



Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters

Yankov, Metodi Plamenov; Moura, Uiara Celine de; Da Ros, Francesco

Published in:
Journal of Lightwave Technology

Link to article, DOI:
[10.1109/JLT.2021.3061632](https://doi.org/10.1109/JLT.2021.3061632)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Yankov, M. P., Moura, U. C. D., & Da Ros, F. (2021). Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters. *Journal of Lightwave Technology*, 39(10), 3154-3161. <https://doi.org/10.1109/JLT.2021.3061632>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters

Metodi Plamenov Yankov, *Member, IEEE*, Uira Celine de Moura, Francesco Da Ros *Member, OSA, Senior Member, IEEE*,

Abstract—In this paper, machine learning is used to create a differentiable model for the input-output power spectral profile relations of C-band erbium-doped fiber amplifiers (EDFAs). The EDFA model is demonstrated to generalize to multiple physical devices of the same make while only trained on experimental data from a single unit. The model is combined with a differentiable model for simulating stimulated Raman scattering (SRS) effects during propagation through the optical fiber to create a differentiable model for a multi-span system with an arbitrary configuration of number of spans, length per span and launch power per span. The cascade system model is used to predict and optimize the power profile output of several such experimental configurations of up to three spans with an arbitrary target power profile. A flat target profile is exemplified experimentally, achieving <3 dB of power excursions for EDFAs exhibiting >10 dB of excursion per device in the cascade. The experimental data used to create the EDFA model is made public and available online.

Index Terms—EDFA, Machine learning, power spectral density, power optimization, stimulated Raman scattering.

I. INTRODUCTION

The exponentially growing data rate demand in the last decades has exerted an enormous amount of pressure on the backbone long haul and regional fiber optic networks. At the same time, energy efficiency and cost of the optical links is getting critical as more and more connections are being deployed. For a long time, spectral uniformity of the performance of wavelength division multiplexed (WDM) optical links has been ensured or at least approached by compensating for the non-frequency flat gain of erbium doped fiber amplifiers (EDFA) with gain flattening filters (GFFs). Since GFFs essentially waste power in the over-performing channels of the WDM signal, this solution is by definition energy-inefficient. On the other hand, 'naked' EDFAs (ones without GFFs) exhibit gain tilt and ripples [1], which result in power excursions accumulating over multi-span links. The gain tilt and ripples are nonlinear functions of the input power spectral density (PSD) and are nontrivial to model analytically, with typical models resulting in significant inaccuracies [2]. Recently, machine learning (ML) has been employed to produce more accurate models for the EDFA gain spectrum [1], [2], [3], [4], [5], [6]. These models are trained and tested on the same device [3], [4], [6] resulting in uncertainties of their generalization capabilities, apply to gain-flattened EDFAs only [1], lack system demonstration where a multitude of EDFAs are employed, or still produce too poor accuracy for practical

applications [4]. Multi-span system demonstrations have been analyzed for only a few channels [7] or limited to the linear region of transmission [6].

The power excursion problem is exacerbated by the stimulated Raman scattering (SRS) effect, resulting in power transfer from channels at high frequencies to channels at low frequencies. The push for high data rates inevitably drives optical links to operate with high power entering the nonlinear region, where the SRS effect becomes pronounced, even for classical C-band links limited to 5 THz of bandwidth. Some solutions exist for countering the SRS effect by launch PSD optimization [7], [8], [9], which do not take into account the gain tilt of practical EDFAs without a GFF. Alternatively, ML can be applied for the end-to-end system [3], [10], [11], which at least in the linear region of transmission can be employed for PSD optimization. Such solutions require a lot of training data to be generated for each link in the network, which is time consuming, and is furthermore susceptible to even minor changes in the link. Fast reconfigurable software defined networks require a component-based model for the link, which supports changes on the fly and allows for fast (power allocation) optimizations.

In this paper, an ML model is developed for the EDFA gain spectrum which is experimentally demonstrated to generalize to multiple physical devices. The model is then used in conjunction with a model for the power evolution in a standard single mode fiber (SSMF) supporting SRS to predict and optimize the PSD at the output of multi-span fiber links. This paper extends our previous contributions on EDFA modeling [12] and system optimization [13] by 1) providing more details and analysis on the EDFA training data and training procedures; 2) providing more details on the experimental procedures; 3) expanding the system characterization analysis to include characterization w.r.t. PSD tilt; and 4) including derivation details of the multi-span system.

II. EDFA CHARACTERIZATION

The task of an EDFA gain model is to predict the output PSD from the input PSD at a given total input power and average gain, or equivalently, at a given total input power and total output power. In this paper, the latter format is chosen. In order not to rely on either simplified analytical models of potentially low accuracy, nor on highly complex and extremely specific physical models, the ML approach applied in this paper is driven by experimentally generated data. The experimental setup for data acquisition is given

The authors are with the Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, meya@fotonik.dtu.dk

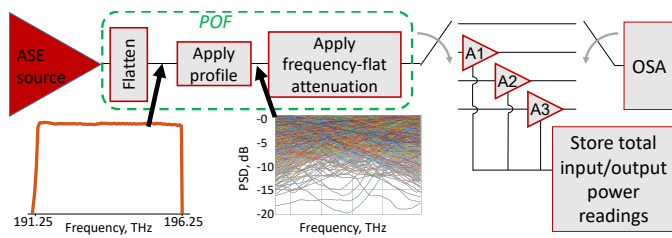


Fig. 1. Experimental EDFA characterization setup. A POF is used to 1) flatten the ASE source; 2) apply the desired profile from the dataset; and 3) attenuate to the desired total power input to the EDFA. The input profile and the EDFAs output profiles are recorded sequentially by an OSA.

in Fig. 1. A power profile at a given total input power is generated by shaping and attenuating an ASE source using a programmable optical filter (POF). The resulting power profile is launched into three different EDFAs of the same make by using a switch, and the resulting PSD is captured by an optical spectrum analyzer (OSA). The EDFAs model is “Keopsys, KPS-STD-BT-C-18-SD-111-FA-FA”. The EDFAs are driven in constant output power mode. For each desired power profile, the specified total input and total output power are swept on a 2D grid. The specified total output power takes values in the set $\{15, 16, 17, 18\}$ dBm. The input power is controlled by imposing an additional frequency flat attenuation in the POF. The attenuation values are coarsely calibrated to produce total power input to the EDFAs in the range $[-9; 9]$ dBm. Histograms of the *measured* total input and output powers are given in Fig. 2. The total output power is controlled using a feedback loop inside the EDFA, resulting in an accuracy better than 0.5 dBm and average measured values of $\{15.0, 16.1, 17.3, 18.2\}$ dBm for the specified $\{15, 16, 17, 18\}$ dBm. Due to the dependence of the total power on the desired PSD output of the POF, the EDFA input power cannot be controlled as accurately as the output. However, this is not a requirement for the model, as long as the *measured* EDFA total input power is used for the characterization and modeling instead of the desired one. This is ensured by extracting the input power reading of the EDFA instrument for each profile and each total input/output power combination. In this study, the EDFA power meters’ precision was ≈ 0.1 dB and their accuracy sufficient to inform an accurate model. It is noted that in the case when the internal EDFA power monitors cannot be trusted, they can be replaced by external ones in order to perform characterization and data collection.

It should also be noted that some combinations cannot be obtained due to physical limitations of the POF. For example, as the input power into the POF is limited by the instrument power handling, a few PSDs with high power excursions resulted in too high filtering loss and did not allow for probing the EDFAs with the highest input power levels.

In summary, a single characterization data point consists of 1) EDFA Id; 2) a unique Id for the target PSD; 3) measured total input power; 4) measured total output power; 5) measured input PSD; and 6) measured output PSD. The measured dataset with all ≈ 270.000 data points (≈ 90.000 per EDFA) is available at [14].

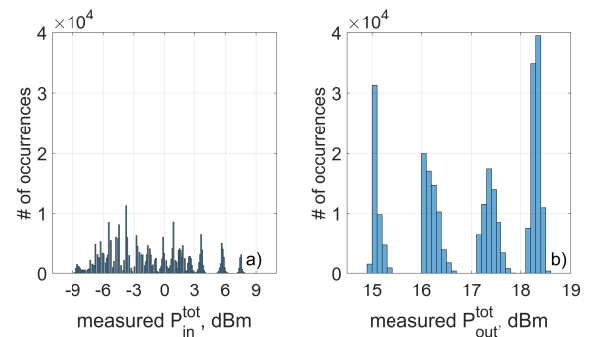


Fig. 2. Histograms of the measured **a)**: total input and **b)**: total output power for all EDFAs for the specified total input power in the range $[-9; 9]$ dBm and total output power in the set $\{15, 16, 17, 18\}$ dBm.

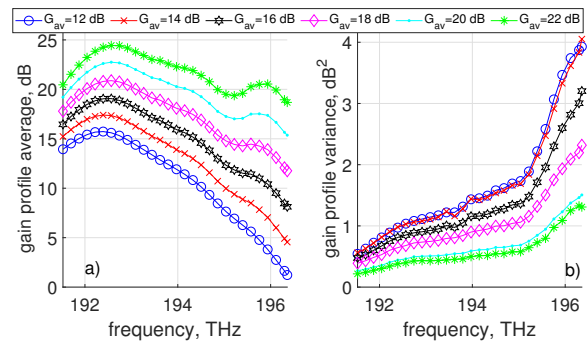


Fig. 3. Characteristics of the employed EDFAs. **a)**: gain profile mean; and **b)**: gain profile variance for all output powers and all amplifiers at given average gain.

A. Characterization dataset

The profile dataset consists of ≈ 2000 distinct profiles, generated by parametrizing a smoothed Gaussian random walk of the form

$$P_{in}(n) = P_{in}(n-1) + w_n, \quad (1)$$

where $P_{in}(n)$ is the power in the log domain at the n -th frequency on the specified grid, w_n is a zero-mean Gaussian distributed variable with a variance σ_W^2 , and $P_{in}(0) \in [-14; 0]$. The power spectral density is defined as $PSD = \{P(1), P(2), \dots, P(n)\}$. The points are generated on a uniform frequency grid with a step of 61.6 GHz, resulting in $N = 83$ grid points in the C-band, limited by the bandwidth of the POF. The profile is then smoothed by an all-ones box filter of a variable length, which controls the level of smoothness. Longer filter results in a smoother profile and vice versa. Filters of lengths 11, 31 and 51 are used in this work. The variance σ_W^2 implicitly controls the maximum excursion in the profile and is chosen from the set $\{0, 0.1, 0.5\}$, resulting in excursions between 0 (corresponding to $\sigma_W^2 = 0$) and 20 dB, defined as $\max_n P_{in}(n) - \min_n P_{in}(n)$. The desired power profiles are then translated to attenuation profiles to be applied by the POF, and are illustrated in the inset in Fig. 1.

The average gain profiles of the EDFAs are given in Fig. 3a). The profiles are obtained by first calibrating the input PSD PSD_{in} and the output PSD PSD_{out} to the total P_{in}^{tot} and P_{out}^{tot} recorded by the amplifier instruments, and then averaging the

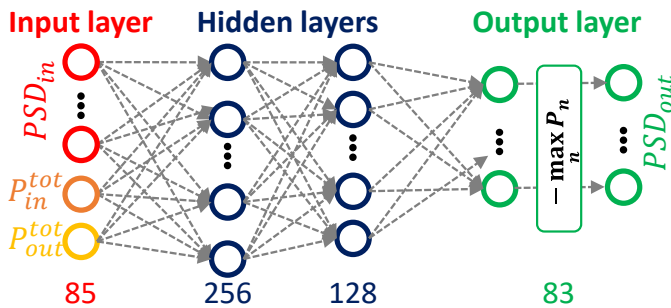


Fig. 4. Neural network topology for the EDFA gain model. The inputs consist of the normalized input PSD together with the total input and output powers. The output consists of the normalized output PSD. Two-layer model is adopted with 256 and 128 nodes per layer, respectively, and a ReLU activation function.

gain profile $PSD_{out} - PSD_{in}$ for all power profiles and all amplifiers at a given average gain, obtained as $G_{av} = P_{out}^{tot} - P_{in}^{tot}$ (with a precision of $\pm 1dB$). For higher average gain, the gain profile is flatter, with excursions of ≈ 5 dB at $G_{av} = 22$ dB. The excursions becomes more severe with reduced average gain and approach 15 dB at $G_{av} = 12$ dB.

The spread of the gain profiles is given in terms of the variance in Fig. 3b). The variance is calculated over the profiles collected as for Fig. 3a), and clearly demonstrates the output profile dependency on the input profile and the average gain. The amplifier gain profile appears more constant for low frequencies and at high gain, and becomes highly input-dependent for low G_{av} and at high frequencies.

B. Machine learning based EDFA model

A general purpose neural network (NN) is adopted in this work to create the ML model. An illustration is provided in Fig. 4. The NN input is 85 dimensional and consists of $\{P_{in}^{tot}, P_{out}^{tot}, P_{in}(1), P_{in}(2), \dots, P_{in}(83)\}$. The NN output is 83 dimensional and consists of $\{P_{out}(1), P_{out}(2), \dots, P_{out}(83)\}$. Both the input and output PSDs are normalized such that $\max_n P_{in}(n) = 0$ and $\max_n P_{out}(n) = 0$. The network is trained using gradient descent with the Adam optimizer in the PyTorch automatic differentiation environment using the mean squared error (MSE) as cost function [15]. The network topology is coarsely optimized to two fully-connected layers with 256 and 128 nodes in the first and second layer, respectively. No improvement was visible with larger network topologies, either in depth or size per layer. The ReLU activation function was applied in the hidden layers.

The dataset is split into 75% training (of which 10% validation) and 25% testing. Modeling performance is estimated in terms of *intra*-MSE (training and testing data from the same EDFA) and *inter*-MSE (training on a single EDFA and testing on the other two). A summary of the results is provided in Fig. 5. An *intra*-EDFA MSE of $\approx 0.02dB^2$ was achieved for any of the EDFAs. Similar error was achieved when data from all EDFAs were used for training and testing (indicated as A1+A2+A3 in the legend of Fig. 5). For comparison, an *intra*-EDFA *root* MSE of $\approx 0.16dB$ was achieved in [3], and *intra*-EDFA *root* MSE of $\approx 0.5dB$ was achieved in [4]. The MSE is increased to up to $\approx 0.045dB^2$ of *inter*-MSE,

with performance relatively flat across average gains and total output power, worst for EDFA A2 operated at high gain or low output power. The *inter*-MSE is mainly attributed to the error at high frequencies, as demonstrated in Fig. 5c). This error can in turn be attributed to fabrication variations between different units, possibly related to the erbium doped fiber length or doping fluctuations.

The MSE is finally studied as a function of the input profile tilt. In order to capture both the direction of the tilt (on the frequency axis) and its severity, the tilt measure adopted in this work is

$$T = 10 \cdot \log_{10} \sum_{n=1}^{n=41} P_{in}^{lin}(n) - 10 \cdot \log_{10} \sum_{n=43}^{n=83} P_{in}^{lin}(n), \quad (2)$$

where $P_{in}^{lin}(n)$ is the power at the n -th frequency in watts. Note that the tilt is not equivalent to the *excursion*, which was defined in Section II-A. A typical example of an input-output relation for a given input profile of tilt $T \approx -8.84$ dB (excursion of ≈ 12 dB) is given in Fig. 6, for which the prediction results in $MSE < 0.032$ dB² for all training configurations.

In Fig. 7, the *inter*-MSE is given as a function of the tilt for the studied total output powers and averaged over the amplifiers and the average gain. Small tilts result in very good prediction with $MSE < 0.02$ dB², increasing to ≈ 0.05 dB² for negative tilts and a more severe MSE of > 0.1 dB² for positive tilts of $T > 9$ dB. The poorer performance at high absolute tilt can be explained by the fact that fewer data points are present in the training dataset. The MSE asymmetry w.r.t. the tilt can be attributed to the positive gain tilt of the amplifiers as shown in Fig. 3a), which amplifies the positive tilt of the profiles for a severe power excursion at the EDFA output often exceeding 20 dB. On the other hand, negative input profile tilts are somewhat compensated by the positive tilt of the amplifier.

III. MULTI-SPAN SYSTEM

The EDFA model from Section II-B can be used as a component in a multi-span system model. As mentioned above, an accurate cascade model can only be realized with component models which generalize to all physical units of the respective device. The two different components in this paper are the EDFA model (trained on A1 and showing the very good generalization properties demonstrated in Section II-B) and a SSMF model (to be detailed below) of a given arbitrary length.

The cascade model considered in this work is summarized in Fig. 8. A PSD for the first span is specified together with the desired $P_{out}^{tot}(k)$ for the k -th EDFA in the system, for $k \in \{1 : K\}$. Without loss of generality, the total input power to the first span is specified such that the average gain of the first EDFA corresponds to the fiber loss of the first span

$$P_{in}^{tot}(1) = P_{out}^{tot}(1) - \alpha \cdot L(1), \quad (3)$$

where α is the fiber loss (in dB/km) and $L(k)$ is the length of the k -th fiber span (in km).

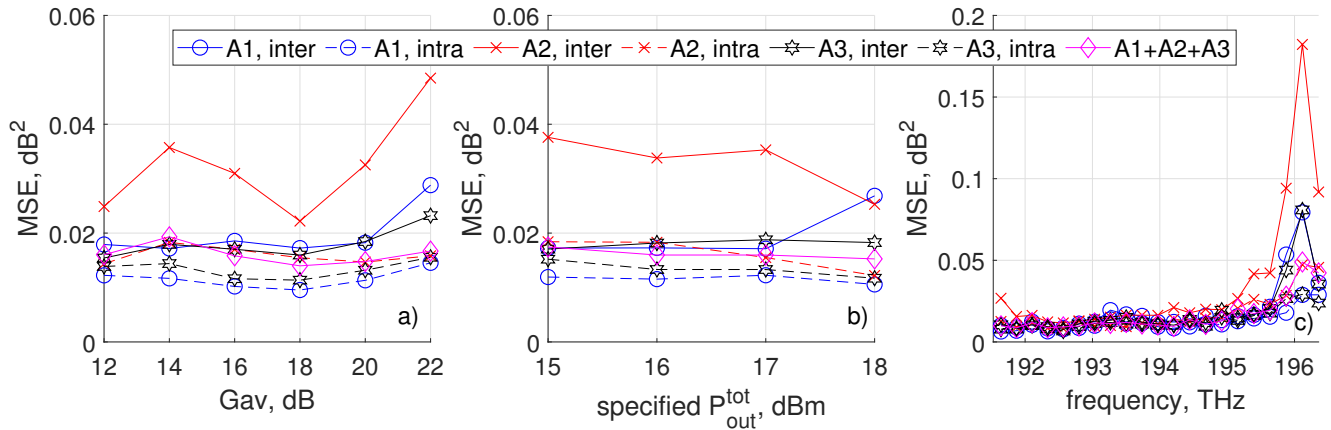


Fig. 5. Summary of the modeling performance. Intra-MSE means trained and tested on the same EDFA (for different profiles). Inter-MSE means trained on one EDFA and tested on the other two (for different profiles). Intra- and inter-EDFA MSE as a function of **a)** the average gain for all studied amplifiers; **b)** the total output power, averaged across average gains; **c)** the channel frequency, demonstrating that the MSE is dominated by the performance at high frequencies.

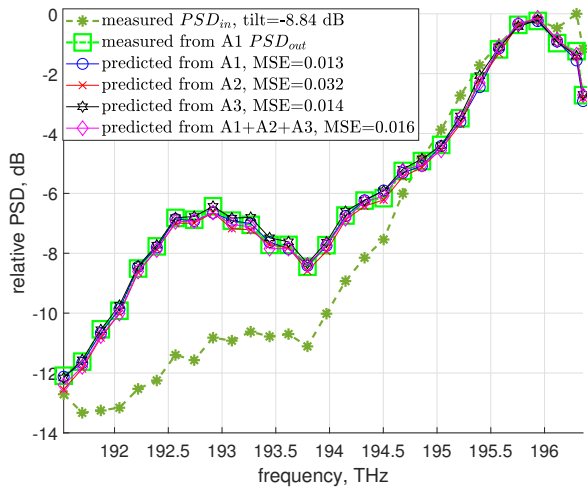


Fig. 6. Example input and output PSDs for amplifier A1. Also the PSDs predicted by models, trained as shown in the legend. Corresponding MSEs also in the legend.

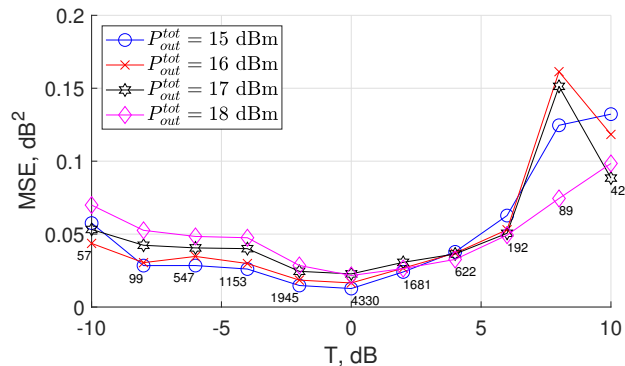


Fig. 7. Inter-EDFA MSE as a function of the input PSD tilt as defined in Eq. (2), averaged over all amplifiers and the average gain. The average number of profiles used to estimate the MSE is given as plot annotations.

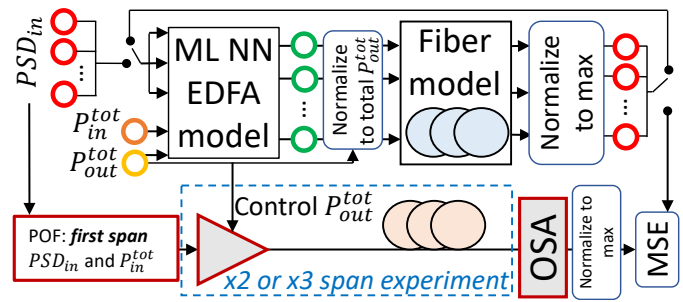


Fig. 8. The multi-span system model used in this work together with the experimental setup. The (normalized) PSD at the input of the first span is specified, then launched into the EDFA model. The EDFA output is normalized to the desired total launch power and launched into the fiber model. The output of the fiber model is normalized to $\max_n P(n) = 0$ and used as input to the next fiber span. After a number of spans, the normalized output is compared to the normalized output of the experimental system, which consists of 2- or 3-span inline EDFA fully-loaded C-band link, loaded with the input PSD by a POF.

As mentioned in Section II-B, the EDFA model produces a PSD, which is normalized such that $\max_n P_{out}(n) = 0$. Before launching the PSD into the fiber model (which is detailed in Section III-A), it is normalized such that $\sum_n P_{out}(n)^{lin} = P_{out}^{tot,lin}(k)$. The PSD output of the fiber is then normalized back to $\max_n P_{out}^{fiber}(n) = 0$, and is used as input to the EDFA of the next span. In this work, the launch power (EDFA output power) is assumed constant for every span, however, the model is flexible to arbitrary launch power per span, as long as it does not fall outside of the training set for the EDFA (in the range [15; 18]), which will result in a poor EDFA and thus system model accuracy.

The model is benchmarked against an experimental inline-amplified multi-span system of either two or three spans, as shown in Fig. 8. The desired PSD is provided to the POF in a similar manner to the case of EDFA characterization in Fig. 1. The POF also *coarsley* controls the total power input to the first EDFA of the experiment. The inline EDFAs are controlled in constant output power mode, as was done for training

the EDFA model. Fiber spools of four different lengths are studied in this work: $L \in \approx \{70, 80, 90, 100\}$ with measured losses of $\alpha^{tot} \in \approx \{15, 16.5, 17.6, 18.7\}$ dB, respectively. In order to maintain generalization of the model regardless of the average loss in the fiber, the fiber lengths are rescaled to effective lengths $\{75.0, 82.5, 88.0, 93.5\}$, respectively, such that the total propagation loss is matched with an average loss of $\alpha = 0.2$ dB/km, which is uniform with distance. The effective lengths can be higher than the true lengths due to additional connector losses and/or true average loss above 0.2 dB/km for those fiber spools. They can also be lower (in the cases of 90 km and 100 km spans), which can be attributed to average loss below 0.2 dB/km. These mismatches are assumed frequency-flat, and thus do not influence the model accuracy. They are therefore not modeled explicitly, but are implicitly accounted for during the normalization.

The output of the last fiber is recorded by an OSA, normalized, and compared to the model output in terms of MSE.

A. Fiber models

Three models for the power evolution in the fiber are studied in this work.

1) *Bulk loss model*: The first model assumes a frequency flat bulk loss (BL), governed by the fiber length

$$P(n, z) = P(n, 0) - \alpha \cdot z, \quad (4)$$

where $P(n, z)$ is the power at the n -th carrier at distance z .

2) *SRS model*: The second model takes into account the SRS effect, which results in a transfer of power from carriers at high frequencies to carriers at low frequencies. Following the notation from [8], the power evolution across distance L_s can be described by

$$P(n, z) = P(n, z - L_s) - \alpha L_s + \sum_{m=1}^{83} \frac{g_R(\omega_m - \omega_n)}{A_{eff}} L_{eff}(L_s) e^{P(m, z - L_s)}, \quad (5)$$

where g_R is the tabulated Raman gain coefficient for a given frequency offset, ω_n is the angular frequency of the n -th carrier, A_{eff} is the fiber effective area, and $L_{eff}(L) = \frac{1 - \exp(-2 \cdot \alpha \cdot L)}{2 \cdot \alpha}$ is the effective power interaction length [8]. Equation (5) is an approximation to the differential equation governing the power evolution in the fiber, and is accurate provided the step size L_s is small enough¹. In this work, L_s was reduced until no improvements in the modeling accuracy were found. In particular, $L_s = 100$ m. We have assumed standard single mode fiber parameters such that $A_{eff} = 82.3 \mu\text{m}^2$, and g_R was taken from [9, Fig. 3] corresponding to the Raman efficiency with pump centered at 1550 nm.

The mismatches in the fiber lengths of the model and the experiment described above are minor, and since they are effected at the end of the span where the power is relatively low, their contribution to the Raman gain is negligible.

¹The contributions of Kerr nonlinearities to the power evolution (e.g. due to spectral broadening) are disregarded in the model as insignificant.

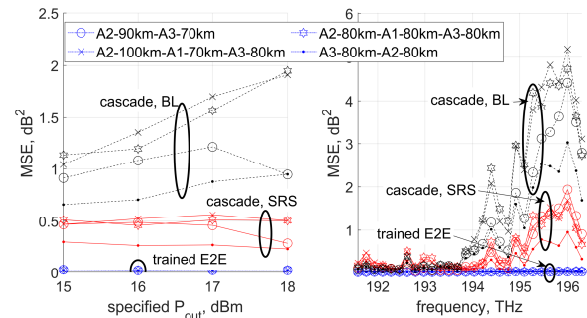


Fig. 9. MSEs for the considered multi-span systems. **a)**: MSE as a function of the launch power for the 3 different fiber models studied in this work; **b)**: the MSE per channel SRS frequency averaged over launch power for the different systems using the SRS-included model.

3) *End-to-end ML model*: Finally, an ML end to end (E2E) model is also studied, which is based on the input-output PSD for the entire system. In this case, a NN with identical configuration as in Fig. 4 is trained to predict the PSD output of the total system, similar to e.g. [10], [11]. This model is expected to provide an improved accuracy. It is however inflexible w.r.t. 1) number of spans; 2) span length; and 3) launch power (especially cases where the launch power is different in every span). In this case, similar to the EDFA training procedure, 75% of the profiles are used for training and 25% for testing.

B. System prediction error

Several random configuration of the available EDFAs and fiber spans are studied. The configurations are specified by the order of the element in the cascade, e.g. **"A2-90km-A3-70km"** corresponding to a link composed of (in order) amplifier A2, 90 km SSMF, amplifier A3, 70 km SSMF (true fiber lengths given above). For each configuration, the ≈ 2000 profiles already used for EDFA characterization are transmitted through the experimental system and their corresponding output PSDs recorded. In order to improve the accuracy of the cascade model, each EDFA component is supplied with the total input power readings obtained from the power meters of each corresponding physical EDFA instead of the *expected* input power, obtained as

$$P_{in}^{tot}(k) = P_{out}^{tot}(k-1) - \alpha \cdot L(k-1). \quad (6)$$

The expected input power does not account for the potential connector losses and fiber length mismatches as described above and degrades the accuracy of the EDFA model when used as a component in a system. These power readings are available for the system measurements data together with the EDFA measurements data in [14].

The MSE for the test profiles is reported in Fig. 9 for the configurations, given in the legend. In Fig. 9a), the MSE is given as a function of the launch power (equivalently, the P_{out}^{tot} of every EDFA) for the three studied models. We note that system output samples with a measurement of < -60 dBm fall below the noise floor of the adopted OSA. These are mostly relevant for input profiles of strong positive tilt

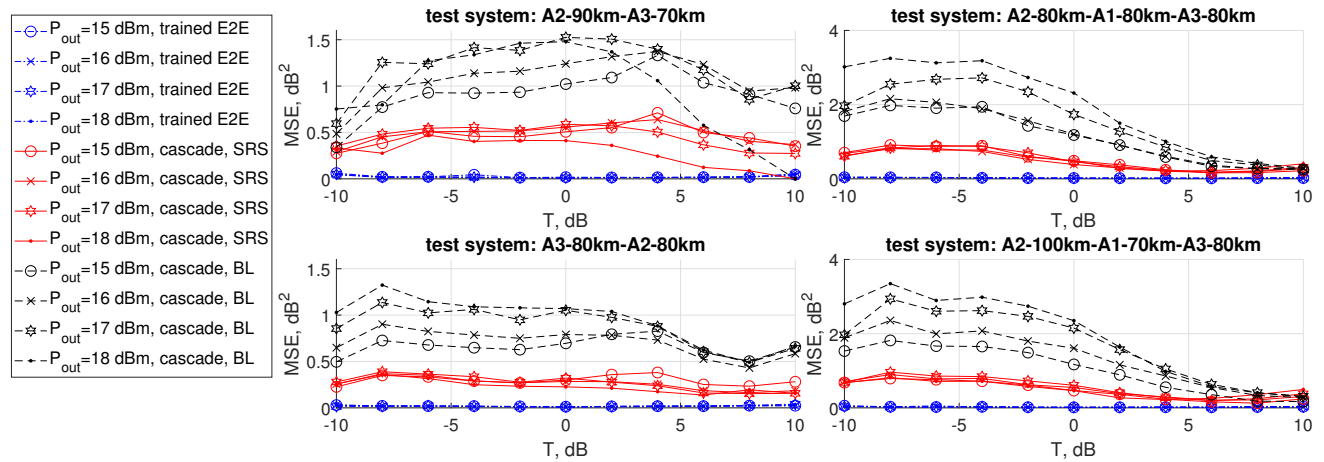


Fig. 10. MSE for the different systems as a function of the input profile tilt. For strong *positive* tilts and for lower launch power, the SRS is negligible and the bulk loss model becomes more accurate.

due to the accumulation of the EDFA gain tilt. In order to take advantage of as much of the data as possible, *individual carrier measurements* below -60 dBm (as opposed to *entire profiles* with individual points in the output PSD < -60 dBm) are excluded from the MSE estimation.

The E2E model provides an accurate modeling performance of $MSE \approx 0.02\text{dB}^2$, which is similar to the intra-MSE of the EDFA model. The SRS model provides an $MSE \approx 0.5\text{dB}^2$, which is relatively constant across the studied systems and launch powers. As mentioned above, this penalty in performance comes at the benefit of robustness, on-the-fly reconfigurability and generalization. The BL model results in significant performance penalty, especially at high powers (where the SRS is more significant) and more spans (where the SRS accumulates).

In Fig. 9b), the MSE is given as a function of the frequency for the SRS model and averaged over the launched power. Similar to the single EDFA case, the system MSE is dominated by the poor modeling performance at high frequencies, which is now worsened by the presence of multiple EDFAs in the system.

In Fig. 10, the MSE is given as a function of the input profile tilt (as defined in (2)) for the studied systems and models. For positive tilts, the BL model generally has an improved accuracy due to the fact that high-frequency channels are transmitted with low power, and the Raman gain is therefore negligible. For negative tilts, and especially with high total power and longer systems, the BL model suffers significant penalty. The SRS model exhibits relatively constant performance across tilts and total powers, slightly worse for long systems and negative tilts. As discussed in Section II-B, the EDFA model treats the EDFA as a black box with its boundaries extending from its input- to its output-power monitor. Therefore, connector and splice losses following the output-power monitor are not accounted for by the model and may result in slight variations in the expected launch power into the following fiber span. These variations, together with potential inaccuracies in estimating the fiber parameters, have an impact

in the accuracy of the Raman gain for the SRS model. Further characterization of localized losses and improved estimation of the fiber parameters could potentially improve the performance at the cost of additional measurements. This is left for future work to confirm.

C. System optimization

The models described above are all fully differentiable, which means that gradient descent can be applied in order to optimize the system w.r.t. e.g. the output PSD, as long as a differentiable cost function is applied. One example of such a cost function is the MSE between a target PSD and an output PSD. In this work, we optimize the system with a target of a flat PSD output. When the PSD is normalized as above, this translates to optimization with a cost function

$$Cost = -\min_n P_{out}^{system}(n). \quad (7)$$

Such an optimization is applied offline for the SRS and the BL models, in order to obtain system input PSDs that flattens the system output PSD. Details on the differentiation process are given in the Appendix. The learning rate for the gradient descent algorithm was optimized to 0.01. It is infeasible to measure to total input and output power of each EDFA at every iteration of the gradient descent algorithm. It is therefore not possible to supply these measurements to the model as done when estimating the model accuracy in Section III-B. For the system optimization, then, the model reverts to the expected total input and output power of each EDFA, as calculated through the fiber loss in Eq. (6)

These optimizations are compared to a naïve optimization strategy which assumes that the system is linear and independent of the input. Examples of optimized profiles are given in Fig. 11 for two different systems, together with their respective system output PSDs, at a launch power of 18 dBm. In Fig. 11a) and Fig. 11c), the optimized profiles (which are specified to the POF and desired at its output) are given, together with the true PSDs that are measured with an OSA at the input of the system. A non-insignificant discrepancy can be seen

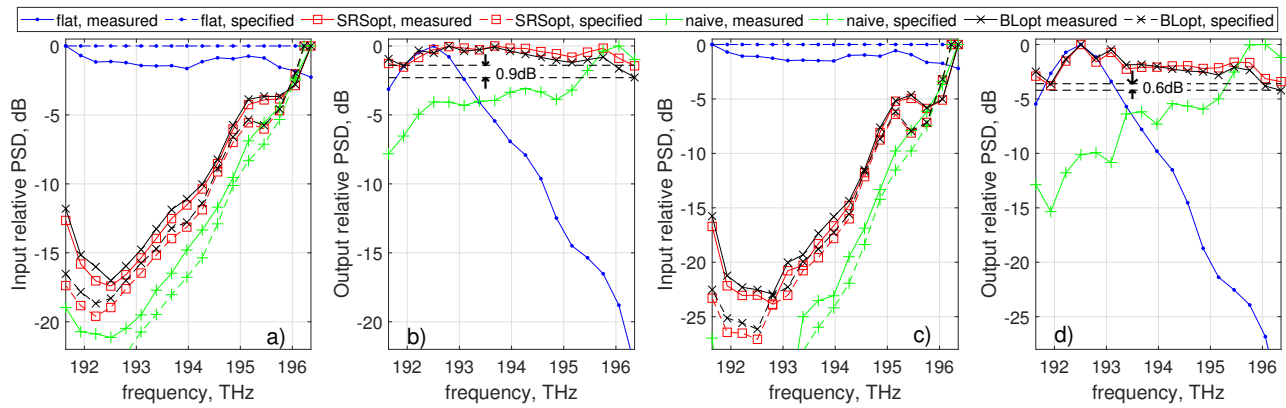


Fig. 11. Example input and outputs for two different systems. **a)**: input of **A2-90km-A3-70km**; **b)**: output of **A2-90km-A3-70km**; **c)**: input of **A2-100km-A3-80km-A1-70km**; **d)**: output of **A2-100km-A3-80km-A1-70km**; Due to finite precision in the POF, there is a mismatch between the true PSD launched into the fiber and the desired, specified PSD, obtained through the optimization. This mismatch is the main reason for the resulting non-flatness at the system output.

already at what is assumed a 'flat' input PSD, which in reality corresponds to a PSD with ≈ 2 dB of excursion. The sub-optimal flattening within the POF is due to limitation in the accuracy of the attenuation profile, as well as possible frequency calibration mismatches between the POF and the OSA. Iterative optimization may lead to improved flatness but that is beyond the scope of this work.

We found that a system optimization converges to < 0.1 dB of excursion for both the SRS and the BL models when evaluated in simulation, regardless of the non-convexity of the cost function in the input PSD, described in [8], [9]. The experimental system output PSDs are given in Fig. 11b) and Fig. 11d) for the systems **A2-90km-A3-70km** and **A2-100km-A3-80km-A1-70km**, respectively. Experimentally, ≈ 2 dB and ≈ 3 dB of excursion is achieved for a 2-span and a 3-span system, respectively when the system is optimized with the SRS model. This penalty is attributed to the above-mentioned non-flat reference. The excursion is increased by ≈ 0.9 dB and ≈ 0.6 dB for the two systems, respectively, if the BL model is applied. The additional excursion is due to the overestimated output power at high frequencies in the model, leading to lower power in the experiment at those frequencies for the same profile.

Finally, the naïve strategy overcompensates for the system tilt, and results in a tilt of the opposite direction due to the fact that the nonlinearities in the fiber and the EDFA are not correctly captured by this optimization.

IV. DISCUSSION

In this work, the POF was used to control both the profile and the total power input to the first EDFA in the cascade, which is challenging due to the limited attenuation range of the POF and potentially affects the accuracy of the applied profile, especially the ones of high excursion values. Adopting an explicit optical attenuator with a flat frequency response for the second task might alleviate some of the stress on the POF and improve the accuracy.

The models developed in this work require total input and output power readings from the EDFA devices, which might not be as accurate for commercially deployed EDFAs. We note

that in such cases, the transmission conditions are static and furthermore, the total fiber+connector losses of every span are characterized much more thoroughly and do not vary significantly over time. The expected total input and output power of each EDFA in the link is then close to reality and can be used to inform the cascade model without relying on readings from the devices themselves.

In this work, system lengths of up to three spans were considered due to the availability of three EDFAs of the same make. Increasing the system length is of obvious interest for exploring the generalization of the EDFA model and the overall accuracy of the cascade system model.

Optimizations for C+L band transmission are of even greater interest due to the more pronounced SRS effect, especially for hybrid EDFA+Raman amplifiers and their design thereof [16]. These applications are well within the capabilities of the cascade models presented in this paper, and are obvious directions for future research.

V. CONCLUSION

A neural network was proposed for modeling the power spectral density (PSD) response of an EDFA operating in C-band. The model was trained using experimental input-output PSD relations for varying total input power and total output power (equivalently, varying average gain). The model was demonstrated to generalize to several devices of the same make, enabling component-based modeling of multi-span systems. Cascade models are the key to the optimization of reconfigurable optical networks. System experiments demonstrated that such cascade system models can accurately predict the power evolution in fiber links provided that the stimulated Raman scattering (SRS) effect is captured by the model. A differentiable, SRS-included fiber model was combined with the differentiable EDFA model in order to optimize the power profile at the output of the experimental system e.g. for flatness. Maximum excursions of up to 2 dB and 3 dB for a 2-span and a 3-span gain-flattening-free system, respectively were demonstrated by the optimization, which overwhelmingly outperforms naïve optimization strategies which assume the system is linear and independent of the input.

FUNDING

This work was supported by the Villum Foundations (Villum Young Investigator VYI grant OPTIC-AI no.29344), the EU H2020 programme (Marie Skłodowska-Curie grant no. 754462) and the Danish National Research Foundation (DNRF) (Centre of Excellence Silicon Photonics for Optical Communications (SPOC), ref.DNRF123).

APPENDIX A

DERIVATIVE OF THE CASCADE SYSTEM MODEL

With the advent of graph-based numerical models for optimization, complicated functions can be differentiated and optimized efficiently by expressing them as a factor graph [15]. In brief, the derivative of the cost function is calculated at each node w.r.t its input and back-propagated to the input of the graph, which represents the optimization variable. The only restriction is that each node is differentiable w.r.t. its inputs. This is exemplified in Fig. 12, for which the derivative of the cost function w.r.t. e.g. x_1 may be calculated as

$$\frac{\partial Cost}{\partial x_1} = \frac{\partial Cost}{\partial f_4} \times \frac{\partial f_4(\frac{\partial f_1}{\partial x_1}, \frac{\partial f_3}{\partial f_2} \times \frac{\partial f_2}{\partial x_1})}{\partial(f_1, f_2)}. \quad (8)$$

The cascade model proposed in this paper contains 4 types of nodes: 1) linear; 2) ReLU; 3) normalization; and 4) SRS application. The linear and ReLU layers are trivial to differentiate. The normalization node is defined as $Norm(y_k) = x_k - \max_l(x_l)$, where x_k is the k -th input to the node and y_k is its output. The derivative is calculated as

$$\frac{\partial Norm(y_j)}{\partial x_k} = 1 \cdot (k == j) - 1 \cdot (x_k == \max_l(x_l)), \quad (9)$$

where '==' is the conditional equality operator. The derivative of the SRS layer defined in (5) can be calculated as

$$\frac{\partial P(n, z)}{\partial P(m, z - L_s)} = 1 \cdot (m == n) + \frac{gR}{A_{eff}} L_{eff} f(L_s) e^{P(m, z - L_s)}. \quad (10)$$

Finally, the derivative of the cost function defined in (7) w.r.t. its input layer is similar to above calculated as

$$\frac{\partial Cost}{\partial x_k} = -1 \cdot (x_k == \min_j(x_j)). \quad (11)$$

Once all derivatives of the cost function w.r.t. the power of every input carrier has been computed, standard gradient descent-type update rules can be applied. The process continues until convergence or until some pre-defined maximum number of iterations.

REFERENCES

- [1] A. Mahajan, K. Christodoulou, R. Martinez, S. Spadaro, and R. Munoz, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2616–2629, 2020.
- [2] 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 (OFC) 2020*, 2020, p. T4B.4.
- [3] M. Ionescu, A. Ghazisaeidi, J. Renaudier, P. Pecci, and O. Courtois, "Design optimisation of power-efficient submarine line through machine learning," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, 2020, p. STh4M.5.

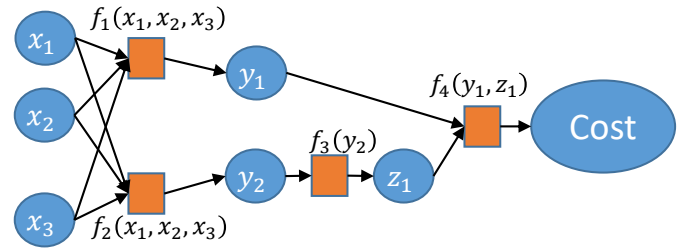


Fig. 12. Example of a graph used to differentiate a cost function w.r.t. the input variables (x_1, x_2, x_3), also known as leaf variables.

- [4] E. d. A. Barboza, A. A. B. Silva, J. C. Pinheiro Filho, M. J. Da Silva, C. J. A. Bastos Filho, and J. F. Martins Filho, "Optical amplifier response estimation considering non-flat input signals characterization based on artificial neural networks," *Journal of Lightwave Technology*, pp. 1–1, 2020.
- [5] 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.
- [6] J. Cho, G. Raybon, E. Burrows, J.-C. Antona, N. K. Fontaine, R. Ryf, H. Chen, S. Chandrasekhar, E. Sula, S. Olsson, S. Grubb, and P. Winzer, "Optimizing gain shaping filters with neural networks for maximum cable capacity under electrical power constraints," in *European Conference on Optical Communications (ECOC)*, 2020.
- [7] G. Borraicini, S. Straullu, A. Ferrari, E. Virgillito, S. Bottacchi, S. Swail, S. Piciaccia, G. Galimberti, G. Grammel, and V. Curri, "Using QoT-E for open line controlling and modulation format deployment: an experimental proof of concept," in *European Conference on Optical Communications (ECOC)*, 2020.
- [8] I. Roberts, J. M. Kahn, J. Harley, and D. W. Boertjes, "Channel power optimization of WDM systems following gaussian noise nonlinearity model in presence of stimulated raman scattering," *Journal of Lightwave Technology*, vol. 35, no. 23, pp. 5237–5249, 2017.
- [9] D. Semrau, E. Sillekens, P. Bayvel, and R. I. Killey, "Modeling and mitigation of fiber nonlinearity in wideband optical signal transmission," *J. Opt. Commun. Netw.*, vol. 12, no. 6, pp. C68–C76, Jun 2020.
- [10] J. Cho, S. Chandrasekhar, E. Sula, S. Olsson, E. Burrows, G. Raybon, R. Ryf, N. Fontaine, J. Antona, S. Grubb, P. Winzer, and A. Chraplyvy, "Maximizing fiber cable capacity under a supply power constraint using deep neural networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, p. W1K.2.
- [11] A. D'Amico, S. Straullu, A. Nespoli, 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.*, vol. 12, no. 6, pp. C1–C11, Jun 2020.
- [12] F. Da Ros, U. C. de Moura, and M. P. Yankov, "Machine learning-based EDFA gain model generalizable to multiple physical devices," in *European Conference on Optical Communications (ECOC)*, available <https://arxiv.org/abs/2009.05326>, 2020.
- [13] M. P. Yankov, U. C. de Moura, and F. Da Ros, "Power evolution prediction and optimization in a multi-span system based on component-wise system modeling," in *European Conference on Optical Communications (ECOC)*, available <https://arxiv.org/abs/2009.05348>, 2020.
- [14] M. P. Yankov and F. Da Ros, "Input-output power spectral densities for three C-band EDFAs and four multi-span inline EDFA fiber optic systems of different lengths," *Technical University of Denmark. Dataset* <https://doi.org/10.11583/DTU.13135754>, 2020.
- [15] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5595–5637, 2017.
- [16] M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in *2019 21st International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2019, pp. 1–4.