Abstract

The propagation of the signal in the optical fiber is governed by the nonlinear Schrödinger equation (NLSE). In high-speed communication systems, nonlinear interference caused by the Kerr nonlinearity of the optical fiber poses a challenge to the system design. Conventional ways of mitigating optical fiber nonlinearity include solving NLSE in the reverse direction to obtain an equalized version of the original transmitted signal. This is commonly referred to as digital back propagation (DBP) [1]. DBP-based algorithms often suffer from high computational burden and methods to reduce the complexity were widely discussed in literature. In this deliverable we review the digital back propagation and neural networks to mitigate the nonlinear effects in optical fibers.

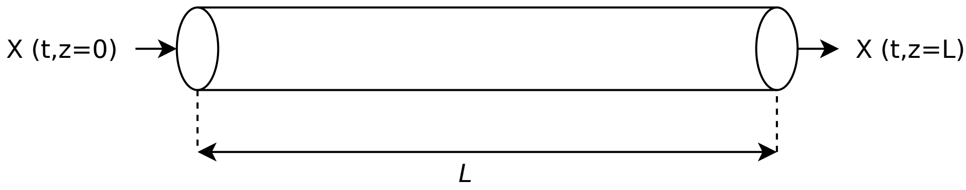


Fig. 1. Optical fiber channel

1 Introduction

Optical fibers are narrow strings of glass that carry optical signals over long distances. The main material in constructing fibers is silica, although plastic optical fibers can be used for shorter distance. The advantages of optical communications system over electrical copper-based systems are multifold. Firstly, a single silica fiber can carry thousands of channels simultaneously, a much larger number than its counterpart. Secondly, the losses for light propagating in fibers are very small, around 0.2 dB/km for modern single-mode silica fibers. Lastly, thanks to high achievable transmission rates, the cost per transported bit can be low.

Increasing the fiber transmission rates to meet the growing demand for data rates has became a major research interest. Nonlinear interference (NLI) is one factor that negatively affects the signal quality, reducing the achievable transmission rates. In order to mitigate the effects of NLI, channel models that accurately describe the propagation of the signal must be obtained [2].

In a signal-mode optical fiber, signals propagate according to nonlinear Schrödinger equation (NLSE) [3]. Conventional way of recovering the transmitted signal consists of solving the NLSE with the received signal as the boundary condition at the input. This is known as digital back-propagation (DBP). However, this scheme suffers from high computational complexity, making it less desirable for real-time application [4–6]. Recently neural networks (NNets) have attracted great interest as alternative means of nonlinearity compensation with promising performance [7–9].

In this deliverable, we study the optical fiber channel model, split step Fourier method and the deep learning for mitigating the fiber nonlinearity. Lastly, we provide a brief literature review of the state-of-the-art methods that utilize NNets for the nonlinearity mitigation in optical fiber communication.

2 Channel Model

The optical fiber channel is non-linear due to the Kerr effect caused by a change in the refractive index of the fiber proportional to the square of the amplitude of the signal. The channel can be described by the nonlinear Schrödinger equation (NLSE) modeling dispersion and the Kerr nonlinearity. Let $Q(t,z)$ be the complex envelope of the propagating signal at time t and distance z. In a single mode fiber (SMF), the NLSE is

$$\frac{\partial Q(t,z)}{\partial z} = -\frac{\alpha}{2}Q(t,z) - \frac{j\beta_2}{2}\frac{\partial^2 Q(t,z)}{\partial t^2} + j\gamma|Q(t,z)|^2 Q(t,z) + N(t,z), \tag{1}$$

where α, β_2 and γ are the attenuation, dispersion, and nonlinearity parameters, and $N(t,z)$ is circular symmetric complex Gaussian noise. We consider here the case of lossy propagation, *i.e.* , $\alpha \neq 0$.

Depending on the value of β_2, two modes of dispersion exist: normal dispersion where $\beta_2 < 0$ and the group velocity decreases with increasing the optical frequency, and the anomalous dispersion where $\beta_2 > 0$. More generally, $\beta(\omega)$ can be written as a Taylor expansion

$$\beta(\omega) = \beta_0 + \frac{\partial \beta}{\partial \omega}(\omega - \omega_0) + \frac{1}{2}\frac{\partial^2 \beta}{\partial \omega^2}(\omega - \omega_0)^2 + \frac{1}{6}\frac{\partial^3 \beta}{\partial \omega^3}(\omega - \omega_0)^3 + ..., \tag{2}$$

where ω_0 is the carrier frequency. Dispersion terms higher than the second order term are called "higher-order dispersion". The zero-order term describes the common phase shift. The first-order term is the inverse group velocity, *i.e.* , the group delay per unit length

$$\beta_1 \equiv \frac{\partial \beta(\omega)}{\partial \omega} = \frac{1}{v_g}.$$

The second-order dispersion coefficient describes the group delay dispersion per unit length, and its unit is s^2/m. It is also the derivative of the inverse group velocity with respect to angular frequency

$$\beta_2 \equiv \frac{\partial^2 \beta(\omega)}{\partial \omega^2} = \frac{\partial}{\partial \omega}\frac{\partial \beta(\omega)}{\partial \omega} = \frac{\partial}{\partial \omega}\frac{1}{v_g}.$$

The higher-order dispersion coefficients are

$$\beta_k \equiv \frac{\partial^k \beta(\omega)}{\partial \omega^k}, \quad k \geq 3.$$

3 Back Propagation

The NLSE (1) cannot generally be solved analytically for non-zero values of α, β_k and γ. However, it is commonly solved numerically using the split-step Fourier method (SSFM). The SSFM splits the channel into a linear and nonlinear components and applies each component recursively along the fiber. Let us rewrite (1) as

$$\frac{\partial Q(t, z+\Delta z)}{\partial z} = (C_L + C_N)Q(t,z), \tag{3}$$

where $\Delta z = L/M$ is the length of each of the M segments of the fiber of total length L, and C_L and C_N are the linear and nonlinear terms

$$C_L = \frac{j\beta_2}{2}\frac{\partial^2}{\partial t^2}$$
$$C_N = j\gamma|Q(t,z)|^2.$$

For simplicity, we assumed $\beta_k = 0$ for $k \geq 3$ and $N(t,z) = 0$. It follows from (3)

$$Q(t, z+\Delta z) \approx e^{\Delta z C_L} e^{\Delta z C_N} Q(t,z). \tag{4}$$

For small values of Δz, the linear and the nonlinear parts can be applied individually. For the linear part of the (3), it is useful to transform the equation first to the frequency domain to simplify the term $\frac{\partial^2}{\partial t^2}$. Solving the two equations individually yields

$$\hat{Q}(\omega, z+\Delta z) = -e^{-\frac{\beta_2 \omega^2 \Delta z}{2}} \hat{Q}(\omega, z), \tag{5}$$

$$Q(t, z+\Delta z) = e^{j\gamma|Q(t,z)|^2 \Delta z} Q(t,z). \tag{6}$$

The solution for the propagated signal along an optical fiber of length L is

$$Q(t, L) = \prod_{j=1}^{M} e^{zC_L} e^{zC_N} Q(t,0). \tag{7}$$

Equation (7) can also be used to obtain the input signal $Q(t,0)$ from the output signal $Q(t,L)$ by simply negating the Δz in C_L and C_N, which is equivalent to negating the values of α, β_2 and γ. This approach is referred to as *digital back-propagation*. Although we discussed the basic SSFM here, in practice more sophisticated versions based on the symmetric SSFM are used.

4 Neural Networks

Neural networks (NNets) are machine learning models often used in practice in the context of the supervised learning. Supervised learning is the task of using labelled data to learn the function that maps each input to its corresponding output. Neural networks can be used in numerous applications in classification and regression. One advantage of using machine learning techniques over conventional methods in communication is that machine learning methods may provide a satisfactory estimation of the channel model. This is especially useful in cases where an exact model is not known, is unavailable or is difficult to produce.

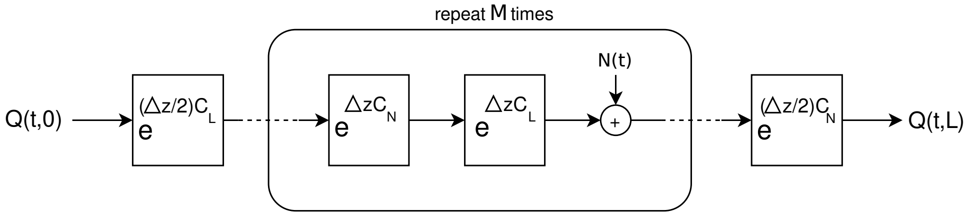


Fig. 2. Split-step Fourier method

4.1 Perceptron

The neural networks are inspired by the biological nervous system studied in animals [10]. Neural networks consist of basic building blocks called perceptrons that are motivated by the brain neurons. Perceptrons are the most basic neural networks, with a single input layer and an output node.

Consider one training example of the form $(\mathbf{x}, y)$, where $\mathbf{x} = [x_1, x_2, ..., x_k]$ is a vector containing k feature variables, and the observed output $y \in \{-1, 1\}$ is a binary class variable. The perceptron output is

$$\hat{y} = \text{sign}(b + \mathbf{w} \cdot \mathbf{x}) = \text{sign}\left(b + \sum_{j=1}^{k} w_j x_j\right) \tag{8}$$

where sign$(\cdot)$ is the sign function that is either $\{+1, -1\}$, $\mathbf{w}$ is a vector containing the weights $[w_1, w_2, ..., w_k]$ and $\mathbf{x}$ is the input feature vector. The notation $\hat{y}$ indicates that $\hat{y}$ is an estimation of the output y.

The prediction error for one training example is $l(\mathbf{x}) = y - \hat{y}$. The output of the perceptron represents and index of a class. The training algorithm is aimed at minimizing the classification or prediction error. For non-zero values of $l(\mathbf{x})$, the weights are updated in the direction of negative gradient. We can define the training goal as the minimization of the loss of with respect to all training examples in the data set $\mathcal{D}$ containing all $(\mathbf{x}, y)$

$$\min_{\mathbf{w}} L = \sum_{(\mathbf{x},y)\in\mathcal{D}} (y - \hat{y})^2 = \sum_{(\mathbf{x},y)\in\mathcal{D}} \left(y - \text{sign}\left(b + \sum_{j=1}^{k} w_j x_j\right)\right)^2. \tag{9}$$

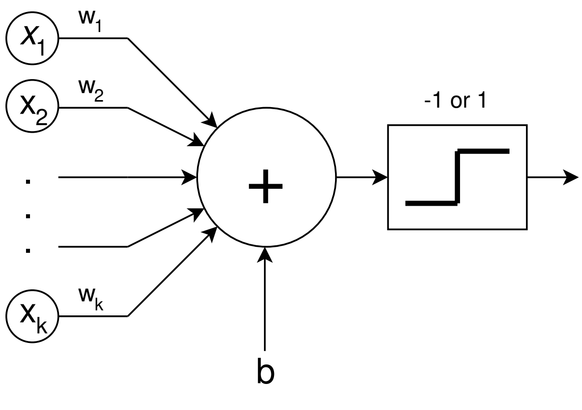


Fig. 3. diagram of a single perceptron

Note that the gradient of sign(x) is zero almost everywhere except at $x = 0$ which is not differentiable. This problem becomes more important when we move from the perceptrons to multi-layered neural networks. As a result, the sign function is replaced with other smoother functions called the *activation function*. Examples include *sigmoid*, *hyperbolic-tangents* and *rectified linear units* (ReLU) with various degrees of smoothness. Figure

4 shows popular nonlinear activation functions. These activation functions are mathematically defined as:

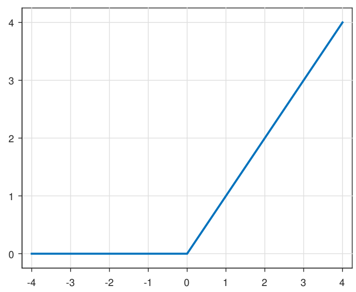
$$\Phi(v) = \frac{1}{1+e^{-v}} \qquad \text{(sigmoid function)}$$

$$\Phi(v) = \frac{e^{2v}-1}{e^{2v}+1} \qquad \text{(tanh function)}$$

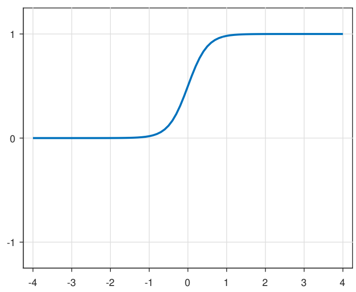
$$\Phi(v) = \begin{cases} 0 & v \leq 0 \\ v & \text{elsewhere} \end{cases} \qquad \text{(ReLU)}$$

$$\Phi(v) = \begin{cases} -b & v \leq -b \\ v/b & -b < v \leq b \\ b & v > b \end{cases} \qquad \text{(hard limiter).}$$

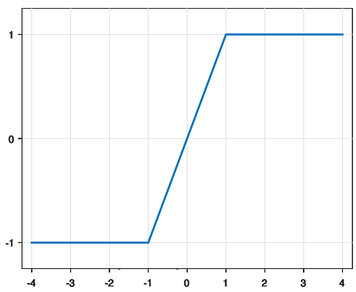
The ReLU and hard limiter activation functions have replaced the sigmoid and tanh activation functions in modern neural networks mainly because their derivatives are simple to compute, simplifying the training multilayered neural networks with these activation functions.



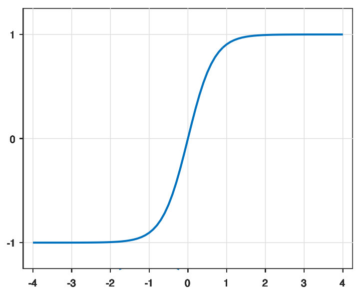
(a) ReLU



(b) Sigmoid



(c) Hard Limiter



(d) hyperbolic tangent

Fig. 4. Number of widely used activation functions

4.2 Feedforward Neural Networks

The perceptron consists of an input and output layer, and the output layer is the only layer that performs computations. The input layer merely passes the data to the output layer. The computations inside each perceptron are visible to the user. Multilayer neural networks on other hands consist of several computational layers, and the computations in the intermediate layers between the input and output are not visible to the user; therefore these layers are referred to as *hidden layers*. The architecture of the *fully-connected feed-forward* neural networks assumes that all nodes in each layer are connected with all nodes in the following layer.

The neural networks are trained using an iterative algorithm called the (stochastic) gradient descent. The update rule for the weights using the gradient descent is

$$\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} - \delta \nabla L(\mathcal{D}, \boldsymbol{W}^{(i)}) \,, \tag{10}$$

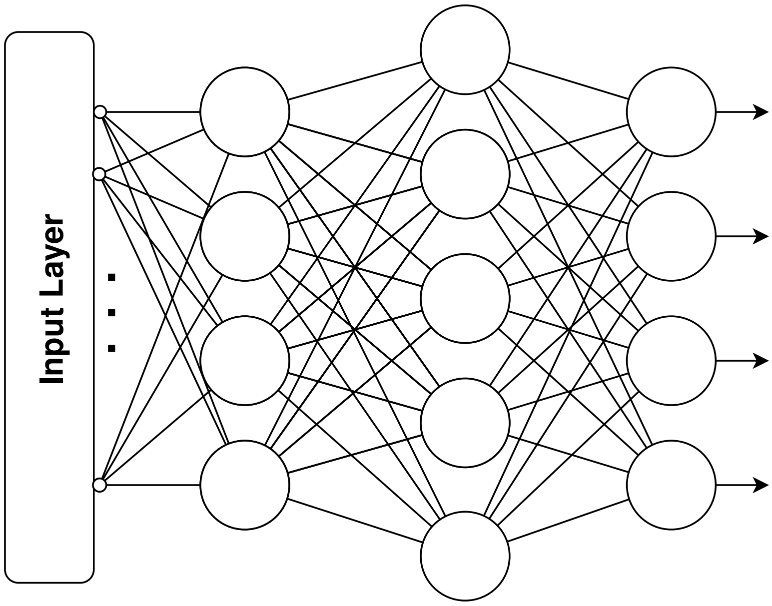


Fig. 5. Diagram of a fully connected neural network.

where $\boldsymbol{W}^{(i)}$ is a vector that contains the weights at the i-th iteration, $\delta > 0$ is the learning rate, and $\nabla L(\mathcal{D}, \boldsymbol{W}^{(i)})$ is the gradient of the loss function $L(\cdot)$. The loss function can be defined in several ways, *e.g.* , as the expected value of the mean-squared loss over all training examples. The weights are updated until the loss function is approximately minimized. A diagram of a fully-connected feed-forward neural network is shown in Fig. 5.

4.3 Review of the Deep Learning in Optical Communication

It is only in recent years that neural networks have shown an accuracy on some tasks that exceeds that of a human. For example, recent results show that neural networks can surpass human performance in image recognition, which was not considered possible a few years ago. Deep learning provides the system with the ability to learn and detect meaningful patterns between the input and output with reduced complexity. Because of that, machine learning techniques were utilized in many applications, and -unsurprisingly,- found numerous applications in optical fiber communications. Some of the state-of-the-art applications in optical communications are found in nonlinearity mitigation, performance monitoring and automatic modulation format detection.

Examples from literature include [11], [7] and [12], where neural networks were used to reduce the complexity of the conventional DBP. Other examples include [13] and [14], where deep learning is used to achieve a better performance than state-of-the-art methods. Papers [15], [16] and [17] consider the end-to-end optical fiber communication system as an auto-encoder.

C. Häger and H. Pfister proposed a learned digital back-propagation (LDBP) scheme to mitigate NLI, where the network design is based on unrolling the SSFM [7]. They concluded that using learned DBP significantly reduces the complexity compared to conventional DBP implementations. The proposed approach can adapt to real-world imperfections that are not included in the analytical model.

In [15], the authors proposed an end-to-end deep learning design of optical communication system. They also proposed a novel training method to improve the robustness to distance variations that offers a significant level of flexibility. They concluded that their design outperforms conventional receiver design in terms of bit error rate (BER) for the tested range of distances.

5 Conclusions

We reviewed the channel model in the optical fiber governed by the nonlinear Schrödinger equation. The NLSE can be solved numerically using the split-step Fourier method, which consists of dividing the distance into a number of small segments, and solving the linear and nonlinear parts of the NLSE individually in each segment. We then presented an overview of the neural networks and deep learning. Deep learning provides a promising alternative to the digital back-propagation. Lastly, a brief literature review of the state-of-the-art methods that utilize NNets for nonlinearity mitigation in fiber communications was discussed.

References

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

[2] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *Journal of Lightwave Technology*, vol. 34, no. 16, pp. 3884–3889, 2016.

[3] G. P. Agrawal, "Chapter 2 - pulse propagation in fibers," in *Nonlinear Fiber Optics (Sixth Edition)* (G. P. Agrawal, ed.), pp. 27 – 55, Academic Press, sixth edition ed., 2019.

[4] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part i: Mathematical tools," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4312–4328, 2014.

[5] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part ii: Numerical methods," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4329–4345, 2014.

[6] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part iii: Spectrum modulation," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4346–4369, 2014.

[7] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, 2018.

[8] D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, "Machine learning techniques in optical communication," in *2015 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2015.

[9] S. Owaki and M. Nakamura, "Simultaneous compensation of waveform distortion caused by chromatic dispersion and spm using a three-layer neural-network," in *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, pp. 1–3, 2017.

[10] C. C. Aggarwal, *Neural Networks and Deep Learning.* Cham: Springer, 2018.

[11] C. Häger and H. D. Pfister, "Deep learning of the nonlinear schrödinger equation in fiber-optic communications," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1590–1594, 2018.

[12] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Unsupervised representation learning of structured radio communication signals," *CoRR*, vol. abs/1604.07078, 2016.

[13] T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2017.

[14] E. Nachmani, Y. Be'ery, and D. Burshtein, "Learning to decode linear codes using deep learning," *CoRR*, vol. abs/1607.04793, 2016.

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

[16] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," *CoRR*, vol. abs/1804.02276, 2018.

[17] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *2018 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2018.