Hybrid Machine Learning EDFA Model

Shengxiang Zhu[1], Craig Gutterman[2], Alan Diaz Montiel[3], Jiakai Yu[1],
Marco Ruffini[3], Gil Zussman[2], Daniel Kilper[1]
(1) University of Arizona, (2) Columbia University, (3) Trinity College Dublin
szhu@email.arizona.edu

Abstract: A hybrid machine learning (HML) model combining a-priori and a-posteriori knowledge is implemented and tested, which is shown to reduce the prediction error and training complexity, compared to an analytical or neural network learning model.
OCIS codes: 060.2320, 060.4510

1. Introduction

The traffic and latency demands in 5G backhaul and data center interconnection (DCI) networks has kept growing. It would be beneficial for optical networks to have dynamic channel reconfigurability to accommodate the changing traffic pattern and minimize the latency and power consumption. Within the fiber links, erbium doped fiber amplifiers (EDFA) are widely used to compensate for the signal loss due to the insertion loss of fiber and reconfigurable optical add-drop multiplexer (ROADM) components. In transparent optical networks, optical signals transmitted through network nodes are not regenerated and as a result the quality of the optical signals needs to be estimated before an optical channel is provisioned. Since optical channels share the same fiber, the add/drop operation of optical channels may have impact on other channels, such as crosstalk, fiber nonlinearity, and channel power excursion. Wavelength dependent gain in EDFAs is a significant source of variation and uncertainty in channel performance estimation. Recently there has been much interest in developing machine learning based models to account for the detailed EDFA behavior [1-3], replacing lookup tables or analytical models [4-9]. However, these machine learning models are built solely from a-posteriori knowledge which is trained from the experimental data, ignoring the existing a-priori knowledge. In this paper, we proposed a hybrid machine learning (HML) model for EDFAs, which combines an analytical model with a neural network machine learning model to achieve higher prediction accuracy while reducing the training complexity, in both the training time and the size of the training sample data sets.

1.1. Analytical models of EDFA – a-priori knowledge

The EDFA gain spectrum can be modelled using a center of mass function based on single channel and fully channel loaded ripple functions, shown in Eq. 1 [5].

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

In this expression, $\hat{g}(\lambda_i)$ is the wavelength dependent gain of wavelength λ_i when a set of wavelengths $\{\lambda_1, \dots, \lambda_n\}$ is input to the EDFA. $g(\lambda_j)$ is the wavelength division multiplexed (WDM) gain spectrum, i.e., the gain spectrum when all WDM input channels are input to the amplifier at the desired signal power and mean gain value. $g_s(\lambda_j)$ is the measured single channel gain at λ_j. An example of measured $g(\lambda_j)$ and $g_s(\lambda_j)$ is shown in Fig. 1. This center of mass equation provides an estimate of the wavelength dependent gain based on simple EDFA measurements, but it does not capture many important features of the EDFA wavelength dependent gain and its behavior over the full operating range.

1.2. Machine learning models of EDFA – a-posteriori knowledge

Recent research on ML based EDFA modeling uses deep neural networks often implemented in TensorFlow. A typical deep neural network contains an input layer, multiple hidden layers, and an output layer. The input layer contains information regarding the input and configuration of the EDFA, e.g. input channel power, gain and tilt settings. The output layer contains the gain or output power spectrum. Since the gain spectrum is a relatively smooth function of wavelength, a few sample channels can be used to quickly find the gain profile [8]. When enough samples are collected and neural network parameters (e.g. activation function, batch size, learning rate) are carefully chosen, the prediction accuracy of ML models were shown to outperform analytical models [6, 7].

2. Hybrid machine learning model

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.
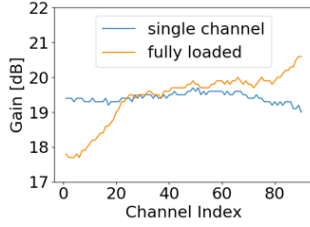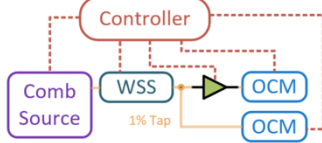
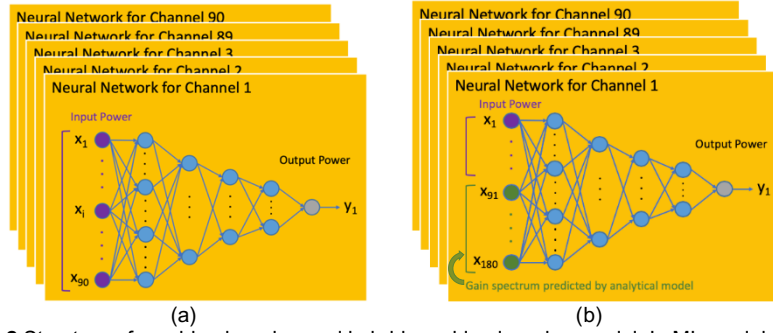**Fig. 1** EDFA gain spectrum

Fig. 3 Data capturing setup

Fig. 2 Structure of machine learning and hybrid machine learning model. In ML model, 90 features (x_1-x_{90}) of input channel power are used and 4 hidden layers have 90 neurons each. The output is the optical channel power. In HML model, 90 features (x_1-x_{90}) of input channel power and another 90 features (x_{91}-x_{180}) of gain spectrum predicted by analytical model are used. The hidden layers have 180, 90, 90, 45 neurons.

For the regular ML model, ninety features are used as the input to the NN, representing the power levels of each of the 90 channels. The resulting parameters for the NN architecture are described as follows. All power levels are converted into decimal power levels and normalized. Each neural network has 4 hidden layers and an output layer with artificial neuron transfer function of ReLU (rectified linear unit), Linear, ReLU, Linear, and ReLU. Each of the first 4 layers contain 90 neurons and the last layer contains 1 neuron. The full NN architecture is shown in Fig. 2(a).

For the hybrid ML model, one hundred eighty features are used as the input to the NN. The first ninety features represent the power levels of each of the 90 channels. The second 90 features are the predicted output power levels determined by the analytical model. All power levels are converted into decimal power levels and normalized. Each neural network has 4 hidden layers and an output layer with artificial neuron transfer functions of ReLU (rectified linear unit), Linear, ReLU, Linear, and ReLU. Each layer contains 180 neurons, 90 neurons, 90 neurons, 45 neurons, and 1 neuron, respectively. The full NN architecture can be seen in Fig. 2(b).

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 model is trained by minimizing the MSE loss function using stochastic gradient descent with backpropagation with $m = 60$ mini batch size and a learning rate alpha = 0.00025.

3. Experimental Setup

As shown in Fig. 3, optical channels are generated by a comb source, which is filtered by wavelength selective switch (WSS) to select the active channels and fine tune the channel power. EDFA is the device under test (DUT) and has setting of 3 dB tilt and 18 dB gain (other EDFA settings are also tested, which have similar results). Automatic gain control (AGC) mode is used by the EDFA to ensure the total gain is 18 dB. Optical channel monitors (OCM) are used to capture the input and output channel powers. A controller is used to remotely control all the devices and automate the data capturing process.

4. Results

In the experiment, the EDFA is configured to have 18 dB gain and the input channels are randomly set to be 'on' or 'off', with -18 dBm +/- 6 dB variation in optical channel power for each "on" channel. Three dimensions of machine learning system performance are evaluated, as shown below.

First is the accuracy of HML model with abundant data samples (12000 samples is used) and unlimited training time (25000 iterations is used). The mean square error (MSE) of the predicted channel power is 0.362 dB (analytical), 0.160 dB (ML), 0.144 dB (HML), in which HML has a 10.5% reduction of error, compared with ML model. In the worst case, HML has 1.15% of *high error* (error > 0.5 dB), which is lower than ML (1.63%). As a result, the ultimate performance of HML is slightly better than ML given unlimited data and time. However, when data and time are limited, HML has a much better performance than ML. In Fig. 4, the normalized frequency density of error shows that HML has much narrower error distribution than ML, when the number of iterations is limited to 5000 (originally 25000, training sample size is the same as original). From Fig. 5(a), the convergence process of the models is shown,
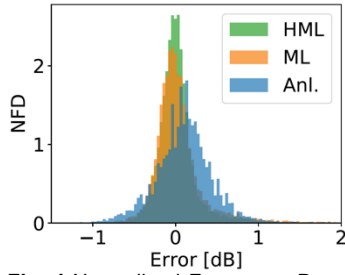
Fig. 4 Normalized Frequency Density of Analytical, ML, HML after 5000 iterations.

(a)   (b)
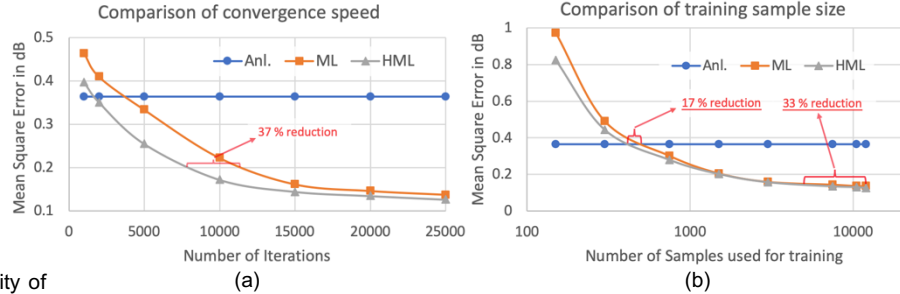
Fig. 5 Comparison of models in convergence speed and size of samples.

where the HML has a faster speed. In fact, considering a target MSE of 0.2 dB, HML requires 37% less time than the ML model (with the same ML configuration).

The third dimension is the number of training samples required. In Fig. 5(b), the MSE with different numbers of samples are shown for the three models. With fewer training samples, HML can achieve the same performance as ML. For example, to achieve the ultimate 0.134 dB MSE by the ML model, HML needs 33% fewer training samples. In another example, in order to surpass the prediction accuracy of the analytical model, HML needs 17% fewer training samples than the ML model.

Greater accuracy in the channel power prediction for EDFAs is expected to improve the accuracy of channel performance or QoT estimation, which is critical for the control and management of optical systems. In a typical optical system, the accuracy of power control and measurement is usually around 0.1 dB, which means that the 0.144 dB prediction error of the HML model is close to the system limitation. Although the HML can largely reduce the training sample size and training time relative to the ML approach, it still requires a longer data capturing process, compared to the analytical model, and further progress is needed to reduce this training process, for example through the use of transfer learning.

5. Conclusion

In this paper, we examined a hybrid machine learning model which takes advantage of an analytical model as input to the machine learning model. Based on experimental measurements, this HML model is shown to increase prediction accuracy of the output optical power spectrum of an EDFA by 10.5%, reduce the training sample size by 33% and reduce the training time by 37%.

6. Acknowledgements

References

[1] Gutterman, Craig L., et al. "Neural network based wavelength assignment in optical switching." Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks. ACM, 2017.

[2] Mo, Weiyang, et al. "Deep-neural-network-based wavelength selection and switching in ROADM systems." Journal of optical communications and networking 10.10 (2018): D1-D11.

[3] Mahajan, Ankush, et al. "Machine Learning Assisted EDFA Gain Ripple Modelling for Accurate QoT Estimation." 2019 European Conference on Optical Communication (ECOC). IEEE, 2019.

[4] Junio, Joseph, et al. "Channel power excursions from single-step channel provisioning." IEEE/OSA Journal of Optical Communications and Networking 4.9 (2012): A1-A7.

[5] Ishii, Kiyo, et al. "Experimental investigation of gain offset behavior of feedforward-controlled WDM AGC EDFA under various dynamic wavelength allocations." IEEE Photonics Journal 8.1 (2016): 1-13.

[6] Zhu, Shengxiang, et al. "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra." 2018 European Conference on Optical Communication (ECOC). IEEE, 2018.

[7] You, Yuren, et al. "OSNR prediction using machine learning-based EDFA models." 2019 European Conference on Optical Communication (ECOC). IEEE, 2019.

[8] Mo, Weiyang, et al. "EDFA wavelength dependent gain spectrum measurement using weak optical probe sampling." IEEE Photonics Technology Letters 30.2 (2017): 177-180.

[9] D'Amico, Andrea, et al. "Machine-learning aided OSNR prediction in optical line systems." 2019 European Conference on Optical Communication (ECOC). IEEE, 2019.