Multi-Span OSNR and GSNR Prediction using Cascaded Learning

Zehao Wang[1,2], Giacomo Borraccini[2], Andrea D'Amico[2], Yue-Kai Huang[2], Ting Wang[2], Dan Kilper[3], Koji Asahi[4], and Tingjun Chen[1]

[1]Duke University, Durham, NC, USA; [2]NEC Laboratories America, Princeton, NJ, USA;
[3]CONNECT Centre, Trinity College Dublin, Ireland; [4]NEC Corporation, Abiko, Japan
[1]{zehao.w, tingjun.chen}@duke.edu, [2]{gborraccini, adamico, kai, ting}@nec-labs.com, [3]dan.kilper@tcd.ie

Abstract: We implement a cascaded learning framework leveraging three different EDFA and fiber component models for OSNR and GSNR prediction, achieving MAEs of 0.20 and 0.14 dB over a 5-span network under dynamic channel loading.

1. Introduction

The rapid growth of AI, telecommunication, and cloud service relies on the high throughput and low latency optical network, leveraging the dense wavelength division multiplexing (DWDM) technique. When DWDM signals transmit through the optical link, different channels will experience wavelength-dependent effects such as erbium-doped fiber amplifiers (EDFAs) gain and fiber nonlinearities. Accurate estimation of the end-to-end optical channel performance such as optical signal-to-noise ratio (OSNR) for background traffic channels and general signal-to-noise ratio (GSNR) for signal channels is important for guaranteed quality of transmission (QoT) to facilitate selection of modulation formats and margin design.

There are two main approaches to model a multi-span optical link for end-to-end prediction. The first method uses an end-to-end (E2E) model to characterize the entire multi-span link and the second method directly cascades either machine learning (ML) or analytical component-level models. For the E2E model, although it shows high accuracy, it will be time-consuming to collect sufficient training data for every new topology and link configuration [1]. Although analytical models such as the Bigo equation and Gaussian Noise (GN) model show high accuracy on Stimulated Raman Scattering (SRS) tilt and fiber nonlinearity [2, 3], EDFA gain and noise figure (NF) modeling still requires ML, especially with dynamic channel loadings. To reduce the error accumulation from component models, parameter refinement (PR) is used to adapt the analytical model to the link using the end-to-end link measurements [4, 5]. However, there is no existing method to adapt ML-based models together with analytical models, where the error accumulation has not been fully solved.

In this paper, we extend the cascaded learning (CL) framework introduced in [6] from the multi-span power spectrum prediction to OSNR and GSNR prediction, to minimize the error accumulation from the cascading effect. We combine separately characterized component-level models for the EDFA gain, NF, and fiber nonlinearity, with fully connected (FC) layers, whose parameters are trained using end-to-end multi-span link measurements. We verify the performance of the CL model under a 5-span link with a total fiber length of 396 km under three different link settings with various channel loading conditions. We also compare the CL model with two common baselines: the E2E model and component cascading with parameter refinement [1, 4, 5]. Experimental results show that the CL model achieves a mean absolute error (MAE) of 0.20/0.14 dB OSNR/GSNR prediction, which is 0.06/0.15 dB and 0.40/1.03 dB smaller compared to the E2E and component cascading model, using only 41 link measurements for training. The CL model also shows adaptation capability over unseen component device settings.

2. Cascaded Learning (CL) for OSNR and GSNR Prediction

We consider a multi-span optical link with K spans and $K+1$ EDFAs, as shown in Fig. 1(a). The input channels include both transceiver (TRX) signal channels and background WDM channels emulated by ASE noise. Given the input spectrum to the first EDFA, $S^1_{in}(\lambda_i)$, the goal is to predict the $GSNR(\lambda_i)$ for the transceiver channels and $OSNR(\lambda_i)$ for the background WDM channels at the output of a K-span optical link, with the actual GSNR and OSNR verified by real-time BER and optical spectrum link measurements. Each EDFA in the link is associated with pre-trained ML-based gain and NF models, with the same input features including EDFA gain and tilt settings, photodiode (PD) power readings, input power spectrum, and channel loading conditions, and with output of gain and NF for each channel. Similarly, each fiber is associated with an ML-based nonlinearity model, with input features of fiber length, input power spectrum, and channel loading conditions, and with output of nonlinearity SNR for each channel. We consider two approaches for the OSNR/GSNR prediction as baselines: end-to-end (E2E) learning and component cascading. The **E2E link model (E2E)** trains a new model based on link measurements including input power spectrum, channel loading settings, total input and output power at each EDFA, and the EDFA gains and tilts. We implement the E2E model using the DNN architecture with three hidden FC layers of 128/64/64 neurons and exponential linear unit (ELU) activation function, shown in Fig. 1(b). The output of the last hidden layer connects to two separate 2×40 neurons' FC layers without activation function, to predict the link OSNR and GSNR separately. For the **component cascading only (CC-Only) model**, the input

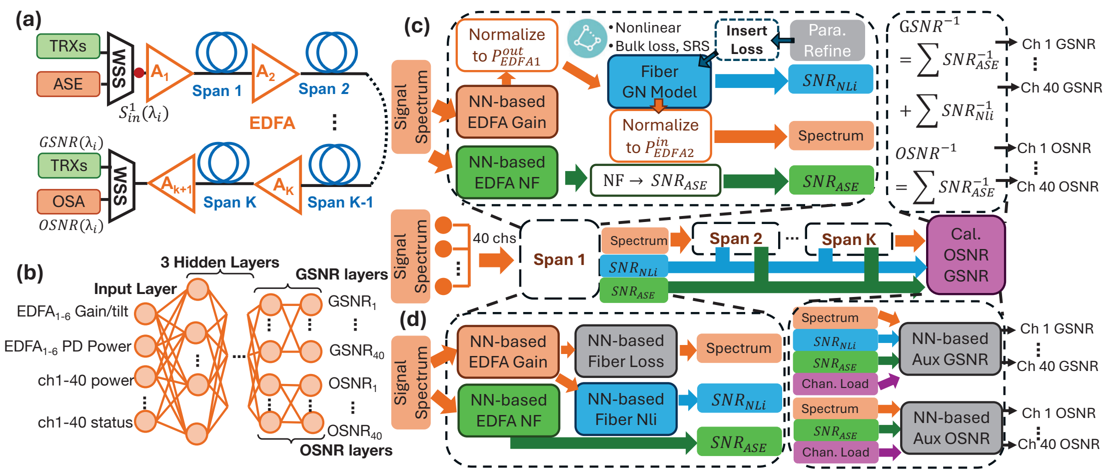

Fig. 1: (a) Multi-span measurement pipeline for background ASE channels' OSNR and 400 GbE channels' GSNR. (b) E2E link model diagram. (c) Component cascading model with the insert loss parameter refinement (CC-PR). (d) Cascaded learning (CL) framework for multi-span OSNR and GSNR prediction.

power spectrum is first put into the EDFA gain and NF model, to predict the first span's SNR_{ASE} and power spectrum after EDFA. The spectrum is normalized and sent into GNPy-based fiber model to calculate the nonlinearity SNR_{NLi} and output spectrum after fiber, where the spectrum is normalized again and sent to the next span. For the **component cascading with parameter refinement (CC-PR) model**, in addition to the CC-Only model, we use the link measurement to refinement to adapt the GN-based analytical fiber models to the link, without adapting the ML-based EDFA models. The refinement checks the averaged difference between OSNR or GSNR on loaded channels, and refines the insert loss at the beginning of each fiber. The proposed **cascaded learning (CL) framework** effectively adapts all the component models to the link, compared to the CC-PR. It has a similar diagram as CC-PR (see Fig. 1(c) and 1(d)), but replaces the GNPy-based fiber model with a pre-trained fiber nonlinear model, and a fiber loss model. The auxiliary GSNR and OSNR models at the end take predicted SNR_{ASE} and SNR_{NLi} after each span, predicted spectrum at the last span, and channel loading as their input and predict the link GSNR/OSNR. They have the same NN architecture, consisting of three FC layers of 40 neurons and an ELU activation function at the first layer. Note that the fiber loss and auxiliary GSNR/OSNR models do not need to be trained individually but will be trained as part of the CL model.

3. Experimental Setup and Results

ML-based Component Models. We train EDFA gain and NF models using separately characterized gain and NF profiles from individual EDFAs with varying input power levels, gains, tilts, and channel loadings, using an ASE source and an OSA. We use GNPy with the measured fiber parameters (e.g., loss coefficients, fiber length, etc.) to generate synthetic data with various channel loadings and launch power levels, which is then used to train the fiber nonlinearity ML model [7]. For transceiver characterizations, we measure the OSNR-BER curve under Tx-Rx DCOs back-to-back connections. The MAEs averaged over six EDFAs and five fibers are 0.08 dB, 0.12 dB, and 0.19 dB for EDFA gain, NF, and fiber nonlinearity models, respectively.

5-span Link Measurements. Fig. 1(a) shows the experimental setup of the 5-span optical link with 40×100 GHz channels and a starting frequency of 192.2 THz. The NEC Phoenix whitebox transponder consisting of four Lumentum 400 GbE CFP2 digital coherent optics (DCO) pluggables, together with a broadband ASE source, connects to a wavelength-selective switch (WSS) for multiplexing. The WSS flattens the spectrum and transmits the 400G dual-polarization 16QAM signals together with ASE-emulated background traffic signals through a 5-span link with six EDFAs and a total fiber length of 396 km. After transmission, the background ASE and 400G channels are demultiplexed by the WSS and sent to the OSA and another whitebox transponder for OSNR and BER measurements, respectively. The BER measurement is averaged 10 times and converted to link-only GSNR by removing the back-to-back transceiver noise [8]. We record the channel loading, input power spectrum at the first EDFA, PD power readings before and after each EDFA, GSNR for 400G channels, and OSNR for ASE channels.

We evaluate the proposed method under various channel loadings and link settings. We consider three types of channel loading conditions: (*i*) full loadings, where all 40 channels are loaded with ASE noise with four consecutive channels substituted with 400G channels successively for 10 times to obtain GSNR measurements; (*ii*) fixed loadings, where the transceiver channel indexes are fixed at $\{5, 20, 21, 35\}$ for the most different nonlinearity effect, and 10 random configurations each for ASE channel number $n \in \{5, 10, 15, 20, 25\}$; and (*iii*) Optical spectrum as a Service (OSaaS) loadings, where the spectrum is evenly divided into four groups with two transceiver

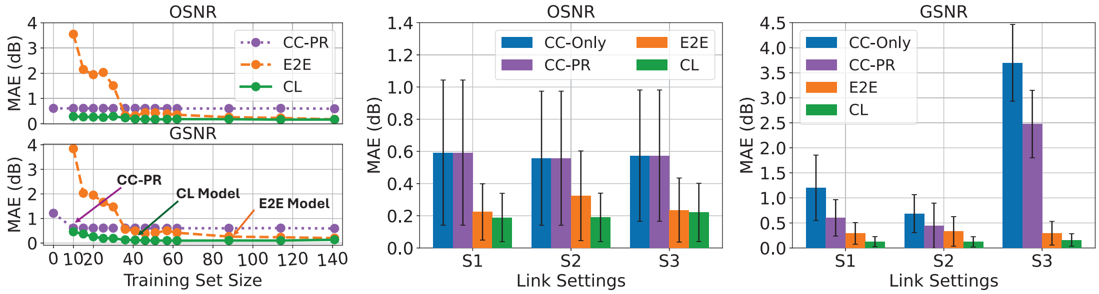

Fig. 2: (a) Different link models' mean absolute error (MAE) with various training data sizes for the 5-span link. (b) MAE of OSNR prediction for different link models under various link settings. (c) MAE of GSNR prediction for different models under various link settings.

channels assigned to two groups. Each group has a total number of loaded channels $n_{\text{total}} \in \{3,5,7\}$ with random loading. Each group is turned on and off sequentially, with 9 different on/off operations including all groups on, one group off, and selected two groups off. The total number of full/fix/OSaaS channel loading measurements is 10/100/162, respectively. We consider three different link settings: (*i*) $S1$ sets the EDFA gain and tilt using the value from the component EDFA measurement, with an average launching power $+1.5$ dBm per channel; (*ii*) $S2$ reduces the input power by 1 dB for all channels compared to $S1$; and (*iii*) $S3$ sets the gain and tilt differently from component EDFA characterization, with a launch power of $+3.5$ dBm per channel for every span.

Multi-span Link Models Training and Refinement. The E2E model is trained using the link measurements by an Adam optimizer with a learning rate of 5e-3 over 800 epochs. For CC-PR, we use PyCMA to optimize the insert loss with 10 iterations. The CL model is trained using a two-step process: First, we freeze the weights of all pre-trained component models and train the rest parts (fiber loss and auxiliary OSNR/GSNR models) using the Adam optimizer with a learning rate of 1e-2 over 400 epochs. Then, all the weights are unfrozen and fine-tuned using the same link measurements, with a learning rate of 1e-3 over 40 epochs.

Training Set Size Selection. Fig. 2(a) shows the model performance under different training set sizes. The minimum training set size for the E2E and CL models is 10 link measurements under full channel loadings, which guarantees each channel has at least one GSNR information. The E2E model suffers from a high MAE under a small training set, but obtains stable OSNR and GSNR prediction performance when the training set is larger than 88 samples. The CL model achieves an MAE of 0.29/0.46 dB for OSNR/GSNR prediction with only 10 link measurements, and the accuracy is further improved with increased training set size. For CC-PR, the fiber insert loss refinement improves the MAE of GSNR prediction from 1.2 dB to 0.6 dB, and has little improvement for OSNR prediction, using 10 link measurements. We empirically select 10/41/88 training samples for CC-PR, CL, and E2E models, and use the remaining 184 link measurements as test sets for all models (see Fig. 2(a)).

5-span Prediction Results. Figs. 2(b) and 2(c) show the MAE of OSNR and GSNR prediction achieved by different models across three link settings. The CL model achieves an MAE of 0.20/0.14 dB for OSNR/GSNR averaged on three link settings, with 0.06/0.15 dB lower MAE and only use half of the link measurement training data compared to the E2E model. The parameter refinement improves GSNR but has little impact on the OSNR prediction. Even though the CC-PR achieves respectable MAE of 0.57/0.52 dB under the $S1$ and $S2$ settings on GSNR, its prediction gets significantly worse at 2.5 dB when the link is configured to the $S3$ setting, as the EDFA gains/tilts are set differently from ones used during component device measurement. The CL model shows better adaptation capability when the EDFA settings are unseen during the component model characterization.

4. Conclusion

We studied a cascaded learning framework for multi-span OSNR and GSNR prediction leveraging pre-trained fiber nonlinearity, EDFA gain, and NF models with minimized link measurements. It achieves an MAE of 0.20/0.14 dB for OSNR/GSNR prediction across a 5-span link with 6 EDFAs using 41 link measurements.

Acknowledgments. This work was supported by NSF grants CNS-2112562, CNS-2211944, and CNS-2330333, Science Foundation Ireland under Grant #13/RC/2077_P2, Horizon Europe research and innovation program under Grant 101139194.

References

1. Y. He *et al.*, "QoT estimation for large-scale mixed-rate disaggregated metro DCI networks by artificial neural networks," in *OFC,* (2024), pp. W3G–2.
2. S. Bigo *et al.*, "Experimental investigation of stimulated raman scattering limitation on WDM transmission over various types of fiber infrastructures," IEEE Photonics Technol. Lett. **11**, 671–673 (1999).
3. P. Poggiolini *et al.*, "The GN-model of fiber non-linear propagation and its applications," J. Light. Technol. **32**, 694–721 (2013).
4. M. P. Yankov *et al.*, "SNR optimization of multi-span fiber optic communication systems employing EDFAs with non-flat gain and noise figure," J. Light. Technol. **39**, 6824–6832 (2021).
5. N. Morette *et al.*, "Machine learning enhancement of a digital twin for wavelength division multiplexing network performance prediction leveraging quality of transmission parameter refinement," J. Opt. Commun. Netw. **15**, 333–343 (2023).
6. Z. Wang *et al.*, "Multi-span optical power spectrum prediction using ML-based EDFA models and cascaded learning," in *OFC,* (2024), pp. M1H–6.
7. G. Borraccini *et al.*, "Optical line physical parameters calibration in presence of edfa total power monitors," in *OFC,* (2024), pp. M3I–5.
8. T. Mano *et al.*, "Modeling the input power dependency of transceiver BER-ONSR for QoT estimation," in *OFC,* (2024), pp. M1H–4.