Machine Learning Based Prediction of Erbium-Doped Fiber WDM Line Amplifier Gain Spectra

Shengxiang Zhu [(1)], Craig L. Gutterman [(2)], Weiyang Mo [(1)],
Yao Li [(1)], Gil Zussman [(2)], Daniel C. Kilper [(1)]

[(1)] College of Optical Sciences, University of Arizona, Tucson, AZ, USA, szhu@email.arizona.edu
[(2)] Electrical Engineering, Columbia University, New York, NY, USA

Abstract *Machine learning based modelling of Erbium-Doped Fiber Amplifiers (EDFA) is used to determine wavelength dependent gain for use in optical transmission systems, and achieves root mean square error (RMSE) of 0.08, 0.18, and 0.27 dB under input ranges of +/- 3, 6, 9 dB.*

Introduction

The use of software defined networking (SDN) in wavelength division multiplexed (WDM) optical systems is enabling enhanced levels of control and adaptation. Greater automation and software control capabilities require more accurate information regarding the physical system characteristics. The optical amplifier gain spectrum for example determines the individual channel powers launched into the transmission fiber, which is important for both the impact of fiber nonlinearity induced signal impairments and the optical signal to noise ratio. The gain spectrum will also impact the optical power excursions in dynamic wavelength routing operations. Accurate channel models, with individual channel power information at each amplifier will improve the potential benefits and performance of new functionality such as software control of wavelength routing or modulation format adaptation.

Optical power divergence is managed through the use of channel power controls in wavelength selective switches in reconfigurable optical add drop multiplexer nodes. Performance margins are then used to account for the uncertainties in power at individual amplifier outputs. The ability to determine the power at each amplifier output will enable reduced margins thereby lowering system costs. Accurate power models will also improve the speed and stability of wavelength provisioning.

In previous research, methods to predict optical channel power divergence and dynamics in transmission systems include neural network based dynamic channel power estimation[1] and machine learning based power divergence prediction[2]. Furthermore, several mathematical models were introduced to predict individual channel output power under changing channel configurations, including a numerical power estimation framework used to predict EDFA output power[3], a detailed analytical model for the wavelength dependent gain impact[4] and a model designed for system applications using a simple characterization method[5], but with limited accuracy.

In this paper, we examine the use of machine learning to determine the channel configuration and input power dependent EDFA gain spectrum. Deep neural networks are built and trained to predict the gain spectrum based on the input power spectrum. As a result, all of the channel power related amplifier effects are captured in the process. The resulting trained neural network for each individual EDFA provides a computational tool for use in quality of transmission (QoT) estimation. Thus, by employing machine learning to characterize each amplifier in a system prior to deployment, accurate wavelength dependent gain models are potentially made available for improved QoT estimation.

WDM channel gain models

Gain dynamics occur in automatic gain controlled (AGC) WDM line EDFAs due to wavelength dependent gain[4], where the gain excursion $\hat{g}$ due to a change in channel powers P_j can be written:

$$\hat{g}(\lambda_i) = \frac{G_{TC}}{G_M}\left[\frac{\sum_j P_j + N_I + N_C}{\sum_j P_j g_j t_j + g_R N_R + g_I N_I}\right] g(\lambda_i) \quad (1)$$

Accurate channel output power estimation therefore requires the gain ripple g_j, tilt t_j, and input noise N_I and gain g_I, amplifier noise N_R and gain g_R, and the amplifier noise compensation factor N_C. Many of these parameters such as g_j and t_j are also dependent on the input channel configuration through the internal amplifier gain. Other effects such as spectral hole burning may also play a role.

In another study, a model was proposed for calculating the gain spectrum based on measurement of the single channel ripple and WDM ripple functions[5], which can be written as:

$$\hat{g}(\lambda_i) = g(\lambda_i) + \frac{\sum_{j=1}^{n}[g_s(\lambda_j) - g(\lambda_j)]}{n} \quad (2)$$

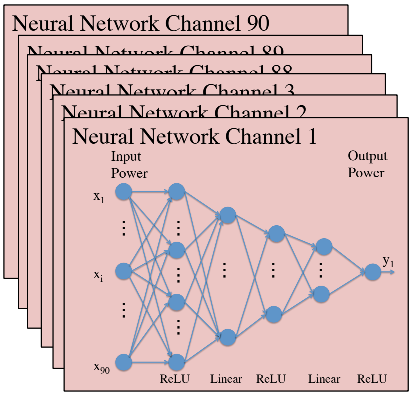

Fig. 1 Architecture of neural networks.

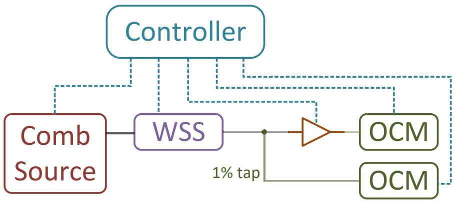

Fig. 2 Experiment setup for data capture.

In this expression, $\hat{g}(\lambda_i)$ is the gain spectrum of wavelength λ_i when a set of wavelengths $\{\lambda_1, \ldots, \lambda_n\}$ is input to the EDFA. $g(\lambda_i)$ represents the characterized WDM gain spectrum, i.e., the gain spectrum when all WDM input channels are active. $g_s(\lambda_i)$ denotes the characterized single channel gain spectrum, i.e., the spectrum of the gain of each channel λ_j when no other channels are present. This center of mass (CM) approach conveniently provides an estimate from easily measured configurations, but misses detailed effects that can be important, particularly in certain corner cases.

Machine learning modeling

A supervised machine learning algorithm is designed to train a neural network model of EDFAs to predict the gain spectrum based on the input power spectrum. The Neural Network (NN) architecture is implemented with TensorFlow.

Ninety features are used as the input to the NN, representing the power levels of each of the 90 channels. A separate NN is created for each output channel. Data is divided into 3 classes: training data, validation data, and test data. The training data is used to train the NN to minimize the Mean Square Error (MSE) loss function. The validation data is used to determine which parameters provide optimal performance after using the training data. The test data is used to evaluate the trained model.

The resulting parameters for the NN architecture are described as follows. All power levels are converted into decimal power levels, normalized, and scaled by a factor of 300. Each

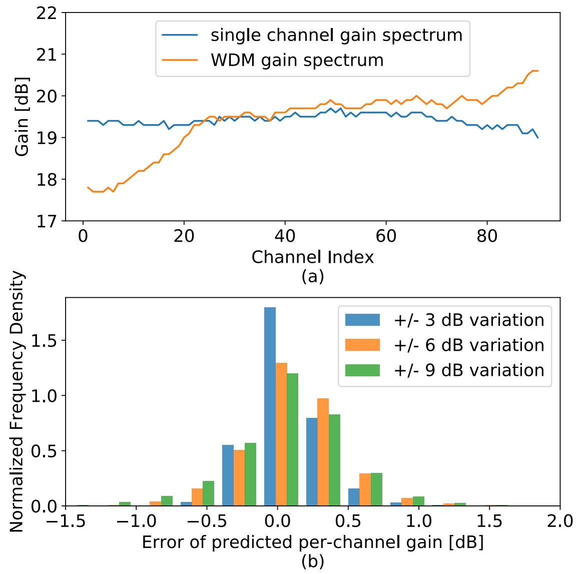

Fig. 3 (a) Measured WDM and single channel gain spectral. (b) Normalized frequency density function of prediction error using the CM analytical model.

neural network has 4 hidden layers with artificial neuron transfer function of ReLU (rectified linear unit), Linear, ReLU, Linear, and ReLU. The full NN architecture can be seen in Fig. 1. The model is trained by minimizing the MSE loss function using stochastic gradient descent with back-propagation with a mini batch size of $m = 60$ and a learning rate, $\alpha = 0.00025$. Training is done over 15000 iterations.

Experiment setup and results

As shown in Fig. 2, a 90-channel comb source is used to generate the WDM input. The wavelength selective switch is used to control the input power spectrum. The EDFA is configured to have 3 dB tilt and work in AGC mode, with target gain set as 18 dB. Two optical channel monitors are used to monitor the input and output power spectrum. The controller is used to communicate with all of these devices and capture data.

First, we characterized the single channel and WDM ripple of the EDFA and built a model using Eq. 2. In Fig. 3(a), the typical single channel and WDM input gain spectra are shown. We can see that for different channel loading the gain spectra are quite different.

Then we used captured samples of measured input and output spectra to evaluate the error of analytical model derived from Eq. 2. In Fig. 3(b), the normalized frequency density (NFD) of the error of analytical model is shown. It is shown that wider dynamic range of the input power leads to higher error in the prediction of the gain spectrum.

To build the NN, we captured data samples of the input and output spectra under different channel loadings. We then compare the root

Tab. 1: RMSE of analytical and ML model

Dynamic range [dB]	Analytical [dB]	Machine learning [dB]
+/- 3	0.247	0.081
+/- 6	0.359	0.180
+/- 9	0.410	0.266

mean square error (RMSE) distribution of the CM analytical model and the ML model. In Tab. 1, the dynamic range is defined as the range over which the input channel power is varied around the target power (-18 dBm) in a uniform random distribution. It is shown in the table that the ML model is able to reduce the RMSE by 67%, 50% and 35% in scenarios where dynamic range is +/- 3 dB, +/- 6 dB and +/- 9 dB.

We then analyze the error distribution of the analytical model and the ML model using the same test set, shown in Fig. 4(a), Fig. 4(c) and Fig. 4(e). In the case of +/- 3 dB dynamic range, the ratio of the errors below 0.5 dB is 99.95% for the ML model and 94.9% for the analytical model. In the +/- 6 dB case, the ratio is 98.37% vs. 85.11%. In the +/- 9 dB case, the ratio is 93.57% vs. 81.07%. The ML model shows better performance in gain spectrum prediction for all dynamic range values.

Since the most severe channel power dynamics occur when there is a handful of open channels, it is of interest to analyze how the ML models perform in these corner cases. Here we look at the corner case with just two channels turned on. In Fig. 4(b), Fig. 4(d) and Fig. 4(f), the error distributions of these corner cases are shown. With dynamic ranges of +/- 3, 6, 9 dB, the ratios of prediction errors below 0.5 dB are 99.38% vs. 95.88%, 98.57% vs. 87.78% and 95.62% vs. 83.18%, respectively. The ML model is able to avoid high errors (over 0.5 dB) for most of the cases, outperforming the analytical model.

To verify the performance of ML modeling under different EDFA configurations, we adjusted the gain setting of the EDFA and test the error of the models. In Fig. 5, the EDFA with 14 dB and 22 dB gain were modeled and tested and the error distributions are shown. The overall RMSE error reduction in +/- 9 dB dynamic range is 10.37% (0.3535 vs. 0.3944) for 14 dB gain and 15.21% (0.3501 vs. 0.4129) for 22 dB.

The EDFA spontaneous emission noise will show up at the downstream amplifier inputs on blocked channels if more than one amplifier is used between nodes (before being blocked again at the next node). In this ML model, noise power outside the dynamic range is ignored. Test results further show that the RMSE reduction is 51.71% (0.2368 vs. 0.4904), 33.18% (0.2346 vs. 0.3511), 9.44% (0.2840 vs. 0.3136), with blocked channel noise levels of -40, -35, -30 dBm per channel, respectively.

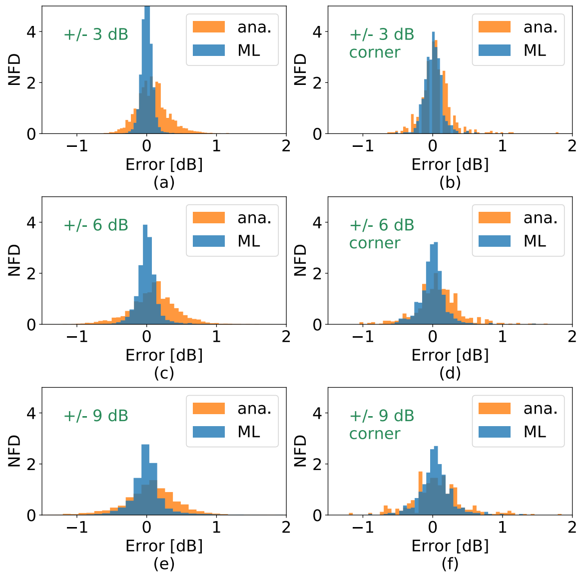

Fig. 4: Error distribution of analytical model and ML model with dynamic range of +/- 3, 6, 9 dB.

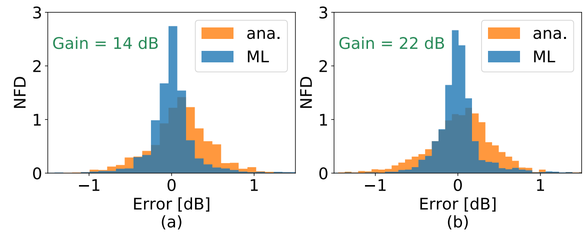

Fig. 5: Error distribution of analytical model and ML model with gain value of 14 dB and 22 dB.

Conclusion

A machine learning based modeling method is analyzed and shown to reduce the gain estimation error compared with analytical models. For future studies, it will be of interest to evaluate the prediction performance in QoT using this model in large scale optical networks.

Acknowledgements

This work was supported by the National Science Foundation ECC-0812072, DGE-16-44869, CNS-1650685, CNS-1650669, and Department of Energy DE-SC0015867.

References

[1] C. L. Gutterman, et al. Big-DAMA workshop 37-42 (2017).
[2] Y. Huang, et al. *Opt. Express* 25.3 (2017).
[3] Fei, Yue, et al. ICC 5222-5227(2015).
[4] J. Junio, et al. J. Opt. Comm. Networks 4.9 A1-A7 (2012).
[5] K. Ishii, et al. OFC W3E.4 (2014).