

# Mapping EDFA Noise Figure and Gain Flatness Over the Power Mask Using Neural Networks

Carmelo J. A. Bastos-Filho, Erick de A. Barboza

Programa de Pós-Graduação em Engenharia de Sistemas, Universidade de Pernambuco  
[carmelofilho@ieee.org](mailto:carmelofilho@ieee.org), [eab@ecomp.poli.br](mailto:eab@ecomp.poli.br)

Joaquim F. Martins-Filho

Departamento de Eletrônica e Sistemas, Universidade Federal de Pernambuco, Recife, Brazil  
[jfmf@ufpe.br](mailto:jfmf@ufpe.br)

Uiara C. de Moura, Juliano R. F. de Oliveira

Centro de Pesquisa e Desenvolvimento em Telecomunicações – CPqD, Campinas, Brazil  
[umoura@cpqd.com.br](mailto:umoura@cpqd.com.br), [jrfo@cpqd.com.br](mailto:jrfo@cpqd.com.br)

**Abstract**— Optical Amplifiers play an important role in reconfigurable optical communications networks. The device characterization within the dynamic operational range is crucial for the proper deployment and usage of such devices. In general, one needs to measure a certain number of operation points to complete the characterization. In spite of this, there is a lot of missing data for some operational deployment cases. We show that one can use simple neural networks to execute a regression task and obtain a continuous characterization curve of Gain Flatness and Noise Figure performance, along the entire Amplifier Power Mask. This regression can be made using a lower number of points than usual. We obtained estimated errors lower than 0.1 dB for Gain Flatness and Noise Figure over the entire operational range.

**Index Terms**— Artificial Neural Networks, Multilayer Perceptron, Optical Amplifiers, Device Characterization.

## I. INTRODUCTION

The ever-increasing traffic demand generated by the new Internet and video-based services has driven the telecommunication market to deploy exhaustively optical communication systems. Erbium-Doped Fiber Amplifiers (EDFAs) have been successfully used in multi-channel optical systems since the last century to compensate for losses generated by the transmission fibers and optical devices [1]. In general, different channels in a multi-channel system can be amplified by different gains due to the variations in the EDFA gain spectrum. This variation occurs mainly due to the intra-band sub-levels distribution within the energy transition band. One can define the Gain Flatness (GF) as the maximum gain excursion within the optical transmission band regarding a predefined gain, whereas the Automatic Gain Control (AGC) accuracy is defined as the difference between the set point gain and the measured gain.

Besides, the gain spectrum can vary depending on the total input signal power. Furthermore, EDFAs insert noise within the transmission band due to Amplified Spontaneous Emission (ASE) [1]. The Optical Signal-to-Noise Ratio (OSNR) is defined as the ratio between the signal power and the noise (ASE) power in a specific point of the system. The degradation of the OSNR generated by a

device is defined as the Noise Figure (NF), which is given in a logarithm scale (dB). GF and NF are important metrics to define the quality of amplifiers, which have impact on the quality of transmission (QoT) in optical communication systems.

In reconfigurable optical networks, the need for measurement of these figures of merit is even more important, since the number of channels and their powers can vary dynamically and, sometimes, unpredictably. Because of this, the deployed optical amplifiers must operate properly in a predefined range of input powers and gains, inserting low noise and providing flat gain for all wavelength division multiplexed (WDM) channels. The previous knowledge of the amplifier performance is essential to provide lightpaths with high QoT to support different modulation formats and bit rates [2].

Moura *et al.* [3] developed an automatic amplifier characterization scheme for optical amplifiers with AGC in order to obtain the dynamic performance of the amplifier in terms of NF, GF spectrum and AGC accuracy. Oliveira *et al.* [4] used the same platform to assess the performance of AGC integrated hybrid amplifiers composed of distributed Raman amplifiers and EDFAs. The same setup was also used in [2] to design the control plane of a cognitive EDFA.

However, the characterization process of an EDFA is slow and the resulting characterization data does not provide continuous nor high-resolution gain and NF curves. In this paper we propose to apply a simple and well-known type of Artificial Neural Network, called Multilayer Perceptron (MLP), to accomplish the regression task over the entire *Power Mask* range with low error.

The rest of this paper is organized as follows. Section II presents details regarding the amplifiers characterization process. Section III presents the concepts regarding Multilayer Perceptron Neural Networks. Section IV presents our proposal for regression of the GF and NF curves using MLP. Section V and VI present the simulation setup and some results, respectively. Finally, we give our conclusions in section VII.

## II. AMPLIFIER CHARACTERIZATION

This section aims to present the amplifier characterization process background. The first subsection presents the *Power Mask* concept. Sub-section B details the characterization process and the last subsection presents how the results of the amplifier characterization are presented.

### A. *Power Mask*

The amplifier operation region defined by its input and output power is defined as its *Power Mask* [5], as shown in Fig. 1. In the *Power Mask*, each diagonal corresponds to a constant gain, varying from a minimum to a maximum (right to left) value along the mask. An input power drop corresponds to a reduction on channel load or a power decrease by some power line impairment.

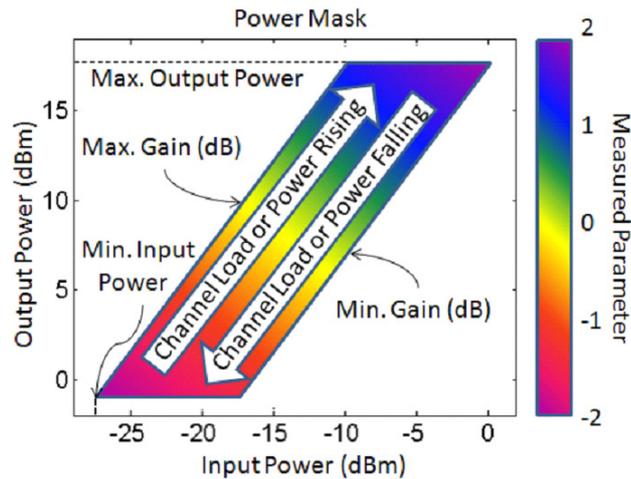


Fig. 1. Power mask definition.

NF, GF, AGC accuracy and others parameters such as the amplifier power consumption and the laser temperature could be measured and presented as an amplifier *Power Mask*. For this, one should measure the values for the chosen metrics varying the input and output power values. The values for the chosen metrics are coded into colors in the right color bar, also shown in Fig. 1. This helps to visualize the behavior of the metrics along the entire *Power Mask*. These measurements are accomplished accordingly to the characterization process described hereafter.

#### B. Characterization process

The characterization process is divided in two stages: the experimental phase and the data processing phase. In the experimental phase, 40 non-modulated and flattened channels spaced at 100 GHz compose the amplifier input signal grid, which consists of a full C-band load (ITU-T grid). This input signal grid is varied, jointly with a set point gain, to characterize the entire *Power Mask*. The granularity of the characterization is defined by two adjustable variables, which are the *step* size and the number of points to be measured. For each point, the input and the output spectra are measured and stored.

The experimental setup is showed in Fig. 2 and is composed by a laser bench and a Wavelength Selective Switch (WSS) to provide a flattened C-band load channels by measuring a sample of the input signal spectrum and adjusting each channel power in a loop process composed by the Optical Spectrum Analyzer (OSA), the computer and the WSS, all of them controlled by a Labview<sup>®</sup> program. An auxiliary amplifier is used to raise the maximum input power inserted in the amplifier under characterization. Another Labview<sup>®</sup> program is used to control the Variable Optical Attenuator (VOA) and the amplifier gain set point in order to sweep the entire power mask region. This same Labview<sup>®</sup> program is used to set the optical switch in order to obtain the input and the output spectra by the OSA. All the connections from the computer to the equipment are depicted in Fig. 2.

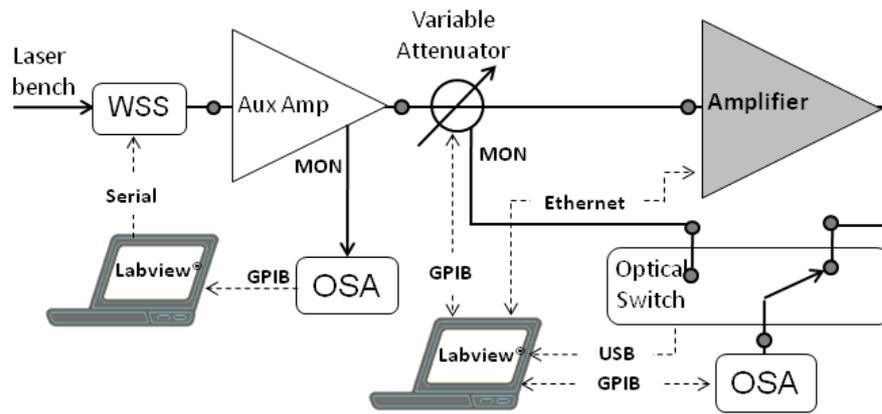


Fig. 2. Experimental setup for Amplifier Characterization.

When all the points are captured, a Matlab<sup>®</sup> program is used for the data processing. In this phase, the spectra are used to calculate the NF of each channel using the following equation [6]:

$$NF = \frac{P_{ASE}}{h\nu G \Delta\nu} + \frac{1}{G}, \quad (1)$$

in which  $P_{ASE}$  is the ASE noise power emitted by the amplifier,  $h$  is the Planck constant,  $\nu$  is the channel frequency,  $\Delta\nu$  is the measured bandwidth of the signal and  $G$  is the channel gain.

Still in the data processing stage, GF and AGC accuracy are also calculated and stored in a table jointly with the NF of the worst channel.

### C. Characterization results

As a result from the characterization process, a table with information of input and output total powers, set point gain, NF, GF and AGC accuracy is generated. A sample of this table, with just a few points, is presented in table I to illustrate an example of the obtained data.

TABLE I. RESULT TABLE OF THE CHARACTERIZATION PROCESS WITH SOME POINTS.

Input Power (dBm)	Output Power (dBm)	Configured Gain (dB)	NF (dB)	GF (dB)	AGC accuracy (dB)
2.289	13.182	11.000	5.976	3.151	-0.106
1.819	12.721	11.000	5.965	3.127	-0.098
1.355	12.244	11.000	5.956	3.157	-0.111
0.832	11.701	11.000	5.970	3.121	-0.131
0.298	11.163	11.000	5.938	3.137	-0.136
-0.147	10.754	11.000	5.959	3.122	-0.099
-0.693	10.164	11.000	5.983	3.171	-0.143
-1.226	9.637	11.000	5.963	3.202	-0.137
-1.678	9.183	11.000	5.974	3.122	-0.139

The information presented in table I can also be plotted graphically as depicted in Fig. 1. Fig. 3 presents examples for (3.a) NF, (3.b) GF and (3.c) AGC accuracy, considering the following power

mask parameters: minimum input power of -25 dBm, maximum output power of 21 dBm, minimum gain of 14 dB, maximum gain of 24 dB and characterization step equal to 1 dB.

One must observe that depending on the step value adopted for the characterization process and the size of the power mask considered for the amplifier, the number of lines in the resulted table can increase substantially. If this information is stored in the amplifier or in a control plane as in [2], a huge amount of information can occupy a valuable part of the memory area, which could be necessary for other purposes in real-world amplifiers.

Moreover, the characterization results are represented and stored as discrete values. As a consequence, it is not easy for a micro-controller used in amplifiers implementations to obtain the NF or GF for points that were not measured during the characterization.

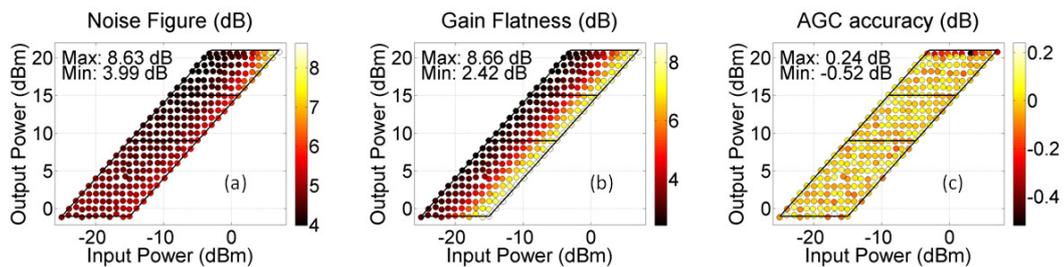


Fig. 3. Power masks with the results of the characterization in terms of (a) NF, (b) GF and (c) AGC accuracy.

### III. MULTILAYER PERCEPTRON NEURAL NETWORKS

An artificial neural network (ANN) can be defined as a structure composed by a set of simple interconnected processing units, called artificial neurons. Each neuron has a set of inputs mapped to one output. The neuron is responsible to perform a weighted sum of the inputs and to set the output according to a nonlinear activation function [7].

The simplest ANN model was proposed by Frank Rosenblatt in 1958 and is called Perceptron [8]. In this model, a set of artificial neurons is connected to just one output unit. Although the Perceptron presents the capability to “learn” through examples, and increase the accuracy of its output along the time, this type of ANN is not able to solve problems that are non-linearly separable. This problem was first presented by Minsky and Papert in [9]. Then, the Multilayer perceptron network (MLP) was proposed as an alternative to solve this problem of the Perceptron. The MLP is a generalization of the Perceptron by organizing Perceptrons in multiple interconnected layers. The traditional MLP is composed by, at least, three layers, and each layer performs a specific task. In the input layer, each neuron represents an input variable of the problem, whereas each neuron in the output layer of the MLP corresponds to a system output. The hidden layer, or the set of hidden layers, is responsible to add the capability to represent nonlinearities in the classification or regression task assigned for the MLP. However, to reach this goal, the neurons must use a non-linear activation function. The sigmoid logistic function is the most common activation function [10].

It is necessary to use an algorithm to find the set of weights that optimizes the performance of the MLP. The problem in the training of the MLP is that the error in the hidden layer is unknown, and this

error is necessary to perform the adjustment of the weights. In 1974, Werbos [11] proposed a generalization of the delta rule that was used by Widrow and Hoff to perform the training of a neural network called ADALINE [12]. The algorithm proposed by Werbos is currently called as backpropagation (BP). The main feature of the BP algorithm is the capability to propagate the error recursively through the layers of the MLP. The algorithm is divided in two steps. In the first one, the values of the neurons (signals) are propagated in the forward direction (from the input to the output layer), and the error is calculated, but the weights are not updated. In the second step, the errors are recursively propagated (from the output to the input layer) and the weights are updated according to the adjust weight rule (generalized delta rule) [10].

In [13], Hornik *et al.* showed that a MLP with as few as one hidden layer using arbitrary squashing functions (*e.g.* sigmoidal logistic) is capable to approximate any function. This implies that any lack of success in the application of MLP for approximation purposes must arise from inadequate learning, insufficient numbers of hidden neurons or the lack of a deterministic relationship between input and target. Some aspects are important to approximate functions with a MLP:

1. **Data pre-processing:** the data need to be processed before its presentation to the MLP. This processing consists, generally, in: normalization of the values, shuffle the entire dataset, and definition of the training, validation and test datasets. The normalization is necessary to avoid discrepancy among the values that will be processed by the MLP; the shuffle will help the MLP to learn from different patterns of the problem, simultaneously; and the division will define the training, validation and test datasets. We used 50%, 25% and 25% of the data for training, validation and test, respectively.
2. **Stop condition:** The training process is executed by presenting the training data to the MLP. The number of times we present all the training examples is called epochs. The proper number of epochs to stop the training is important to avoid a premature convergence, or, in the opposite, to avoid a state of extreme memorization (loss of generalization ability). The stop condition can be defined according to the validation error. In most of the problems, when the MLP starts to memorize, the validation error starts to increase, so it is the best time to stop the training. However, there are problems that the error does not increase, but stay with the same value for a long period, so the training can be stopped in this moment of stability.
3. **Number of neurons in the hidden layer:** as stated in [12], the ability of the MLP to solve non-linear problems depends on the hidden layer. Therefore, the number of neurons impacts on the network performance depending on the nonlinearity degree of the input decision space.

#### IV. MAPPING NF AND GF USING MLP

As stated in the section II, the granularity of characterization depends on the size of the step used to sweep the variables involved in this process. If one needs a high precision characterization, lots of measurements will be necessary and the characterization process will take a long time. On the other hand, if one increases the step, the process is faster, but there will be a bigger interval between the data points, which may lead to a bad approximation of the mapping function.

As shown in Fig. 4, the values of NF and GF depend on the values of the input power ( $P_{in}$ ) and output power ( $P_{out}$ ). Thus, our proposal is to use a MLP to interpolate the results of the characterization in order to create a general approximation function to express the dependency between the inputs ( $P_{in}$  and  $P_{out}$ ) and the outputs (NF and GF). Our hypothesis is that the mapping of the NF and GF using MLPs may avoid the necessity of a small *step* to obtain a characterization higher resolution. This means that, the usage of a MLP as an auxiliary tool in the characterization process can reduce the time spent to characterize an amplifier in the experimental stage, but reaching a very good precision, *i.e.* the error generated by the regression process is lower than the error of the physical layer measurement. This interpolation process can be applied to any type of EDFA, including the ones with more than one stage. This occurs because, in principle, a MLP can approximate any differentiable nonlinear function.

In order to develop this auxiliary tool, a MLP that will receive  $P_{in}$  and  $P_{out}$  as inputs, and then will return the NF and GF as outputs was designed. The MLP scheme is presented in Fig. 4. The MLP has three layers with two neurons in the input layer, two neurons in the output layer, and the number of neurons in the hidden layer will be defined in the analysis that will be described hereafter. The neurons in the input and output layers use the identity function as their activation function. On the other hand, the neuron in the hidden layer uses the sigmoidal logistic function to enable the capability of the MLP to solve nonlinear problems.

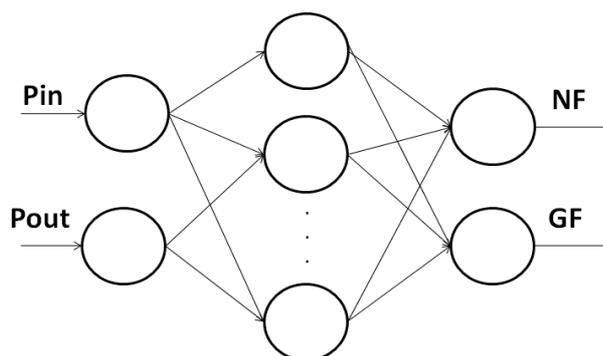


Fig. 4. The architecture of the MLP used to map  $NF$  and  $GF$  as a function of  $P_{in}$  and  $P_{out}$ .

The error used in all the stages of the process (training, validation and test) was the average absolute error considering both the GF and NF. The number of training epochs was defined after an analysis of the behavior of the training error and validation error.

## V. EXPERIMENTAL SETUP

In order to evaluate the precision of the MLP for different granularities of the characterization, we used *Power Masks* with different characterization *steps*. All the *Power Masks* were generated from a mask with the gain varying between 11 dB and 24 dB and a minimum input and a maximum output powers of -30 dBm e 14 dBm, respectively. We used a gain step of 0.5 dB with 41 points of characterization for each gain, which results in a total of 1107 points. The *step* values, with respect to input power, used in the experiments are: 1 dB, 1.5 dB, 2 dB, 2.5 dB, 3 dB, 3.5 dB, 4 dB and 6 dB. We used the gain step equal to 0.5 dB for all experiments.

We divided the data in order to evaluate the ability of the proposed MLP to find values that are not included in the training set. As an example, the training set is composed by the operation points that are equally spaced by 1 dB from their neighbor points in the *Power Mask* with *step* equal to 1 dB (called 1 dB *Power Mask*). This means that just half of the points of the 0.5 dB *Power Mask* were used for training of the MLP with *step* of 1 dB, *i.e.* 567 points. The points that are not included in the training dataset were used to compose the validation dataset and the test dataset. This does not mean that all the points that are not in the training set will be used in the validation and test sets. The number of points used for validation and test is a half the part of the total number of points in the training dataset. Table II shows the sizes of the sets for each case. One must observe that a lower number of points is necessary as the step within the power mask increases.

TABLE II. THE NUMBER OF POINTS USED FOR EACH MASK TO TRAINING, VALIDATE AND TEST.

<i>Power Mask</i>	Number of Training Points	Number of Validation Points	Number of Test Points
1 dB	567	257	257
1.5 dB	369	184	184
2 dB	277	138	138
2.5 dB	222	111	111
3 dB	186	93	93
3.5 dB	159	79	79
4 dB	139	69	69
6 dB	93	46	46

All the data points were shuffled and normalized between 0.15 and 0.85 before the training phase of the MLP. Each result presented in the next section was obtained after 30 independent trials of the complete training process of the MLP.

## VI. RESULTS

Fig. 5 shows the convergence curve of the training process using 1 dB *Power Mask* and 4 neurons in the hidden layer. The 1 dB *Power Mask* was first chosen because it presents the largest training dataset, which means that it is necessary a high number of epochs for the MLP to learn the patterns. As one can observe, the validation error did not increase with a high number of epochs, and with

5,000 epochs the process reached a stable state. Because of this, we used 5,000 epochs in the next experiments as the stop condition.

Fig. 6 and Fig. 7 present the Box-Plot of the test error for different numbers of neurons in the hidden layer with 1dB *Power Mask* and 3 dB *Power Mask*, respectively. One can observe that just some few neurons are necessary to accomplish the learning process in both cases. Hence, we used 4 neurons to avoid outliers.

Fig. 8 shows the results of the Box-Plot of the test error as a function of the step for 4 neurons in the hidden layer and 5,000 epochs varying the *step* of the *Power Mask*. As one can observe, the error of the MLP slightly increases as the step increases, but maintains an error around 0,1 dB for a maximum *step* of 3 dB.

Fig. 9 and Fig. 10 depict an example of the GF and NF curves generated by using the MLP, varying the input power and output power with resolution 0.1 dB.

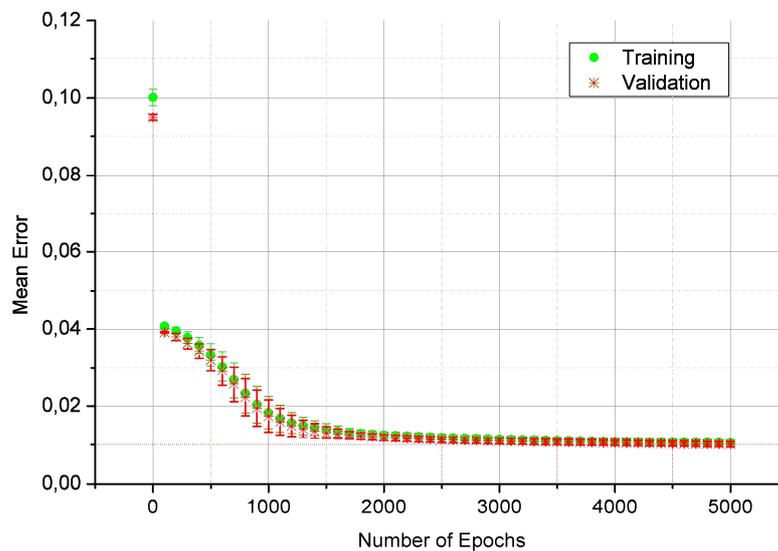


Fig. 5. Convergence curve for the validation and the test processes as a function of the number of epochs.

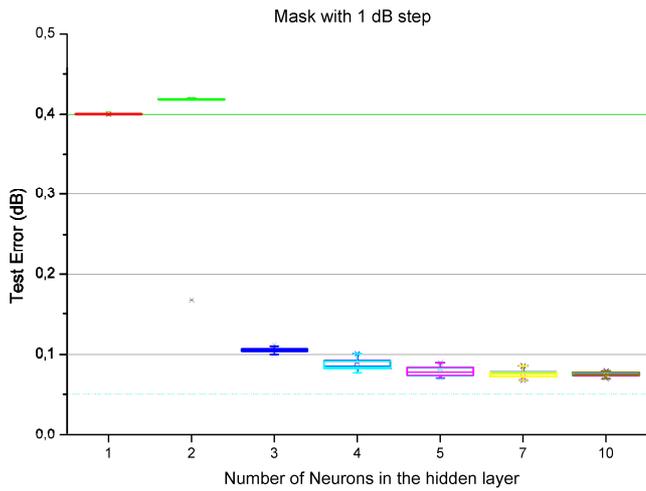


Fig. 6. Box-Plot of the test error as a function of the number of neurons in the hidden layer for 1dB Power Mask.

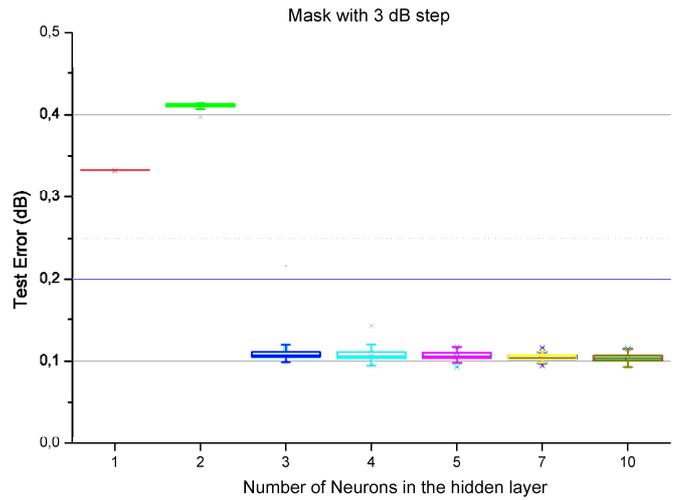


Fig. 7. Box-Plot of the test error as a function of the number of neurons in the hidden layer for 3dB Power Mask.

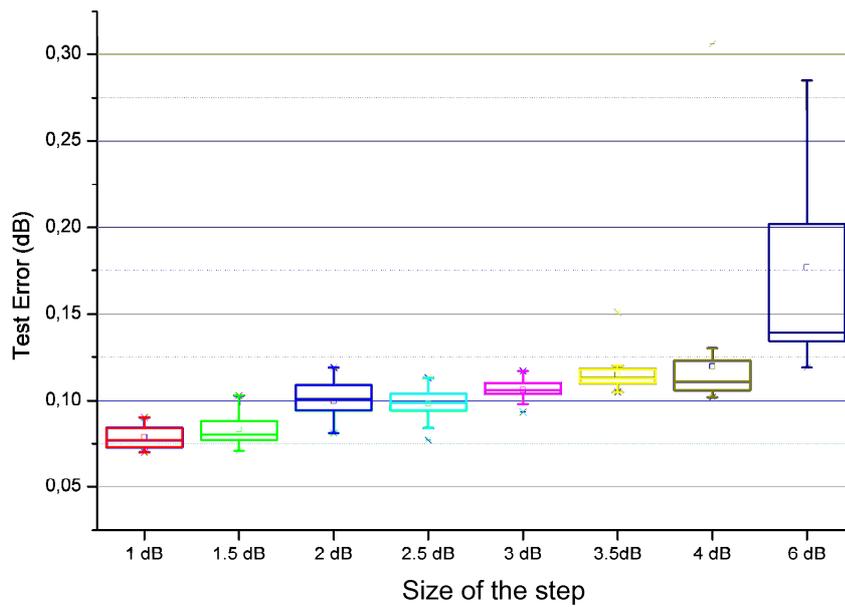


Fig. 8. Box-Plot of the test error as a function of the step for 4 neurons in the hidden layer.

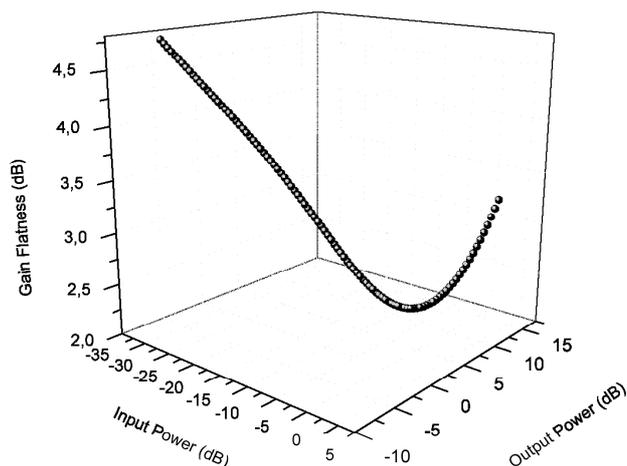


Fig. 9. Example of the GF curve generated by using the MLP, varying the input power and output power with resolution 0.1 dB.

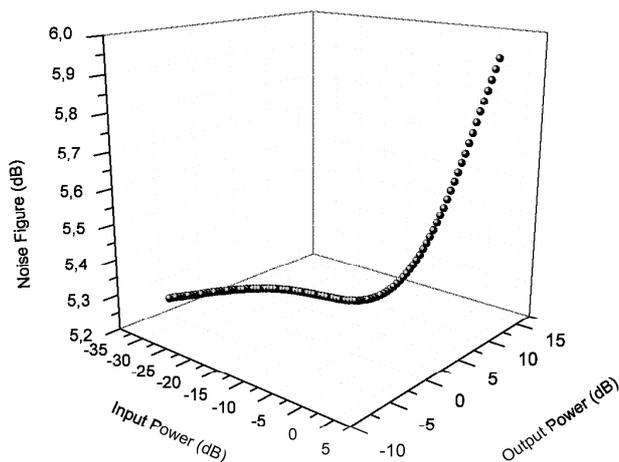


Fig. 10. Example of the NF curve generated by using the MLP, varying the input power and output power with resolution 0.1 dB.

## VII. CONCLUSIONS

Optical Amplifiers are widely deployed in optical communications. Because of this, a proper characterization within the operational range is necessary for practical usage. Currently, the characterization is performed by measuring a certain number of points of Noise Figure (NF) and Gain Flatness (GF) and the amplifier will work in one of these points. We showed that one can use Multi-Layer Perceptron to map the NF and GF as a function of the input and output powers applied to the amplifier. We also demonstrated that MLPs may avoid the necessity of a small *step* to obtain a high resolution characterization. This means that the usage of a MLP as an auxiliary tool in the characterization process can reduce the time spent to characterize an amplifier in the experimental stage by just measuring operation points with a gain interval of 3 dB, which results presenting errors as low as of 0.1 dB.

The continuous curves for the amplifier characterization, obtained by the proposed MLP scheme, will have impact on the implementation of cognitive (self adaptive) control plane EDFA, for dynamic and self-reconfigurable optical networks.

## REFERENCES

- [1] E. Desurvire, "Erbium-Doped Fiber Amplifiers, Device and System Developments", Wiley, 2002.
- [2] J. Oliveira, et al, "Demonstration of EDFA Cognitive Gain Control via GMPLS for Mixed Modulation Formats in Heterogeneous Optical Networks," in OFC Conference, OSA Technical Digest (online) (Optical Society of America, 2013), paper OW1H.2.
- [3] U. C. Moura, et al, "Caracterizador Automatizado de Máscara de Potência de Amplificadores Ópticos para Redes WDM," SBrT12, 2012, Brasília (In Portuguese).
- [4] J. Oliveira, et al, "Hybrid Distributed Raman/EDFA Amplifier with Novel Automatic Gain Control for Reconfigurable WDM Optical Networks," Momag'12, 2012, João Pessoa.
- [5] G. J. Cowle, "Challenges and opportunities for optical amplifiers in metro optical networks," in Proceedings of SPIE, vol. 7621, 2010, p 76210B.
- [6] P. C. Becker, et al, "Erbium-doped fiber amplifiers: fundamentals and technology." Academic Press, 1999.
- [7] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bulletin of mathematical biology, vol. 5, n. 4, pp. 115-133, 1943.
- [8] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," Psychological review, vol. 65, n. 6, p. 386, 1958.

- [9] M. Minsky and P. Seymour, *Perceptrons*, Massachusetts: MIT press, 1969.
- [10] S. O. Haykin, "Neural Networks and Learning Machines", 3rd Edition, Kindle edition, Amazon, 2008.
- [11] P. Werbos, *Beyond regression: New tools for prediction and analysis in the behavioral sciences*, Massachusetts: Harvard University, 1974.
- [12] B. Widrow, M. E. Hoff and others, "Adaptive switching circuits," Defense Technical Information Center, 1960.
- [13] K. Hornik, M. Stinchcombe e H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.