Spectral Power Profile Optimization of Field-Deployed WDM Network Enabled by Remote EDFA Modeling

RASMUS T. JONES[1], KYLE R. H. BOTTRILL[2], NATSUPA TAENGNOI[2], PERIKLIS PETROPOULOS[2], AND METODI P. YANKOV[1*]

[1] DTU Electro, Technical University of Denmark, Lyngby, Denmark
[2] Optoelectronics Research Centre, University of Southampton, Southampton, United Kingdom
[*] Corresponding author: meya@dtu.dk

Compiled September 6, 2023

We propose a technique for modeling erbium-doped fiber amplifiers (EDFAs) in optical fiber networks, where the amplifier unit is located at a distant node outside the lab. We collect data on an optical point-to-point link with the amplifier as the only amplification stage. Different amplifier operating points are modeled using probe signals and by adjusting the settings of the amplifier through a control network. The data are used to train a machine learning algorithm integrated within a physical EDFA model. The obtained mathematical model for the amplifier is used to model all amplifiers of a network and links with multiple amplification stages. In order to confirm the modeling accuracy, we thereafter predict and optimize launch power profiles of two selected links in the network of 439.4 km and 592.4 km length. A maximum/average channel optical signal to noise ratio prediction error of 1.41/0.68 dB and 1.62/0.83 dB is achieved for the 2 multi-span systems, respectively, using the EDFA model trained on the single span system with margin-optimized launch power profiles. Up to 2.2 dB of margin improvements are obtained w.r.t. unoptimized transmission.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Availability and reliability of optical networks require a design margin due to aging effects of system components and model uncertainties [1]. The performance of different wavelength-division multiplexing (WDM) channels for a link and across the network is not uniform and the effective margin is based on the worst performing channel, leaving potential for improved performance and resource allocation by methods allowing a system to operate at lower margin [2–5]. The disaggregation trend requires a new tier of flexible and elastic networks with finer and dynamic control over provided spectral bands and data rates [6–8], further enabling systems with low margin design [9–11]. For new networks planned with low margin and existing networks operated at lower margin an accurate quality of transmission (QoT) indicator, such as the signal-to-noise ratio (SNR), guaranteeing reliable operation is imperative. Further, such a QoT model is used to optimize system components during planning or operation parameters towards a system operating with lower margin. Several methods have been proposed for QoT estimation, relying on physical models of network components [12], data driven machine learning (ML) models [13–23], and integration of the two [24]. More recently, also partially loaded systems have been addressed [25, 26]. The uncertainty in component parameters and the effect on the QoT estimate has been investigated in [27, 28]. Also methods addressing the uncertainty in component parameters have been proposed [29], e.g. by estimating the fiber type and fiber parameters of a deployed network [30]. If available, components can also be characterized in the lab before deployment by fitting a physical model or learning a model from data, as shown for erbium-doped fiber amplifiers (EDFAs) by [25, 31–35] and also in our previous work [36, 37]. For operational networks, methods leveraging monitoring data have been proposed in [38, 39]. With an accurate QoT estimate in place, techniques to optimize the channel input powers to a link are of interest. In [40], system optimization of the channel input powers is demonstrated using a Gaussian noise (GN) model including stimulated Raman scattering (SRS) effects with step-wise convex properties, leading to minimum margin improvement in SNR. In [41], it is shown that including optical amplification effects such as the wavelength dependent gain and noise contributions are critical for optimization of wide-band systems. Joint optimization of input power and amplifier gain is shown in [42]. In [43, 44] an heuristic algorithm is used to optimize SNR via the channel powers in a testbed with commercial, real-time equipment. Optimization of inline amplifier settings for a S+C+L multi-band system is demonstrated in [45]. A sequentially loaded and re-optimized system is experimentally

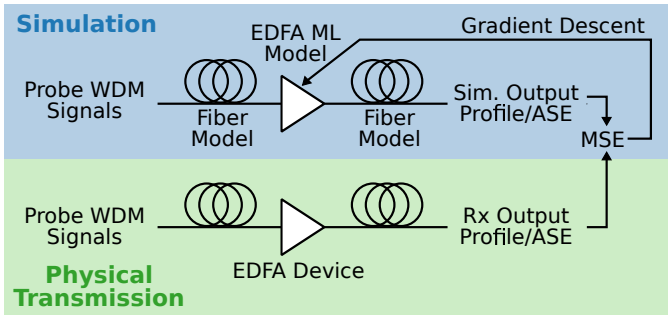


Fig. 1. Remote modeling of EDFA device.

investigated in [46].

The work we present here is an extension of our conference contribution [47], and based on our previous work in [36, 37], where an ML algorithm is trained to learn the channel input and output power relation of EDFA devices from data acquired in the lab with direct access to the EDFA input and output ports. The devices are then placed in a multi span system setup and a corresponding model is built with a cascade of fiber and EDFA ML models. A set of channel input powers are obtained via optimization of the system margin in SNR and validated on the physical system. In contrast to our previous work, in this work we show how to train a novel EDFA ML model on a system without direct access to the EDFA device itself and thus taking a further step towards practical systems. The proposed EDFA ML model is enhanced by data aided ML but guided by physical properties as proposed in Meseguer et al. [35] and Saleh et al. [48]. We show that such hybrid composition allows the EDFA model to be trained while the device is placed in the field, Fig. 1. Further, assuming sufficient similarity between the different EDFAs deployed in the network, we only train on a single EDFA device and use it to model all other EDFA devices in the network. With an analytical fiber model and a trained EDFA model, we compose a differentiable mathematical model for a subset of links in the network. The subsequent optimization of the launch power profiles shows improvement in optical signal-to-noise ratio (OSNR)/SNR margin in a field deployed system compared to a flat launch power profile. As the system model is fully trained before the system optimization, our method does not require any feedback from the system as in [43], nor long system shutdown periods for characterization. The structure of this paper is as follows: Section II describes the models used to model fiber and EDFA devices, in particular we propose a novel hybrid EDFA model combining physical properties with an ML algorithm. Section III describes the optical network on which the data acquisition and experiments are carried out on.

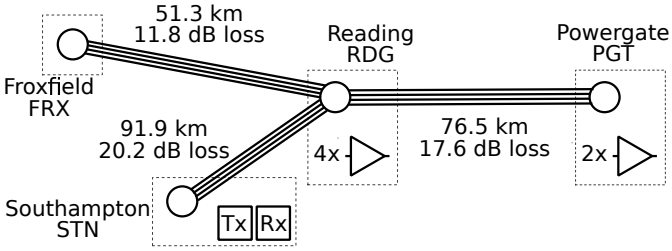


Fig. 2. Network topology. All edges are comprised of 4 individual fiber links.

Section IV describes the data acquisition and remote training technique for modeling an offsite EDFA device. For compactness of this paper and limiting the required experimental measurements, the results in Section V and Section VI are described with a progression from shorter to longer links and increasing system complexity from OSNR to SNR measurements, where the latter includes a full transceiver (TRX) setup. In Section V we show experimentally the accuracy of the trained EDFA model by optimizing the OSNR. In Section VI we show experimentally how the SNR margin is improved using the proposed EDFA model. In Section VII we conclude our findings.

2. SYSTEM COMPONENT MODELS

In this Section the models of the optical system components are described. The overall goal is to predict the signal power and the accumulated noise power for all WDM channels of a link as a function of its input power profile. For this, a cascade of WDM system component models are used: A fiber model, an EDFA model and a TRX penalty model. As this work is an extension of our previous work, we refer the reader to [37], for the description of the fiber model. In short, the fiber model is comprised of a GN model including attenuation and SRS effects. The EDFA model is novel and described in the sub-section below. The TRX penalty model is data aided and described in Appendix B. With an arbitrary configuration of these models and with resulting accurate predictions of the signal output power and the accumulated noise output power, the wavelength dependent OSNR and SNR of the WDM system is obtained as follows:

$$
\begin{aligned}
OSNR_k &= \frac{P^{out}_{S,k}}{P^{out}_{N,k}}, \\
[SNR_k]_{dB} &= \left[\frac{P^{out}_{S,k}}{P^{out}_{N,k} + P_{GN,k}}\right]_{dB} - [\Delta SNR([P^{out}_{S,tot}]_{dBm}, \lambda_k)]_{dB}, \\
P^{out}_{S,tot} &= \sum_{i=1}^{N_{ch}} P^{out}_{S,i},
\end{aligned}
\tag{1}
$$

where here $P^{out}_{S,k}$ and $P^{out}_{N,k}$ refer to the channel signal and amplified spontaneous emission (ASE) noise power of the system output, $P_{GN,k}$ is the nonlinear interference noise term modeled by the GN model, $\Delta SNR(P^{out}_{S,tot}, \lambda_k)$ is the total channel power dependent TRX penalty term at WDM channel k with center wavelength λ_k which is described in Appendix B Eq. (12) . All variables are in linear units and the $[\cdot]_{dBm/dB}$ is used for the corresponding versions in dBm/dB.

A. EDFA Model

The EDFA model described in this Section is from Meseguer et al. [35] which is based on Saleh et al. [48] A short and comprehensive description of the model in Saleh et al. can be found in

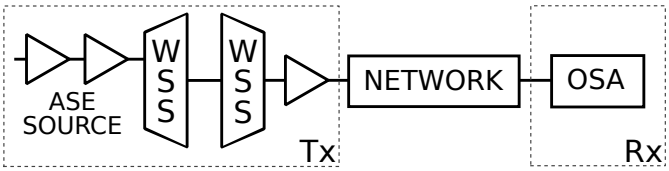
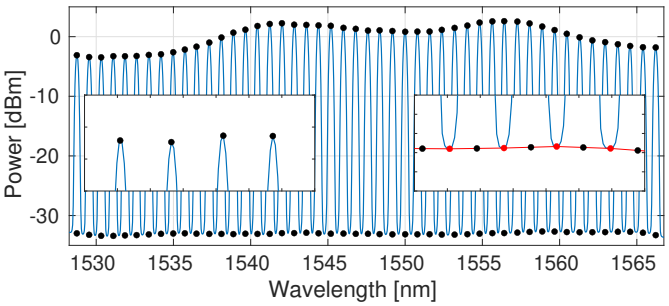
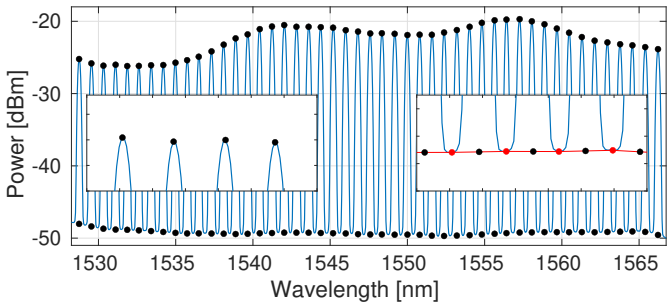


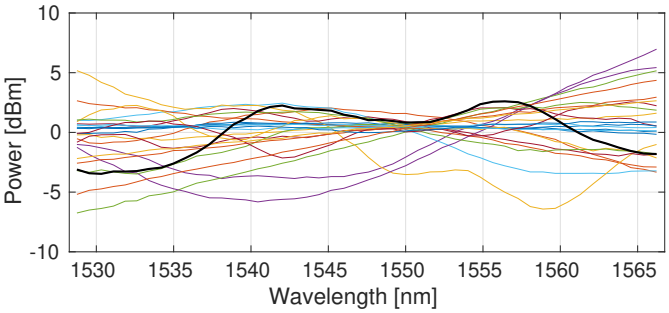
Fig. 3. Experimental setup of EDFAs, wavelength selective switches (WSSs) and an optical spectrum analyzer (OSA) for data acquisition and OSNR optimization measurements.



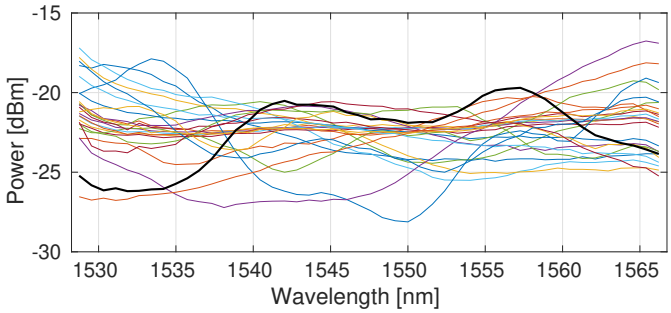
(a) Example input spectra.



(b) Example output spectra.



(c) Examples of input power profiles.



(d) Examples of output power profiles.

Fig. 4. Measurement examples of link scenario STN-RDG*-STN with the EDFA APC setting set to 18 dBm. **(a)** Shaped ASE input spectra with extracted power and noise profiles (black dots), and **(b)** its corresponding system output. The left inlets in **(a)** and **(b)** are zoomed in versions of the power profile of the first four channels, and the right inlets are zoomed in versions of the noise profile of the last four channels, here the noise profile at the respective channel center (black dots) is interpolated from the noise floor (red dots). **(c)** and **(d)** show in black the input and output power profile from the corresponding example spectra in **(a)** and **(b)**. In color 24 additional power profile samples of the data set.

the Appendix of Perin et al. [41] The model by Saleh provides the wavelength dependent gain profile of an EDFA given an arbitrary input power profile. The EDFA model channel output powers of signal and noise are given by:

$$\begin{aligned} P^{out}_{S,k} &= P^{in}_{S,k} G_k(\mathbf{P}^{in}_S), \\ P^{out}_{N,k} &= P^{in}_{N,k} G_k(\mathbf{P}^{in}_S) + P_{ASE,k}, \\ \mathbf{P}^{in}_S &= [P^{in}_{S,1}, \ldots, P^{in}_{S,N_{ch}}]^\mathsf{T}, \end{aligned} \tag{2}$$

where $P^{in}_{S,k}$, $P^{out}_{S,k}$, $P^{in}_{N,k}$ and $P^{out}_{N,k}$ refer to the channel input/output powers for signal and noise of the EDFA, respectively, $G(\cdot)_k$ is the gain profile function, and $P_{ASE,k}$ is the channel dependent ASE noise power at WDM channel k with channel center wavelength λ_k. In order to determine the gain profile function, an EDFA property is used that relates the gain profile for an arbitrary input power profile to the gain profile for an *equivalent flat* input power profile, where *equivalent* refers to the flat power profile having an equal total power to the arbitrary power profile. Assuming a gain function for flat input power profiles exists with dependence on the total input power, such as $G_{EF}(P^{in}_{S,tot})$: $\mathbb{R} \to \mathbb{R}^{N_{ch}}$, then the gain profile for an arbitrary input power profile is given by:

$$\begin{aligned} G_k(\mathbf{P}^{in}_S) &= G_{EF,k}(P^{in}_{S,tot} + N_{ch}\Delta P_{S,EF}), \\ \Delta P_{S,EF} &= \frac{\sum_{i=1}^{N_{ch}} \lambda_i \Delta P_{S,i} (G_{EF,k}(P^{in}_{S,tot}) - 1)}{\sum_{i=1}^{N_{ch}} \lambda_i (G_{EF,k}(P^{in}_{S,tot}) - 1)}, \\ P^{in}_{S,tot} &= \sum_{i=1}^{N_{ch}} P^{in}_{S,i}, \\ \Delta P_{S,k} &= P^{in}_{S,k} - \frac{P^{in}_{S,tot}}{N_{ch}}, \end{aligned} \tag{3}$$

where $G_{EF,k}(\cdot)$ is the *equivalent flat* gain function, $\Delta P_{S,EF}$ is an adjustment factor for the *equivalent flat* power, $P^{in}_{S,tot}$ is the total input power, and $\Delta P_{S,k}$ is the channel dependent deviation from the channel average input power. In Meseguer et al. [35], the *equivalent flat* gain function is determined by a look up table based on measurements of flat input power profiles and their corresponding gain profiles at the input and output ports of the EDFA device. Similar to the gain function, assuming that a noise figure function for flat input power profiles exists with dependence on the total input power, such as $NF_{EF}(P^{in}_{S,tot})$: $\mathbb{R} \to \mathbb{R}^{N_{ch}}$, then the ASE channel noise power contribution is given by:

$$P_{ASE,k} = NF_{EF,k}(P^{in}_{S,tot}) G_k(\mathbf{P}^{in}_S) h f_k \Delta f_{ref}, \tag{4}$$

where $NF_{EF,k}(\cdot)$ is the noise figure function, f_k is the frequency at wavelength λ_k and Δf_{ref} the reference bandwidth. In Meseguer et al. [35], it is reported that the noise figure function is also determined by a look up table based on measurements.

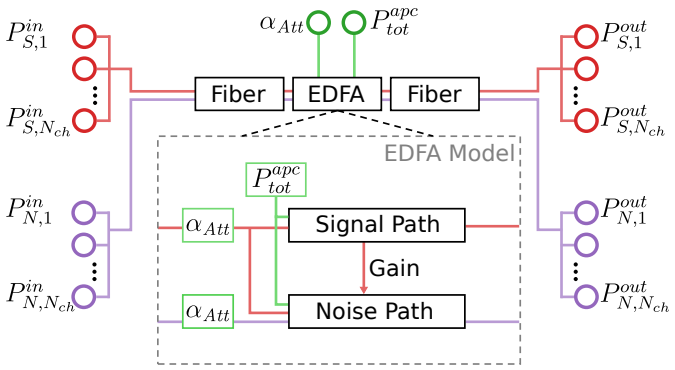


Fig. 5. Schematic flow chart of how the data of Eq. (7) is connected. The signal path (red) is independent of the noise path (purple), yet the noise path depends on the signal path.

B. Hybrid Machine Learning EDFA Model

In this Section, we describe the novel hybrid EDFA ML model. The EDFA model of the previous Section is enhanced by replacing the look up tables for the total power dependent gain profile and noise figure functions with neural networks, resulting in a fully differentiable model. Further, for both functions we introduce the EDFA automatic power control (APC) setting as an additional input parameter. Assuming the EDFA is in constant power control mode, this enables the model to include a range of operational points in terms of the EDFA total output power level:

$$\begin{aligned} G_{EF}(P^{in}_{S,tot}) &\leftrightarrow G_{EFNN}(P^{in}_{S,tot}, P^{apc}_{tot}), \\ NF_{EF}(P^{in}_{S,tot}) &\leftrightarrow NF_{EFNN}(P^{in}_{S,tot}, P^{apc}_{tot}), \\ G(\mathbf{P}^{in}_{S}) &\leftrightarrow G(\mathbf{P}^{in}_{S}, P^{apc}_{tot}) \end{aligned} \tag{5}$$

where $G_{EFNN}(\cdot)$ and $NF_{EFNN}(\cdot)$ are both modeled as neural networks, and for completeness also the gain function for arbitrary input power profiles, $G(\cdot)$, is adjusted to include the APC setting as parameters. The neural networks are further defined as:

$$\begin{aligned} [G_{EFNN}(P^{in}_{S,tot}, P^{apc}_{tot})]_{dB} &= NN([P^{in}_{S,tot}]_{dBm}, [P^{apc}_{tot}]_{dBm}), \\ [NF_{EFNN}(P^{in}_{S,tot}, P^{apc}_{tot})]_{dB} &= NN([P^{in}_{S,tot}]_{dBm}, [P^{apc}_{tot}]_{dBm}), \end{aligned} \tag{6}$$

where $NN : \mathbb{R}^2 \rightarrow \mathbb{R}^{N_{ch}}$ and the dBm/dB annotations refer to that the neural networks are trained in the logarithmic domain for both input and output. The training methodologies of the neural networks and the complete description of the dataset used for training is provided in Section 4 A. The proposed adjustments to the EDFA model allow to model a range of operation pump powers and to train the *equivalent flat* gain function (now a neural network) as part of the complete EDFA model end to end, Eq. (2), relieving the requirements for the measurements to have flat input power profiles. Central to this work, this allows the remote modeling by embedding the EDFA model within fiber models, where there is no explicit control over the profile input to the EDFA. Then, the *equivalent flat* gain function can be trained on system measurements end to end. The same applies to the noise figure function of the EDFA model.

3. NETWORK

The remote training and the following link optimization is carried out on a network of fiber links (the UK's Dark Fibre Facility) depicted in Fig. 2. The fiber links are composed of standard, single mode fiber with distances and total losses annotated in Fig. 2. The network is composed of 4 nodes at Southampton (STN), Reading (RDG), Froxfield (FRX) and Powergate (PGT), with 6 EDFAs in total: 4 in RDG and 2 in PGT. The accessing node is located at STN. Each line between the nodes in Fig. 2 indicates an independent fiber connection, allowing us to use the 4-node network to obtain longer links by re-using some of the routes. Four different link scenarios of the network are used. The link scenario STN-RDG*-STN (where * denotes an amplification stage), is used for remote training of an EDFA model on a single EDFA device in RDG, described in Section 4. The obtained model is then used for all other EDFA devices in the following link scenarios. The link scenario STN-RDG*-FRX-RDG*-STN is used to verify that the EDFA model accurately models the second amplification stage although it was trained on the first amplifier, described in Section 5. Finally, we show that the link scenarios STN-RDG*-PGT*-RDG*-FRX-RDG*-STN and STN-RDG*-PGT*-RDG*-PGT*-RDG*-FRX-RDG*-STN are accurately modeled and optimized towards a flat output SNR, described in Section 6. The EDFAs in the network are of the model CEFA-644-00 from Lea Photonics with integrated fixed gain flattening filters, aiming at flattening the gain spectrum for a specific total gain. Due to the non-uniformity of fiber span lengths in a regional network, the amplifier must operate at a variety of total gains, leading to non-flatness across the network. The control network mentioned in the abstract refers to the infrastructure used to remotely access and control the various components within the optical fiber network, including the EDFA devices, attenuators and switches. The control network is accessed from our lab via a virtual private network (VPN), and the various components are controlled using hypertext transfer protocol (HTTP) requests. With this setup, data can be collected remotely and the performance of the entire network can be optimized from a single access point. In the control network, we only control and monitor the EDFA operating point, which is much less complex and less expensive than a full optical management system that includes optical spectrum analyzers (OSAs). However, this also means that the methodology might not be applicable for networks with more complicated requirements.

4. REMOTE TRAINING

This Section describes the data acquisition process and the subsequent training of the EDFA model, with an experimental setup as shown in Fig. 3. If the EDFA unit is not deployed or a similar unit is available in the lab, as shown in [25, 36, 37], a calibration in the lab is preferred over the remote training method proposed in this Section.

A. Data Acquisition

For the data acquisition, the system emulates a C-band, 48-channel, 12.5 GBd WDM signal on a 100 GHz grid with a setup of EDFAs and wavelength selective switches (WSSs), and link scenario STN-RDG*-STN. The bandwidth of the WDM channels was chosen to align with technical limitations of the WSS, which can deliver a minimum channel size of 12.5 GHz, and with the limited bandwidth of the coherent receiver of 10 GHz. The input to the system is signals with ASE shaped spectra of 1000 random power profiles at total launch power of 20 dBm, an example is shown in Fig. 4a. The spectra at the output of the system are measured using an OSA, an example is shown in Fig. 4b. The power and noise profiles of the input and output

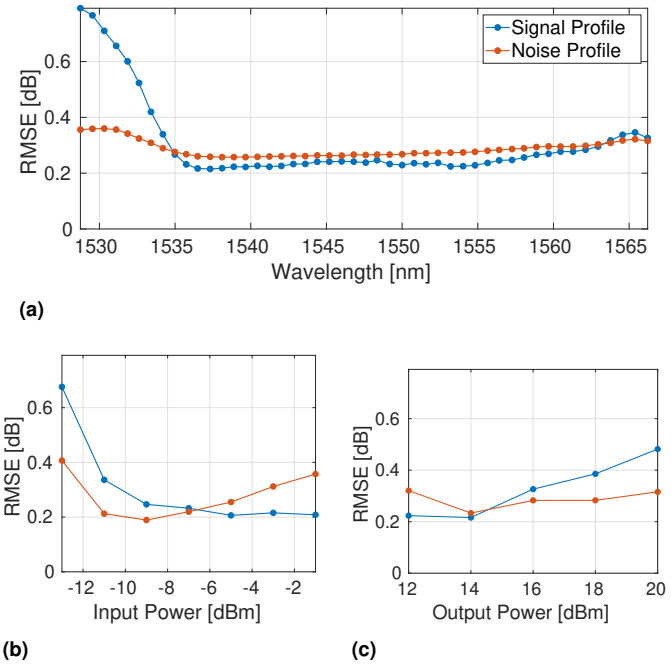


Fig. 6. Root mean squared error on the test set for power (blue) and noise (red) profile in respect to **(a)** the wavelength, **(b)** the EDFA input power and **(c)** the EDFA APC setting.

spectra are extracted (black dots in Fig. 4a and 4b), the noise profiles at the WDM channel centers are interpolated from the noise floor (red dots in inlets). Fig. 4c and 4d show the extracted input and output power profiles, respectively, as obtained from 25 individual profile samples (including those in Fig. 4a and 4b). From the power profiles and the noise profiles the OSNR is calculated. An attenuator leading into the EDFA device in RDG, allows to sweep the input power to the EDFA. Every profile is measured at 7 different EDFA input power levels and at 5 different EDFA APC value settings. The EDFA is in constant power control mode and the total output power ranges from 12 dBm to 20 dBm. With a total of 1000 power and noise profiles the resulting dataset is of 35000 samples. The corresponding system model is shown in Fig. 5, with the dataset defined as:

$$\{(\mathbf{P}_S^{in,i}, \mathbf{P}_S^{out,i}, \mathbf{P}_N^{in,i}, \mathbf{P}_N^{out,i}, \alpha_{Att}^i, P_{tot}^{apc,i})\}_{i=1}^{N_D}, \tag{7}$$

where N_D=35000 is the number of data samples, $\mathbf{P}_S^{in,i}$, $\mathbf{P}_S^{out,i}$, $\mathbf{P}_N^{in,i}$ and $\mathbf{P}_N^{out,i}$ are the system input power profiles, output power profiles, input noise profile and output noise profiles, respectively, and α_{Att}^i and $P_{tot}^{apc,i}$ are the attenuation applied to the input of the EDFA and the EDFA APC value settings, respectively. The dataset is split 10%, 20% and 70% into training, validation and test set, respectively. The split is carried out such that a profile with all its permutations in α_{Att} and P_{tot}^{apc} appears only in one of the sets. This results in 100 training profiles with 3500 data samples, 200 validation profiles with 7000 data samples, and 700 testing profiles with 24500 data samples. A low percentage of training data is unusual for ML applications, yet during development we realised that the amount of required training data was lower than expected, which allowed to split the dataset as shown and thus put more emphasis on the test error and the generalization capabilities of the model. A similar property was noted for the look up table version of the *equivalent flat* EDFA model in [35]. Further, this increases the training speed as less data is processed per training step. Finally, this property is important for obtaining training data in cases, where long periods of link availability for characterization are infeasible. The fiber model parameters, such as the lumped loss and effective nonlinear coefficient, were estimated from data acquired skipping the amplification stage in RDG and a fit of the fiber model including the SRS effect, a similar technique has been reported in [25].

B. Training

The EDFA model described in Section 2 B includes two neural networks, $G_{EFNN}(\cdot)$ and $NF_{EFNN}(\cdot)$, which are fully connected in structure and consist of an input layer with 2 neurons, two hidden layers with 32 neurons each, and an output layer with 48 neurons equal to the number of channels N_{ch}. The model and data setup is shown in Fig. 5. During training, the fiber model and the physical part of the EDFA model are continuously simulated such that the neural network learns the isolated effects of the EDFA gain and added noise. First, the signal path of the model is used to train the neural network on the signal path of the EDFA model. Thereafter, the noise path of the system model is used to train the second neural network on the noise path of the EDFA model. The neural network parameters of the signal and noise path are denoted as θ_S and θ_N, respectively. The models are trained consecutively via stochastic gradient descent (SGD). The signal path of the EDFA model is optimized by minimizing a mean squared error (MSE) loss function as follows:

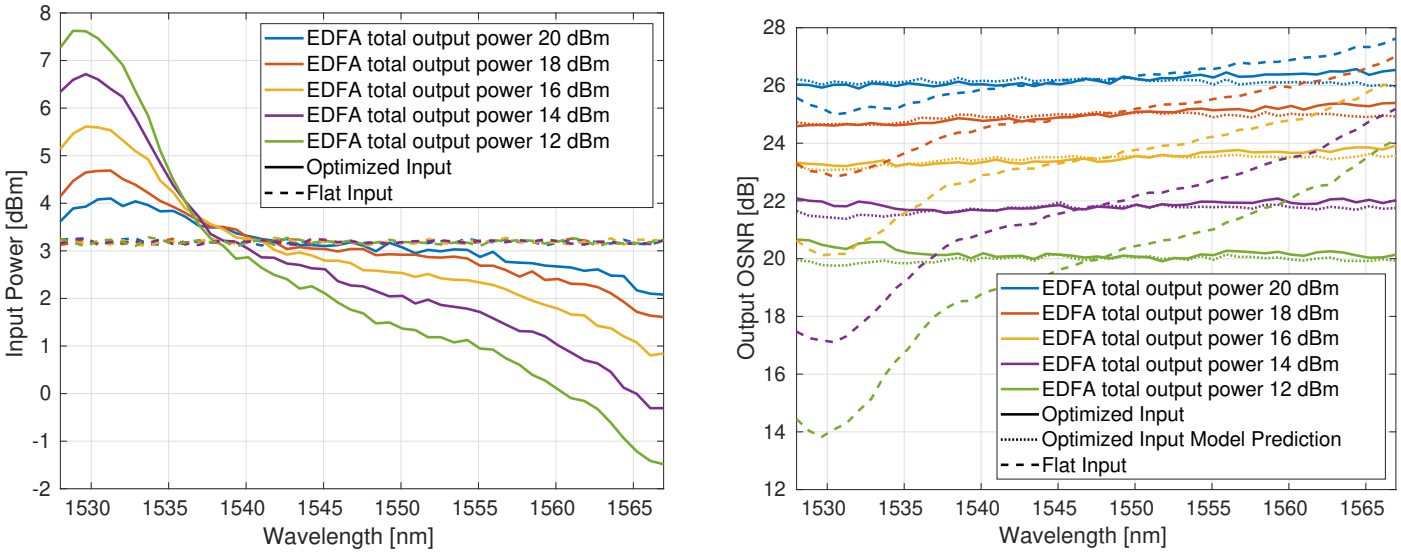
$$\begin{aligned} &loss_S(\theta_S) = \\ &\frac{1}{N_B N_{ch}} \sum_{i=1}^{N_B} \sum_{k=1}^{N_{ch}} \left([\mathcal{M}_{S,k}^i(\theta_S)]_{dBm} - [P_{S,k}^{out,i}]_{dBm} \right), \\ &\widehat{\theta}_S = \underset{\theta_S}{\operatorname{argmin}}\, loss_S(\theta_S), \end{aligned} \tag{8}$$

where $\mathcal{M}_{S,k}^i(\theta_S)$ is the prediction of the model along the signal path at the channel wavelength λ_k for data sample i, and N_B=32 is the batch-size of data samples per SGD iteration. After convergence, the signal path parameters are frozen during training of the noise path, with $\theta_S = \widehat{\theta}_S$.

The noise path of EDFA model is optimized by minimizing a MSE loss function as follows:

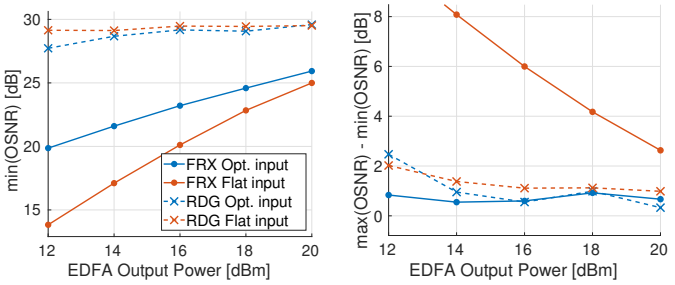
$$\begin{aligned} &loss_N(\theta_N) = \\ &\frac{1}{N_B N_{ch}} \sum_{i=1}^{N_B} \sum_{k=1}^{N_{ch}} \left([\mathcal{M}_{N,k}^i(\widehat{\theta}_S, \theta_N)]_{dBm} - [P_{N,k}^{out,i}]_{dBm} \right), \\ &\widehat{\theta}_N = \underset{\theta_N}{\operatorname{argmin}}\, loss_N(\theta_N), \end{aligned} \tag{9}$$

where $\mathcal{M}_{N,k}^i(\widehat{\theta}_S, \theta_N)$ is the prediction of the model along the noise path at the channel wavelength λ_k for data sample i and frozen signal path neural network weights $\widehat{\theta}_S$. After convergence, the obtained noise path of the EDFA model, with $\theta_N = \widehat{\theta}_N$, completes the EDFA model which can now be used for predictions. We note that the training in Eq. (8) and (9) is carried out with the Adam optimizer [49] and the automatic differentiation within the software framework PyTorch [50]. After optimization, the root mean squared error (RMSE) on the test dataset is 0.34 dB and 0.29 dB for the predicted signal and noise profiles, respectively. Fig. 6 shows the RMSE at different (a) wavelengths, (b) EDFA total input power levels and (c) EDFA total output powers levels, for the predicted signal and noise profiles. The model



(a) (b)

Fig. 7. **(a)** Flat (dashed) and optimized (solid) input power profiles for different total output power of the second EDFA, **(b)** output OSNR for flat (dashed) and optimized (solid) input power profiles for different total output power of the second EDFA, and OSNR model predictions (dotted) with optimized input power profile.



(a) (b)

Fig. 8. **(a)** Minimum OSNR and **(b)** difference of maximum and minimum OSNR across wavelength for different total output powers of the second EDFA. Link configuration going to FRX and back in solid lines and going to RDG and back in dashed lines.

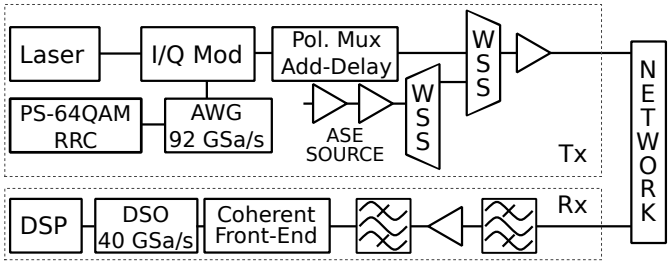
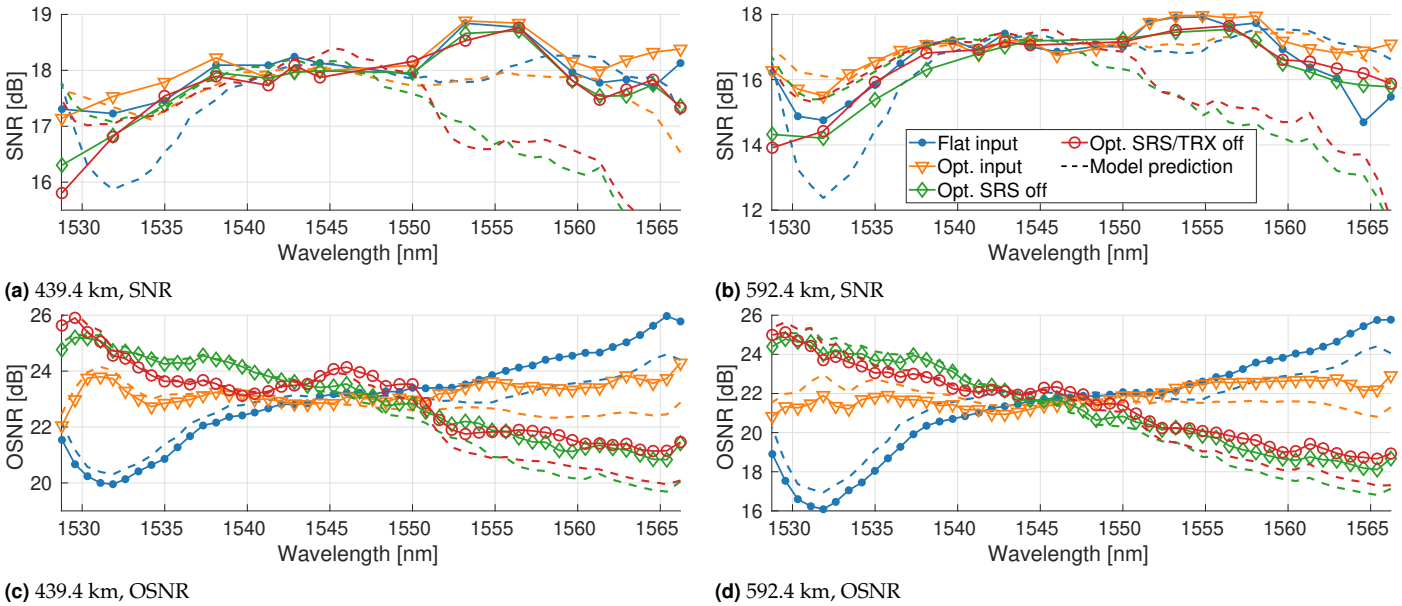


Fig. 9. Experimental system setup.

prediction shows a larger error at short wavelengths, low EDFA input powers and large EDFA total output powers. In order to quantify the errors, we present a study in Appendix A. In this Appendix, a model as described in Section 2 B is trained on simulated data not in a remote setting. The error at short wavelengths is reproduced on the simulated data, such that we conclude that this error, as reported in the remote modeling scenario of this Section is due to the insufficient accuracy of the *equivalent flat* approximation of the EDFA model. We attribute the larger error at low EDFA input powers and large EDFA total output powers to artifacts introduced by the transmission system. Instead of the model proposed in Section 2 B, it is possible to train an EDFA model comprised exclusively of neural networks as in our previous work [37], yet, in a remote modeling setup as conducted here, we found that such a pure ML model would overfit to residual effects of the fiber and other components. In contrast, the proposed hybrid EDFA model of Section 2 B is regularized in its function space by the properties of its integrated physical model. A comparison of different EDFA models, physical and ML aided, and the effect of the dataset size on the performance of the modeling is left for future research. We note that, for such fair comparisons another experiment should be designed where data are collected for a single EDFA unit in lab and remote scenarios. The lab trained models could then be tested in the remote setting and compared with the performance of the remotely trained models. Unfortunately, with the available data from the presented study, such comparison is not possible.

5. OSNR OPTIMIZATION

In order to verify whether the trained EDFA model can be used for modeling another EDFA device in the system, we show that the link STN-RDG*-FRX-RDG*-STN can be modeled and optimized for a flat output OSNR with one additional EDFA at RDG. Similar to our work in [37], we build a cascade of fiber and EDFA models and optimize the input power profile of the system, with the cost function $Cost = -\min_{\lambda_k}(\mathrm{OSNR}(\lambda_k))$, targeting a flat OSNR. Once a differentiable EDFA model is available, input profile optimization is applied using the techniques detailed in [36, 37]. The optimization uses gradient descent to update the input power profile. The gradients are obtained via automatic differentiation within the software framework PyTorch [50]. As in our previous work [36, 37] the cost function



(a) 439.4 km, SNR **(b)** 592.4 km, SNR

(c) 439.4 km, OSNR **(d)** 592.4 km, OSNR

Fig. 10. Performance in SNR **(top)** and OSNR **(bottom)** of the link scenario STN-RDG*-PGT*-RDG*-FRX-RDG*-STN (439.4 km, **left**), and of the link scenario STN-RDG*-PGT*-RDG*-PGT*-RDG*-FRX-RDG*-STN (592.4 km, **right**). All four figures show measurements (solid lines with markers) and model predictions (dashed lines) of the system performance. Markers differentiate the SNR and OSNR measurements with a flat (dot marker), optimized (triangle marker), optimized without SRS (diamond marker), optimized without SRS and TRX (circle marker) input power profiles.

is motivated by a combination of factors. For a fixed launch power and due to the EDFA gain competition and nonlinear fiber effects, the OSNR/SNR must be upper bounded. Then, an increase in OSNR/SNR in one channel leads to a decrease in OSNR/SNR in another. Assuming the optimization is not getting stuck in a local optimum, the optimizer should iterate towards a constant value across channels. In contrast to the training system setup, no artificial attenuation is applied before any EDFA device and the noise power profile at the input to the model is chosen to be the average of the noise power profiles of the dataset in Eq. (7). The WSSs are used to apply a flat or optimized power profile across the WDM channels. The signal is launched into the link at 20 dBm total power, the APC setting of the first APC device is fixed and set to 20 dBm, whereas the APC setting of the second EDFA device is swept from 12 to 20 dBm in steps of 2 dB. For each APC setting (second EDFA device) the input power profiles are optimized in a separate process resulting in 5 optimized input power profiles. After optimization, the obtained input power profiles are transmitted through the system with corresponding APC settings and the output OSNR is extracted from an OSA. The OSNR is calculated from the power profiles and the interpolated noise profiles as described in Section 4 A. For comparison, also flat input power profiles are transmitted through the link with corresponding OSNR measurements. Fig. 7a shows in dashed and solid lines the flat and optimized input power profiles, respectively, at different APC settings of the second EDFA device. Fig. 7b shows in dashed and solid lines the OSNR for flat and optimized input power profiles, respectively, at different APC settings of the second EDFA device. Additionally, Fig. 7b shows the model prediction in dotted lines for the optimized input power profiles. The overall RMSE of the 5 predicted and measured OSNR values for optimized input power profiles is 0.23 dB. Besides the low error on these predictions, the fact that the model was used to optimize the input power profiles towards a flat output OSNR given a range of system settings verifies our assumption that the trained EDFA model can also be used to model other devices within the system. The results further show that the EDFA gain flatness is dependent on the APC setting. The EDFA is in constant power control mode and the APC setting ranges from 12 dBm and 20 dBm. For decreasing APC setting the slope of the OSNR plots increases for a flat input profile. Next, we compare the performance of the optimized input profiles compared to their flat counterparts in terms of the minimum OSNR and the difference between the maximum OSNR and minimum OSNR across wavelength. The former metric shows the worst performing channel and the latter describes the flatness of the OSNR with 0 being all flat. Two link scenarios are investigated, STN-RDG*-STN and STN-RDG*-FRX-RDG*-STN. Fig. 8a shows the minimum OSNR and Fig. 8b shows difference of maximum and minimum OSNR in respect to the EDFA to total output power levels. For the longer link scenario the optimized profile outperforms the flat profile in terms of worst performing channel and flatness of OSNR. For the shorter link scenario there is no such benefit, which is explained by the limited amount of tilt the spectra undergo during propagation through the link. We conclude that for longer links and a larger number of amplification stages, larger benefits are expected from the optimized profiles and our method.

6. SNR OPTIMIZATION

For the SNR optimization, the setup from the OSNR optimization is adjusted and a data signal is introduced, as shown in Fig. 9. The data channel uses 64QAM with probabilistic amplitude shaping (PAS) [51], low-density parity check forward error correction (FEC) with overhead of 33%, root-raised-cosine pulse shaping with 0.3 roll-off at a baud rate of 15.33 GBd. Due to the limited bandwidth of the coherent receiver of 10 GHz, without loss of generality, a relatively small symbol rate of 15.33 GBd

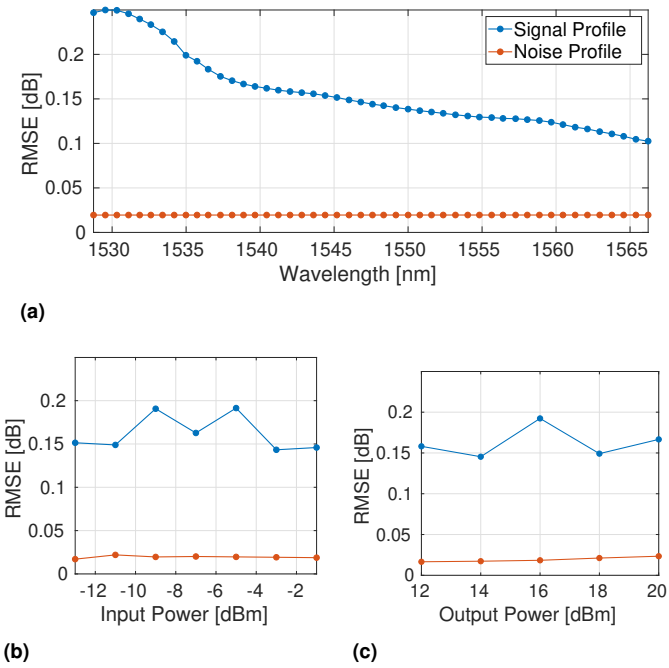


Fig. 11. RMSE on the test set for power (blue) and noise (red) profile in respect to **(a)** the wavelength, **(b)** the EDFA input power and **(c)** the EDFA APC setting on simulated data.

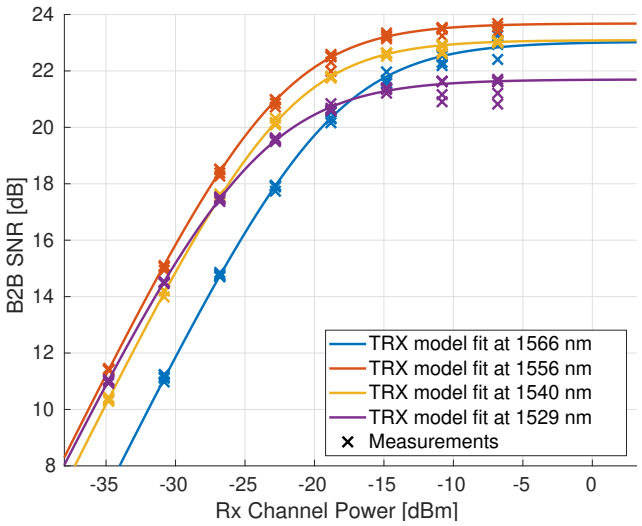


Fig. 12. Receiver channel power in respect to B2B SNR at four different channels, model fit (solid lines) and system measurements (cross markers).

is chosen, which also allows an integer oversampling of the arbitrary waveform generator (AWG) and thus lower implementation complexity. Polarization multiplexing is emulated with a standard delay-and-add technique. The entropy of the constellation is chosen to 5.5 bits/symbol, which is near-optimal for the range of expected link SNRs below 18 dB. The decoding threshold for this coding and modulation scheme was measured in back-to-back (B2B) to be 12.5 dB of SNR. The data signal is multiplexed into the system on the wavelength under test. After the WSSs apply the flat or optimized power profile across the WDM channels, the signal is launched into the network at a total power of 18 dBm. Similarly, all EDFAs are set to an APC setting of 18 dBm. This was found optimal for the scenarios under consideration. At the receiver, the data channels are extracted with two optical filters and an EDFA, followed by a coherent front-end and a digital storage oscilloscope (DSO). The digital signal processing (DSP) is pilot-based as reported in our previous work [37]. In this work, the ASE interfereres follow a Gaussian distribution, resulting in cross-channel nonlinearities accurately described by the standard GN model. Furthermore, the PAS-based constellation exhibit similar high-order moments as a Gaussian [52], allowing to apply the standard GN model for the fiber with sufficient accuracy. Two links of the network are chosen, for the demonstration of SNR optimization. The *shorter* link of 439.4 km length and the *longer* link of 592.4 km length, denoted as STN-RDG*-PGT*-RDG*-FRX-RDG*-STN and STN-RDG*-PGT*-RDG*-PGT*-RDG*-FRX-RDG*-STN, respectively. Overall, in the shorter and longer link the signal is amplified 4 and 6 times, respectively, by independent EDFAs. In contrast to Section 5, a TRX penalty model is added to the mathematical model of the relevant link scenario, allowing prediction and optimization of the input power profile launched into the link in regards to the SNR. The cost function changes accordingly to $Cost = -\min_{\lambda_k}(\mathrm{SNR}(\lambda_k))$, targeting a flat SNR. The optimization algorithm is equivalent to the OSNR optimization, but the model also includes the differentiable TRX penalty model. Optimization is performed for mathematical models of varying levels of complexity: 1) including all impairments; 2) excluding SRS and Kerr fiber nonlinearities; 3) excluding nonlinearities and the TRX penalties. Together with the flat input profile, this leads to 4 different candidate power profiles. The received SNR of up to 24 out of the 48 WDM channels is measured. In Fig. 10, the SNR and OSNR measurements and model predictions of both link scenarios are shown. Solid lines with blue dot, orange triangle, green diamond and red circle markers show the SNR (top) and OSNR (bottom) measurements, for a flat, optimized, optimized excluding nonlinear (NL) and optimized excluding NL/TRX power profile. On the left for the shorter link and on the right for the longer link. Dashed lines show the model predictions in the corresponding colors. Our method shows an improvement in both SNR and OSNR, especially in the shorter and longer wavelength regions. When the system is optimized on the mathematical model including all impairments, the minimum SNR values (around 1532.5 nm) are improved by < 0.1 dB and ≈ 0.8 dB, for the short and long link, respectively. In addition to the minimum SNR, we also note the improvement of the margin in the long wavelength region by ≈ 0.8 dB and ≈ 2.2 dB for the short and long link, respectively. The minimum OSNR values are improved by ≈ 2.1 dB and ≈ 4.8 dB, for the short and long link, respectively. The large improvements in OSNR are not translated to SNR which suggests that the TRX setup is not fully accounted for by the TRX penalty model. The results further show that exclusion of the NL and TRX effects from the mathematical model does not translate to improvement in the minimum SNR and OSNR in these optimized input profiles. In particular, the SRS effects are important, as shown by the tilt of the OSNR curve, the conversion of optical energy from the low to the high wavelengths is not modeled and is not compensated for during the optimization of the input power profile. For more accurate predictions and improved optimization performance, the impact of additional effects should be considered to be modeled, such as polarization dependent loss and filtering effects. The accuracy of our remote modeling approach is lim-

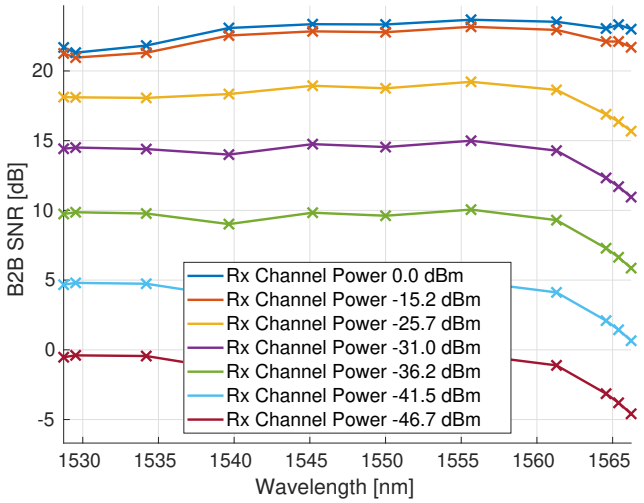


Fig. 13. Wavelength in respect to B2B SNR at seven different received channel powers interpolated from the per channel model fits.

ited by the assumption of similarity between different EDFAs in the network. As the number of devices in a point-to-point link increases, the accuracy decreases, potentially leading to decreased downstream QoT prediction and thus reduced system performance optimization. Our methodology can be improved by applying the remote training procedure sequentially for other devices in the network, although this will increase the modeling time. Additionally, unknown devices or defects in the system can also impact the accuracy of the EDFA modeling, as an accurate model of all components surrounding the EDFA is required for the methodology to be effective.

7. CONCLUSION

We have shown a method that allows for modeling of an EDFA device without requiring direct access to its network node. A novel ML aided EDFA model is trained on data obtained from signals transmitted on a point-to-point link with the amplifier as the only amplification stage. Assuming sufficient similarity between the different EDFA devices deployed in the network, only a single EDFA device is characterized and used to model all other EDFA devices in the network. We verify our assumption by using the obtained model to optimize the launch power spectral profiles into links including up to 6 amplified spans and up to 592.4 km of length. The EDFA model allowed for a prediction error below 1.62 dB in OSNR and 1.85 dB in SNR when transmitting the optimized profiles over the longest link, supporting up to 2.2 dB of margin improvements.

APPENDIX

A. EDFA MODEL PERFORMANCE

The EDFA model by Meseguer et al. [35] and the ML extension described in Section 2 B is an approximation of the analytical EDFA model presented by Saleh et al. [48] In this Appendix, a short study is presented on how the EDFA ML model performs on simulated data obtained from the analytical EDFA model. The analytical EDFA model is parameterised as presented in the appendix of Perin et al. [41] Similar to Section 4 A, we acquired through simulation 1000 data samples per profile, at 7 different EDFA total input power levels and at 5 different EDFA APC power settings. With the same training, validation and test split (10%/20%/70%), the EDFA ML model from Section 4 B is trained and tested. The obtained test RMSE is 0.16 and 0.02 dB for signal and noise profiles, respectively, and Fig. 11 shows the RMSE with respect to wavelength, EDFA total input power levels and EDFA APC power settings. We note that the presented simulation study is free of measurement noise, which explains that the observed error is of a different magnitude than in Section 4 B. Qualitatively, the results show that the approximation by the EDFA model presented in this paper cause modeling errors for shorter wavelengths. In the simulation, low EDFA input powers and high EDFA APC power settings do not contribute to modeling errors. We conclude that the *equivalent flat* approximation by Meseguer et al. leads to errors at short wavelengths, which also explains the error at short wavelength for the remote modeling case in Section 4 B and the error in OSNR/SNR prediction at short wavelengths throughout this work.

B. TRANSCEIVER PENALTY MODEL

This Section describes the TRX penalties model which is required to model the SNR of the system including limitations of the optical front-end and DSP. We assume that the TRX penalties are dependent on the received channel power and can be fully characterized by B2B measurements. We carried out B2B SNR measurements of the 48 WDM channel system described in Section 6 and depicted in Fig. 9. For the measurements, a subset of 11 of the 48 WDM channels is chosen and the received channel power is swept by using an attenuator. We propose the following parameterized softplus expression that agrees well with the measurements in Fig. 12. We note that it can be expected that for commercially available pluggable transceivers, the penalty is lower and a good model of the penalties may be simpler. The measurements for each channel m, $m = 1...11$, are fit to a reflected softplus function:

$$\begin{aligned} \text{softplus}(x,\beta) &= \frac{1}{\beta}\log(1.0+\exp(\beta x)), \\ TRX(x) &= -\text{softplus}(-(x+x_0),\beta)+y_0, \end{aligned} \tag{10}$$

where softplus$(\cdot)$ is the softplus function with a β roll off parameter, and $TRX(\cdot)$ is the softplus function point reflected in the origin with two additional shift parameters x_0, y_0 along the abscissa and ordinate, respectively. For every set of measurements at wavelength λ_m the parameters β, x_0 and y_0 are fit such that:

$$TRX_m(P^i_{S,m}) \approx SNR^i_m, \tag{11}$$

where $P^i_{S,m}$ and SNR^i_m are the received B2B channel power and SNR from 5 independent measurements $\{(P^i_{S,m}, SNR^i_m)\}_{i=1}^5$. The fit is in logarithmic domain meaning power and SNR terms are in dBm and dB, respectively. Fig. 12 depicts measurements and corresponding fits for 4 out of the 11 measured WDM channels. For a wavelength not part of the measured 11 channel subset, an interpolation across wavelength results in a fully continuous TRX model across power and wavelength, $\overline{TRX}(P_S,\lambda)$. Examples of the subsequent interpolation are shown in Fig. 13. As shown, the SNR decreases with decreasing per-channel received power, and it saturates at large per-channel received power $P_{sat} > -7$ dBm. Then, the SNR penalty in dB is defined as:

$$\begin{aligned} &\Delta SNR(P_S,\lambda) = \\ &\overline{TRX}(P_{sat},\lambda) - \overline{TRX}(P_S,\lambda). \end{aligned} \tag{12}$$

As mentioned in Section 6, the system accuracy is highly-dependent on the TRX penalty model. The TRX model presented is only dependent on the received channel signal power. In practice, the TRX penalties also depend on the received OSNR and received channel noise power, and they should ideally be included in the penalty model. This is left for future research.

ACKNOWLEDGMENT

These experiments were carried out on EPSRC's National Dark Fibre Facility. The work received financial support by the UK's EPSRC through research projects EP/S028854/1 and EP/S002871/1, and Innovation Fund Denmark project RANON, ref. 1047-00013.

REFERENCES

1. Y. Pointurier, "Design of low-margin optical networks," J. Opt. Commun. Netw. **9**, A9–A17 (2017).
2. V. Curri, A. Carena, A. Arduino, G. Bosco, P. Poggiolini, A. Nespola, and F. Forghieri, "Design strategies and merit of system parameters for uniform uncompensated links supporting Nyquist-WDM transmission," J. Light. Technol. **33**, 3921–3932 (2015).
3. P. Soumplis, K. Christodoulopoulos, M. Quagliotti, A. Pagano, and E. Varvarigos, "Network planning with actual margins," J. Light. Technol. **35**, 5105–5120 (2017).
4. C. Delezoide, K. Christodoulopoulos, A. Kretsis, N. Argyris, G. Kanakis, A. Sgambelluri, N. Sambo, P. Giardina, G. Bernini, D. Roccato *et al.*, "Marginless operation of optical networks," J. Light. Technol. **37**, 1698–1705 (2019).
5. C. Delezoide, "Method for a comprehensive evaluation of margins in optical networks," in *45th European Conference on Optical Communication (ECOC 2019),* (IET, 2019), pp. 1–4.
6. J.-L. Augé, "Can we use flexible transponders to reduce margins?" in *Optical Fiber Communication Conference,* (Optical Society of America, 2013), pp. OTu2A–1.
7. M. Jinno, "Elastic optical networking: Roles and benefits in beyond 100-Gb/s era," J. Light. Technol. **35**, 1116–1124 (2016).
8. K. Christodoulopoulos, C. Delezoide, N. Sambo, A. Kretsis, I. Sartzetakis, A. Sgambelluri, N. Argyris, G. Kanakis, P. Giardina, G. Bernini *et al.*, "Toward efficient, reliable, and autonomous optical networks: the ORCHESTRA solution," J. Opt. Commun. Netw. **11**, C10–C24 (2019).
9. K. Kaeval, H. Griesser, K. Grobe, J.-P. Elbers, M. Tikas, and G. Jervan, "Channel performance estimations with extended channel probing," in *Photonic Networks; 21th ITG-Symposium,* (VDE, 2020), pp. 1–5.
10. O. Karandin, F. Musumeci, O. Ayoub, A. Ferrari, Y. Pointurier, and M. Tornatore, "Quantifying resource savings from low-margin design in optical networks with probabilistic constellation shaping," in *2021 European Conference on Optical Communication (ECOC),* (IEEE, 2021), pp. 1–4.
11. K. Kaeval, S. L. Jansen, F. Spinty, K. Grobe, H. Griesser, T. Fehenberger, M. Tikas, and G. Jervan, "Characterization of the optical spectrum as a service," J. Opt. Commun. Netw. **14**, 398–410 (2022).
12. I. F. de Jauregui Ruiz, A. Ghazisaeidi, T. Zami, S. Louis, and B. Lavigne, "An accurate model for system performance analysis of optical fibre networks with in-line filtering," in *45th European Conference on Optical Communication (ECOC 2019),* (IET, 2019), pp. 1–4.
13. L. Barletta, A. Giusti, C. Rottondi, and M. Tornatore, "QoT estimation for unestablished lighpaths using machine learning," in *Optical Fiber Communication Conference,* (Optical Society of America, 2017), pp. Th1J–1.
14. J. Wass, J. Thrane, M. Piels, R. Jones, and D. Zibar, "Gaussian process regression for WDM system performance prediction," in *Optical Fiber Communication Conference,* (Optica Publishing Group, 2017), pp. Tu3D–7.
15. L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for OSNR estimation," Opt. express **27**, 19398–19406 (2019).
16. J. Yu, W. Mo, Y.-K. Huang, E. Ip, and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," J. Opt. Commun. Netw. **11**, C48–C57 (2019).
17. J. Cho, S. Chandrasekhar, E. Sula, S. Olsson, E. Burrows, G. Raybon, R. Ryf, N. Fontaine, J.-C. Antona, S. Grubb *et al.*, "Supply-power-constrained cable capacity maximization using multi-layer neural networks," J. Light. Technol. **38**, 3652–3662 (2020).
18. A. D'Amico, S. Straullu, A. Nespola, I. Khan, E. London, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Using machine learning in an open optical line system controller," J. Opt. Commun. Netw. **12**, C1–C11 (2020).
19. J. Pesic, M. Lonardi, T. Zami, N. Rossi, and E. Seve, "Transfer learning using ANN for G-OSNR estimation in WDM network topologies," in *Photonic Networks and Devices,* (Optical Society of America, 2020), pp. NeM3B–3.
20. M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," J. Opt. Commun. Netw. **13**, B92–B101 (2021).
21. N. Morette, I. F. de Jauregui Ruiz, and Y. Pointurier, "Leveraging ML-based QoT Tool Parameter Feeding for Accurate WDM Network Performance Prediction," in *Optical Fiber Communication Conference,* (Optical Society of America, 2021), pp. Th4J–4.
22. Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," J. Opt. Commun. Netw. **13**, B60–B71 (2021).
23. H. Maryam, T. Panayiotou, and G. Ellinas, "Learning quantile QoT models to address uncertainty over unseen lightpaths," Comput. Networks **212**, 108992 (2022).
24. E. Seve, J. Pesic, and Y. Pointurier, "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds," J. Opt. Commun. Netw. **13**, C21–C30 (2021).
25. N. Morette, I. F. de Jauregui Ruiz, H. Hafermann, and Y. Pointurier, "On the Robustness of a ML-based Method for QoT Tool Parameter Refinement in Partially Loaded Networks," in *Optical Fiber Communication Conference,* (Optica Publishing Group, 2022), pp. M3F–1.
26. K. Kaeval, F. Slyne, S. Troia, E. Kenny, J.-J. Pedreño-Manresa, S. K. Patri, K. Grobe, D. C. Kilper, M. Ruffini, and G. Jervan, "Exploring Service Margins for Optical Spectrum Services," in *2022 European Conference on Optical Communication (ECOC),* (IEEE, 2022), p. Mo3B.4.
27. M. Lonardi, J. Pesic, T. Zami, and N. Rossi, "The perks of using machine learning for QoT estimation with uncertain network parameters," in *Photonic Networks and Devices,* (Optica Publishing Group, 2020), pp. NeM3B–2.
28. O. Karandin, A. Ferrari, F. Musumeci, Y. Pointurier, and M. Tornatore, "Low-Margin Optical-Network Design with Multiple Physical-Layer Parameter Uncertainties," in *2022 European Conference on Optical Communication (ECOC),* (IEEE, 2022), p. Mo3B.2.
29. E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," J. Opt. Commun. Netw. **10**, A298–A306 (2018).
30. E. Seve, J. Pesic, C. Delezoide, A. Giorgetti, A. Sgambelluri, N. Sambo, S. Bigo, and Y. Pointurier, "Automated fiber type identification in SDN-enabled optical networks," J. Light. Technol. **37**, 1724–1731 (2019).
31. 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.
32. S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman, and D. C. Kilper, "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra," in *2018 European Conference on Optical Communication (ECOC),* (IEEE, 2018), pp. 1–3.
33. M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in *2019 21st International Conference on Transparent Optical Networks (ICTON),* (IEEE, 2019), pp. 1–4.
34. S. Zhu, C. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman,

and D. Kilper, "Hybrid machine learning EDFA model," in *Optical Fiber Communication Conference,* (Optica Publishing Group, 2020), pp. T4B–4.
35. A. C. Meseguer, J.-C. Antona, A. Bononi, J. Cho, S. Grubb, P. Pecci, O. Courtois, and V. Letellier, "Highly accurate measurement-based gain model for constant-pump EDFA for non-flat WDM inputs," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (IEEE, 2021), pp. 1–3.
36. M. P. Yankov, U. C. de Moura, and F. Da Ros, "Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters," J. Light. Technol. **39**, 3154–3161 (2021).
37. M. P. Yankov, P. M. Kaminski, H. E. Hansen, and F. Da Ros, "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).
38. C. Delezoide, P. Ramantanis, and P. Layec, "Leveraging field data for the joint optimization of capacity and availability in low-margin optical networks," J. Light. Technol. **38**, 6709–6718 (2020).
39. C. Delezoide, P. Ramantanis, and P. Layec, "Investigating Q-drops and Their Probable Causes," in *2022 European Conference on Optical Communication (ECOC),* (IEEE, 2022), p. Mo3B.5.
40. I. Roberts, J. M. Kahn, and D. Boertjes, "Convex channel power optimization in nonlinear WDM systems using Gaussian noise model," J. Light. Technol. **34**, 3212–3222 (2016).
41. J. K. Perin, J. M. Kahn, J. D. Downie, J. Hurley, and K. Bennett, "Importance of amplifier physics in maximizing the capacity of submarine links," J. Light. Technol. **37**, 2076–2085 (2019).
42. R. Hashemi, H. Beyranvand, and H. Rabbani, "Joint channel power and amplifier gain optimization in coherent DWDM systems," Opt. Commun. **475**, 126212 (2020).
43. V. V. Garbhapu, A. Ferrari, I. F. de Jauregui Ruiz, D. Le Gac, G. Charlet, and Y. Pointurier, "Network-Wide SNR-based Channel Power Optimization," in *2021 European Conference on Optical Communication (ECOC),* (IEEE, 2021), pp. 1–4.
44. A. Ferrari, V. V. Garbhapu, D. Le Gac, I. F. de Jauregui Ruiz, G. Charlet, and Y. Pointurier, "Demonstration of AI-Light: an Automation Framework to Optimize the Channel Powers Leveraging a Digital Twin," in *Optical Fiber Communication Conference,* (Optica Publishing Group, 2022), pp. M3Z–14.
45. S. E. Landero, I. F. de Jauregui Ruiz, A. Ferrari, D. Le Gac, Y. Frignac, and G. Charlet, "Link Power Optimization for S+ C+ L Multi-band WDM Coherent Transmission Systems," in *2022 Optical Fiber Communications Conference and Exhibition (OFC),* (IEEE, 2022), pp. 1–3.
46. X. Yang, A. Ferrari, N. Morette, D. Le Gac, S. Escobar Landero, G. Charlet, and Y. Pointurier, "Experimental Impact of Power Re-Optimization in a Mesh Network," in *2022 European Conference on Optical Communication (ECOC),* (IEEE, 2022), p. Mo3B.3.
47. R. T. Jones, K. R. H. Bottrill, N. Taengnoi, P. Petropoulos, and M. P. Yankov, "Spectral Power Profile Optimization of Field-Deployed WDM Network by Remote Link Modeling," in *2022 European Conference on Optical Communication (ECOC),* (IEEE, 2022), p. We1A.4.
48. A. Saleh, R. Jopson, J. Evankow, and J. Aspell, "Modeling of gain in erbium-doped fiber amplifiers," IEEE Photonics Technol. Lett. **2**, 714–717 (1990).
49. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
50. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32,* H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds. (Curran Associates, Inc., 2019), pp. 8024–8035.
51. G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation," IEEE Transactions on Commun. **63**, 4651–4665 (2015).
52. T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On Probabilistic Shaping of Quadrature Amplitude Modulation for the Nonlinear Fiber Channel," J. Light. Technol. **34**, 5063–5073 (2016).