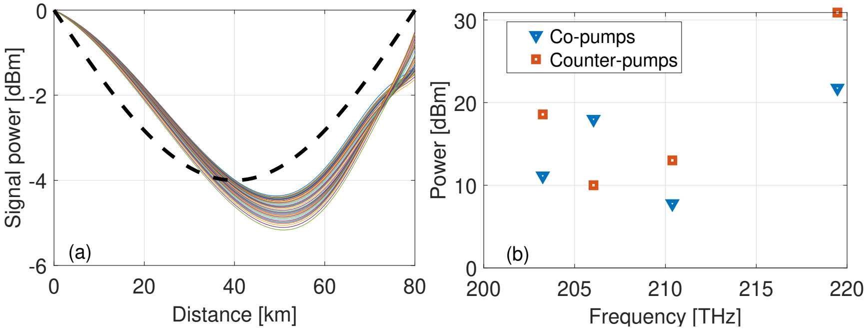


Fig. 5.21: CNN model results for designing 2D symmetric power profile. (a) Resulting 2D power profile. (b) Resulting pump power values at different frequencies.

This profile is considered the ideal target symmetric 2D profile for all channels. The predicted pump power values using the CNN model for the symmetric input target are applied to the amplification setup and the resulting power evolution value is generated. Fig. 5.21 (a) shows the resulting power evolution profile for the symmetric target input, represented with a dashed black curve. The resulting pump power values are depicted in Fig.5.21 (b). The cost value, which is the maximum asymmetry value among all frequency channels is 33.2 %.

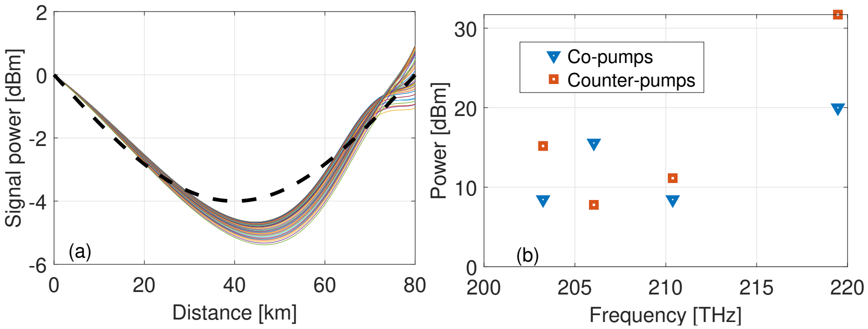


Fig. 5.22: CNN-assisted DE framework results for designing 2D symmetric power profile. (a) Resulting 2D power profile, together with the target profile shown as the dashed black curve. (b) Resulting pump power values at different frequencies.

In the next step, we use the CNN model prediction to initialize the DE population and perform the CNN-assisted DE framework. Fig. 5.22 (a) shows the generated 2D power profile after applying the optimum pump power values predicted by the CNN-assisted DE. The maximum asymmetry value for this profile is 14%, providing almost 19% improvement over the stand-alone CNN model. The pump power values for this profile are reported in Fig. 5.22 (b). Additionally, with a visual comparison of profiles in Fig. 5.22 (a) and Fig. 5.21 (a), we can conclude that the CNN-assisted DE is considerably better than the CNN for designing 2D symmetric profiles.

It is worth noting that the asymmetry cost value for the power evolution in Eq.5.11 is according to a power evolution profile in a linear [mW] scale. Therefore, a potential approach to visually illustrate the symmetry of the results could be to plot the resulting profiles in linear [mW] scale together with their reversed version in distance.

Fig.5.23 (a) and (b) show both profiles resulting from the CNN and the CNN-assisted DE (labeled as CNN+DE) with their corresponding reversed version in distance. According to this, we can also conclude that the pump power value prediction using the CNN-assisted DE is more accurate than the CNN model. The pump power values for both CNN and the CNN-assisted DE are reported in table 5.6.

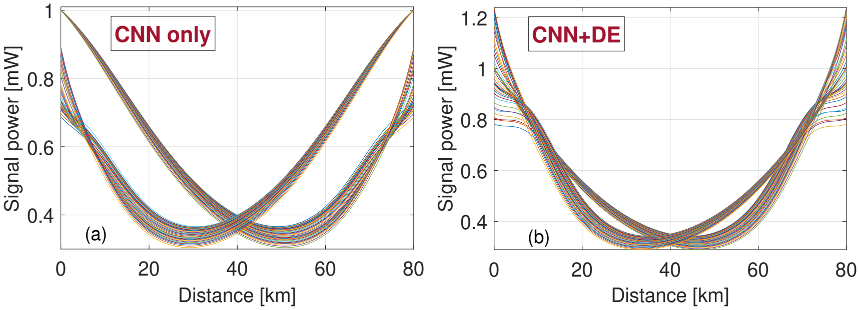


Fig. 5.23: Resulting 2D power profiles for (a) CNN, and (b) CNN-assisted DE, together with their reversed version over the fiber distance.

Tab. 5.6: Predicted pump power values using the CNN and the CNN-assisted DE framework for symmetric power evolution profile

Model	CNN only	CNN-assisted DE
$p_1[mW]$	150	100
$p_2[mW]$	6	7
$p_3[mW]$	63	36
$p_4[mW]$	13	7
$p_5[mW]$	1230	1480
$p_6[mW]$	20	13
$p_7[mW]$	10	6
$p_8[mW]$	72	33

In Fig. 5.24 the asymmetry values for both the CNN and the CNN-assisted DE are calculated for each channel and shown as a function of frequency. For both approaches the average asymmetry is represented using a dashed line with the same color as the main asymmetry values. According to this figure, the online DE framework improves the CNN asymmetry by more than 20%. In addition, the maximum asymmetry value for CNN is 33.2% which takes place at 195.9 THz and its minimum value is 28.5%, taking place at 192.1 THz. For the CNN-assisted DE, the maximum asymmetry is 14% at 195.7 THz, while the minimum value is 7.2%, taking place at 193.8 THz.

To investigate the impact of the CNN on the DE initialization (as illustrated in the previous section), we perform a set of simulations to

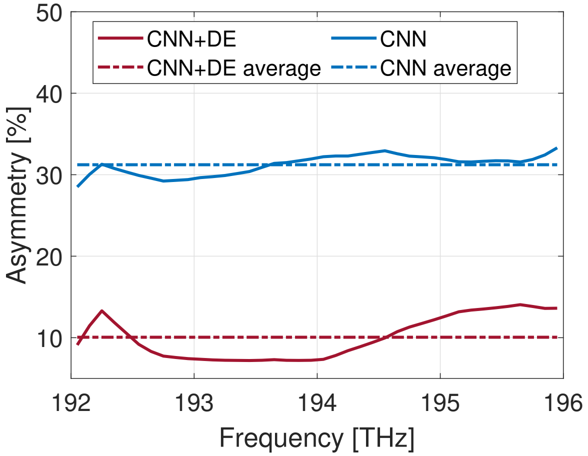


Fig. 5.24: Asymmetry values for CNN model and CNN-assisted DE framework over the signal frequency.

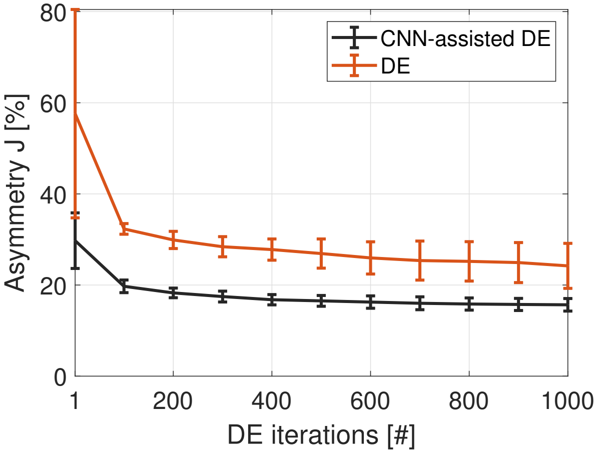


Fig. 5.25: Average cost (plot line) and standard deviation (error bars) for the CNN-assisted DE and the DE frameworks with random initialization for designing 2D symmetric profile.

compare DE and the CNN-assisted DE frameworks in terms of their convergence speed, the accuracy of the final results, and the proneness to a local minimum for designing a 2D symmetric power profile. Regarding this, we run both DE and the CNN-assisted DE ten times for designing a 2D symmetric profile and each time record the evolution of the asymmetry over the number of the DE iteration. Fig. 5.25 shows the line plot as the average error (asymmetry) value and with the error standard deviation (error bars) over the number of the DE iteration for both DE and the CNN-assisted DE. It illustrates that CNN initialization improves the DE error by more than 10%. Moreover, it is evident that the error bars for DE are much wider than those for the CNN-assisted DE, which confirms the greater reliability of the DE framework when initialized by the CNN. Hence, the probability of the DE converging to a local minimum is much higher when it is randomly initialized.

5.5 Summary

We have proposed an online framework for adjusting Raman pump power values to design 2D flat and 2D symmetric target profiles. This framework incorporates a DE technique that leverages the amplifier setup within an apply-feedback loop to dynamically optimize the pump power values. However, the DE optimization process can be slow and susceptible to local minima when starting with a random pump power initialization. To address this issue, we employ a pre-trained CNN model to provide the DE algorithm with an initial estimate of the pump power values. The CNN-based initialization has demonstrated improvements in the DE performance, both in terms of convergence speed and the accuracy of pump power predictions. This framework has proven to be effective for designing 2D flat and 2D symmetric profiles. Moreover, it offers the flexibility to be used in a new setup to design an arbitrary 2D profile with desired objectives.

6

Experimental validation of the spatial-spectral power evolution design using ML-enabled Raman amplifiers

6.1 Introduction

In the previous chapter, we presented and numerically verified an online machine learning (ML) framework, based on differential evolution (DE) that can be effectively used to design various target 2D power profiles. The proposed framework mainly aims at designing these 2D profiles by directly adjusting the pump power values in a Raman amplifier setup. In this chapter, we present a Raman amplifier setup and experimentally verify the performance of the CNN model, the DE framework, and the CNN-assisted DE framework for designing 2D power profiles. More specifically, the experimental validation consists of two phases as follows

- We validate the proposed ML-based frameworks by designing target 2D power profiles that are achievable using the amplifier setup under test. Achievable 2D profiles include 2D profiles that are generated by exciting the amplifier setup with randomly generated pump power values.

- We experimentally test the DE framework in a scenario where the target is to shape the 2D signal power evolution to satisfy multiple spatial-spectral objectives simultaneously. Two cost functions are aimed to be minimized during the online optimization process by tuning the Raman pump power values directly using the setup. Those cost functions are 1) the maximum deviation from a spectrally flat-gain profile at the end of the fiber (over the entire C-band), and 2) the maximum spectral power excursion along the fiber distance. This is a multi-objective optimization problem with non-differentiable cost functions with respect to the free parameters. Moreover, this optimization process is challenging to be solved with a gradient-based neural network (NN) model such as presented in [66].

6.2 Experimental setup

Fig.6.1 shows the schematic of the proposed experimental setup for validation of the ML-based framework to optimize the pump power values. We have investigated a standard single-mode fiber (SSMF) span with a length of 50 km. The Raman pump module consists of four counter-propagating pump lasers. The pump frequency values are fixed and shown with their maximum power value p_{max} in table 6.1, with the ability to amplify the entire C-band.

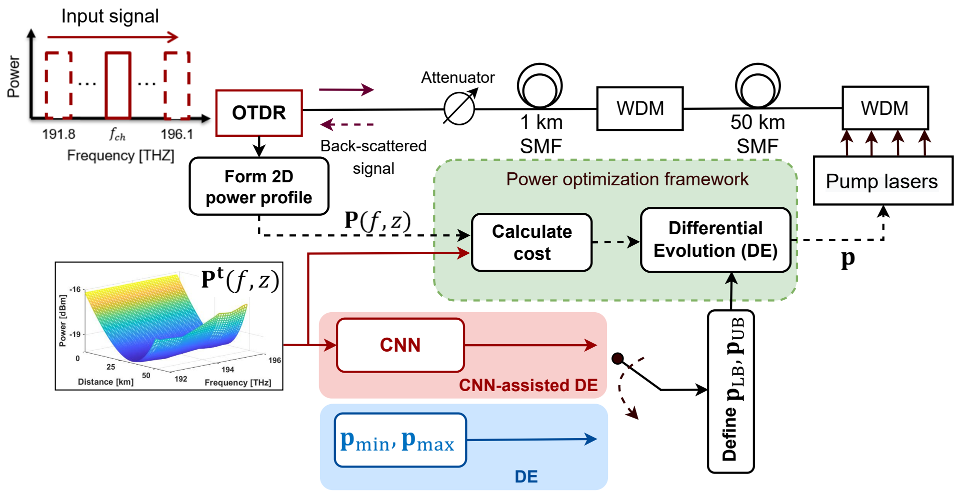


Fig. 6.1: The experimental setup and the block diagram of the framework used to optimize the pump powers values for designing a 2D target power profile $\mathbf{P^t}(f,z)$.

In the proposed setup, a frequency-tunable optical time-domain reflectometer (OTDR) is used to measure the signal power evolution over the spectrum and along the fiber distance. The signal bandwidth of the OTDR can cover the C-band with frequencies between 191.8 THz and 196.1 THz. In the OTDR we divide the C-band to $n_{ch} = 44$ number of channels with 100 GHz spacing. There are three elements between the OTDR and the fiber under test; a tunable attenuator, mainly used to control the signal power flow into the OTDR, a 1km SMF fiber used to cover the *dead-zone* of the OTDR, and a wavelength division multiplexer (WDM) which is employed to filter out the pump frequencies in the range between 203.9 and 211.1 THz. The reason for using this WDM is to filter the pump signals such that they do not enter the OTDR. A WDM coupler is also used at the end of the fiber to combine the signals and the pumps.

Tab. 6.1: The pump frequency values with their corresponding maximum pump power available.

Pump	p_1	p_2	p_3	p_4
$Frequency$ [THz]	210.6	209.0	207.5	206.1
p_{max} [dBm]	20.57	21.9	21.15	19.,94

The OTDR lunch signal power for each channel is -12 dBM.The attenuator adds a 4 dB loss to each channel power resulting in -16 dBm power per channel at the input of the 50 km fiber span. Once the signal is introduced into each channel, its back-scatter is measured, which is used by the OTDR to measure the actual signal power evolution inside the span. This process is performed sequentially, meaning that the signal power evolution is measured for one channel at a time until the total C-band is covered. It is worth noting that due to relatively low signal launch power, the signal-signal interaction due to the stimulated Raman scattering (SRS) effect can be neglected. Therefore the signal power evolution profile in the setup will look very much the same as the case where a full-load frequency comb signal is introduced to the fiber.

To reduce the noise and the power fluctuations in the measurements and to have a more clear trace for each channel, the signal pulse width of the OTDR is set to 3 μs, and the distance resolution is set to 8.2 m. Additionally, after the signal power is measured over the distance for each channel, a Savitzky-Golay smoothing filter [67] with window size $w = 19$

and polynomial order $n = 2$ in distance to reduce the further signal fluctuations. The window size of the smoothening filter is equivalent to 19×8.2=155.8 m in distance. According to the previous numerical results provided in Chapters 4 and 5, a 2D power evolution profile with a distance resolution of 500 m is sufficient for the CNN training and evaluation. Therefore, the smoothed traces in the experimental setup are down-sampled in distance using linear interpolation to achieve 500 m resolution. Consequently, a 2D power profile $\mathbf{P}(f, z)$ of size 44×100 is created. $\mathbf{P}(f, z)$ is used as the input to the cost calculation block, where its maximum absolute error (MAE) value with respect to the pre-defined target 2D profile $\mathbf{P}^{\mathbf{t}}(f, z)$ is calculated. After the cost calculation, the power optimization framework updates the pump powers and applies a new set of pump powers to the setup, in order to reduce the MAE in the next iteration. The process of applying pump powers, recording the resulting 2D profile, the cost calculation, and pump power value update continue until convergence criteria such as a minimum cost value without considerable variation in pump power values or a maximum number of iterations is achieved.

As proposed in Chapter 5, the main building block of the proposed pump power adjustment framework is the DE algorithm. This framework was numerically validated in Chapter 5 for designing 2D power profiles of practical interest. The DE aims to tune the pump powers in a closed-loop apply-feedback scenario to dynamically reduce the error between the target 2D power profile $\mathbf{P}^{\mathbf{t}}(f, z)$ and the resulting power evolution profile $\mathbf{P}^{(}f, z)$ in each iteration (for more details regarding the DE optimization please see Chapter 5).

The DE population can be initialized either with a random sampling of the pump power values within the minimum and maximum pump power ranges or based on an initial guess provided by an inverse mapping model such as the CNN. Considering this, in the experimental setup we have used two different approaches for the initialization of the DE by defining the lower-bound $\mathbf{p}_{LB}$ and upper-bound $\mathbf{p}_{UB}$ values for the pump powers. In the first approach, which we previously referred to as the CNN-assisted DE, the pre-trained CNN model is used first to predict the pump power values $\mathbf{p}'$ for the given target 2D power profile $\mathbf{P}^{t}(f, z)$. Next the lower-bound and upper-bound values are defined by considering $\mathbf{p}_{LB} = \mathbf{p}' - \boldsymbol{\Delta}_p \cdot \mathbf{p}'$ and $\mathbf{p}_{UB} = \mathbf{p}' + \boldsymbol{\Delta}_p \cdot \mathbf{p}'$, respectively. Similar to the numerical analysis in the previous chapter, we set $\boldsymbol{\Delta}_p = [0.5, 0.5, 0.5, 0.5]$, allowing all pump power values to deviate within 50% of the predicted values by the CNN model.

In the second scenario as we also have previously referred to as DE, the lower-bound and the upper-bound values are defined without having prior information and they essentially are the minimum $p_{min} = -5$ dBm and the maximum pump power p_{max} (according to table 6.1) values available, respectively.

6.3 Designing achievable 2D profiles

In this section, we apply both CNN-assisted DE and DE approaches to design 2D profiles that are achievable by the proposed experimental setup. The achievable test set consists of the 2D profiles that are generated by applying randomly selected pump powers to the system and measuring their resulting 2D power profile. The resulting 2D power profiles are then used to evaluate the pump power adjustment framework's performance. Therefore, first, we generate a test data set consisting of 500 samples of randomly selected pump power values and their corresponding 2D power profile measured by the experimental setup.

Before applying the proposed optimization frameworks, first we need to train and evaluate the performance of the CNN model. Regarding this, we have trained the CNN model using a training data set generated by a random selection of pump power values. It is worth noting that we also have checked that the samples in the training data set do not overlap the samples in the test data. The training data set consists of 4100 generated 2D power profiles and their corresponding pump power values, and 300 2D power profiles are used for the model validation in the training process.

After training the CNN model, we evaluate its accuracy using the 500 2D profiles in the test data set. In order to provide a visual intuition of the model's accuracy, we have plotted the scatter plot of the true versus the predicted pump power values for all four pumps in the setup in Fig.6.2. According to this figure, the CNN model shows a better performance in predicting the high pump power values (>10 dBm), which its main reason is the low influence of the pumps with low power values in shaping the 2D signal power evolution. For different pump power values the R^2 test score is measured and reported in table 6.2. As introduced in Chapter 4, the R^2 score is a metric that determines the goodness of fit for regression models and varies between 0 and 1, where a high value represents a better prediction [65].

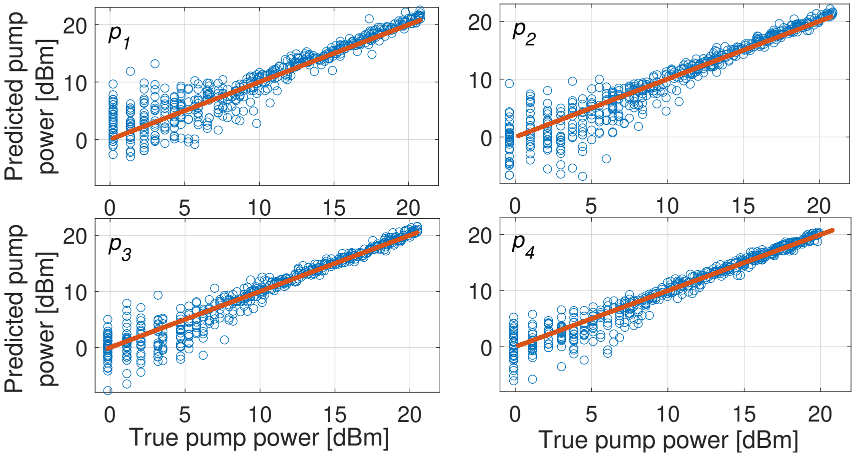


Fig. 6.2: The scatter plot of the true versus predicted pump power values using the CNN model on test data. Each blue dot corresponds to a test 2D profile and the orange solid line represents the ideal prediction.

Tab. 6.2: R^2 test scores for the CNN model prediction.

Pump	p_1	p_2	p_3	p_4
R^2	0.86	0.87	0.91	0.93

According to the values reported in table 6.2, we can see that in general, the CNN model has a good performance in predicting all four pump power values with slightly better performance in predicting the pump power values p_3 and p_4 (pumps operating at 207.5 THz and 206.1 THz), compared to p_1 and p_2 (pumps operating at 210.6 THz and 209.0 THz). The main reason that the CNN model is more accurate in predicting p_3 and p_4 values is due to their higher impact on the signal power evolution control in the C-band, as the peak of their corresponding Raman gain efficiency lies inside the proposed signal bandwidth (between 191.8 THz and 196.1 THz). However, the peak of the Raman gain efficiency for two pumps p_1 and p_2 lies slightly outside the signal bandwidth.

A more rigorous approach to evaluate the CNN model performance could be to apply its predicted pump power values to the amplifier setup, measure the resulting 2D profiles, and calculate the MAE between the

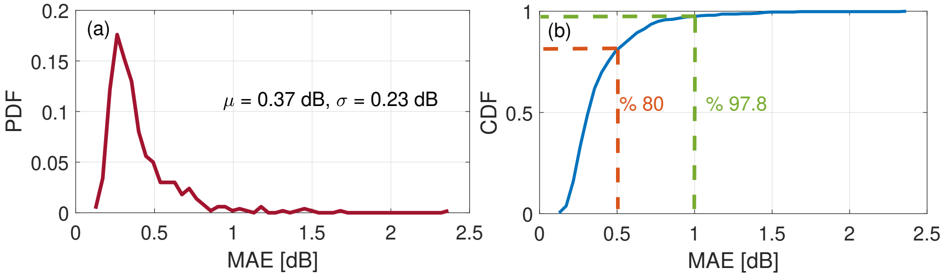


Fig. 6.3: The CNN model performance on test profiles. (a) PDF of the MAE, (b) CDF of the MAE.

target 2D profiles and the resulting profiles. Fig. 6.3 (a) shows the PDF of the MAE for the test data set resulting in an average MAE value $\mu = 0.37$ dB with the standard deviation $\sigma = 0.23$ dB. Additionally Fig. 6.3 (b) shows the cumulative distribution function (CDF) of the MAE for the test data set demonstrating that 80% of the test profiles result in MAE less than 0.5 dB, and almost 97.8% of the profiles (11 2D profiles out of 500 2D profiles) show MAE less than 1 dB.

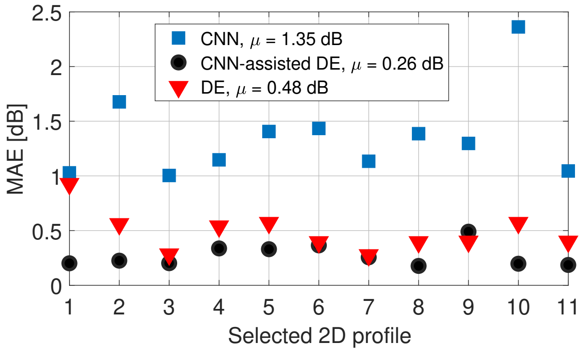


Fig. 6.4: The CNN model performance on test profiles.

Based on the results depicted in Fig.6.3, the CNN model demonstrates statistically low MAE values. However, it does not perform accurately for eleven 2D profiles with an MAE>1 dB. To address this issue, we employ the CNN-predicted pump power values as an initial guess and utilize the CNN-assisted DE framework to perform online optimization directly on the experimental setup. Additionally, we conduct a separate set of experiments for designing these eleven selected 2D profiles, employing the DE framework online without initialization using the CNN model. Subsequently, we compare the performance of the CNN-assisted DE framework with that of the DE framework alone.

In Fig.6.4, the MAE value for all eleven 2D profiles (starting with indexes from 1 to 11), is shown for the CNN model, the DE, and the CNN-assisted DE frameworks. According to this figure, the CNN model results in MAE>1 dB for all eleven 2D profiles with an MAE average equal to 1.35 dB. However, the CNN-assisted DE shows a considerable improvement for all profiles resulting in MAE<0.5 dB with an average of MAE equal to 0.26 dB. On the other side, the DE framework without CNN initialization, results in MAE<1 dB for all cases with an average equal to 0.48 dB. The obtained results demonstrate a promising improvement achieved by the CNN-assisted DE framework compared to both the standalone CNN model and the DE framework with random initialization.

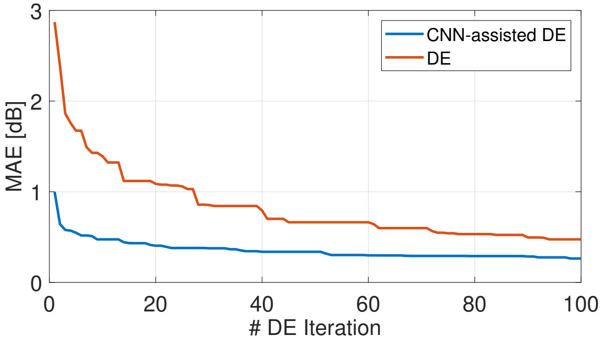


Fig. 6.5: The CNN model performance on test 2D power profiles.

Besides the better accuracy of using the CNN-assisted DE iteration over the DE framework, another performance investigation can be made by looking at the speed of convergence by measuring the evolution of the error over the optimization time. Considering this, we have recorded the MAE values over the number of the DE iterations for both CNN-assisted DE and the DE for the selected eleven 2D target profiles. Fig.6.5 shows the average MAE value for all eleven 2D profiles using both frameworks over the number of the DE iterations. This figure asserts that the CNN-assisted DE can converge to the optimal parameters with less number of DE iterations. More specifically, the CNN-assisted DE archives average 0.5 dB error after 5 iterations while the DE achieves this error after 90 iterations. The faster convergence can be a very beneficial characteristic for saving the experiment time, especially when different target 2D profiles are aimed to be designed.

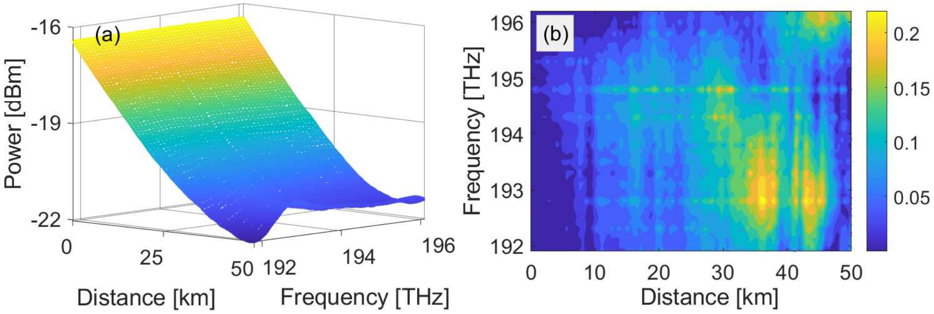


Fig. 6.6: CNN-assisted DE result for the 10th selected 2D profile in Fig.6.4.(a) Target 2D profile, (b) Heatmap of the absolute error (in dB) between the target and the predicted 2D profiles over the frequency and distance domains.

Moreover, a visual representation of the CNN-assisted DE framework result (10th selected 2D profile in Fig.6.4) is shown in Fig.6.6, providing the target 2D profile Fig. 6.6 (a) and the resulting heatmap of its absolute error with the predicted 2D profile over the frequency and fiber distance Fig. 6.6 (b). For this case, the resulting MAE value between the target and the predicted 2D profile is 0.22 dB.

6.4 Designing objective-based 2D profiles

In the previous section. we experimentally evaluated the performance of the CNN model, CNN-assisted DE, and the DE in designing 2D power profiles that are already generated by exciting the amplifier setup with random pump power values. In this section, we consider a different scenario in which the target is to design a 2D profile that fulfills one or multiple objectives rather than minimizing the MAE with the target 2D profile. More specifically, in the following scenario, we are aiming to experimentally design 2D profiles with flat gain levels at the end of the fiber, meanwhile, having minimum spectral excursion along the fiber.

This scenario can be modeled as a multi-objective optimization process, which is quite complex to be solved with the heuristic tuning of the pump power values. While we prove that the DE framework can solve it online directly by employing the experimental setup, it is worth noting that the CNN-assisted DE in Fig.6.1 cannot be practically applied for this case since there is no 2D target profile $\mathbf{P^t}(f, z)$ to be used as the input to the CNN model. Therefore, we only investigate the DE approach depicted in Fig.6.1 aiming to minimize two cost functions.

The primary goal is to achieve a 2D power evolution in the fiber distance that remains spectrally flat. To accomplish this, we introduce the first cost function $J_0(\mathbf{p})$, also known as the maximum spectral power excursion, which is formulated as follows and needs to be minimized.

$$J_0(\mathbf{p}) = \max_z[\max_f(\mathbf{P}(f, z|\mathbf{p})) - \min_f(\mathbf{P}(f, z|\mathbf{p}))]. \tag{6.1}$$

By minimizing $J_0(\mathbf{p})$, we can obtain a 2D profile with the least spectral excursion across the entire fiber distance. To better understand the value of J_0, Fig. 6.7 illustrates a power evolution profile for all channels over distance, showing the maximum spectral excursion corresponding to the specified J_0 value. In this particular scenario, the maximum excursion takes place at the end of the fiber.

To address the second objective of attaining a spectrally flat target gain level, denoted as $\mathbf{g}^t(f)$, at the end of the span, we introduce the cost function $J_1(\mathbf{p})$. This cost function measures the maximum absolute deviation between the achieved gain $\hat{\mathbf{g}}(f, \mathbf{p})$ and the target gain level $\mathbf{g}^t(f)$ at the end of the fiber. Mathematically, it is formulated as follows:

$$J_1(\mathbf{p}) = \max_f |\hat{\mathbf{g}}(f, \mathbf{p}) - \mathbf{g}^t(f)|. \tag{6.2}$$

where L is the span length, and the on-off gain $\hat{\mathbf{g}}(f, \mathbf{p})$ is defined as:

$$\hat{\mathbf{g}}(f, \mathbf{p}) = \mathbf{P}(f, z = L|\mathbf{p}) - \mathbf{P}(f, z = L|\mathbf{p}_{off}). \tag{6.3}$$

where $\mathbf{P}(f, z|\mathbf{p}_{off})$ is the 2D power profile when all pumps are turned off. In Fig. 6.8, a target gain level is targeted and the J_1 value is specified according to a sample 2D profile's achieved gain at the end of the fiber.

Our objective is to minimize both $J_0(\mathbf{p})$ and $J_1(\mathbf{p})$ simultaneously by finding the optimal set of pump power values $\mathbf{p}^*$. In order to simplify the optimization process and have better control over the impact of each objective on the final result, we transform the multi-objective optimization into a weighted sum, as described in Chapter 4. To achieve this, we assign weights to each objective, as hyperparameters. Afterward, we

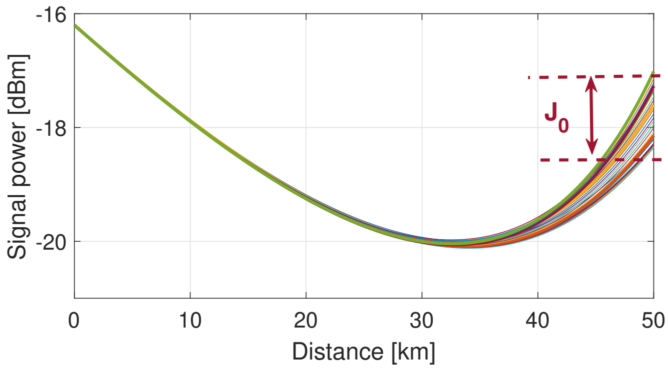


Fig. 6.7: Spatial representation of J_0,

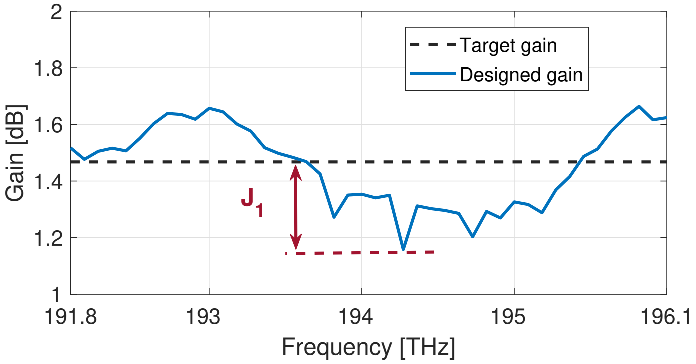


Fig. 6.8: Spectral representation of J_1.

multiply each objective by its corresponding weight and combine them using the following equation:

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} [m_0 J_0(\mathbf{p}) + m_1 J_1(\mathbf{p})], \tag{6.4}$$

$$\text{such that } \mathbf{p}_{LB} \leq \mathbf{p}_{pump} \leq \mathbf{p}_{UB},\ m_0, m_1 > 0,\ m_0 + m_1 = 1.$$

The weights m_0 and m_1 are used to determine how much the cost values J_0 and J_1 affect the optimization process. In this analysis, we choose to set both m_0 and m_1 as 1/2, indicating that we consider both cost functions to have an equal impact on the optimal point. To get an intuition on the highest approximate gain level provided by all four amplifiers, we adjust them all to their maximum values and measure the resulting gain, which results to be 4.9 dB. Therefore, it will not be realistic to expect the DE to design power profiles with a gain of more than 4.9 dB since the physics of the system limits the framework's ability to achieve this target.

In the proposed scenario, we aim to design five equally-spaced gain levels between 0.48 dB and 4.4 dB. Approaching this, we apply the DE framework for each case to solve the optimization process proposed by Eq. 6.4. Once the optimal pump power values are achieved, we apply them to the experimental setup to generate the resulting 2D power profile. Fig. 6.9 (a) shows all five power evolution profiles over the distance which are designed with different target gain levels, represented with different color codes. Moreover, the corresponding spectral gain is shown for the five resulting profiles in Fig. 6.9 (b). The resulting spectral gain profiles are represented in solid curves with the same corresponding color codes presented in Fig.6.9 (a). The dashed line represents the target gain level for each profile.

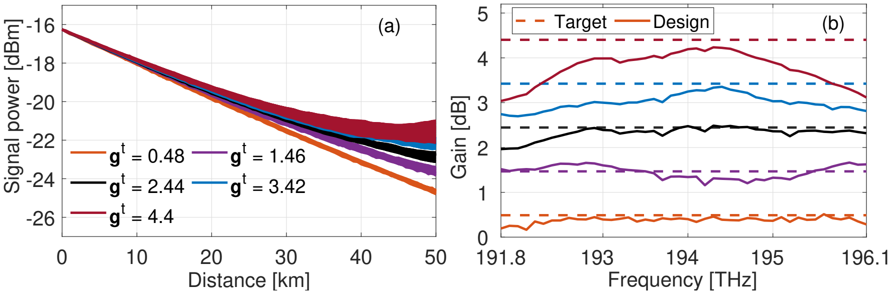


Fig. 6.9: DE results of pump power optimization by solving Eq.(6.4) with different target gain levels. (a) Spatial representation of the resulting power evolution profiles over the distance. (b) Spectral representation of target and designed gain levels at span end.

Fig.6.10 shows the average weighted sum cost values (i.e. $1/2J_0 + 1/2J_1$) for all five 2D power profiles, over the number of the DE iterations. According to this plot, the DE framework, on average, converges to the minimum after almost 40 iterations.

Additionally, in Fig.6.11, both cost values J_0 and J_1 are shown for different target gain levels, demonstrating less than 1 dB gain deviation

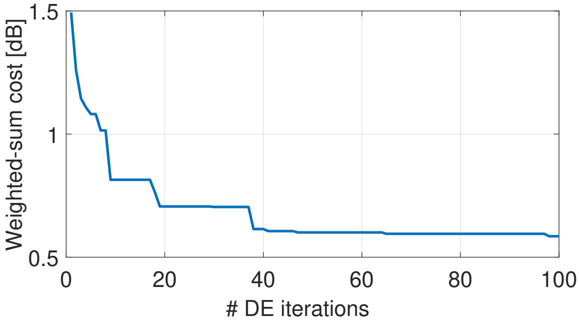


Fig. 6.10: Average cost value over the number of DE iterations for all five 2D target profiles.

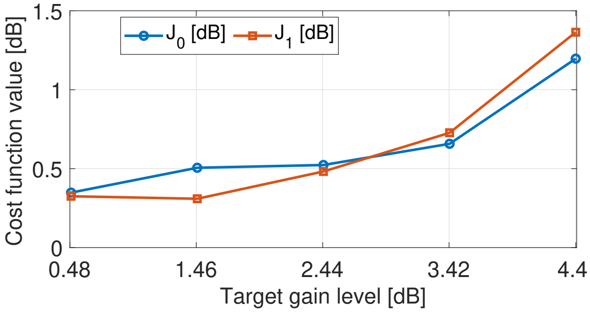


Fig. 6.11: Cost values J_0 and J_1 for different 2D power profiles with flat targeted gain levels at the end of the fiber.

for target gain values less than 4 dB. However, we can see a noticeable increase in cost values for the target gain value $\mathbf{g}^t = 4.4$ dB. This is mainly because of the upper-bound limitation of the pump power values provided by the setup, rather than the failure of the DE framework in fine-tuning the pump power values. Regarding this statement, we have shown the resulting pump power values for all four pumps at different target gain levels in Fig.6.12. According to this figure, it can be particularly seen that the value of the pump power p_4 increases with the target gain level and reaches the maximum at $\mathbf{g}^t = 3.42$, with a tendency to go higher for the target gain level $\mathbf{g}^t = 4.4$. However, since the pump does not provide higher power values, therefore, the framework will be physically limited by the setup and as a result, its performance will be affected.

We expect that the model performance would improve if the pumps could provide more power values. To prove the correctness of this statement, we perform a set of simulations to numerically emulate the

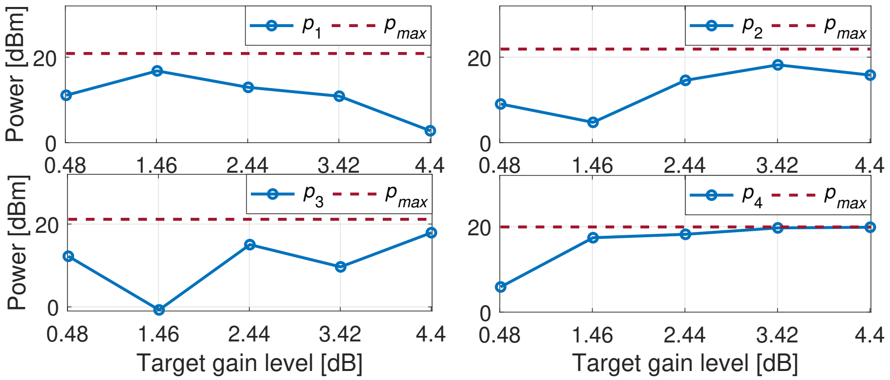


Fig. 6.12: Resulting pump power values in the experimental analysis for different target gain levels.

experimental setup provided in Fig.6.1. The numerical setup has the same configuration as presented in Fig.6.1, except that all pumps provide higher power values of up to 23 dBm each. In addition, since the range of space to explore with the new pump power ranges has a higher volume compared to the experimental analyses, we set the number of the DE iterations to 300 (higher than 100 iterations in the experimental analysis).

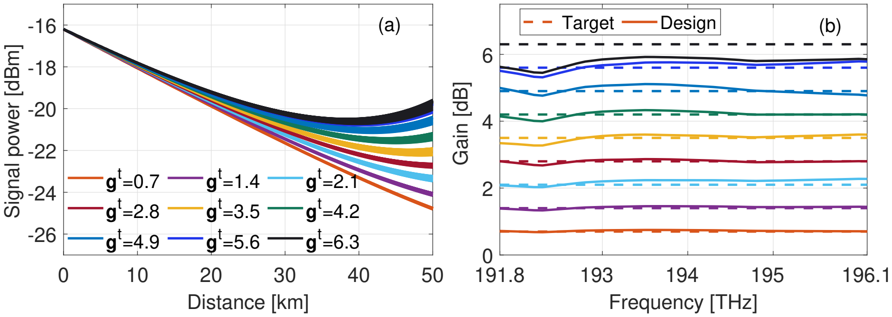


Fig. 6.13: Simulation results of pump power optimization by solving Eq.(6.4) with different target gain levels. (a) Spatial representation of the resulting power evolution profiles over the distance. (b) Spectral representation of target and designed gain levels at the end of the fiber.

Once the amplifier setup is emulated, the DE framework is applied to solve the optimization process in Eq.6.4. In this simulation analysis, we target nine 2D profiles with different equally-spaced flat gain levels at

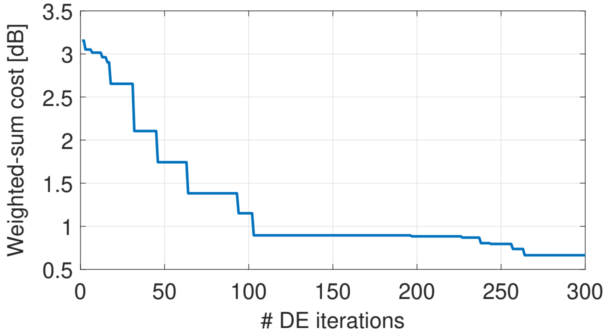


Fig. 6.14: Average cost value over the number of DE iterations for all five 2D target profiles.

the end from $\mathbf{g}^t = 0.7$ dB and up to $\mathbf{g}^t = 6.3$ (slightly higher than 60% merit of Raman pumping as proposed in [68]). Fig.6.9 (a) shows all of the nine 2D power profiles generated with numerical simulations with their corresponding target gain level values $\mathbf{g}^t$. Fig.6.9 (b) illustrates the designed gain levels (solid curves) with their corresponding target flat gain levels (b). Moreover, Fig.6.14 illustrates the average weighted sum cost value for all nine 2D profiles over the number of DE iterations, asserting that the framework converges after almost 260 iterations.

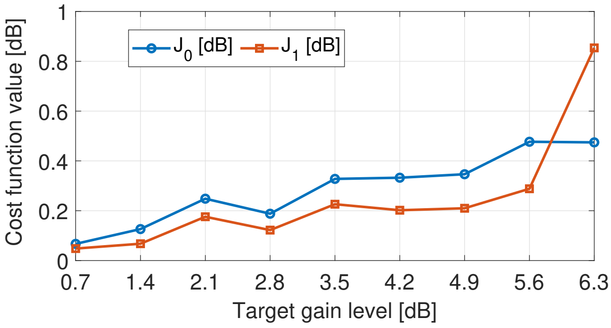


Fig. 6.15: Cost values J_0 and J_1 for different 2D power profiles with flat targeted gain levels at the end of the fiber.

The resulting cost values J_0 and J_1 are shown in Fig.6.15 over the different target gain levels, with a value less than 1 dB for all target profiles. More particularly, the cost values are slightly increasing for different gain levels from $\mathbf{g}^t = 0.7$ to $\mathbf{g}^t = 5.6$. However, the spectral gain deviation J_1 shows more than almost 0.6 dB increase for $\mathbf{g}^t = 6.3$ (from J_1 =0.24 dB to J_1 =0.85 dB). Similar to the experimental setup

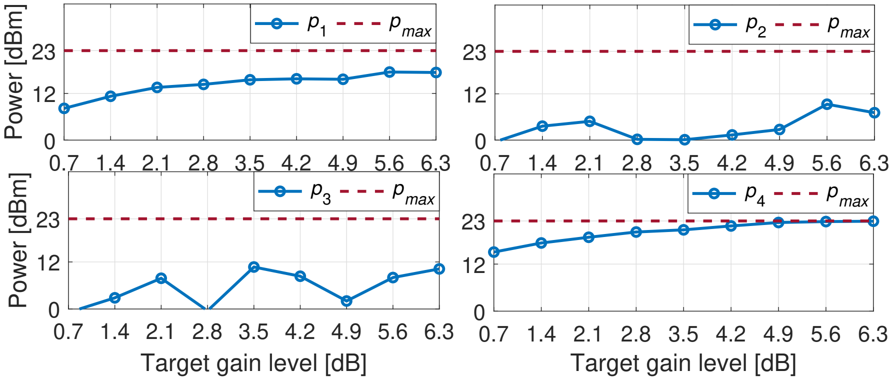


Fig. 6.16: Resulting pump power values in the simulation analyses for different target gain levels.

analysis, we believe the main reason is that the pump power p_4 has already reached its maximum available power level at 23 dBm with no possibility of an increase for gains higher than 5.6 dB. To clarify this, we have plotted all pump power values for different target profiles over the target gain levels in Fig.6.16. This plot confirms that for all cases the pump p_4 has an increasing trend as it reaches the maximum power value at the gain level $\mathbf{g}^t = 5.6$ dB with no possibility to increase further (higher than 23 dBm) for the target gain level $\mathbf{g}^t = 6.3$ dB.

6.5 Summary

We experimentally validated the CNN model, the DE, and the CNN-assisted DE frameworks for designing 2D power evolution profiles in Raman amplifiers. The results demonstrate that the CNN model achieves an average Mean Absolute Error (MAE) of less than 0.4 dB for the test 2D profiles, with an accuracy rate of 97.8%. For the remaining 2.2% of profiles where the CNN model is inaccurate, the CNN-assisted DE approach is applied to fine-tune the pump power values. This refinement process leads to an average improvement of over 1 dB compared to the results obtained solely from the CNN model. In another scenario, the DE framework is utilized to design 2D profiles with two distinct objectives: 1) achieving flat gain levels and 2) minimizing spectral power deviation. The results demonstrate the effectiveness of the proposed frameworks in designing 2D profiles through online tuning of the pump power values within a test amplifier setup.

7

Conclusion and future work

Distributed Raman amplifiers (DRAs) have emerged as a solution for enhancing signal quality and increasing the capacity of optical communication systems. This amplification scheme, besides its practical advantages, encounters challenges in solving the optimization of the pump power and their frequency value due to the increasing number of parameters in current optical networks. Furthermore, to align with future autonomous system applications where the quality of service (QoS) is continuously monitored, the utilization of precise optimization techniques can save time and prevent sub-optimal fine-tuning of various components, especially DRAs. Machine learning (ML) models, due to their ability to solve complex optimization challenges, have shown to be a promising tool in improving the performance of different optical communication components. In this thesis, we numerically and experimentally verified the ability of ML and optimization techniques to optimize the Raman amplifiers to shape the signal power evolution in frequency and fiber distance. We optimized the Raman pump parameters in different scenarios where a desired signal power evolution profile was aimed to be achieved in an amplification setup under test. This chapter provides a conclusion of the results and contributions presented in this work, along with an outline of future research perspectives.

7.1 Conclusion

The following results are considered as the outcome of this project:

- **Inverse DRA design model**: An offline inverse system model based on a Convolutional Neural Network (CNN) model is investigated, which primarily learns the mapping between the signal power profiles, defined in frequency and fiber distance, and their corresponding Raman pump power and wavelength values. This model has shown statistically good performance by providing low mean and standard deviation of maximum absolute error (MAE) values for the 2D test profiles that are achievable using the amplifier setup under the test.

- **Online optimization framework for DRA design**: Once the CNN model is trained, its parameters are fixed, and besides its statistically good performance, it lacks the potential for further improvement. Additionally, the CNN is not accurate enough for certain practical power profiles, such as 2D flat and 2D symmetric, which present challenges. Therefore, an online framework that can be combined with the physical model to fine-tune the pump parameters is practically advantageous. This framework saves time by efficiently adjusting DRAs parameters with a high degree of freedom and eliminates the need for heuristic tuning processes. More specifically, in this project, the proposed framework consists of the Differential Evolution (DE) technique, demonstrating its capability in optimizing pump power values for various 2D power profiles. The online DE framework is further integrated with the inverse CNN model to reduce uncertainty in the optimization results, mitigate sensitivity to the initialization process, and achieve faster convergence. This combination enhances the overall performance of the optimization process, leading to more reliable and efficient parameter tuning.

- **Experimental validation of the ML-based framework for DRA design**: The proposed CNN model, DE framework, and the combination of both, are experimentally evaluated in a Raman amplifier setup, verifying the effectiveness of the proposed models to be used in practical scenarios. The DE framework is used further to tune the pump power values in a scenario where two different objectives regarding the spectral gain and the power evolution profile shape are aimed to be fulfilled. This framework has shown the flexibility to be used for designing different profiles with various objectives.

7.2 Future work

Throughout this project, various aspects of Raman amplification in optical communication systems have been explored and interesting results have been obtained. Nevertheless, there is still potential for further research, with some details requiring additional work.

In Chapter 4, the primary objective of the numerical analysis was to address the inverse DRA design problem while assuming to have a fiber with fixed parameter values. However, in practical scenarios, different fiber types with diverse parameters, such as attenuation profile, effective area, and length, may be employed. In our proposal in Chapter 4, section 4.3, we introduced a length-aware CNN model to design 2D profiles for different fiber lengths. We anticipate that this model can be further expanded to incorporate additional fiber variables, such as attenuation profile and Raman gain efficiency. Having a single inverse DRA design model which includes the fiber parameters, can be beneficial, as it can reduce the complexity and increase the model's generalization capabilities. A previous study has utilized a Neural Networks (NN) model, so-called fiber-agnostic NN [69], to design gain profiles using Raman amplifiers for different fiber types, which aligns with our approach explained here.

Considering the impact of variations in fiber parameters, such as in attenuation profile, Raman gain coefficient, length, and effective area, on the signal power evolution due to the stimulated Raman scattering (SRS) effect, a potential approach to investigate can be the application of Raman amplifiers to characterize the fiber parameters by predicting their values or their perturbations. More particularly, we have started investigating an inverse model in which pump configuration and the gain profile are provided and the objective is to predict the fiber parameters or their perturbation values. With this approach, fiber parameters involved in SRS can be estimated with a few gain or power profile measurements after exciting the fiber with Raman amplifiers. The application of Raman amplifiers for characterizing fibers is currently in its preliminary stages of development.

Throughout the analyses conducted in this thesis, the signal bandwidth was constrained to the C-band. However, Raman amplifiers have demonstrated their potential in extending the available bandwidth beyond the C-band, such as C+L and S+C+L. Taking into account the

ML-based approaches in this project, a potential future work could involve extending this analysis to design signal power profiles in a broader bandwidth using Raman amplifiers.

Moreover, the quality of the designed 2D flat and 2D symmetric profiles in Chapter 5 can be assessed numerically and then experimentally in an optical communication system by evaluating parameters such as the signal-to-noise ratio (SNR) or the effectiveness of nonlinearity mitigation techniques like Optical Phase Conjugation (OPC). This investigation will provide valuable insights into the performance and applicability of the designed power profiles in practical optical communication systems.

The ML-based techniques proposed in this project were initially trained and evaluated separately for both numerical and experimental scenarios. It is important to highlight that the data acquisition process for the simulation analysis was significantly faster than the experimental analysis. Considering the time-consuming nature of the experimental analysis, leveraging the numerically trained models through *transfer learning* techniques becomes particularly advantageous. By applying transfer learning, the knowledge gained during the numerical training can be effectively transferred and utilized in the experimental setup. This approach enables the ML models to benefit from prior learning and accelerates their performance in the experimental context, improving efficiency and effectiveness.

At last, we have started investigating the application of using Raman amplifiers for compensation of Microresonator frequency combs (Kerr combs) in an optical communication scenario. Kerr soliton frequency combs have gained significant attention in recent years due to their potential applications in precision metrology, spectroscopy, and optical communications [70, 71]. For instance, a single-soliton frequency comb that exhibits a $sech^2$ envelope in the frequency domain will have a non-flat power at the end of the fiber. Having a power compensation technique to maintain a flat power level at the receiver side can be practically advantageous. Raman amplifiers, with their design flexibility, offer a potential solution for compensating for these frequency comb signals. In particular, we have explored a simulation scenario where different flat power levels are attained for a single-soliton frequency comb by optimizing the pump power and wavelength values of multiple Raman amplifiers. However, further investigations are still required to deepen our understanding of this area.

Bibliography

[1] C. Headley and G. P. Agrawal, *Raman Amplification in Fiber Optical Communication Systems*. Academic Press, 2005 (cited on pages 1, 2, 5–9, 23, 49).

[2] *Cisco annual internet report (2018–2023) white paper*, 2023. [Online]. Available: `https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html` (cited on page 1).

[3] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: The previous 20 and the next 20 years [invited]," *Opt. Express*, vol. 26, no. 18, pp. 24 190–24 239, Sep. 2018. DOI: `10.1364/OE.26.024190`. [Online]. Available: `https://opg.optica.org/oe/abstract.cfm?URI=oe-26-18-24190` (cited on page 1).

[4] B. J. Puttnam, G. Rademacher, and R. S. Luís, "Space-division multiplexing for optical fiber communications," *Optica*, vol. 8, no. 9, pp. 1186–1203, Sep. 2021. DOI: `10.1364/OPTICA.427631`. [Online]. Available: `https://opg.optica.org/optica/abstract.cfm?URI=optica-8-9-1186` (cited on page 1).

[5] R. H. Stolen, "Fundamentals of Raman amplification in fibers," in *Raman Amplifiers for Telecommunications 1: Physical Principles*, M. N. Islam, Ed. New York, NY: Springer New York, 2004, pp. 35–59, ISBN: 978-0-387-21583-9. DOI: `10.1007/978-0-387-21583-9_2`. [Online]. Available: `https://doi.org/10.1007/978-0-387-21583-9_2` (cited on pages 2, 5).

[6] R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010. DOI: `10.1109/JLT.2009.2039464` (cited on page 2).

[7] J. Bromage, "Raman amplification for fiber communications systems," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 79–93, 2004. DOI: `10.1109/JLT.2003.822828` (cited on pages 2, 5, 6).

[8] B. D. E. Desurvire D. Bayart and S. Bigo, *Erbium-Doped Fiber Amplifiers and Device and System Developments*. New York: Wiley, 2002 (cited on pages 2, 5).

[9] D. Zibar, M. Piels, R. Jones, and C. G. Schäeffer, "Machine learning techniques in optical communication," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442–1452, 2016. DOI: `10.1109/JLT.2015.2508502` (cited on page 2).

[10] D. Zibar, F. Da Ros, G. Brajato, and U. C. de Moura, “Toward intelligence in photonic systems,” *Opt. Photon. News*, vol. 31, no. 3, pp. 34–41, Mar. 2020. DOI: `10.1364/OPN.31.3.000034`. [Online]. Available: `https://www.optica-opn.org/abstract.cfm?URI=opn-31-3-34` (cited on page 2).

[11] W. S. Pelouch, “Raman amplification: An enabling technology for long-haul coherent transmission systems,” *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 6–19, 2016. DOI: `10.1109/JLT.2015.2458771` (cited on pages 5, 23).

[12] R. Sutherland, *Handbook of Nonlinear Optics* (Optical Science and Engineering). Taylor & Francis, 2003, ISBN: 9780824742430. [Online]. Available: `https://books.google.dk/books?id=ccXo3WrHp2UC` (cited on page 5).

[13] M. D. Pelusi and B. J. Eggleton, “Optically tunable compensation of nonlinear signal distortion in optical fiber by end-span optical phase conjugation,” *Opt. Express*, vol. 20, no. 7, pp. 8015–8023, Mar. 2012. DOI: `10.1364/OE.20.008015`. [Online]. Available: `http://www.opticsexpress.org/abstract.cfm?URI=oe-20-7-8015` (cited on pages 5, 24).

[14] J. D. Ania-Castanon, “Quasi-lossless transmission using second-order Raman amplification and fibre Bragg gratings,” *Opt. Express*, vol. 12, no. 19, pp. 4372–4377, Sep. 2004. DOI: `10.1364/OPEX.12.004372`. [Online]. Available: `http://www.opticsexpress.org/abstract.cfm?URI=oe-12-19-4372` (cited on pages 6, 24, 43, 44).

[15] J. D. Ania-Castañón, V. Karalekas, P. Harper, and S. K. Turitsyn, “Simultaneous spatial and spectral transparency in ultralong fiber lasers,” *Phys. Rev. Lett.*, vol. 101, p. 123 903, 12 Sep. 2008. DOI: `10.1103/PhysRevLett.101.123903`. [Online]. Available: `https://link.aps.org/doi/10.1103/PhysRevLett.101.123903` (cited on pages 6, 24, 25, 43, 44, 50, 51).

[16] M. A. Iqbal, M. A. Z. Al-Khateeb, L. Krzczanowicz, I. D. Phillips, P. Harper, and W. Forysiak, “Linear and nonlinear noise characterisation of dual stage broadband discrete Raman amplifiers,” *Journal of Lightwave Technology*, vol. 37, no. 14, pp. 3679–3688, 2019. DOI: `10.1109/JLT.2019.2919429` (cited on page 6).

[17] K. Rottwitt, J. Bromage, A. J. Stentz, L. Leng, M. E. Lines, and H. Smith, “Scaling of the Raman gain coefficient: Applications to germanosilicate fibers,” *J. Lightwave Technol.*, vol. 21, no. 7, p. 1652, Jul. 2003. [Online]. Available: `https://opg.optica.org/jlt/abstract.cfm?URI=jlt-21-7-1652` (cited on page 6).

[18] D. Hollenbeck and C. D. Cantrell, “Multiple-vibrational-mode model for fiber-optic Raman gain spectrum and response function,” *J. Opt. Soc. Am. B*, vol. 19, no. 12, pp. 2886–2892, Dec. 2002. DOI: `10.1364/JOSAB.19.002886`. [Online]. Available: `https://opg.optica.org/josab/abstract.cfm?URI=josab-19-12-2886` (cited on page 7).

[19] C. R. S. Fludger, “Linear noise characteristics,” in *Raman Amplifiers for Telecommunications 1: Physical Principles*, M. N. Islam, Ed. New York, NY: Springer New York, 2004, pp. 91–120, ISBN: 978-0-387-21583-9. DOI: `10.1007/978-0-387-21583-9_4`. [Online]. Available: `https://doi.org/10.1007/978-0-387-21583-9_4` (cited on pages 8, 11, 12).

[20] M. P. Yankov, F. D. Ros, U. C. de Moura, A. Carena, and D. Zibar, "Flexible Raman amplifier optimization based on machine learning-aided physical stimulated Raman scattering model," *J. Lightwave Technol.*, vol. 41, no. 2, pp. 508–514, Jan. 2023. [Online]. Available: `https://opg.optica.org/jlt/abstract.cfm?URI=jlt-41-2-508` (cited on page 10).

[21] U. C. de Moura, A. M. R. Brusin, A. Carena, D. Zibar, and F. D. Ros, "Simultaneous gain profile design and noise figure prediction for Raman amplifiers using machine learning," *Opt. Lett.*, vol. 46, no. 5, pp. 1157–1160, Mar. 2021. DOI: `10.1364/OL.417243`. [Online]. Available: `http://www.osapublishing.org/ol/abstract.cfm?URI=ol-46-5-1157` (cited on pages 11, 24).

[22] K. Rottwitt, A. Stentz, T. Nielsen, P. Hansen, K. Feder, and K. Walker, "Transparent 80 km bi-directionally pumped distributed Raman amplifier with second order pumping," in *Proc. European Conference on Optical Communication (ECOC'99*, 1999 (cited on page 13).

[23] J. Bouteiller, K. Brar, and C. Headley, "Quasi-constant signal power transmission," *2002 28TH European Conference on Optical Communication*, vol. 3, pp. 1–2, 2002 (cited on pages 14, 24).

[24] T. Ellingham, J. Ania-Castanon, R. Ibbotson, X. Chen, L. Zhang, and S. Turitsyn, "Quasi-lossless optical links for broad-band transmission and data processing," *IEEE Photonics Technology Letters*, vol. 18, no. 1, pp. 268–270, 2006. DOI: `10.1109/LPT.2005.862001` (cited on pages 14, 25, 44).

[25] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer New York, 2016, ISBN: 9781493938438. [Online]. Available: `https://books.google.dk/books?id=kOXDtAEACAAJ` (cited on page 17).

[26] C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018, ISBN: 9783319944647. [Online]. Available: `https://books.google.dk/books?id=AsTswQEACAAJ` (cited on pages 17, 41).

[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, `http://www.deeplearningbook.org` (cited on pages 18, 28).

[28] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *CoRR*, vol. abs/1511.08458, 2015. arXiv: `1511.08458`. [Online]. Available: `http://arxiv.org/abs/1511.08458` (cited on page 18).

[29] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6. DOI: `10.1109/ICEngTechnol.2017.8308186` (cited on pages 18, 19).

[30] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021 (cited on page 18).

[31] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: `10.1109/5.726791` (cited on page 20).

[32] L. Deng, “The MNIST database of handwritten digit images for machine learning research,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012 (cited on page 20).

[33] L. Bottou, F. E. Curtis, and J. Nocedal, “Optimization methods for large-scale machine learning,” *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018. DOI: `10.1137/16M1080173`. eprint: `https://doi.org/10.1137/16M1080173`. [Online]. Available: `https://doi.org/10.1137/16M1080173` (cited on page 21).

[34] J. Larson, M. Menickelly, and S. M. Wild, “Derivative-free optimization methods,” *Acta Numerica*, vol. 28, pp. 287–404, May 2019. DOI: `10.1017/s0962492919000060` (cited on page 21).

[35] T. Eltaeib and A. Mahmood, “Differential evolution: A survey and analysis,” *Applied Sciences*, vol. 8, no. 10, 2018, ISSN: 2076-3417. DOI: `10.3390/app8101945`. [Online]. Available: `https://www.mdpi.com/2076-3417/8/10/1945` (cited on pages 21, 46, 47).

[36] R. Storn and K. Price, “Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces,” *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, Dec. 1997 (cited on page 21).

[37] U. C. de Moura *et al.*, “Multi–band programmable gain Raman amplifier,” *J. Lightwave Technol.*, vol. 39, no. 2, pp. 429–438, Jan. 2020. [Online]. Available: `http://jlt.osa.org/abstract.cfm?URI=jlt-39-2-429` (cited on pages 23, 24, 50).

[38] J. Chen and H. Jiang, “Optimal design of gain-flattened Raman fiber amplifiers using a hybrid approach combining randomized neural networks and differential evolution algorithm,” *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–15, 2018. DOI: `10.1109/JPHOT.2018.2817843` (cited on page 23).

[39] D. Zibar, A. M. Rosa Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena, “Inverse system design using machine learning: The Raman amplifier case,” *Journal of Lightwave Technology*, vol. 38, no. 4, pp. 736–753, 2020. DOI: `10.1109/JLT.2019.2952179` (cited on pages 23, 24, 27, 50).

[40] G. Marcon, A. Galtarossa, L. Palmieri, and M. Santagiustina, “Model-aware deep learning method for Raman amplification in few-mode fibers,” *Journal of Lightwave Technology*, vol. 39, no. 5, pp. 1371–1380, 2021. DOI: `10.1109/JLT.2020.3034692` (cited on pages 23, 24, 27, 50).

[41] V. Perlin and H. Winful, “Optimal design of flat-gain wide-band fiber Raman amplifiers,” *Journal of Lightwave Technology*, vol. 20, no. 2, pp. 250–254, 2002. DOI: `10.1109/50.983239` (cited on page 23).

[42] F. Da Ros, U. de Moura, R. Luis, G. Rademacher, B. Puttnam, A. Rosa Brusin, A. Carena, Y. Awaji, H. Furukawa, and D. Zibar, “Optimization of a hybrid edfa-Raman C+L band amplifier through neural-network models,” in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3 (cited on page 24).

[43] A. Carena, V. Curri, and P. Poggiolini, “On the optimization of hybrid Raman/Erbium-Doped Fiber Amplifiers,” *IEEE Photonics Technology Letters*, vol. 13, no. 11, pp. 1170–1172, 2001. DOI: `10.1109/68.959353` (cited on page 24).

[44] V. Perlin and H. Winful, "On trade-off between noise and nonlinearity in WDM systems with distributed Raman amplification," in *Optical Fiber Communication Conference and Exhibit*, 2002, pp. 178–180. DOI: `10.1109/OFC.2002.1036288` (cited on page 24).

[45] L. F. Mollenauer and K. Smith, "Demonstration of soliton transmission over more than 4000 km in fiber with loss periodically compensated by Raman gain," *Opt. Lett.*, vol. 13, no. 8, pp. 675–677, Aug. 1988. DOI: `10.1364/OL.13.000675`. [Online]. Available: `http://ol.osa.org/abstract.cfm?URI=ol-13-8-675` (cited on pages 24, 49).

[46] S. T. Le, J. E. Prilepsky, P. Rosa, J. D. Ania-Castañón, and S. K. Turitsyn, "Nonlinear inverse synthesis for optical links with distributed Raman amplification," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1778–1786, 2015 (cited on pages 24, 49).

[47] P. Rosa, S. T. Le, G. Rizzelli, M. Tan, and J. D. Ania-Castañón, "Signal power asymmetry optimisation for optical phase conjugation using Raman amplification," *Opt. Express*, vol. 23, no. 25, pp. 31 772–31 778, Dec. 2015. DOI: `10.1364/OE.23.031772`. [Online]. Available: `http://www.opticsexpress.org/abstract.cfm?URI=oe-23-25-31772` (cited on pages 24, 43, 44, 49, 56).

[48] P. Rosa, S. T. Le, G. Rizzelli, M. Tan, and J.-D. Ania-Castañón, "Signal power asymmetry optimisation for optical phase conjugation using random DFB laser Raman amplification," in *Asia Communications and Photonics Conference 2015*, Optical Society of America, 2015, AM3D.5. DOI: `10.1364/ACPC.2015.AM3D.5`. [Online]. Available: `http://www.osapublishing.org/abstract.cfm?URI=ACPC-2015-AM3D.5` (cited on pages 24, 43, 44, 56).

[49] I. Phillips *et al.*, "Exceeding the nonlinear-Shannon limit using Raman laser based amplification and optical phase conjugation," in *Optical Fiber Communication Conference*, Optical Society of America, 2014, p. M3C.1. DOI: `10.1364/OFC.2014.M3C.1`. [Online]. Available: `http://www.osapublishing.org/abstract.cfm?URI=OFC-2014-M3C.1` (cited on page 24).

[50] S. Jansen, D. van den Borne, G. Khoe, H. de Waardt, P. Krummrich, and S. Spalter, "Phase conjugation for increased system robustness," in *2006 Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference*, 2006, p. 3. DOI: `10.1109/OFC.2006.215408` (cited on page 24).

[51] M. Tan, M. A. Ai-Khateeb, M. A. Iqbal, and A. D. Ellis, "Distributed Raman amplification for combating optical nonlinearities in fibre transmission," in *2018 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, IEEE, 2018, pp. 1–2 (cited on page 24).

[52] P. Rosa, G. Rizzelli, M. Tan, P. Harper, and J. D. Ania-Castañón, "Characterisation of random DFB raman laser amplifier for wdm transmission," *Opt. Express*, vol. 23, no. 22, pp. 28 634–28 639, Nov. 2015. DOI: `10.1364/OE.23.028634`. [Online]. Available: `http://opg.optica.org/oe/abstract.cfm?URI=oe-23-22-28634` (cited on pages 25, 44).

[53] A. E. Bednyakova, M. Fedoruk, P. Harper, and S. Turitsyn, "Hybrid gain-flattened and reduced power excursion scheme for distributed Raman amplification," *Opt. Express*, vol. 21, no. 24, pp. 29 140–29 144, Dec. 2013. DOI: `10.1364/OE.21.029140`. [Online]. Available: `http://opg.optica.org/oe/abstract.cfm?URI=oe-21-24-29140` (cited on pages 25, 44, 49).

[54] U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, 1995. DOI: `10.1137/1.9781611971231`. eprint: `https://epubs.siam.org/doi/pdf/10.1137/1.9781611971231`. [Online]. Available: `https://epubs.siam.org/doi/abs/10.1137/1.9781611971231` (cited on page 26).

[55] U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, 1995. DOI: `10.1137/1.9781611971231`. eprint: `https://epubs.siam.org/doi/pdf/10.1137/1.9781611971231`. [Online]. Available: `https://epubs.siam.org/doi/abs/10.1137/1.9781611971231` (cited on page 26).

[56] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, ISSN: 1476-4687. DOI: `10.1038/nature14539`. [Online]. Available: `https://doi.org/10.1038/nature14539` (cited on page 28).

[57] T. Tieleman and G. Hinton, *Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude*, COURSERA: Neural Networks for Machine Learning, 2012 (cited on page 31).

[58] A. Ferrari *et al.*, "GNPy: An open source application for physical layer aware open optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 12, pp. C31–C40, 2020 (cited on pages 31, 49).

[59] D. A. Chestnut, C. D. de Matos, P. Reeves-Hall, and J. Taylor, "Copropagating and counterpropagating pumps in second-order- pumped discrete fiber Raman amplifiers.," *Optics letters*, vol. 27 19, pp. 1708–10, 2002 (cited on page 43).

[60] C. Blum, R. Chiong, M. Clerc, K. De Jong, Z. Michalewicz, F. Neri, and T. Weise, "Evolutionary optimization," in *Variants of Evolutionary Algorithms for Real-World Applications*, R. Chiong, T. Weise, and Z. Michalewicz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–29, ISBN: 978-3-642-23424-8. DOI: `10.1007/978-3-642-23424-8_1`. [Online]. Available: `https://doi.org/10.1007/978-3-642-23424-8_1` (cited on page 44).

[61] D. Whitley, "An overview of evolutionary algorithms: Practical issues and common pitfalls," *Information and Software Technology*, vol. 43, no. 14, pp. 817–831, 2001, ISSN: 0950-5849. DOI: `https://doi.org/10.1016/S0950-5849(01)00188-4`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0950584901001884` (cited on page 44).

[62] J. Ania-Castanon, S. Kobtsev, A. Pustovskikh, and S. Turitsyn, "Simple design method for gain-flattened three-pump Raman amplifiers," in *The 15th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, vol. 2, 2002, 500–501 vol.2. DOI: `10.1109/LEOS.2002.1159400` (cited on page 49).

[63] T. J. Ellingham, J. D. Ania-Castanon, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn, “Quasi-lossless optical links for broad-band transmission and data processing,” *IEEE Photonics Technology Letters*, vol. 18, no. 1, pp. 268–270, 2006. DOI: `10.1109/LPT.2005.862001` (cited on pages 50, 51).

[64] S. Das and P. N. Suganthan, “Differential evolution: A survey of the state-of-the-art,” *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011. DOI: `10.1109/TEVC.2010.2059031` (cited on page 52).

[65] A. Colin Cameron and F. A. Windmeijer, “An r-squared measure of goodness of fit for some common nonlinear regression models,” *Journal of Econometrics*, vol. 77, no. 2, pp. 329–342, 1997, ISSN: 0304-4076. DOI: `https://doi.org/10.1016/S0304-4076(96)01818-0`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0304407696018180` (cited on pages 61, 78).

[66] U. C. de Moura, F. D. Ros, A. M. R. Brusin, A. Carena, and D. Zibar, “Experimental characterization of Raman amplifier optimization through inverse system design,” *Journal of Lightwave Technology*, vol. 39, no. 4, pp. 1162–1170, 2021. DOI: `10.1109/JLT.2020.3036603` (cited on page 74).

[67] A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures.,” *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964. DOI: `10.1021/ac60214a047`. eprint: `https://doi.org/10.1021/ac60214a047`. [Online]. Available: `https://doi.org/10.1021/ac60214a047` (cited on page 75).

[68] V. Curri and A. Carena, “Merit of Raman pumping in uniform and uncompensated links supporting NyWDM transmission,” *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 554–565, 2016. DOI: `10.1109/JLT.2015.2477599` (cited on page 87).

[69] U. C. de Moura, D. Zibar, A. Margareth Rosa Brusin, A. Carena, and F. Da Ros, “Fiber-agnostic machine learning-based Raman amplifier models,” *Journal of Lightwave Technology*, vol. 41, no. 1, pp. 83–95, 2023. DOI: `10.1109/JLT.2022.3210769` (cited on page 91).

[70] I. Pupeza, C. Zhang, M. Högner, and J. Ye, “Extreme-ultraviolet frequency combs for precision metrology and attosecond science,” *Nature Photonics*, vol. 15, no. 3, pp. 175–186, Mar. 2021, ISSN: 1749-4893. DOI: `10.1038/s41566-020-00741-3`. [Online]. Available: `https://doi.org/10.1038/s41566-020-00741-3` (cited on page 92).

[71] H. Guo, M. Karpov, E. Lucas, A. Kordts, M. H. P. Pfeiffer, V. Brasch, G. Lihachev, V. E. Lobanov, M. L. Gorodetsky, and T. J. Kippenberg, “Universal dynamics and deterministic switching of dissipative kerr solitons in optical microresonators,” *Nature Physics*, vol. 13, no. 1, pp. 94–102, Jan. 2017, ISSN: 1745-2481. DOI: `10.1038/nphys3893`. [Online]. Available: `https://doi.org/10.1038/nphys3893` (cited on page 92).