OSNR prediction for optical links via learned noise figures

Sarah Kamel(1), Hartmut Hafermann(1), Dylan Le Gac(1), Ludovic Dos Santos(2), Balázs Kégl(2), Yann Frignac(1), Gabriel Charlet(1)

(1) Huawei Technologies France, Paris Research Center, Optical Communication Technology Lab, hartmut.hafermann@huawei.com
(2) Huawei Noah's Ark Lab, Paris, France

Abstract *We predict the per-channel OSNR of optical links with up to 23 EDFAs via a machine learning model based on learned noise figures from experimental data. For a 20 span link, the error margin to cover 99% of cases is less than 0.35 dB.*

Introduction

Optical signal-to-noise ratio is an important quality of transmission metric for the operation of optical networks. Accurate OSNR predictions facilitate the reduction of design margins and wasted capacity. The main determinant of OSNR is the amplified spontaneous emission (ASE) noise generated, for example, by Erbium-doped fiber amplifiers (EDFAs). In particular for a varying channel count, the output power and ASE noise of an EDFA are complicated nonlinear functions of its inputs. The lack of accurate amplifier models has motivated neural network (NN) modeling[1].

There are various possibilities for implementing OSNR prediction using machine learning. The link OSNR has been predicted directly via a NN that takes power, gain and noise figures of all N_{EDFA} individual EDFAs[2] as input. The number of inputs hence scales linearly with the link length, while the number of trainable parameters scales even with the square of the number of inputs, which becomes prohibitive for long links. In addition, the hyperparameters of the NN have to be optimized for each link length, which is an issue when training occurs under time constraints.

One can also build models that predict both power and ASE of individual EDFAs[3] or the entire link[4],[5] and compute the OSNR from these predictions. Cascading individual NN models in general leads to severe error amplification which rapidly degrades prediction accuracy for longer links. Errors in the ASE predictions are amplified as ASE enters the OSNR in the denominator.

On the other hand, assuming amplifier noise figures are known, the OSNR contribution of each amplifier is a function of its input power. Denoting by $\text{OSNR}_{c,n}$ the OSNR of channel c due to the noise contribution from amplifier n, the link OSNR can be computed as follows[6]:

$$\text{OSNR}_c = \left(\sum_{n=1}^{N_{\text{EDFA}}} \frac{1}{\text{OSNR}_{c,n}} \right)^{-1}, \tag{1}$$

$$\text{OSNR}_{c,n} = \frac{P_{c,n}^{\text{in}}}{h\nu_c B_{\text{ref}} \text{NF}_{c,n}[P_{c,n}^{\text{in}}]}. \tag{2}$$

$\text{NF}[P_{c,n}^{\text{in}}]$ indicates the dependence of noise figures on the channel loading configuration[3]. B_{ref} is the OSNR reference bandwidth (12.5GHz at 1550nm), h is Planck's constant and ν_c is the channel center frequency.

We propose two NN architectures for OSNR prediction based on above equations. Specifically, the noise figures are replaced by trainable parameters or NNs and learned from the data.

Experiment

Fig. 1 shows a schematic of the optical link, which consists of four sections with five spans each. The first three sections consist of six commercial EDFAs, while the last one has five. The spans in the first and last section are made of 80 km standard single mode fiber and in the second and third of 100 km pure silica core fiber. A Wavelength Division Multiplexing (WDM) comb of $N_{\text{c}} = 40$ channels of 50 GHz width on a 150 GHz grid in the C band (1525 – 1572 nm) is generated by spectrally shaping ASE noise of an EDFA via two wavelength-selective switches (WSS) to obtain a good extinction ratio and OSNR at the link entrance. We operate the EDFAs at fixed gain and tilt while randomly changing the number and position of active channels by means of the WSS. For the acquisition of training data, an optical spectrum analyzer (OSA) records the power spectrum before and after every EDFA via an optical switch connected to the indicated locations (dashes lines). Its output is integrated over chan-

978-1-6654-3868-1/21/$31.00 ©2021 IEEE

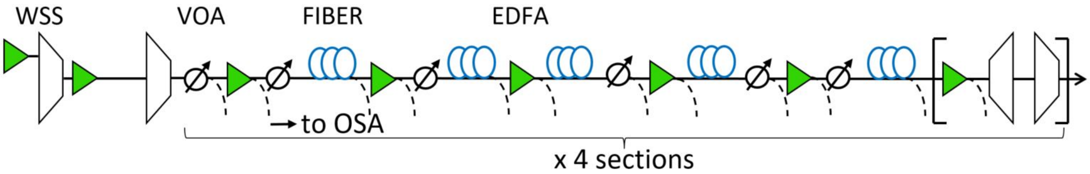

Fig. 1: Optical link made up of 4 sections comprising 20 spans and 23 EDFAs. Dashed lines indicate locations of power measurement (prediction) during training (operation). Square brackets indicate that the last section misses one amplifier.

nel bins during post-processing. The ASE noise per channel is obtained by interpolating the noise level in the empty 50 GHz slots adjacent to it. OSNR is recorded at the end of each link section.

Machine learning models and results

All machine learning models are trained using mean square error (MSE) loss suitable for regression. We however evaluate our models in terms of an *error margin*, which is more pertinent to system design. We define EMX as the X-percentile of absolute values of the ratios between predicted and true values in dB, $P_X(\{|10\log_{10}(Y_c^{\mathsf{pred},i}/Y_c^{\mathsf{true},i})|,\ i = 1,\dots,N_\mathsf{s}; c = 1,\dots,N_\mathsf{ch}\})$, where N_s and N_c are the number of data samples and channels, respectively. It can be interpreted as the margin required to cover the deviation of X percent of all predictions from the true value. We henceforth use EM99 and EM100 or MEM (maximum error margin).

We assume that during link operation, channel power can be measured at the input of each section, but not at the input of each EDFA. Hence to obtain the channel power $P_{c,n}^{\mathsf{in}}$ at each EDFA input which enters Eq. (2), one has to rely on predictions. For this purpose we train a separate model for each link section to predict $P_{c,n}^{\mathsf{in}}$ at each EDFA in the respective section taking only the section input power as input. We quantify the validation accuracy in terms of EM99 with 95% confidence intervals obtained from 5-fold cross validation. For the four link sections, we obtain 0.67 $\pm$ 0.02, 0.84 $\pm$ 0.03, 1.03 $\pm$ 0.14 and 1.53 $\pm$ 0.55 dB, respectively.

We now focus on models for OSNR prediction. Model 1 is obtained by combining equations (1) and (2), while placing learnable parameters *in lieu* of the noise figures, i.e. $\mathsf{NF}_c = \theta_c^{\mathsf{NF}}$. Here a possible dependence of the noise figure on the channel loading configuration is neglected. The noise figures are learned in complete analogy to training a neural network via mini-batch stochastic gradient descent with batchsize N_b using the loss function

$$\mathcal{L}[\theta] = \frac{1}{N_\mathsf{b}}\frac{1}{N_\mathsf{c}}\sum_{s=1}^{N_\mathsf{b}}\sum_{c=1}^{N_\mathsf{c}}(\mathsf{OSNR}_{c,s}^{\mathsf{pred}} - \mathsf{OSNR}_{c,s}^{\mathsf{true}})^2.$$

In model 2, we wish to account for the dependence of the noise figures on the channel loading configuration. While this can be accomplished in different ways, we found that best results are obtained by rewriting the noise figures as $\mathsf{NF}_{c,n}[P_{c,n}^{\mathsf{in}}] = \mathsf{NF}_{c,n} + \Delta\mathsf{NF}_{c,n}[P_{c,n}^{\mathsf{in}}]$ in Eq. (2) and modeling $\Delta\mathsf{NF}_{c,n}[P_{c,n}^{\mathsf{in}}]$ through a NN (one for each EDFA) taking $P_{c,n}^{\mathsf{in}}$ as input. Here $\mathsf{NF}_{c,n} = \theta_{c,n}^{\mathsf{NF}}$ are also learned. Another possibility would be to replace the entire right-hand side in Eq. (2) by a NN that models per-channel OSNR per amplifier instead, but we found this to give worse results. Note that for both models, backpropagation is performed through the loss *and* Eqs. (1) and (2) (through active channels only). To show the advantage of exploiting the above equations we compare results to those of a reference, which directly models the per channel OSNR as a function of the power predictions: $\mathsf{OSNR}_c = \mathsf{NN}_\theta[P_{c,n}^{\mathsf{in}}]$. It takes the input powers $P_{c,n}^{\mathsf{in}}$ of all EDFAs and all channels as input. Note that we do not use ASE noise as input, so that the reference uses the same information as the other models.

We report results for links with the number of spans ranging from 5 to 20, i.e., with up to 23 EDFAs. Each training sample consists of the input power per channel for each section and the OSNR at the end of the respective link. Except for the first EDFA in each section, $P_{c,n}^{\mathsf{in}}$ is a prediction of the respective power prediction model. The training data consists of $N_\mathsf{s} = 774$ samples whereof 80% are used for training and 20% for testing. Unless otherwise stated we use 100 epochs with $N_\mathsf{b} = 128$ and the Adam optimizer at a learning rate of 0.01. We report validation results with their 95% confidence intervals obtained from 10-fold cross validation.

For model 1, the noise figure parameters can converge to unphysical values, leading to strong variations across channels. This behavior is effectively suppressed by initializing all to the same typical value of 5.5 dB (in linear units). For 100 epochs, the values typically remain close to the initial value. For a large number of epochs (several hundred), unphysical values may be observed as a sign of overfitting. While unneces-

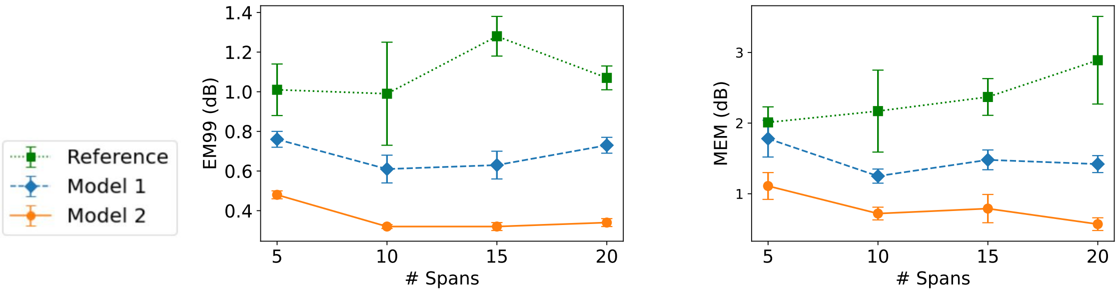

Fig. 2: Error margins of the three OSNR prediction models required to cover 99% (EM99) and all predictions (MEM).

sary here, such tendency could be suppressed in general by adding a penalty term to the loss. Model 2 was trained for four different NN architectures with ReLU activations given by the combinations of 1 and 2 hidden layers (HL) and hidden sizes (HS) 64 and 128. The results for all cases are similar, allowing to choose one architecture *independently* of the number of sections. Here we report results for HL=2 and HS=128. For the reference model the number of inputs $N_{\text{inp}} = N_{\text{c}} \times N_{\text{EDFA}}$ changes with link length, so that the hyperparameters have to be optimized for each case. We restrict ourselves to a maximum of 2 HL and simulate four combinations of HS: $(N_{\text{inp}}, N_{\text{inp}}/2)$, $(N_{\text{inp}}, 2N_{\text{c}})$, $(N_{\text{inp}}/2, N_{\text{inp}}/2)$, $(N_{\text{inp}}/2, 2N_{\text{c}})$. The optimal combination varies with link length. The reference model typically does not converge within 100 epochs. Since we do not observe overfitting, we train for 500 epochs to obtain better accuracy.

Fig. 2 summarizes the main results. Model 2 improves the margin of model 1 by at least 0.25 dB in terms of EM99 and 0.5 dB in terms of MEM, which indicates that the dependence of noise figures on the loading configuration is significant. The reference model is consistently outperformed by both models. Model 2 reduces the MEM of the reference by 1 to 2 dB. Note also the small confidence intervals, indicative of a consistent prediction accuracy across different training runs.

Fig. 3 shows the distribution of the ratio

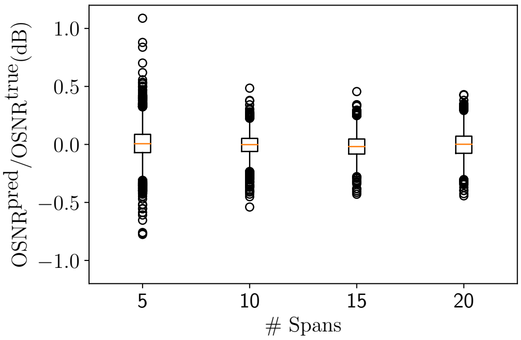

Fig. 3: Distribution of the ratio of predicted to true OSNR values for model 2 and different link lengths.

$\text{OSNR}^{\text{pred}}/\text{OSNR}^{\text{true}}$ for the best model 2. One can see that the margin required to cover a fixed percentage of the distribution is essentially independent of the link length. This is in contrast to cascaded predictions prone to error amplification and is due to joint learning on two levels: i) the power predictions are jointly learned per section and ii) all noise figures for all amplifiers are also learned jointly. Consistent with Fig. 2, the distribution of the ratio is larger for the shortest 5-span link, which might be due to the fact that in this case all input power predictions stem from the same underlying model and are correlated, while power predictions from different link sections are independent. Note that such correlations can be learned by the underlying OSNR prediction model, so that the learned noise figures should be interpreted as effective parameters rather than physical values.

Conclusion

We introduced neural architectures to predict OSNR from uncertain power predictions, based on the analytical relation between OSNR and learned amplifier noise figures. The models use the per-channel power per EDFA but not ASE noise power as input. The accuracy is essentially independent of the link length in contrast to approaches which cascade predictions. Contrary to a black box approach, parametrizing noise figures through NNs significantly improves accuracy and reduces the number of trainable parameters, training time and maximum errors. The NN architecture can be chosen independent of link length. For the best model 2 we further achieve a margin reduction between 1 and 2 dB compared to the reference and a reduction in training time by a factor of 20. For a link of 20 spans a margin of only $<$0.35 dB is required to cover the OSNR prediction uncertainty in 99% of cases. The comparison to the reference model illustrates that machine learning models can greatly benefit from incorporating available physical knowledge.

References

[1] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks", *IEEE/OSA Journal of Optical Communications and Networking*, vol. 13, no. 4, B60–B71, 2021. DOI: 10.1364/JOCN.417434.

[2] S. Yan, F. N. Khan, A. Mavromatis, D. Gkounis, Q. Fan, F. Ntavou, K. Nikolovgenis, F. Meng, E. H. Salas, C. Guo, C. Lu, A. P. T. Lau, R. Nejabati, and D. Simeonidou, "Field trial of machine-learning-assisted and SDN-based optical network planning with network-scale monitoring database", in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1–3. DOI: 10.1109/ECOC.2017.8346091.

[3] Y. You, Z. Jiang, and C. Janz, "OSNR prediction using machine learning-based EDFA models", in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–3. DOI: 10.1049/cp.2019.1044.

[4] A. D'Amico, S. Straullu, A. Nespola, I. Khan, S. Abdelfattah, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, S. Abrate, and V. Curri, "Machine-learning aided OSNR prediction in optical line systems", in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–4. DOI: 10.1049/cp.2019.0758.

[5] J. Cho, S. Chandrasekhar, E. Sula, S. Olsson, E. Burrows, G. Raybon, R. Ryf, N. Fontaine, J.-C. 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, pp. 1–3.

[6] P. Poggiolini, G. Bosco, A. Carena, R. Cigliutti, V. Curri, F. Forghieri, R. Pastorelli, and S. Piciaccia, "The LOGON strategy for low-complexity control plane implementation in new-generation flexible networks", in *2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2013, pp. 1–3. DOI: 10.1364/OFC.2013.OW1H.3.