Transfer Learning-based ROADM EDFA Wavelength Dependent Gain Prediction Using Minimized Data Collection

Zehao Wang[1], Dan Kilper[2], and Tingjun Chen[1]
[1]*Duke University, Durham, NC, USA;* [2]*CONNECT Centre, Trinity College Dublin, Ireland*
[1]*{zehao.w, tingjun.chen}@duke.edu,* [2]*dan.kilper@tcd.ie*

Abstract: We implement and test transfer learning-based gain models across 16 ROADM EDFAs, which achieve less than 0.17/0.30 dB mean absolute error for booster/pre-amplifier gain prediction using only 0.5% of the full target EDFA dataset.

1. Introduction

The high bandwidth and low latency requirements of next-generation wireless (e.g., 5G/6G) and wide-area (e.g., data center interconnect, DCI) networks rely on optical infrastructure employing wavelength-division multiplexing (WDM) technologies that are deployed at different scales in metro, regional, and long-haul networks. A key component to compensate optical link loss is the erbium-doped fiber amplifier (EDFA), whose output spectrum depends on its wavelength dependent gain profile and can impact end-to-end system performance such as the link optical signal-to-noise ratio (OSNR) and quality of transmission (QoT). However, it is challenging to estimate the wavelength dependent gain spectrum of an EDFA as it depends on many factors such as the input power and gain setting, optical channel loading configurations, and internal component parameters.

Recent work has focused on leveraging machine learning (ML) techniques such as deep neural networks (DNNs) to build EDFA gain models [1], and on the generalization of the ML-based EDFA models to multiple EDFAs of the same make by training datasets collected from all devices [2]. Although the DNN-based EDFA gain model can achieve high gain spectrum prediction accuracy, it requires the collection of comprehensive EDFA gain spectrum measurements for *each* EDFA, for example, collecting a full set of gain spectrum measurements from a *single* EDFA covering different gain settings and diverse channel loading configurations can take up to 51 hours [3]. A promising solution to overcome this challenge is *transfer learning (TL)* [4], which allows for building a new target ML model based on a pre-trained source model that shares similar model knowledge using very few data samples collected from the target domain.

In this paper, we investigate TL-based EDFA gain models, and show that using only 0.5% of the new data collected from the target EDFA (13 measurements), the transferred target model can achieve similar gain prediction accuracy compared to the source model with the full training set (2,678 measurements). We demonstrate three different scenarios that can benefit from TL with largely reduced EDFA data collection process: (*i*) TL between EDFAs of the same type (booster or pre-amplifier); (*ii*) TL between different EDFA gain settings, and (*iii*) TL between different EDFA types.

2. EDFA Gain Spectrum Measurements Dataset

We use an open dataset [3] consisting of gain spectrum measurements collected from 16 EDFAs within 8 commercial-grade Lumentum ROADM-20 units deployed in the PAWR COSMOS testbed [5], each with 2 EDFAs (booster/pre-amplifier). Fig. 1 shows the dataset structure, where the gain spectrum of each booster or pre-amplifier EDFA is measured at target gain settings of 15/18/21 dB and 15/18/21/24/27 dB, respectively, in the high gain mode with 0 dB gain tilt. For each EDFA at a given gain setting, the dataset contains a total number of 3,168 gain spectrum measurements across 95×50 GHz channels between 191.325 THz–196.075 THz in the C-band, which are recorded in the machine actionable `json` file format. The gain spectrum measurements of each EDFA are also collected under a diverse set of **Fixed** and **Random** channel loading configurations. In particular, the **Fixed** category includes a set of carefully selected fully/half loaded channels (*Fixed Baseline*), goalpost channels (*Fixed Goalpost*), and single/double (adjacent) channels (*Fixed Extra*). The **Random** category includes a set of randomly loaded channels with small (*Random Baseline*) and large (*Random Extra*) numbers of channel counts. The EDFA total input/output power and individual optical channel power readings are recorded with 0.01 dB and 0.1 dB resolution, respectively.

3. DNN-based EDFA Gain Model with Transfer Learning (TL)

We construct the *source DNN-based EDFA gain model* for each booster and pre-amplifier with three gain settings (15/18/21 dB) for predicting its gain spectrum. Fig. 2 shows the DNN model architecture, which consists of an input layer, four hidden layers with 256/128/128/128 neurons, and an output layer, where the neurons are initialized by the

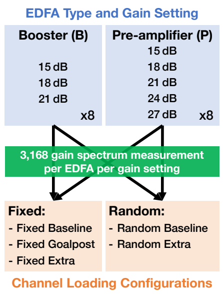

Fig. 1: Structure of the open EDFA gain spectrum measurements dataset [3].

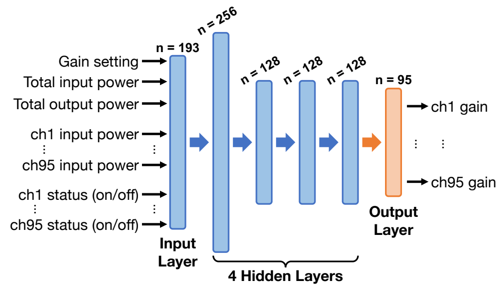

Fig. 2: Architecture of the deep neural network (DNN) model used for EDFA gain prediction and transfer learning (TL) with all fully connected layers.

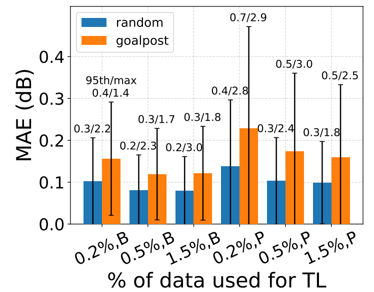

Fig. 3: MAE of EDFA gain prediction accuracy with varying target to source data size ratios ($N_{\text{tgt}}/N_{\text{src}}$) used for TL.

Kaiming normalization. The input features to the DNN model include the EDFA gain setting, total input/output power, input power spectrum, and a binary vector indicating the channel loading configuration. The output layer predicts the EDFA gain spectrum. For the input and hidden layers, we apply batch normalization and use the exponential linear unit (ELU) activation function. The source DNN model for an EDFA is trained using the mean square error (MSE) across all loaded channels (with a gradient clipping threshold of 3.0) as the loss function, with a learning rate of 0.001 over 600 epochs. We split the EDFA gain measurement dataset at each gain setting into training/test sets with a split ratio of 0.85/0.15: the source model uses $N_{\text{src}} = 2,678$ gain measurements at each gain setting as the training set, and the remaining 490 measurements as the test set. Specifically, the test set includes all *Fixed Goalpost* (270 measurements) and 20% of the *Random Baseline* (220 measurements) EDFA gain spectrum measurements, which represent a diverse set of channel loading configurations with randomly selected channels and groups of close-by channels.

To transfer a DNN-based source model to a *target model*, we apply the following procedure for TL. First, we freeze the input layer and all four hidden layers of the DNN (see Fig. 2), which are treated as the feature extractor of the DNN model, and reinitialize the weights of the output layer using Kaiming normalization. Then, the DNN model is re-trained using the same MSE loss function with a step size of 0.05 over 150 epochs. Finally, all layers are unfrozen and fine tuned with a step size of 0.001 over 20 epochs, while the batch normalization parameters are kept unchanged. Based on the pre-trained source model for one EDFA (using $N_{\text{src}} = 2,678$ measurements), we use only $N_{\text{tgt}} = 13$ "new" measurements collected from the target EDFA at each gain setting to construct the target model, which is then tested using a test set with the same random and goalpost channel loading configurations as that used by the source model.

4. Results

Using the dataset and DNN-based EDFA model described above, we first investigate for a given (pre-trained) source model, how much new data is needed from a target EDFA. We consider all cases where each booster/pre-amplifier EDFA serves as the source model, which is then transferred to each of the 7 other booster/pre-amplifier EDFAs using different sizes of target EDFA datasets. Fig. 3 shows the mean absolute error (MAE) and the 95$^{\text{th}}$/max error of the EDFA gain prediction accuracy averaged across all possible source-target model pairs for the random and goalpost test sets. We consider three different numbers of gain spectrum measurements at each gain setting from the target EDFA for TL: 5 ($N_{\text{tgt}}/N_{\text{src}} = 0.2\%$, fully loaded channels), 13 ($N_{\text{tgt}}/N_{\text{src}} = 0.5\%$, full/half loaded channels), and 40 ($N_{\text{tgt}}/N_{\text{src}} = 1.5\%$, fully/half/single/double loaded channels). The results show that the average EDFA gain prediction accuracy of the target model with $N_{\text{tgt}}/N_{\text{src}} = 0.5\%$ outperforms that achieved with a 0.2% target-source data size ratio, but is comparable to that achieved with a 1.5% ratio. Therefore, we empirically select $N_{\text{tgt}} = 13$ in the rest of the evaluations, which reduces the target data size by 200$\times$ while achieving an MAE of <0.2 dB across all EDFAs.

Below, we consider three TL scenarios: (*i*) TL between EDFAs of the same type, (*ii*) TL between gain settings of the same EDFA, and (*iii*) TL between EDFA types. Our results show that TL is able to achieve high EDFA gain prediction accuracy with very small target data size ($N_{\text{tgt}}/N_{\text{src}} = 0.5\%$) and therefore can largely reduce the measurement time.

TL between EDFAs of the Same Type Fig. 4 shows the MAE matrices across 8 EDFAs of the same type (booster/pre-amplifier) under the random and goalpost test sets, with three gain settings (15/18/21 dB) and a target data size of $N_{\text{tgt}} = 13$ ($N_{\text{tgt}}/N_{\text{src}} = 0.5\%$). In each MAE matrix, (*i*) entry $(i,i), i = 1,\dots,8$, corresponds to the component-level DNN-based EDFA model without TL, and (*ii*) entry $(i,j), j \neq i$, corresponds to the transferred EDFA model with the i^{th} and j^{th} EDFA being the source and target model, respectively. It can be seen that TL achieves better average gain prediction accuracy for booster EDFAs, and suffers from lower accuracy under goalpost channel loading configurations. Overall, the MAE for target booster/pre-amplifier model is within 0.17/0.30 dB across the entire random and goalpost test sets.

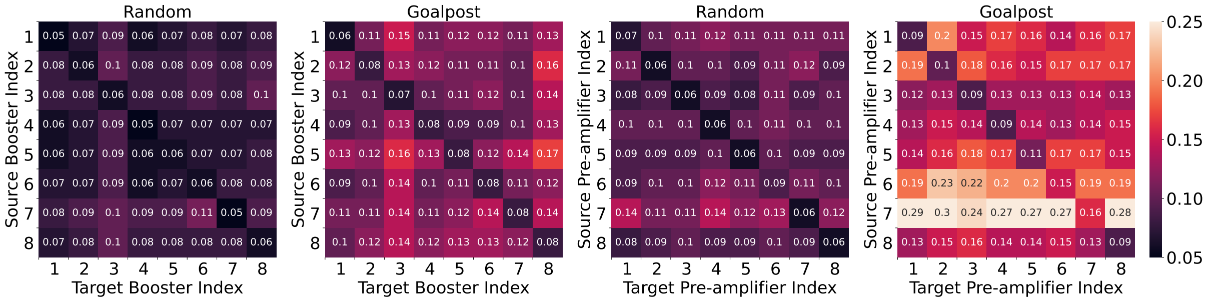

Fig. 4: Mean absolute error (MAE) matrix of ML-based EDFA gain prediction averaged across the random and goalpost test sets, where entry (i,i) corresponds to the DNN-based EDFA model without transfer learning (TL), and entry $(i,j), i \neq j$ corresponds to the TL-based EDFA model with the i^{th} and j^{th} EDFA being the source and target model, respectively.

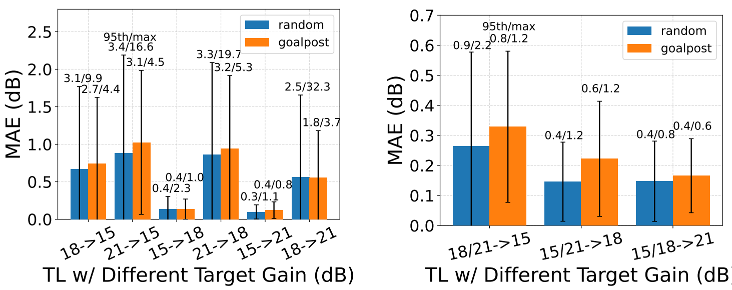

Fig. 5: TL from one source gain setting (left) or two source gain settings (right) to another target gain setting on the same booster EDFA.

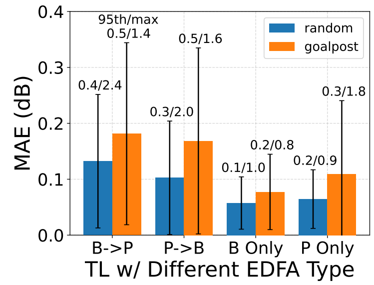

Fig. 6: TL between different EDFA types (B: booster, P: pre-amplifier).

We expect that the performance of the target model can be further improved by including (a small number of) gain measurements under the random/goalpost channel loading configurations in the target data.

TL between Gain Settings of the Same EDFA Fig. 5 shows the MAE and the 95$^{\text{th}}$/max error of the EDFA gain prediction accuracy of TL from one source gain setting (left) or two source gain settings (right) to another target gain setting of the same booster EDFA, using the same random and goalpost test sets. The results show that TL using a single source gain setting can result in an MAE of up to 0.8/1.0 dB (21 dB$\rightarrow$15 dB) under the random/goalpost test set. In such a case, the MAE for the random/goalpost test sets can be largely reduced to 0.26/0.32 dB with the additional domain knowledge from a second gain setting of 18 dB (i.e., 18/21 dB$\rightarrow$15 dB). Similar MAE performance for TL between different gain settings are also observed on the other booster and pre-amplifier EDFAs.

TL between EDFA Types Fig. 6 shows the MAE and 95$^{\text{th}}$/max error of the EDFA gain prediction accuracy when transfer for a source booster model to target pre-amplifier model (B$\rightarrow$P) or vice versa (P$\rightarrow$B), compared to the DNN-based model without TL (B$\rightarrow$B and P$\rightarrow$P). The MAE achieved by the target model is all within 0.2 dB, and TL introduces an MSE degradation of only 0.07/0.05 dB and 0.09/0.05 dB for the booster/pre-amplifier EDFA compared to that achieved by the source model under the random and goalpost test sets, respectively.

5. Conclusions

Using an open ROADM EDFA gain spectrum dataset, we investigated TL-based EDFA gain models that can achieve an MAE of less than 0.3 dB using only 0.5% of the full dataset. We showed that the EDFA gain models can be transferred between different EDFAs of the same type, different gain settings on the same EDFA, and different EDFA types.

Acknowledgments. The work was supported by NSF grants CNS-1827923, OAC-2029295, CNS-2112562, CNS-2211944, Science Foundation Ireland under Grant #13/RC/2077_P2, a Google Research Scholar Award, and an IBM Academic Award.

References

[1] Y. You, Z. Jiang, and C. Janz, "Machine learning-based edfa gain model," in *2018 European Conference on Optical Communication (ECOC),* (IEEE, 2018), pp. 1–3.

[2] F. Da Ros, U. C. De Moura, and M. P. Yankov, "Machine learning-based edfa gain model generalizable to multiple physical devices," in *2020 European Conference on Optical Communications (ECOC),* (IEEE, 2020), pp. 1–4.

[3] Z. Wang, D. C. Kilper, and T. Chen, "A comprehensive open dataset for EDFA gain spectrum measurements," https://wiki.cosmos-lab.org/wiki/Datasets (2022).

[4] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. on knowledge data engineering **22**, 1345–1359 (2009).

[5] T. Chen *et al.*, "A software-defined programmable testbed for beyond 5G optical-wireless experimentation at city-scale," IEEE Netw. **36** (2022).