Optical Fiber Communication Conference (OFC) © 2024 Optica Publishing Group

Multi-Span Optical Power Spectrum Prediction using ML-based EDFA Models and Cascaded Learning

Zehao Wang[1,2], Yue-Kai Huang[2], Shaobo Han[2], Ting Wang[2], Dan Kilper[3], and Tingjun Chen[1]

[1]Duke University, Durham, NC, USA; [2]NEC Laboratories America, Princeton, NJ, USA;
[3]CONNECT Centre, Trinity College Dublin, Ireland
[1]{zehao.w, tingjun.chen}@duke.edu, [2]{kai, shaobo, ting}@nec-labs.com, [3]dan.kilper@tcd.ie

Abstract: We implement a cascaded learning framework using component-level EDFA models for optical power spectrum prediction in multi-span networks, achieving a mean absolute error of 0.17 dB across 6 spans and 12 EDFAs with only one-shot measurement. © 2023 The Author(s)

1. Introduction

Today's telecommunication and cloud service providers rely on optical networks for the exchange of vast volumes of data at high speed and low latency, leveraging reconfigurable optical add-drop multiplexer (ROADM) units and flex-grid dense wavelength division multiplexing (DWDM) techniques. When DWDM signals traverse through an optical link with multiple spans, the signal carried by individual wavelengths can experience different propagation characteristics due to many factors including the wavelength dependent gain of erbium-doped fiber amplifiers (EDFAs) and losses induced by the fiber and other network components. Accurate estimation of end-to-end optical link performance such as the power spectrum evolution and optical signal-to-noise ratio (OSNR) is important for guaranteed quality of transmission (QoT) as well as network planning, maintenance, and configuration [1].

There are two main methods to model a multi-span optical link with various components for predicting the end-to-end link performance. The first method is via the direct cascade of component-level models, each corresponding to different link components, with proper calibrations. These include a range of mathematical, statistical, and machine learning (ML-) based models. The second method is to treat the entire multi-span link as a single entity and characterize the link using an end-to-end (E2E) model. Recent works have shown that characteristics of individual components such as the EDFA gain profiles can be effectively modeled using ML [2, 3] and used to predict optical power signal spectrum in multi-span networks [4]. However, these models perform poorly and can incur accumulated prediction errors via direct cascade. In addition, although it has been shown that an end-to-end model based on ML can improve the accuracy of QoT prediction in multi-span networks, it can be time consuming to collect sufficient training data for every new topology and link configuration [5].

In this paper, we propose a cascaded learning (CL) framework that combines the pre-trained component-level EDFA gain models with end-to-end training using minimal data to improve the power spectrum prediction in multi-span networks. Specifically, we insert fully connected (FC) layers after each pre-trained EDFA model, whose parameters are trained using end-to-end measurements collected from the multi-span link. We verify the performance of two CL-based models in two multi-span links with a total fiber length of 234 km: CL-Full that consists of all pre-trained EDFA models, and CL-Trimmed that consists of only the first and last EDFA models. Experimental results show that the CL-based models achieve a mean absolute error (MAE) in terms of the power spectrum prediction accuracy of less than 0.17 dB in topologies with varying numbers of spans and EDFAs, using only a single (i.e., *one-shot*) end-to-end measurement under fully loaded WDM channel configuration.

2. Proposed Cascaded Learning (CL) Framework

We consider a multi-span ROADM network with K spans and $2K$ EDFAs, as shown in Fig. 1(a), where the k-th span connects the k-th **B**ooster EDFA (B_k) and **P**re-amplifier EDFA (P_k). The input and output power spectrum of the k-th booster/pre-amplifier EDFA is denoted by $S_{\text{in}}^{B_k/P_k}(\lambda_i)$ and $S_{\text{out}}^{B_k/P_k}(\lambda_i)$, respectively. Each EDFA is associated with a component-level gain spectrum model based on a deep neural network (DNN) with six layers, as shown in Fig. 1(b), which takes input features including the EDFA gain setting, total input/output power readings, input power spectrum, and channel loading configurations, and predicts the output power spectrum. Given the input power spectrum at the first EDFA (B_1), $S_{\text{in}}^{B_1}(\lambda_i)$, our goal is to predict $S_{\text{out}}^{P_K}(\lambda_i)$ after the signal propagates through the K-span link. We consider two main approaches for predicting optical signal spectrum in multi-span networks as the baseline: (i) **Direct cascade** of pre-trained component-level EDFA gain models with measured losses for calibration, where the predicted output power spectrum of each EDFA model is calibrated with the measured fiber or insertion loss, and then used as the input power spectrum to the next EDFA model; and (ii) **End-to-end (E2E) learning**, which trains a new model based on the end-to-end measurements across the multi-span link, including the power spectrum and total power at the input of the first EDFA and output of the last EDFA. We implement an E2E model using the same DNN architecture as that used by the component-level EDFA gain model, with two additional FC layers (each with 95 neurons) before the final output layer.

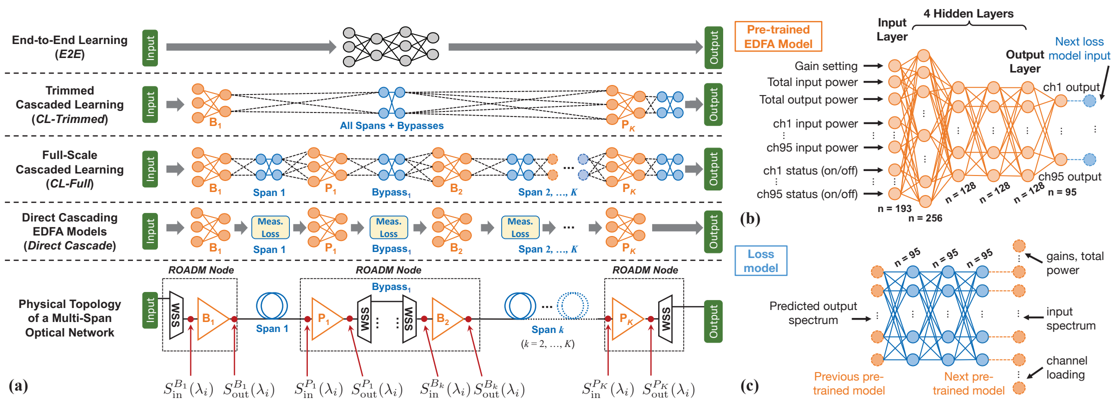

Fig. 1: (a) Different methods for multi-span optical power spectrum prediction with booster (B) and pre-amplifier (P) EDFAs. (b) DNN-based component-level EDFA gain spectrum model. (c) Loss model consisting of three FC layers.

Table 1: Mean/95-th percentile absolute error of predictions for different topologies (spans in *italic font* indicate field fibers).

Absolute prediction error: Mean/95-th percentile (dB)	**Direct Cascade**	**E2E**	**CL-Full**	**CL-Trimmed**
Topo. #1 (6-span, 234 km): 40 km–40 km–40 km–*32 km*–*32 km*–50 km	1.00/2.81	0.14/0.37	0.16/0.43	0.16/0.43
Topo. #2 (4-span, 234 km): 40 km–*72 km*–*72 km*–50 km	0.36/1.17	0.13/0.35	0.16/0.41	0.17/0.47

Our proposed CL framework, depicted in Fig. 1(a), can significantly improve the multi-span optical spectrum prediction accuracy using a small amount of end-to-end measurement data. In particular, we introduce a loss model that can be inserted after each EDFA model for characterizing the fiber and other insertion losses through the multi-span network. This loss model, shown in Fig. 1(c), consists of three fully connected (FC) layers, each with 95 neurons (equal to the number of WDM channels) and an exponential linear unit (ELU) activation function at the first layer. The input to the loss model is connected from the output of the previous EDFA model, and the output of the loss model serves as the input power spectrum fed into the next EDFA model, together with other input features including the gain setting, total input/output power, and channel loading configuration. Note that the loss model does not need to be trained individually but will be trained as part of the CL-based model. We consider two CL-based models: (*i*) **Full-scale CL (CL-Full)**, where all the $2K$ pre-trained EDFA gain models are cascaded based on the physical link topology with one loss model inserted after each EDFA gain model; and (*ii*) **Trimmed CL (CL-Trimmed)**, which uses only two pre-trained EDFA gain models corresponding to the first and last EDFAs (i.e., B_1 and P_K) in the multi-span link, with one loss model inserted after each EDFA gain model.

3. Experimental Setup and Results

Pre-trained EDFA gain models We use the publicly available COSMOS EDFA dataset [3] to generate the component-level models for 12 EDFAs (6 boosters and 6 pre-amplifiers) situated within the Lumentum ROADM-20 units. In particular, each EDFA gain model is trained using 8,196 spectrum measurements across three gains settings (15/28/21 dB) and diverse channel loading configurations, with a learning rate of 1e-3 over 600 epochs.

E2E measurements in multi-span networks To evaluate the performance of the CL framework and different models, we conduct experiments and collect E2E link measurements using the PAWR COSMOS testbed, which is a city-scale programmable optical-wireless testbed deployed in West Harlem, New York City [6]. We set up two multi-span topologies: one 6-span link with 12 EDFAs and one 4-span link with 8 EDFAs. The span configurations for the two topologies are summarized in Table 1, with the same total lengths of 234 km (including 64 km Manhattan field fibers). For each topology, all booster and pre-amplifier EDFAs are set to the high gain mode with 18 dB gain and zero gain tilt. We use a comb source to generate 95×50 GHz WDM channels in the C-band, and use the MUX wavelength selective switch (WSS) of the first ROADM unit to apply different channel loading configurations and to flatten the input spectra at the first EDFA, $S_{in}^{B_1}(\lambda_i)$. After traversing through all spans, the signal is dropped at the DEMUX WSS after the last EDFA. We record the power spectrum ($S_{\rm in}(\lambda_i)$ and $S_{\rm out}(\lambda_i)$) as well as the total input and output power ($P_{\rm in}$ and $P_{\rm out}$) of the booster/pre-amplifier EDFA before and after the k-th span (see Fig. 1). We consider two types of channel loading configurations with different numbers of loaded channels, denoted by n: (*i*) fixed channel loading, including the fully loaded (WDM, with $n = 95$), half loaded (upper/lower halves, even/odd channels, with $n \in \{47, 48\}$), and single/two-adjacent channel loading ($n \in \{1, 2\}$); and (*ii*) random channel loading with n randomly selected channels for $n \in \{1, 2, \ldots, 94\}$. The total number of E2E measurements is 19 and 940 for the fixed and random channel loading configurations, respectively.

Model training The E2E model is directly trained using the end-to-end measurements by an Adam optimizer with a learning rate of 1e-3 over 1,200 epochs. The CL-based models are trained using a two-step process (similar to conventional transfer learning): First, the weights of all pre-trained EDFA models are frozen and the weights of

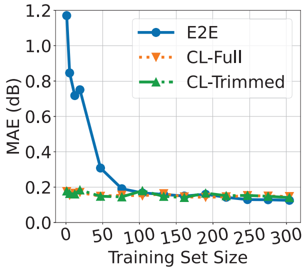

Fig. 2: Model performance with varying training data sizes for the 6-span link.

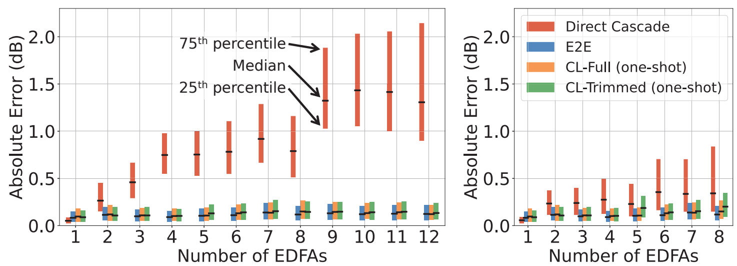

Fig. 3: Absolute error distribution of the power spectrum prediction for different models with varying number of EDFAs (*left*: up to 6-span, *right*: up to 4-span).

the inserted loss models are trained using the end-to-end measurements and Adam optimizer with a learning rate of 5e-3 over 250 epochs. Then, all the weights are unfrozen and fine-tuned using the same end-to-end measurements, with a learning rate of 1e-4 and 1e-3 for the CL-Full and CL-Trimmed models, respectively, over 80 epochs.

Training set size selection To explore the tradeoff between the performance and required amount of training data, we train each model using varying training set sizes selected from 301 measurements: 19 samples from all fixed channel loading and 282 samples from 30% of random channel loading configurations. The remaining 70% of the measurements under random channel loading configurations (658 samples) are used for testing. Note that this includes model training using a single measurement based on the *one-shot* measurement under the WDM channel loading configuration. Fig. 2 shows the prediction accuracy achieved by different models measured by the MAE with varying training set sizes, using measurements collected from the 6-span link. It can be seen that both the CL-Full and CL-Trimmed models achieve an MAE of <0.2 dB with only one-shot measurement under fully loaded WDM channel configuration, while the E2E model requires a much larger training set to achieve the same MAE. This shows that even a one-shot measurement is sufficient to train the inserted FC layers in the CL-based models that leverage information contained in pre-trained EDFA models. We empirically select one measurement under the fully loaded WDM channel configuration (one-shot) to train the CL-Full and CL-Trimmed models, and 160 measurements to train the E2E model, which are then used for the remaining experimental evaluation.

Results We evaluate the CL-based models with the two baseline models (Direct Cascade and E2E) using measurements collected from two multi-span topologies. Table 1 summarizes the mean and 95-th percentile absolute error achieved by different models for two topologies. It shows that the E2E model performs the best when it is trained on the large measurement dataset, achieving an MAE of 0.13–0.14 dB. Compared to the E2E model, the proposed CL-Full and CL-Trimmed models trained using only one-shot measurement achieve an MAE of 0.16–0.17 dB, which is comparable to that achieved by the E2E model. The CL-based models achieve a 95-th percentile error that is slightly larger than that achieved by the E2E model. In addition, we split each K-span topology with $2K$ EDFAs into sub-topologies that include the first M EDFAs $(M = 1, \dots, 2K)$ EDFAs to evaluate the model performance with varying numbers of EDFAs in the multi-span link. For example, for $M = 6$ EDFAs in the 6-span link, we treat the first three spans with three booster and three pre-amplifier EDFAs as an end-to-end system and use the input spectrum before the first EDFA and output spectrum after the 6-th EDFA for training and testing models. Fig. 3 shows the absolute prediction error distribution with varying numbers of EDFAs in the multi-span link, and the 25-th, 50-th (median), and 75-th percentiles are indicated in the boxplot. It can be seen that Direct Cascade suffers from accumulated prediction errors with an increased number of EDFAs, while the E2E and CL-based models maintain a consistently improved prediction accuracy. Compared to the E2E model, the CL-based models achieve a similar performance while requiring only $160\times$ smaller training set size. Note that although the CL-Trimmed model has $2.3\times$ parameters compared to the E2E model, it takes $2\times$ less training time.

4. Conclusion

We proposed a CL-based framework for multi-span optical power spectrum prediction leveraging pre-trained EDFA gain models and minimum end-to-end measurements. Compared to the Direct Cascade and E2E Learning methods, we showed that the CL-based models can achieve an MAE of 0.17 dB for the power spectrum prediction across a 6-span link with 12 EDFAs using only one-shot end-to-end measurement.

Acknowledgments. The work was supported by NSF grants CNS-1827923, OAC-2029295, CNS-2112562, CNS-2211944, CNS-2330333, and Science Foundation Ireland grant #13/RC/2077_P2.

References

1. Y. Pointurier, "Design of low-margin optical networks," J. Opt. Commun. Netw. **9**, A9–A17 (2017).
2. A. Mahajan *et al.*, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," J. Light. Technol. **38** (2020).
3. Z. Wang *et al.*, "Open EDFA gain spectrum dataset and its applications in data-driven EDFA gain modeling," J. Opt. Commun. Netw. **15**, 588–599 (2023).
4. Z. Wang *et al.*, "Optical signal spectrum prediction using machine learning and in-line channel monitors in a multi-span ROADM system," in *ECOC,* (2022).
5. G. Liu *et al.*, "Hierarchical learning for cognitive end-to-end service provisioning in multi-domain autonomous optical networks," J. Light. Technol. **37** (2019).
6. T. Chen *et al.*, "A software-defined programmable testbed for beyond 5G optical-wireless experimentation at city-scale," IEEE Netw. **36** (2022).