Scalable Machine Learning Models for Optical Transmission System Management

Zehao Wang[1,3], Agastya Raj[2], Giacomo Borraccini[3], Shaobo Han[3], Yue-Kai Huang[3], Ting Wang[3], Marco Ruffini[2], Dan Kilper[2], and Tingjun Chen[1]

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

Abstract: Optical transmission systems require accurate modeling and performance estimation for autonomous adaption and reconfiguration. We present efficient and scalable machine learning (ML) methods for modeling optical networks at component- and network-level with minimized data collection.

1. Introduction

Modern services such as data center and telecommunication networks require a flexible, fast-adaptive, and reliable optical transmission system, particularly in metro reconfigurable optical add-drop multiplexer (ROADM) networks [1]. In flex-grid dense wavelength division multiplexing (DWDM) systems, signals traversing an optical path with multiple spans experience varying propagation characteristics across wavelengths due to erbium-doped fiber amplifiers (EDFAs), optical fibers, and wavelength selective switches (WSSs). Accurate modeling and estimation of end-to-end (E2E) optical transmission performance, such as power spectrum and quality of transmission (QoT), are crucial for effective system management, including margin design and modulation format selection.

There are two primary approaches for modeling optical devices and paths. The first relies on analytical models, such as the Gaussian Noise (GN) model, which can accurately capture fiber nonlinearities [2,3]. However, accurate analytical models for multi-stage commercial EDFA gain and noise figure (NF), especially under dynamic channel loadings, remain unavailable. The second method uses machine learning (ML), which excels in characterizing optical devices, particularly EDFAs under varying channel loading conditions [4,5]. Despite its promise, ML-based methods face several challenges, including the lack of benchmark datasets for fair comparisons between ML methods, extensive measurement and data collection required for device model training, and the accumulation of estimation error as the optical path length increases. While ML-based models can achieve high accuracy in EDFA gain spectrum prediction, they require comprehensive and time-intensive measurements for individual devices. Directly applying a pre-trained EDFA model to other devices results in reduced prediction accuracy and poor scalability due to the extensive measurements required for each new device. As illustrated in Fig. 1(a), a desired approach would enable models to adapt to new devices with minimal measurements while maintaining high prediction accuracy. Similarly, for measurement-constrained field-installed fiber links, the desired link models would take minimal link measurements while maintaining a reasonable accuracy, when compared to directly cascading the component models or the end-to-end (E2E) models.

In this paper, we present recent advances in scalable ML methods for modeling the EDFA gain spectrum, validated using open datasets of measurements collected from commercial-grade EDFAs [5]. To minimize the data collection required for unseen devices, transfer learning- (TL-) based methods are proposed and evaluated [6,7]. For optical path modeling, recently proposed cascaded learning (CL) and attention models demonstrate scalability by leveraging the component device and single-span models. These methods require only a few E2E measurements over the installed link path, while achieving prediction accuracy comparable to E2E path models [8,9].

2. Open EDFA Datasets

Fig. 1(b) shows the COSMOS and OpenIreland platforms used for data collection and experiments, both of which are city-scale, reconfigurable optical-wireless testbeds deployed in West Harlem, New York City, and Dublin, Ireland, respectively [10]. These testbeds facilitate the cross-validation of various optical and wireless techniques, as well as enabling transatlantic experimental collaborations. Although different ML-based methods show high accuracy over optical amplifiers, it is difficult to compare models as they are evaluated over different measurement setups and resolutions, where an open optical dataset is urgent for a fair comparison between different methods. In [11], we present an open EDFA gain profile dataset consisting of 16 EDFAs within 8 commercial-grade Lumentum ROADM-20 units deployed in the COSMOS testbed [5]. Shown in Fig. 1(c), each ROADM consists of one booster and one pre-amplifier EDFA, where the gains are set with 0 dB tilt and high gain mode. The gain of each booster and pre-amplifier EDFA is measured for three and five gain settings, respectively. For each gain setting, the dataset consists of 3,168 gain spectrum measurements across 95×50 GHz channels within the C-band, recorded

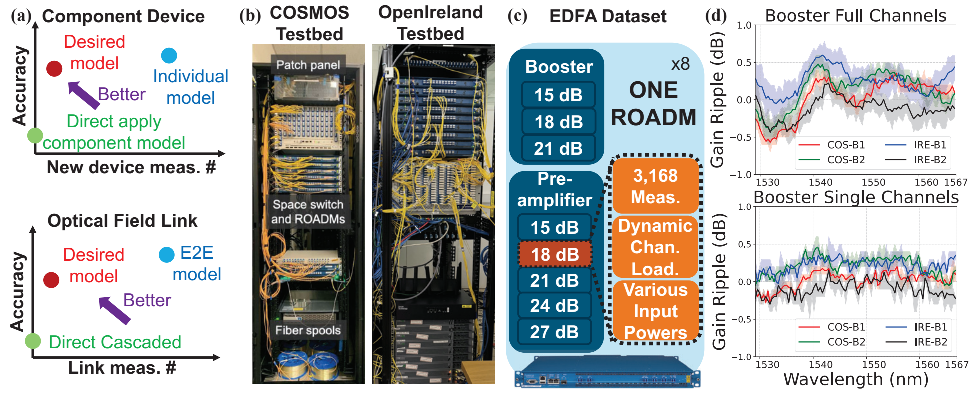

Fig. 1: (a) Tradeoffs between model accuracy and required number of measurements for different component device and optical path models; (b) COSMOS (**COS**) and OpenIreland (**IRE**) testbeds; (c) EDFA gain profile dataset collected using the COSMOS testbed; (d) example **B**ooster EDFA gain ripple measurements under full (upper) and single (lower) channel loading configurations.

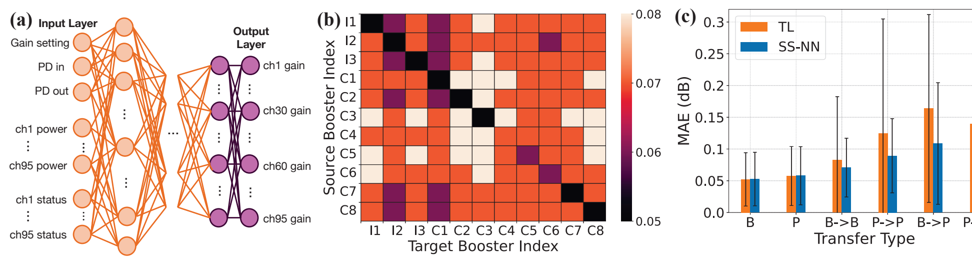

Fig. 2: (a) DNN architecture for modeling EDFA gain spectrum, where the purple layers are re-trained during TL; (b) Mean absolute error (MAE) matrix for SS-NN prediction with cross-testbed validation using data collected from the OpenIreland (**Ix**) and COSMOS (**Cx**) testbed; (c) MAE of EDFA gain prediction by source DNN model for Booster (**B**)/Pre-amplifier (**P**) and TL-based models, averaged across C1–C8.

with machine-actionable `json` file format. The total input/output power and channel power are recorded by photodetectors (PDs) and built-in optical channel monitors (OCMs) with 0.01 dB and 0.1 dB precision, respectively. Fig. 1(d) shows the full and single channel loading gain spectrum ripple for four booster EDFAs inside COSMOS and OpenIreland. We can observe the gain ripple varies for the same EDFA under different channel loadings, and consists of up to 0.5 dB gain ripple differences among different EDFA devices across testbeds. However, the gain spectra still show a similar pattern as the EDFAs belonging to the same vendor, which enables the TL technique, transferring from an existing ML-based EDFA model to a new device with only a few measurements.

3. Scalable EDFA Gain Models using Transfer Learning (TL)

Fig. 2(a) shows the architecture of a DNN-based EDFA gain spectrum model. The input features include the EDFA gain setting, total input/output power, input power spectrum, and channel status, and the output is the gain spectrum. The *source EDFA model* is trained over 2,678 (85%) of the measurements and tested on the remaining 550 (15%) measurements at each gain setting, using the open EDFA dataset.

Transfer learning (**TL**) leverages the knowledge of pre-trained models, enabling rapid adaptation with very few data samples collected from the target domain. To transfer a pre-trained source domain model to a *target* domain on a new EDFA device, a two-step procedure is applied for TL [7]. First, the input and first several layers are frozen and treated as the feature extractor of the DNN model. Then, the inserted adapter layer is re-trained using only 13 (0.5%) measurements at each gain setting. Finally, all layers are unfrozen and fine-tuned using a smaller learning rate. In [6], we also proposed a TL-based semi-supervised, self-normalizing neural network (**SS-NN**) method that only requires a one-shot (0.04%) full channel loading measurement on the new device for transferring with additional ROADM internal variable optical attenuator (VOA) values included as input features. Fig. 2(b) shows the mean absolute error (MAE) achieved by the SS-NN model evaluated over 11 booster EDFAs across the two testbeds, where entry (i,i) corresponds to the DNN-based EDFA model without TL, and entry $(i,j), i \neq j$, corresponds to the TL-based EDFA model transferred from the i^{th} source to the j^{th} target model. Fig. 2(c) shows the MAE with standard deviation for DNN-based source EDFA models and different TL-based target models averaged over 8 booster and pre-amplifier EDFAs in the COSMOS testbed. The source DNN-based EDFA models achieve 0.05 dB and 0.06 dB MAE for booster and pre-amplifier EDFAs, respectively. The transferred models introduce an MAE degradation of 0.02–0.06 dB and 0.04–0.1 dB on the same and different EDFA types, while requiring only less than 0.5% new measurements on the target device.

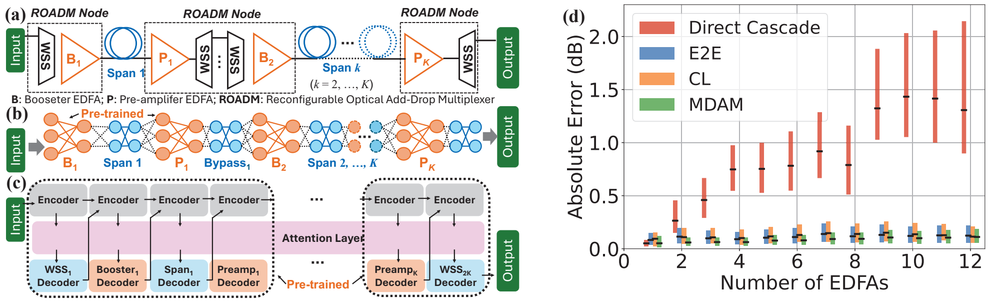

Fig. 3: (a) Physical topology of a multi-span optical path; (b) Cascaded learning (**CL**) model leveraging pre-trained EDFA models; (c) Multiple-decoder attention model (**MDAM**) leveraging single-span models; (d) Absolute error distribution of the power spectrum prediction for different models for up to six spans with 12 EDFAs.

4. Optical Path Model using Cascaded Learning (CL)

Compared to component-level devices, measurements over field-installed fiber links are more constrained, e.g., 10 minutes response requirement for data center traffic requests [12]. Unlike device-level modeling, domain knowledge from other optical paths cannot be effectively transferred, as distinct optical topologies exhibit significant differences. To address this challenge, the proposed cascaded learning assembles the device-level or single-span models with significantly reduced end-to-end link measurement requirements. For a multi-span ROADM network with K spans and $(K+1)$ ROADM nodes, as shown in Fig. 3(a), the research problem is formulated as: given the input power spectrum at the first EDFA, the goal is to predict power spectrum after the signal propagates through the K-span optical path, without any add and drop channel within the path. There are two baselines for optical signal spectrum predictions: (*i*) **Direct cascade** of pre-trained component-level EDFA gain models with measured fiber and insert losses [13]; and (*ii*) End-to-end (**E2E**) which trains a new model based on the end-to-end link measurements across the multi-span path.

Our proposed cascaded learning (**CL**) framework, shown in Fig. 3(b), leverages the pre-trained EDFA models with three fully connected (FC) layers inserted after each EDFA model for characterizing the fiber and other insertion losses through the multi-span network. The input to the FC layers is connected from the output of the previous EDFA model, and the output of the FC layers serves as the input power spectrum fed into the subsequent EDFA model, together with other input features such as the channel loading conditions. Note that the inserted FC layers do not need to be trained independently but as part of the CL-based model. We also proposed a multi-decoder attention model (**MDAM**) that leverages a single pre-trained span model, as shown in Fig. 3(c)) [8]. Specifically, the encoder is a 3-layer long short-term memory (LSTM) model, and the decoders are shallow NN with two FC layers. The attention layer calculates the summary of previous hidden states. A span model including a shared encoder and four decoders for each component (Booster, Preamp, fiber, and WSS) is trained using 3,168 power spectrum measurements collected in the OpenIreland testbed. Then, the pre-trained decoders are replicated K times, along with the shared encoder, and retrained using measurements collected from the target COSMOS optical path. The CL and MDAM methods are evaluated and compared with two baseline models (Direct Cascade and E2E) using 6-span measurements collected on the COSMOS testbed. The CL, MDAM, and E2E methods use one-shot, 48, and 160 end-to-end link measurements for training and are evaluated over 658 measurements. Fig. 3(d) shows the distribution of the absolute prediction error achieved by different methods with varying numbers of EDFAs in the multi-span path, with the 25-th, 50-th (median), and 75-th percentiles. It can be seen that Direct Cascade suffers from accumulated prediction errors with an increased number of EDFAs, while the other methods maintain a consistently improved prediction accuracy. Compared to the E2E model, the CL and attention model perform similarly while reducing the number of required end-to-end link measurements by $160\times$ and $3.3\times$, respectively.

5. Conclusion

We presented the use of TL to build scalable ML methods for modeling the EDFA gain spectrum, validated using a comprehensive, open EDFA dataset. We also highlighted recent advances in scalable ML frameworks for modeling the optical path power profile using CL and MDAM models with pre-trained EDFA or fiber span models.

Acknowledgments. This work was supported by NSF grants CNS-1827923, CNS-2112562, CNS-2211944, and CNS-2330333, Science Foundation Ireland under grants #13/RC/2077_P2, 22/FFP-A/10598, and 18/RI/5721, and Horizon Europe research and innovation program.

References

1. Z. Wang, *et al.*, IEEE J. Light. Technol. **42** (2024).
2. A. Ferrari, *et al.*, IEEE/OSA J. Opt. Commun. Netw. **12** (2020).
3. G. Borraccini, *et al.*, IEEE/Optica J. Opt. Commun. Netw. **16** (2024).
4. A. D'Amico, *et al.*, IEEE OJ-COMS. **2**, 564–574 (2021).
5. Z. Wang, *et al.*, IEEE/Optica J. Opt. Commun. Netw. **15** (2023).
6. A. Raj, *et al.*, in *Proc. ECOC 2023,* p. TuC5.6.
7. Z. Wang, *et al.*, in *Proc. OFC 2023,* p. Th2A.1.
8. A. Raj, *et al.*, in *Proc. ECOC 2024,* p. W2A.93.
9. Z. Wang, *et al.*, IEEE/Optica J. Opt. Commun. Netw. **17** (2025).
10. T. Chen, *et al.*, IEEE Netw. **36** (2022).
11. Z. Wang, *et al.*, "COSMOS EDFA Dataset," `https://wiki.cosmos-lab.org/wiki/Datasets` (2022).
12. H. Nishizawa, *et al.*, IEEE/Optica J. Opt. Commun. Netw. **16** (2024).
13. Z. Wang, *et al.*, in *Proc. ECOC 2022,* p. We3B.5.