Enforcing Passivity of Macromodels via Spectral Perturbation of Hamiltonian Matrices

Generalization Properties of Machine Learning-based Raman Models

U. C. de Moura[1], D. Zibar[1], A. M. Rosa Brusin[2], A. Carena[2], and F. Da Ros[1]

1. DTU Fotonik, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
2. DET, Politecnico di Torino, Corso Duca degli Abruzzi, 24 - 10129, Torino, Italy
uiamo@fotonik.dtu.dk

Abstract: We investigate the generalization capabilities of neural network–based Raman amplifier models. The new proposed model architecture, including fiber parameters as inputs, can predict Raman gains of fiber types unseen during training, unlike previous fiber–specific models.

1. Introduction

Future autonomous optical networks will require reliable quality of transmission (QoT) and optical performance monitoring (OPM) estimators [1]. Optical amplifiers are the main source of noise in optical communications systems, with different gain and noise characteristics along the frequency. Therefore, QoT/OPM estimators should incorporate accurate and fast amplifier models. Highly–accurate machine learning (ML)–based optical amplifier models have been recently proposed in the literature for erbium-doped fiber amplifiers (EDFAs) [2, 3] and Raman amplifiers (RAs) [4, 5]. These models allows effective signal power optimization [6, 7] and QoT estimation [8].

Unlike EDFAs, RAs can provide gain in a variety of optical fibers and frequency bands, and are thereby highly relevant for the next generation of ultra-wideband optical communication systems. Their achievable gain is directly linked to the optical fiber [9], therefore the ML–based RA models proposed so far are specific for a given optical fiber type. This choice would require countless models, one for each considered fiber [10], and massive dataset to train them. Therefore, generalizable RA models that can predict RA performance for fiber types unseen during training are critically needed.

In this paper, first we confirm the poor generalization capabilities of the artificial neural network (NN)–based RA models presented in [10]: trying to predict the gain profile of a fiber different from the one used to train the model results in a poor prediction accuracy. We therefore propose a new NN model that include the fiber parameters as additional inputs and consider a training dataset with samples from multiple fiber types. The chosen fiber parameters define the RA gain and are easily-accessible from the fiber data–sheets, keeping the approach practical. Performance-wise, this refined model yields the same accuracy as a fiber-specific model, i.e., when the fiber–under–test (FUT) is part of the training. Furthermore, good generalization is achieved for a FUT unseen during training, as long as such FUT has parameters within the range of the fibers used to train the model.

2. Experimental setup and the general RA models

Fig.1(a) shows the experimental setup used to generate the datasets to train the NN models. An amplified spontaneous emission (ASE) source is used to generate the C–band signal. The RA consists of an optical FUT and a commercial Raman pump module combining four lasers with fixed wavelengths and adjustable powers (shown in Fig.1(a)). Pumps and signals are combined through a wavelength division multiplexer (WDM) in a counter-propagating scheme. At the RA output, an optical spectrum analyzer (OSA) captures the power spectra to calculate the on-off gain profiles.

A different dataset with 3000 examples is generated for each FUT case in Fig.1(b). Each example in the dataset consists of a pump power configuration, with values drawn from uniform distributions, and its corresponding measured RA on–off gain. The gain profile is down–sampled to 40 channels at 100-GHz (ITU-T grid). The datasets are divided into model selection/training (50%) and test (50%).

A total of 13 NN models are built: 6 *specific models*, independently trained for each FUT case, and 7 *general models*, trained with a combined dataset considering different fiber cases. Two general models are considered: [All], trained considering all fibers cases in Fig.1(b), and [All–one out], trained considering all fiber cases, except one. The NN architecture is illustrated in Fig.1(c). The specific models consider only P_{1-4} as the NN input. The general models, on the other hand, are applied also to learn how the fiber characteristics affect the RA gain profiles. Therefore, some relevant fiber parameters are considered as additional inputs. They are the fiber length (L), the signal (α_s) and pump (α_p) attenuation coefficients, and the effective area (A_{eff}) (values shown in Fig.1(b)). To

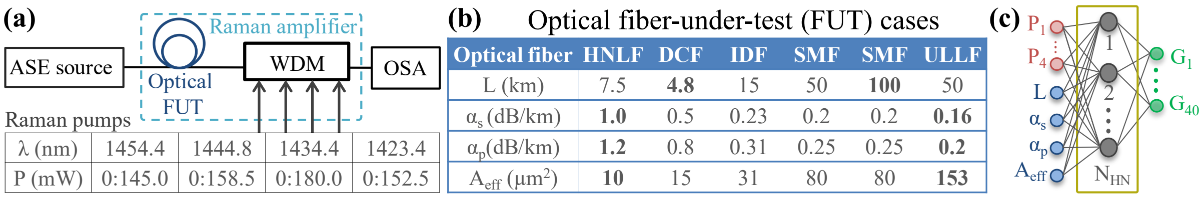

Fig. 1. (a) Raman amplifier experimental setup; (b) evaluated optical fiber cases and their corresponding parameters in terms of length (L), signal (α_s) and pump (α_p) attenuation coefficients, and effective area (A_{eff}) (minimum/maximum values are bold); and (c) neural network architecture.

keep the approach practical, these parameters were chosen due to their importance on determining the gain profiles and also because they are commonly provided by the fiber's suppliers on data–sheets. The Raman gain coefficient requires additional and more complex characterization, and it is, therefore, not considered in this work.

Model selection and training are performed using 10-fold cross validation and extreme learning machine [11], respectively, following the same procedure described in [12]. The parameters of the optimized models are summarized in Table 1, where f_{act} is the activation function, N_{HN} is the number of hidden nodes, λ is the regularization parameter to calculate the last layer's weights and σ_{NN} is the standard deviation of the normal distribution used to assign the hidden layers weights. Model averaging using 20 parallel NNs is used to reduce the impact of the randomly initialized weights.

3. Results and discussions

The statistics of the gain prediction performance are summarized in Fig.2. The figure of merit is the maximum error between target and prediction gains over the signal channels, referred to as E_{MAX}. The plots in Fig.2 show mean (dots) and standard deviation (bars) E_{MAX}. Each FUT case in Fig.2(a) to (f) has four plots. Plots 1-2 are the specific models trained on a given FUT case (marked (a) to (f)). They are tested over the corresponding FUT case used for training [Self] (plots 1) and over all the other fibers [Others] (plots 2). Plots 3 are for the general model [All] jointly trained over all fibers and tested over each FUT case. Plots 4 are for the general model [All–one out] trained over all fiber cases, except one, and tested over the corresponding removed fiber case, i.e. a fiber type unseen by the model during training.

As observed in previous work [10], the specific models are highly accurate when tested over the same FUT they were trained on (plots 1). In such cases, mean E_{MAX} values are $\lessapprox$0.2 dB. However, when these specific models are used to predict the gain of a different FUT (Others, plots 2), the performance can be highly degraded, with mean E_{MAX} >2 dB. This is expected as the Raman gain shape and level are highly dependent on the properties of the optical fiber [9]. A model without knowledge of the fiber characteristic cannot generalize to a different fiber type.

By considering the fiber parameters as input, the general model [All] (plots 3) has performance similar to the 0.2 dB of [Self]. These results show that the impact of the fiber properties on the Raman gain can be learned to provide highly accurate predictions for different fibers types. This general model, however, requires massive measurements to gather the datasets. The ability of a model to generalize, i.e. predicting the gain profile for a unseen FUT, is investigated considering the general models jointly trained over 5 fiber types [All–one out]. The sixth fiber type, i.e. the "one out", is used only for testing and is the corresponding FUT in Fig.2(a-f).

The results for [All–one out] show a degradation in performance with the worst generalization for the fiber cases corresponding to maximum or minimum values of one of their fiber parameters (Fig.1(b)). [All–HNLF7.5], for e.g., has the highest $\alpha_{s/p}$ and the lowest A_{eff} values (Fig.1(b)). Thus, when predicting [HNLF7.5] gain profiles using the [All–HNLF7.5] model, the inputs are outside the training range, and NNs are known for their poor ability to extrapolate information. In this condition, the NN generalization performance is highly dependent on the

Table 1. **Neural network parameters.**

	Specific models						General models	
FUT →	[HNLF7.5]	[DCF4.8]	[IDF15]	[SMF50]	[SMF100]	[ULLF50]	[All]	[All–one out]
f_{act}	sine	tanh	logsig	sine	tanh	tanh	sine	sine
N_{HN}	700	100	400	50	100	60	900	*
σ_{NN}	1e-3	1e-3.5	1e-2.5	1e-3	1e-3.5	1e-4	1e-2.5	1e-2.5
λ	1e-7.5	1e-5.5	1e-2	1e-8	1e-6	1e-8.5	1e-8	1e-8

* 500 (one out = [HNLF7.5]), 400 ([DCF4.8]), 700 ([IDF15] and [ULLF50]), 600 ([SMF50]), 800 ([SMF100]).

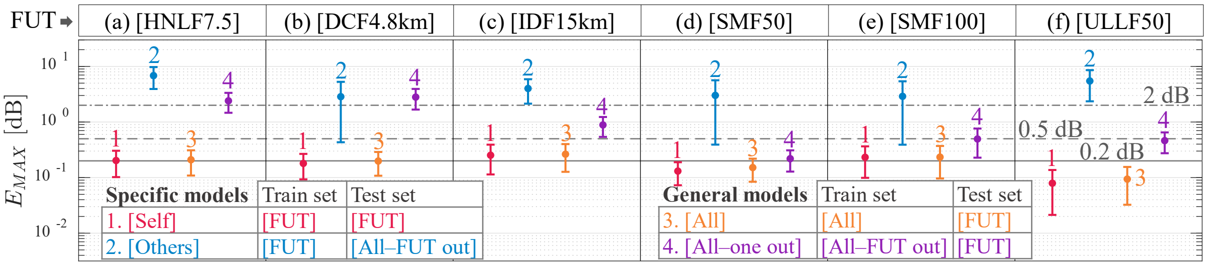

Fig. 2. Maximum errors (E_{MAX}) bar plots with mean (circle) and standard deviation (bars) values for all models tested over different fiber cases (from (a) [HNLF7.5] to (f) [ULLF50]).

FUT, ranging from accurate (as for [All–SMF100] and [All–ULLF50], with mean $E_{MAX} \approx 0.5$ dB) to poor (as for [All–DCF4.8] and [All–HNLF7.5], with mean $E_{MAX} \approx 2$ dB).

Although [IDF15km] parameters have no minimum/maximum values in Fig.1(b), [All–IDF15km] has also a relatively poor generalization performance with mean $E_{MAX} \approx 1$ dB. We believe that this is because the [All–IDF15km] model is trained not considering values close to [IDF15km]'s fiber L and A_{eff} (see Fig.1(b)). [All–SMF50], on the other hand, share all its parameters with other fiber cases ([SMF100] and [ULL50]]), which leads to the best generalization performance for the general model, with mean E_{MAX} of 0.22 dB, i.e. close to the [Self] model performance. This highly accurate result for a totally new FUT show that a good generalization can be achieved as long as the new FUT is inside the model training range and have sufficient examples close to it.

4. Conclusions

In this paper, we proposed a refined NN-based Raman amplifier model considering the optical fiber parameters as additional inputs, and using data from several fiber types during training. The model was able to learn the complex mapping between the Raman pumps and fiber parameters to the Raman gain profile, yielding similar performance to fiber–specific models when the fiber–under–test is part of the training dataset. More importantly, such a model showed good accuracy (e.g. mean(E_{max}) $\approx$ 0.22 dB) as long as the unseen fiber–under–test is inside the model training range of fiber parameters and has sufficient examples around it. These results confirm that the generalization capabilities of the proposed model can reduce the number of models required and the size of the training dataset compared to the fiber-specific models.

Acknowledgements
We thank OFS Fitel Denmark for providing the fibers used in this work. This project has received funding from the European Research Council through the ERC-CoG FRECOM project (grant agreement no. 771878), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462 and the Villum Foundations (VYI OPTIC-AI grant no. 29344).

References

1. K. Christodoulopoulos et al., "Toward efficient, reliable, and autonomous optical networks: the ORCHESTRA solution [Invited]," J. Opt. Commun. Netw., vol. 11, no. 9, pp. C10-C24 (2019).
2. S. Zhu et al., "Hybrid Machine Learning EDFA Model," OFC, 2020 T4B.4.
3. F. Da Ros et al., "Machine learning-based EDFA Gain Model Generalizable to Multiple Physical Devices", ECOC 2020, Tu1A-4.
4. A. M. Rosa Brusin et al., "Introducing Load Aware Neural Networks for Accurate Predictions of Raman Amplifiers" J. Lightwave Technol., vol. 38, no. 23, pp. 6481-6491 (2020).
5. X. Ye, et al., "Experimental Prediction and Design of Ultra-Wideband Raman Amplifiers Using Neural Networks," OFC 2020, W1K.3.
6. M. Ionescu et al., "Design Optimisation of Power-Efficient Submarine Line through Machine Learning," CLEO 2020, STh4M.5.
7. M. Yankov et al., "Power Evolution Prediction and Optimization in a Multi-span System Based on Component-wise System Modeling", ECOC 2020, Mo2K-4.
8. A. Mahajan et al., "Modeling EDFA Gain Ripple and Filter Penalties With Machine Learning for Accurate QoT Estimation," J. Lightwave Technol., vol. 38, no. 9, pp. 2616-2629 (2020).
9. M. A. Iqbal, et al., "Linear and Nonlinear Noise Characterisation of Dual Stage Broadband Discrete Raman Amplifiers," J. Lightwave Technol., vol. 37, no. 14, pp. 3679-3688 (2019).
10. U. C. de Moura et al., "Experimental characterization of Raman amplifier optimization through inverse system design," J. Lightwave Technol., early-access (2020).
11. G.-B. Huang, et al., "Extreme learning machines: a survey," Int. J. Mach. Learn. Cyb., vol. 2, no. 2, pp. 107–122 (2011).
12. D. Zibar et al., "Inverse System Design Using Machine Learning: The Raman Amplifier Case," J. Lightwave Technol., vol. 38, no. 4, pp. 736–753 (2020).