

Inverse design of a Raman amplifier in frequency and distance domains using convolutional neural networks

*Original*

Inverse design of a Raman amplifier in frequency and distance domains using convolutional neural networks / Soltani, M.; da Ros, F.; Carena, A.; Zibar, D.. - In: OPTICS LETTERS. - ISSN 0146-9592. - ELETTRONICO. - 46:11(2021), pp. 2650-2653. [10.1364/OL.422884]

*Availability:*

This version is available at: 11583/2972735 since: 2022-11-01T22:16:06Z

*Publisher:*

Optica Publ. Group

*Published*

DOI:10.1364/OL.422884

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

Optica Publishing Group (formely OSA) postprint/Author's Accepted Manuscript

“© 2021 Optica Publishing Group. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this paper are prohibited.”

(Article begins on next page)

# Inverse design of Raman amplifier in frequency and distance domain using Convolutional Neural Networks

MEHRAN SOLTANI <sup>1,\*</sup>, FRANCESCO DA ROS <sup>1</sup>, ANDREA CARENA <sup>2</sup>, AND DARKO ZIBAR <sup>1</sup>

<sup>1</sup>DTU Fotonik, Technical University of Denmark (DTU), DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Dipartimento di Elettronica e Telecomunicazioni (DET), Politecnico di Torino, Corso Duca degli Abruzzi, 24 - 10129, Torino, Italy

\*Corresponding author: msolt@fotonik.dtu.dk

Compiled April 28, 2021

We present a Convolutional Neural Network (CNN) architecture for inverse Raman amplifier design. This model aims at finding the pump powers and wavelengths required for a target signal power evolution, both in distance along the fiber and in frequency. Using the proposed framework, the prediction of the pump configuration required to achieve a target power profile is demonstrated numerically with high accuracy in C-band considering both counter-propagating and bidirectional pumping schemes. For a distributed Raman amplifier based on a 100 km single-mode fiber, a low mean set (0.51, 0.54 and 0.64 dB) and standard deviation set (0.62, 0.43 and 0.38 dB) of the maximum test error are obtained numerically employing 2 and 3 counter, and 4 bidirectional propagating pumps, respectively.

© 2021 Optical Society of America

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

## 1. INTRODUCTION

In long-haul optical communications, amplifiers play a crucial role in compensation of the link losses. Designing a desired optical amplification scheme is challenging as the requirements in noise figure (NF) and gain profile can be strict. Distributed Raman amplifiers (DRAs) have been extensively researched recently due to their low NF and moreover, providing flexibility in power evolution control by changing the power and wavelength of the pumps [1], which makes them attractive for wideband wavelength division multiplexed (WDM) scenarios [2].

One of the main challenges with inverse DRA design is caused by the nonlinear interactions between pumps and signals making it difficult to choose the relative positions and powers of the pump waves with a required gain spectra at the end of the fiber link [3]. Several machine learning solutions have been proposed for this problem in the literature. In [4], a neural network (NN) model averaging along with a fine tuning technique is employed to learn the relationship between the desired gain

profile at the end of fiber and the corresponding pump powers and wavelengths in a SMF link. Furthermore, [5] proposes an Autoencoder (AE) scheme by embedding a differentiable Raman amplifier model in the training procedure of a NN to predict the pump parameters for a specific family of gains in a FMF link.

Alternatively to designing the power spectral density at the amplifier output (frequency domain), being able to control the power evolution over the transmission span (spatial domain) offers a variety of advantages. Uniform distribution of the power along the span results in a quasi-lossless transmission which minimizes the amplified spontaneous emission (ASE) noise level [6–8]. Such a power distribution would also help several of the Kerr nonlinearity mitigation techniques currently being investigated, e.g. transmission based on the nonlinear Fourier transform theory which assumes lossless transmission [9] or nonlinearity mitigation using mid-link optical phase conjugation which requires a symmetric power distribution [10]. A significant research effort has been devoted both numerically and experimentally into achieving a desired signal power evolution over a narrow frequency bandwidth but with limited work presented on full C-band. So far, no optimization method has been proposed for addressing the power evolution design jointly in spatial and frequency domain.

In this paper, we propose a supervised deep CNN architecture for inverse DRA design in a SMF link to find the pump powers and wavelengths based on the two dimensional signal power profile in frequency and distance along the fiber. The proposed method employs two networks trained in an end-to-end procedure: A CNN network as the feature extraction followed by a multi-layer NN as the regression for predicting the pump power and wavelength values based on the extracted features. The proposed method reduces the high spatial redundancy of the signal power in frequency and distance and also extracts informative features for prediction of the pump parameters. For the evaluation of this framework, numerical simulations are demonstrated in C-band utilizing both counter-propagating and bidirectional pumping schemes.

The remainder of the paper is organized as follows. In section II, systems level CNN-based architecture and the training configuration for inverse DRA design is described. Section III demonstrates the numerical simulation results of the proposed method for counter and bidirectional propagating schemes in C-band. Finally, section IV concludes the paper.

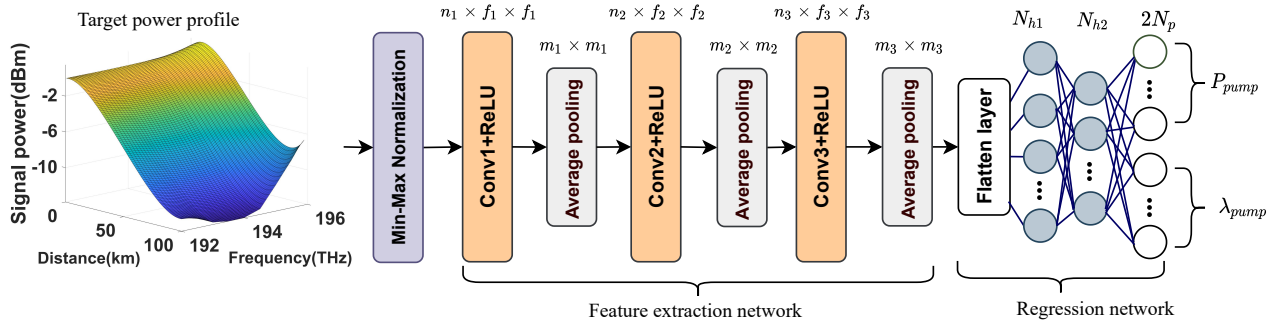


Fig. 1. Diagram of the CNN architecture for the inverse DRA design based on target power profile in frequency and distance

## 2. SYSTEM LEVEL INVERSE DRA DESIGN

### A. CNN architecture for inverse DRA design

Considering a system with a forward mapping denoted as  $Y = f(X)$ , the inverse design problem aims at modeling the mapping  $f^{-1}(\cdot)$  in order to find the input  $X$  providing a desired target output  $Y$ . In many cases, the forward mapping is straightforward and can be solved numerically, while the inverse mapping may be very complex or even unknown. Regarding this, machine learning methods, especially deep learning algorithms, have shown a promising performance in learning the approximate inverse mapping between  $X$  and  $Y$  only based on samples  $(X^s, Y^s)$  generated from the forward model  $f(\cdot)$  [4]. For simplicity in our further analysis, we will refer to  $X$  and  $Y$  as the sampled versions of the input and output spaces.

In inverse DRA design for a SMF link based on signal power evolution, the forward mapping can be described as  $\mathbf{P}_s(f, z) = f(\mathbf{P}_{\text{pump}}; \lambda_{\text{pump}})$  where  $\mathbf{P}_s(f, z) = [p_{ij}]_{N_{ch} \times N_z}$  is the two dimensional signal power in which  $p_{ij}$  is signal power at  $i$ -th frequency channel and  $j$ -th distance index in a WDM system with  $N_{ch}$  number of channels and  $N_z$  distance points,  $f(\cdot)$  is a system of nonlinear differential equations for Raman amplification scheme [1],  $\mathbf{P}_{\text{pump}} = [P_1, \dots, P_{N_p}]^T$  is the pump power vector with  $T$  denoting the transpose operator, and  $\lambda_{\text{pump}} = [\lambda_1, \dots, \lambda_{N_p}]^T$  is the pump wavelength vector.

To be able to simultaneously predict the pump powers and the wavelengths, we present a deep learning algorithm to model the inverse mapping  $[\mathbf{P}_{\text{pump}}; \lambda_{\text{pump}}] = f^{-1}(\mathbf{P}_s(f, z))$ . The dimensionality of the problem does not allow to simply apply the frameworks of [4, 5], since in order to make the input compatible with these methods, the power profile for each sample should be converted to an array of length  $N_{ch} \times N_z$  which results in an extremely large and complex network. For instance, if a system have  $N_{ch} = 40$  channels and 100 km span length with distance resolution of 1 km,  $N_z = 100$ , the number of the nodes of the input layer will be  $N_{ch} \times N_z = 4000$ . The mapping between such a high dimensional input and the pumping configuration requires a network with high number of trainable parameters which not only takes too much time to be trained, but also will be prone to problems like overfitting and local minimum. Additionally, using the approach of [4, 5] would be unnecessarily complex as it would not take advantage of the inherent correlation between the input data. Clarifying this, each point in a WDM frequency-distance space resembled as a pixel of a two dimensional image, has a high spatial redundancy. This means that adjacent points have a lot amount of information in common, however, fully-connected NNs are not capable of reducing these

spatial redundancies. Concerning these two main problems, we found CNNs more attractive since they have been designed to process data coming on the form of multiple arrays, like images, and moreover, they can successfully capture the Spatial and Temporal dependencies in a two dimensional form data through the application of relevant filters [11] and weight sharing.

For a CNN-based demonstration of inverse DRA design, we consider the distance-frequency power evolution matrix  $\mathbf{P}_s(f, z)$  as a two dimensional input to the network aiming to predict the pump configuration leading to the target  $\mathbf{P}_s(f, z)$ . Diagram of the CNN-based method is illustrated in Fig. 1. The proposed framework is made up of two stages trained End-to-End, a *feature extraction*,

$$R(\cdot; \theta_R) : \mathbb{R}^{N_{ch} \times N_z} \rightarrow \mathbb{R}^{q \times r \times n_3}, q \times r \times n_3 < N_{ch} \times N_z \quad (1)$$

and a *regression network*,

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

with trainable parameters  $\theta_R$  and  $\theta_F$ , respectively. First, a pixel-wise min-max normalization is performed on the input data as a pre-processing step. The minimum and maximum values selected for each frequency-distant points are equal to the the minimum and maximum values of the points in the training set, respectively. Afterwards, the normalized profile is passed to the feature extraction network  $R(\cdot; \theta_R)$  which consists of three CNN layers with  $n_1$ ,  $n_2$  and  $n_3$  filters of size  $f_1 \times f_1$ ,  $f_2 \times f_2$  and  $f_3 \times f_3$ , respectively. Moreover, each CNN layer is followed by rectified linear unit ( $\text{ReLU}(x) = \max(0, x)$ ) as the activation function which can speed up the training process due to its simplicity in gradient calculation. Furthermore, spatial pooling is carried out by three average-pooling layers inserted in between successive CNN layers with the window size of  $m_1 \times m_1$ ,  $m_2 \times m_2$  and  $m_3 \times m_3$ , respectively. The function of pooling layers is to progressively reduce the spatial size of the input feature maps resulting in lower amount of parameters and computations in the network. It is worth noting that the reduction of the scale of representation by each pooling layer is equal to its window size. Consequently, each layer of this network generates informative and compact representations of the input through nonlinear mappings. The output of the last CNN layer is a three-dimensional representation of the input profile consisting of  $n_3$  different two-dimensional representations 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  $r = N_z / (m_1 \times m_2 \times m_3)$ . This 3D-representation is converted thereafter to a vector of length  $q \times r \times n_3$  by the flatten layer and then is passed to the regression network  $F(\cdot; \theta_F)$ . The objective of this network, modeled

as a deep fully-connected, is to map the extracted features to the pumping setup. This network has four layers including the flatten layer of size  $q \times r \times n_3$ , two hidden layers of size  $N_{h1}$  and  $N_{h2}$  and the last layer of size  $2N_p$ , representing the pumping configuration vector. The values of  $N_{h1}$  and  $N_{h2}$  are optimized depending on the proposed pump configuration.

## B. Training and evaluation

Since the proposed approach relies on supervised learning, a data-set  $D = \{Y_k, X_k | k = 1, \dots, K\}$  needs to be generated where  $K$  is the number of samples,  $Y_k = [\mathbf{P}_{k,pump}; \lambda_{k,pump}]$  and  $X_k = \mathbf{P}_{k,s}$  are the pumping configuration vector and the corresponding 2D signal profile of the  $k$ -th sample, respectively. In this paper we focus on a data-set generated by solving the Raman amplifier differential equations [1], denoted as *Raman solver*, for different pump powers and wavelengths. For each sample data, each value of the pump parameters denoted as the  $m$ -th value of the vector  $Y$  have been selected based on a uniform distribution

$$y_m \sim U[y_m^{min}, y_m^{max}] \quad (3)$$

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  $Y$ , respectively. After the data generation, same as the most supervised learning approaches, we divide the data into separate training, testing and validation sets. Also, we make sure that the training set contains the minimum and the maximum values of each dimensions of the input signal power evolution matrix  $X_k$  to have a good generalization property [4]. The overall model of the inverse design network can be described as:

$$Y = R(F(X; \theta_F); \theta_R) \quad (4)$$

in which both  $R$  and  $F$  are jointly trained to minimize the average cost function  $C$  between the original target value  $Y_l$  and the approximated value  $\hat{Y}_l = R(F(X_l; \theta_F); \theta_R)$  of the training set:

$$\theta_R^*, \theta_F^* = \underset{\theta_R, \theta_F}{\operatorname{argmin}} \frac{1}{L} \sum_{l=1}^L C(\hat{Y}_l, Y_l) \quad (5)$$

where  $L$  is the number of training samples and for each sample,  $C$  is the mean square error (MSE) value between the target and the approximated pump set-up values  $Y_l$  and  $\hat{Y}_l$ , respectively:

$$C = \frac{1}{2N_p} \sum_{i=1}^{2N_p} (Y_l^i - \hat{Y}_l^i)^2 \quad (6)$$

The parameters of the network are updated in an iterative approach by means of gradient descent algorithm and back-propagation [11]. Furthermore, advanced optimization algorithm RMSprop [12] is employed for updating the parameters as it provides a fast and robust convergence for each parameter.

Once the training of the network has been completed, we fix the set of the learnt parameters of the network  $\theta = \{\theta_R, \theta_F\}$  to evaluate its performance. To this end, we put the network into a schematic as illustrated in Fig. 2. In this scheme, for each input power profile, the corresponding pump powers and wavelengths have been predicted using the network in Fig. 1 and then passed to the Raman solver  $RS(\cdot)$  to compute the power profile based on the predicted pumping setup. Afterwards, maximum absolute difference between the predicted and the input power profile is calculated in frequency ( $f$ ) and distance ( $z$ ) domain as the final prediction error for each sample.

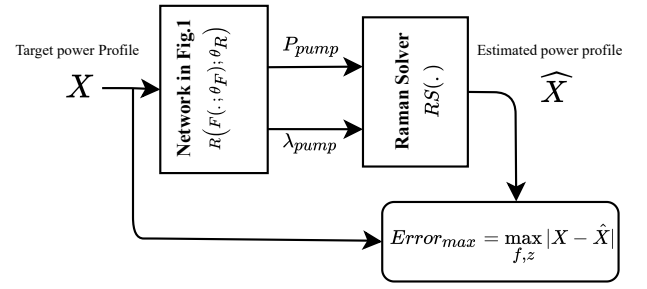


Fig. 2. Evaluation diagram of the proposed method

## 3. SIMULATION RESULTS

In this section, we investigate the CNN-based framework presented in the previous section for the design of Raman amplifiers in C-band. The data-set are generated using the Raman Solver provided by GNPpy [13], an open source application developed recently for analyzing optical networks.

We consider a single span and analyze the evolution of the power profile jointly over the distance and the entire C-band (between 192 and 196 THz). Also, three propagation cases are deployed for the evaluation of the proposed method: 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 1. The superscripts (-) or (+) on the power ranges specify the counter or co-propagation of the corresponding pump, respectively.

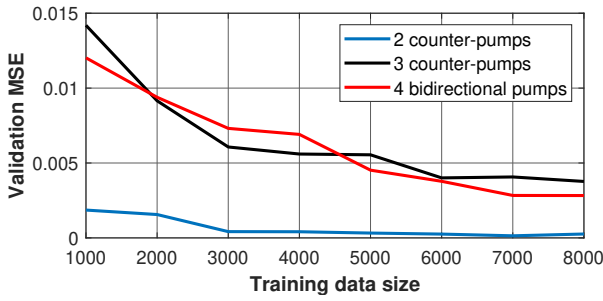
Table 1. Power and wavelength ranges for each DRA case

Param	2 pumps	3 pumps	4 pumps
$P_1$ [mW]	[40 – 400] <sup>-</sup>	[30 – 300] <sup>-</sup>	[30 – 300] <sup>-</sup>
$P_2$ [mW]	[40 – 400] <sup>-</sup>	[30 – 300] <sup>-</sup>	[30 – 300] <sup>-</sup>
$P_3$ [mW]	-	[30 – 300] <sup>-</sup>	[30 – 300] <sup>+</sup>
$P_4$ [mW]	-	-	[30 – 300] <sup>+</sup>
$\lambda_1$ [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]

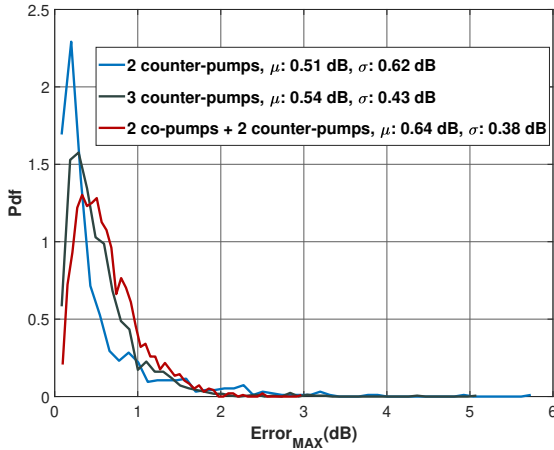
We divided the C-band into 40 channels with 100 GHz spacing. Input signal power per-channel is set to 0 dBm which results in a total WDM signal power of 16 dBm. Furthermore, a standard silica fiber with the following parameters is assumed: span length  $L_{span} = 100$  km, signal data attenuation  $\alpha_s = 0.2$  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.

In order to determine the size of the training data, for each pumping case, different sizes from 1000 to 8000 have been investigated. Models trained on training data-sets with different size, have been evaluated by the MSE on a validation data-set with 1000 points generated separately. Fig. 3 shows the validation MSE as a function of the size of the training data-set. Based on the validation MSE and also the training time, we realized that for 2 and 3 counter and 4 bidirectional cases, best training data sizes are 5000, 6000 and 7000 samples, respectively. Increasing





**Fig. 3.** MSE of the validation data-set for different pumping schemes as a function of the training data-set size



**Fig. 4.** Probability density function of the  $Error_{max}$ . Mean  $\mu$  and standard deviation  $\sigma$  are shown for 2, 3 and 4 pumps.

the training size will not result in a remarkable improvement.

Regarding the parameters of the *feature extraction* network, 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. For the number of filters of each layer, we tried 32 and 64 numbers and observed that 64 filters extremely increase the training time with no improvement in performance. Moreover, filter size of  $3 \times 3$  showed a better validation MSE over a bigger  $5 \times 5$  filter. We also figured out that for the window size of the average-pooling layers, a commonly used  $2 \times 2$  window has a better MSE over a window of size  $3 \times 3$ . Furthermore, regarding the regression network parameters, we evaluated the validation MSE by setting the  $N_{h1}$  and  $N_{h2}$  based on the set  $\{20, 40, 80, 100\}$  and realized that 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 will minimize the validation loss. For all pumping schemes, the batch size in training phase has been set to 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. Moreover, due to the different value ranges between pump powers and wavelengths at the output layer, the network is trained on the min-max normalized pump configuration vector. The resulting normalized pump configuration vector can be linearly mapped to the desired interval of powers and wavelengths based on the specified ranges for these parameters.

Final evaluation of the trained models is performed based on the scheme illustrated in Fig. 2 on test data-sets with 2000 samples generated for each pumping case. Fig. 4 indicates the probability density function (pdf) for the maximum absolute error ( $Error_{max}$ ) of the reconstructed power profile beside its mean ( $\mu$ ) and standard deviation ( $\sigma$ ) for all cases. It has been demonstrated that the  $\mu$  value for 2 counter-propagating, 3 counter-propagating and 4 bidirectional propagating cases is almost 0.51 dB, 0.54 dB and 0.61 dB, respectively, and also the  $\sigma$  value for these cases is almost 0.62 dB, 0.43 dB and 0.38 dB, respectively. We can therefore assert that the proposed method is highly accurate for designing Raman amplifiers based on the signal power profile over a wide band and along the span.

#### 4. CONCLUSION

A CNN framework is presented for inverse DRA design based on desired signal power profile in frequency and distance domain. The proposed method consists of two networks trained end-to-end: 1) a *feature extraction* with 3 CNN layers employed to extract informative features of the 2D signal power profile and 2) a *regression* aiming to predict the pump powers and wavelengths values based on the extracted features. Numerical simulations show that the proposed framework provide high accuracy in terms of predicting the pump parameters for both counter and bidirectional propagating pumps in C-band.

**Acknowledgments.** This work was financially supported by the European Research Council (ERC-CoG FRECOM grant no. 771878), the Villum Foundation (OPTIC-AI grant no. 29334), and the Italian Ministry for University and Research (PRIN 2017, project FIRST).

**Disclosures.** The authors declare no conflicts of interest.

#### REFERENCES

1. C. Headley and G. P. Agrawal, *Raman Amplification in Fiber Optical Communication Systems* (Academic Press, 2005).
2. U. C. de Moura, M. A. Iqbal, M. Kamalian, L. Krzczanowicz, F. Da Ros, A. M. R. Brusin, A. Carena, W. Forsysiak, S. Turitsyn, and D. Zibar, *J. Light. Technol.* **39**, 429 (2020).
3. V. E. Perlin and H. G. Winful, *J. Light. Technol.* **20**, 250 (2002).
4. D. Zibar, A. M. Rosa Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena, *J. Light. Technol.* **38**, 736 (2020).
5. G. Marcon, A. Galtarossa, L. Palmieri, and M. Santagiustina, *J. Light. Technol.* **39**, 1371 (2021).
6. J. D. Ania-Castanon, *Opt. Express* **12**, 4372 (2004).
7. T. J. Ellingham, J. D. Ania-Castanon, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn, *IEEE Photonics Technol. Lett.* **18**, 268 (2006).
8. J. Bouteiller, K. Brar, and C. Headley, "Quasi-constant signal power transmission," in *2002 28TH European Conference on Optical Communication*, vol. 3 (2002), pp. 1–2.
9. S. T. Le, J. E. Prilepsky, P. Rosa, J. D. Ania-Castañón, and S. K. Turitsyn, *J. Light. Technol.* **34**, 1778 (2015).
10. 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.
11. Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**, 436 (2015).
12. T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude," COURSE: Neural Networks for Machine Learning (2012).
13. A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. L. Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, *J. Opt. Commun. Netw.* **12**, C31 (2020).

## FULL REFERENCES

1. C. Headley and G. P. Agrawal, *Raman Amplification in Fiber Optical Communication Systems* (Academic Press, 2005).
2. U. C. de Moura, M. A. Iqbal, M. Kamalian, L. Krzczanowicz, F. Da Ros, A. M. R. Brusin, A. Carena, W. Forysiak, S. Turitsyn, and D. Zibar, "Multi-band programmable gain Raman amplifier," *J. Light. Technol.* **39**, 429–438 (2020).
3. V. E. Perlin and H. G. Winful, "Optimal design of flat-gain wide-band fiber raman amplifiers," *J. Light. Technol.* **20**, 250 (2002).
4. 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," *J. Light. Technol.* **38**, 736–753 (2020).
5. G. Marcon, A. Galtarossa, L. Palmieri, and M. Santagiustina, "Model-aware deep learning method for Raman amplification in few-mode fibers," *J. Light. Technol.* **39**, 1371–1380 (2021).
6. J. D. Ania-Castanon, "Quasi-lossless transmission using second-order Raman amplification and fibre Bragg gratings," *Opt. Express* **12**, 4372–4377 (2004).
7. 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 Technol. Lett.* **18**, 268–270 (2006).
8. J. . Bouteiller, K. Brar, and C. Headley, "Quasi-constant signal power transmission," in *2002 28TH European Conference on Optical Communication*, vol. 3 (2002), pp. 1–2.
9. 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," *J. Light. Technol.* **34**, 1778–1786 (2015).
10. 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.
11. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
12. T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude," COURSE: Neural Networks for Machine Learning (2012).
13. A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. L. Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "GNPy: an open source application for physical layer aware open optical networks," *J. Opt. Commun. Netw.* **12**, C31–C40 (2020).