Machine learning enhancement of a digital twin for WDM network performance prediction leveraging Quality of Transmission parameter refinement

NATHALIE MORETTE[1], HARTMUT HAFERMANN[1,*], YANN FRIGNAC[1], AND YVAN POINTURIER[1]

[1] Optical Communication Technology Lab, Paris Research Center, Huawei Technologies France
[*] Corresponding author: hartmut.hafermann@huawei.com

Compiled April 14, 2023

Digital twins capable of Quality of Transmission (QoT) estimation and prediction are powerful tools for efficient design and operation of optical networks. In this paper, we employ machine-learning techniques to enhance both the QoT estimation and prediction capabilities of an optical network digital twin. By leveraging a method to infer or refine the unknown lumped loss distributions and amplifier gain spectra for accurate characterization of the current optical network state, accuracy of estimation and prediction are substantially improved. For QoT prediction, we further develop a novel neural-network architecture for EDFAs, which after factory training on a single device and with fully loaded configurations only, generalizes to partially loaded configurations seen after deployment and to different gain and tilt settings and other physical devices of the same type. We combine refined parameters and a novel method that predicts the *difference* in per-channel power when individual services are added or removed or multiple services are lost due to a fiber cut. The impact of error amplification due to cascading of individual components' models are shown to be reduced, yielding predictions that are substantially more accurate than simply cascaded predictions. As the network ages, the digital twin can be updated by retraining, while using only information available in the current network state.

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

1. INTRODUCTION

Digital twins hold the promise of enabling efficient design and operation of complex optical networks. They may be used as tools for offline system simulation and aid the design of low-margin optical networks [1]. They can also serve as a proxy for online network (re-)optimization, resource (re-)allocation, service upgrade and quick failure resolution. The development of digital twins is driven by advances in machine learning (ML) techniques which facilitate modeling and decision taking based on large amounts of data [2]. In addition, ML enables the description of network components which are inherently difficult to model using traditional approaches based on device physics. A prominent example are Erbium-doped fiber amplifiers (EDFAs), for which several neural network (NN)-based models have been developed [3–7].

Accurate Quality of Transmission (QoT) estimation and prediction is essential to the design and operation of coherent optical fiber networks. Digital twins should therefore support these two basic functions. The main source of error in QoT estimation stems from missing or inaccurate knowledge of QoT input parameters. Recently, ML-based methods to refine QoT inputs have been developed [4, 8–10]. ML-based approaches to predict and optimize power evolution and signal-to-noise ratio (SNR) [11–16] or QoT [17] have been developed independently.

The design of digital twins often follows a modular approach. Models for the various components (amplifiers, fibers, etc.) and their predictions are cascaded according to link composition. This modularity provides great flexibility. ML models for the amplifiers can be trained as part of the manufacturing process. When possible, link components can be modeled using simpler physics-based models which are described by a handful parameters and are typically easier to train. An example is the power evolution in fiber governed by a differential equation to account for stimulated Raman scattering (SRS) [18]. Lumped losses can be taken into account explicitly, since the power evolution is computed at any point between modules in the model cascade. As a result, constituent models can simply be replaced and parameters can be adjusted when network properties change during the network lifetime due to, e.g., fiber repairs or device replacements. The model can hence be kept up to date without the necessity of retraining. An issue with this approach, however, is error amplification of the cascaded predictions. The error is amplified due to inaccurate inputs from previous predictions in

the cascade. This problem has not received much attention in the literature so far, mainly due to the experimental limitation to short links. Significant error accumulation over 10 EDFAs was reported [19], but not specifically addressed. In most other studies, the number of cascaded ML models of individual amplifiers [12–15] or entire optical multiplexing sections (OMSs) [16] was limited to two or three. While visible, error amplification was not too severe. The error typically increases exponentially, imposing stringent requirements on the prediction accuracy of individual models in the cascade. This may be difficult to attain in practice, because typically several models are cascaded and accuracy limitations stem from the inherent and unavoidable prediction errors of a trained (ML) model, or differences in the training and test data distributions.

An alternative is a single monolithic ML model (a deep neural network or more advanced architectures), which describes (part of) an optical network or link end-to-end and predicts QoT from its input. Hence error amplification is not an issue. Since the model has to generalize to new settings as network properties such as amplifier gain and tilt settings change during network operation, however, the configuration space (the possible combinations of all settings and inputs) is exponentially large. The amount of data required typically increases exponentially with the number of devices in the link. In addition, the model has to be retrained as the network ages and changes its properties. This is challenging as training data is very limited during network operation. As the model treats the whole link as a black box it also lacks explainability. Due to the size and complexity of optical networks, practical approaches would likely still require cascading models for subunits of the network, as in [16].

In this paper, we follow the modular approach and build a machine-learning enhanced digital twin of an optical link with up to 15 cascaded fiber spans and 18 amplifiers. We leverage a method to accurately characterize the current network state in terms of lumped losses and amplifier gain spectra, which we call QoT parameter refinement, or input refinement (IR) for short [10]. IR can accurately estimate power profiles and improve SNR estimates independent of spectral occupation [20].

For QoT prediction, we replace estimated static gain spectra of the EDFAs by NN-based ML models (EDFA-ML) to account for their load dependence [20]. The models are trained simulating offline factory training and subsequent deployment in the field. They generalize over arbitrary nominal amplifier gain and tilt settings. We propose a novel NN architecture which outperforms generic NNs, converges faster, and better generalizes to the data distribution in the field.

To tackle error amplification, we propose a new method that predicts *changes* in the per-channel power relative to the current state when the network undergoes a transition to a new state by addition or removal of multiple services. Through the novel combination of ML-based EDFA models with IR (EDFA-ML + IR) we demonstrate more accurate SNR and power evolution predictions, resulting in significant accuracy gains for QoT prediction.

The paper is organized as follows. In Sec. 2, we recall the QoT parameter refinement approach from [10]. Sec. 3 describes the new NN architecture of the constituent ML models for the EDFAs and their performance evaluation. Sec. 4 introduces our novel framework for making ML-based predictions after service addition or removal, which leverages the QoT parameter refinement. In Sec. 5 we evaluate the new method for QoT estimation and prediction in various scenarios. We show that it provides an accurate and robust static characterization even for partially loaded networks. We conclude in Sec. 6.

This work is an extension of [10, 20], of which part of the results are reproduced in Sec. 5B with additional details. Sec. 3 reports unpublished details on the EDFA ML models. The combination of EDFA-ML with IR in Secs. 4 and 5C is entirely new.

2. QOT PARAMETER REFINEMENT

For our digital twin to serve as proxy for QoT estimation, we require a quantitative, static characterization of the current network state. Among the necessary QoT parameters, the ones that we refine are in particular the amplifier gain spectra and lumped (connector) losses. Due to cost considerations, these parameters are not routinely monitored in deployed networks. Furthermore, uncertainties in the parameters inherent to real-life conditions increase over the network life, leading to increased QoT margins. This raises the problem of obtaining a complete characterization of the network state from incomplete or inaccurate measurement data [8, 10]. IR is a method to reconstruct or refine missing or uncertain QoT parameters, which is briefly described here.

For concreteness, we consider the generic OMS depicted in Fig. 1 as the building block connecting the network nodes of an optical network. Wavelength selective switches (WSSs) at the OMS entrance and end allow to add or remove services. We only assume availability of monitoring data that is realistically monitored in deployed networks. In particular, we assume that optical power monitors (OPMs) are available to provide power spectral information at the output of the first and last EDFA only. Amplifier nominal gain and tilt settings and noise figures are assumed to be known.[1] We further assume the *total* input and output power of each in-line amplifier to be monitored. Hence total span loss, including lumped and fiber losses, can be deduced. Fiber loss and Raman gain spectra can be inferred from datasheets assuming knowledge of the fiber length and fiber type from the network specifications. The difference between total span losses and estimated fiber loss are ascribed to total lumped losses $\delta_i + \delta_i'$, while their distribution between span input and output (the individual values of δ_i and δ_i') is in general unknown. Uncertainties in the fiber attenuation coefficient and Raman gain are hence absorbed by the lumped losses.

The method is based on a digital twin of an OMS (cf. Fig. 1), which is an ML model in which the EDFA gain spectra $G_i(\lambda)$ and the input and output lumped losses δ_i and δ_i' of each span i serve as parameters (as opposed to a generic NN architecture).

In the IR process described in Ref. [10], refined estimates of these QoT parameters are obtained solely based on the knowledge of the per-channel output power at OPM level and the per-channel SNR, which serve as the ground truth. This is made

[1] Here we assume noise figures to be constant, but more elaborate models that relate amplifier gain spectra to per-channel noise figures can be easily incorporated.

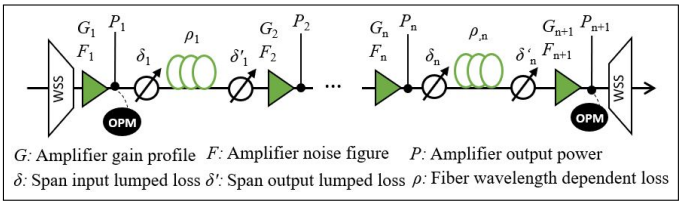

Fig. 1. Schematic of an optical multiplexing section with wavelength-selective switches (WSS) and an optical power monitor (OPM) that measures per-channel power.

possible by taking the link physics into account. In a forward pass, propagation of the power spectrum through the fiber accounts for SRS according to the model [18]:

$$\frac{\partial}{\partial z}P(z,\lambda_j) = -\sum_{k=j+1}^{N}\frac{f_j}{f_k}g_R(|f_j-f_k|)P(z,\lambda_k)P(z,\lambda_j) + \sum_{k=1}^{j-1}g_R(|f_j-f_k|)P(z,\lambda_k)P(z,\lambda_j) - \rho(f_j)P(z,\lambda_j). \quad (1)$$

Here f_j and λ_j are the frequency and wavelength of channel j, respectively, and according with the literature frequencies are ordered such that $f_j < f_k$ for $j > k$. $\rho(f_j)$ denotes the wavelength-dependent fiber attenuation and $g_R(|f_k - f_j|)$ is the normalized and polarization-averaged Raman gain spectrum. The effect of filtering is left out of this study but can be straightforwardly addressed as in [21]. SNR evolution is obtained based on the amplifier noise figures $F_i(\lambda)$ and the Kerr effect through nonlinear interference estimates based on the EGN model [22].

By performing forward passes through the OMS with current values of the lumped losses and amplifier gain spectra, predictions of the per-channel output power and SNR of the OMS are obtained. The QoT parameters are iteratively refined through suitable gradient descent updates, with the goal of minimizing the prediction error relative to the ground truth. The method estimates the per-channel power evolution through the OMS as a byproduct. It is important to note that even though in a forward pass in IR models are cascaded, the boundary condition provided by the OPM at the OMS end allows to obtain accurate gain spectra and per-channel power evolution anywhere inside the OMS. IR allows to estimate the QoT parameters without cross-compensation between effects, such as the refined amplifier gain absorbing an inaccurate Raman tilt. Each OMS is optimized separately. More details can be found in [10]. We note that our conclusions on the EDFA-ML+IR method are independent of the details of how IR is performed. Any reasonably accurate method to refine the lumped losses and amplifier gain spectra will work.

We stress that the method is extremely data efficient: only information in the current network configuration, in particular the current per-channel SNRs and OPM level power spectra are used during learning. This is remarkable considering the number of unknowns: For an OMS with N_{sp} spans and N_s services, we have N_{sp} EDFAs and hence $N_{sp} \times N_s$ unknown gain values and $2N_{sp}$ unknown lumped losses, while there are $2N_s$ per-channel output powers and SNRs and $2N_{sp}$ total span input and output powers. Hence for more than two spans, the problem is in general underdetermined. However it empirically converges to QoT parameters which are close to the ground truth as will be shown below.

As a consequence of the undetermined character of the problem, it is not possible to replace individual amplifiers by more elaborate neural network models for QoT *prediction*, unless these are trained using additional data sources. This is discussed in the next section.

3. NEURAL NETWORK EDFA MODELS

For accurate QoT prediction it is essential to account for the load dependence of the EDFA gain ripples. The difficulties of formulating accurate analytical models has motivated NN-based modeling [3, 5, 7].

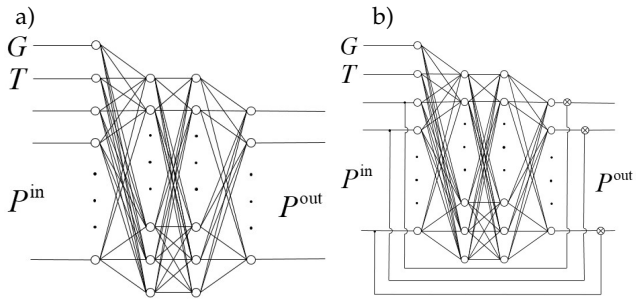

Fig. 2. Neural network architectures used for the EDFA models. a) Generic neural network. b) The proposed RatioNet architecture with multiplicative skip connections.

We train machine learning models for the two types of commercial amplifiers deployed in the OMSs of our testbed. These have a maximum gain of 21 and 25 dB (denoted EDFA21 and EDFA25, respectively) and are equipped with gain flattening filters and automatic gain control. The latter adjusts the actual average gain to be within 0.1 dB of the nominal gain.

We acquire training datasets for a single physical device of each type. The trained models predict the amplifier response for other physical devices of the respective type. We assume that nominal gain and tilt settings are not known until deployment. To simulate offline factory training, we create fully and partially loaded datasets *ex situ* with the respective EDFAs removed from the OMS and whose distribution is independent of the one an amplifier will see in the field, but covers a similar dynamic range. In the following, we mainly use fully loaded configurations for training. This is because separate treatment of unloaded channels makes uniform sampling less straightforward, and including particular partial load patterns introduces biases. There is no reason that for a sufficiently large dynamic range, models should not also generalize to partial loads. Here we propose an architecture which generalizes to partial loads particularly well despite being trained on full load data with limited dynamic range.

The training dataset for EDFA21 consists of 7745 random configurations of 75 GHz channels, where every second slot is blocked for noise floor measurement to compute OSNR and noise figures. The configurations can be viewed as fully loaded with $N_{ch} = 40$ channels of 150 GHz width over the 6 THz C-band. Because of EDFA level broadening [23], small-scale details of the spectral occupation have a small impact. Per-channel power is obtained by integrating the trace of an Optical Spectrum Analyzer over the 150 GHz channels. The integral of the trace is normalized based on a total power measurement with an uncertainty of 0.1 dB. Per-channel power obtained from repeated measurement of the same trace and with different resolution is consistent with an uncertainty of less than 0.1 dB. Due to experimental constraints, a dynamic range of around 10 dB is available at the WSS to set the attenuation of each channel. The maximum and minimum output power of the device imposes constraints on the allowed gain values for a given total input power. We therefore generate a loading case by first randomly selecting an input power for each channel in the interval $[-19.5, -9.5]$ dBm, so that total input power varies between -3.5 and 6.5 dBm. The nominal gain value is selected in the interval of values allowed by the device constraints. The tilt value is similarly chosen randomly within $[0, 3]$ dB, using a uniform distribution on linear scale. The resulting total output power of the dataset is skewed

Model ↓ / Dataset →	Full Load	Partial Load	OMS	Hidden layers	Hidden units
RatioNet (P)	**0.07 / 0.18 / 0.44**	**0.23 / 0.75 / 4.72**	**0.35 / 0.85 / 1.22**	2	128
Generic NN (G)	0.07 / 0.18 / 0.44	0.30 / 1.10 / 6.26	0.37 / 0.90 / 1.32	2	128
Generic NN (P)	0.11 / 0.35 / 1.10	0.38 / 1.34 / 9.36	0.56 / 1.57 / 2.94	4	512

Table 1. Prediction accuracy in terms of RMSE / 99-percentile / maximum error f the distribution in dB for three NN architectures: RatioNet (Fig. 2b) and a generic NN (Fig. 2a) trained on per-channel power (P) or gain (G). Results are reported for Full Load or Partial Load on the same amplifier after training of the full load dataset. The prediction error for the OMS dataset is averaged over the 4 EDFA21 devices in OMS $\overline{AB}$. The 2229 partially loaded configurations are measured *in situ* in the OMS.

towards higher output powers and cut off at the maximum allowed output power of 22.5 dB. The resulting gain distribution is Gaussian-like with a mean of around 19 dB. Similarly, we generate a second dataset of 4995 samples of partially loaded configurations for testing, where each channel has a 50% probability of being loaded. Loaded channels have a dynamic range of 10 dB in input power as before. Analogous datasets are generated for EDFA25. All data are presented to the NNs on linear scale for training. Inputs and outputs are normalized, since otherwise a ~ 20 dB gap needs to be bridged between them. This helps in particular shallower and narrower networks given that the tanh activation is bounded in $[-1, 1]$.

We train models based on two different NN architectures, as shown in Fig. 2. Both architectures share the same inputs and outputs. Each has N_{ch} inputs for per-channel power and N_{ch} outputs for per-channel power or gain, respectively. Two additional inputs are reserved for the nominal gain G and tilt T. Fig. 2a) shows a generic NN (multilayer perceptron) to predict per channel output power from the corresponding per channel input. The architecture in 2b) is similar, but has additional shortcut or skip connections, which connect each input directly to the corresponding output. The skip connections are inspired from residual networks or ResNets [24], but the motivation here is different: In ResNets, the input is added to the output of the skipped layers. The skipped layers effectively learn the *difference* between input and output of the skipped block, i.e., the residual and hence the name. On the contrary, the skip connections in our architecture are multiplicative, i.e. the input is multiplied to the output. As a result, the skipped layers effectively learn the *ratio* between output and input, i.e., the gain. We therefore refer to this architecture as a RatioNet. The motivation is that per-channel output power is difficult to model by an NN since neighboring outputs can differ by orders of magnitude in amplitude. This has motivated training independent NNs for each channel [7], which however uses computational and memory resources inefficiently. Gain, on the other hand, varies slowly from channel to channel. In this architecture, the constraint that – neglecting ASE – channels with zero input should have zero output is automatically fulfilled. Note that the skip connections participate in backpropagation. This way, low-power channels have a reduced impact on the network weights. This is different from a generic NN that directly predicts per-channel gain, which we also consider below.

We train all NN models with tanh activation functions (which perform better than ReLu). The training is performed with a batch size of 128 and using the Adam optimizer at a learning rate of LR=0.001. Since we ultimately test the models on an independent data set measured in the cascade, we use the entire dataset for training and validation (80%/20% split).

The RatioNets are trained with one or two hidden layers with 32, 64, 128 or 256 hidden units per layer. Models with different hyperparameters perform similarly. Similar behavior is also observed for the generic NNs trained on gain as the output. In the following, we use a single layer with 128 hidden units for both model types and train them for 1000 epochs. Training time ranges between one to two hours per model.

For the generic NN trained on output power, we observe significantly slower convergence. We trained models with 1 to 4 hidden layers and 32, 64, 128, 256 or 512 hidden units per layer and for 5000 epochs. Networks with wider layers perform significantly better. The dependence on the number of layers is typically weak for more than one layer. Contrary to the previous models, predictions for unloaded channels are non-zero. We set predictions for these channels to zero and exclude them from the statistics. In the following we report errors for the best performing models with 4 hidden layers and 512 units per layer.

Table 1 summarizes different prediction error metrics for the three model types, all trained on the same full load dataset. We report RMSE, as well as the 99-percentile of the error distribution and the maximum error. On the full load dataset, the RMSE of RatioNet is as low as 0.07 dB, while that of the generic NN trained on *power* is significantly higher (0.11 dB). Also the maximum error is significantly larger (1.10 vs. 0.44 dB). The generic NN trained on *gain* converges faster in training (~200 vs. ~1000 epochs) but performs very similarly[2]. RatioNet performs slightly better than the generic NN trained on gain, but errors are the same when rounded to the second decimal. The advantage of RatioNet becomes apparent for generalization to unseen test distributions. For the partially loaded dataset, RMSE increases to 0.23 dB for RatioNet and to 0.30 dB for the generic NN trained on gain, showing that RatioNet is better at generalizing to partially loaded configurations despite being trained on fully loaded ones only. It shows the advantage of introducing skip connections *inside* the model. The relatively large maximum error stems from a loading configuration for which a single channel has very low output power. The prediction has even lower power such that, even though the absolute error is small, the relative error is large. The third dataset labeled OMS contains 2229 partially loaded configurations (90 distinct input spectral loads with ~25 lumped loss configurations each) measured *in situ* in the OMS (see Sec. 5). Errors are averaged over the four out of six intermediate amplifiers of the same type. Also here RatioNet performs better.

For all datasets the performance of the generic NN trained on power is the worst, reflecting the fact that gain is easier to model than power. The advantage of RatioNet lies in the ability to better generalize to partially loaded configurations. We expect performance of RatioNet to be further improved by increasing the dynamic range to cover lower input per-channel

[2]Output power is obtained by multiplying gain predictions with input power; the relative errors for gain and power predictions are hence identical.

Full Load	Partial Load	OMS
0.11 / 0.39 / 0.92	0.13 / 0.46 / 5.04	0.22 / 0.61 / 1.16

Table 2. Prediction accuracy for RatioNet in terms of RMSE / 99-percentile of error distribution / maximum error in dB after training on the partial load dataset.

power to better account for partial loads. Further increasing frequency granularity may help to some extent, even though the prediction error increases only moderately (from 0.18 to 0.25 for the 99-percentile on the full load dataset) when granularity is reduced by a factor of two by aggregating two neighboring channels and restoring their relative input power ratio at the output. The overall performance is quite satisfactory given that here the networks generalize to samples which lie outside of the training distribution, as well as to different physical devices of the same type *and* to unseen nominal gain and tilt values. RMSE is comparable to results previously reported in the literature [3, 7]. These models will be used for QoT prediction in Sec. 5.

Finally we remark that satisfactory results for training on the partial load dataset were only obtained with RatioNet. Table 2 shows the performance. It is comparable to training on the full load dataset. The performance on the OMS is even slightly better than training on the full load data, likely because the partial load dataset has a larger dynamic range due to the presence of unloaded channels. Due to potential biases in a relatively small partial load dataset, even though less problematic for RatioNet, the approach is less controlled. We therefore use RatioNet trained on the full load dataset in the following.

4. ML-BASED QoT PREDICTION LEVERAGING QoT PARAMETER REFINEMENT

In this section we describe how to enable the digital twin for QoT prediction when the network changes state through addition or removal of services. We distinguish two basic methods of leveraging IR for this purpose. In both methods, the static amplifier gain spectra estimated by IR are replaced by NN-based models (introduced in Sec. 3) when computing power evolution through the OMS. This is necessary because the EDFA gain spectra are highly load dependent [4], in particular for sparsely loaded systems. In the first method, power evolution is predicted simply by cascading the EDFA-ML and fiber models and QoT is estimated based on predicted power evolution. The resulting predictions still leverage IR through the refined lumped losses. We refer to this method as EDFA-ML with input refined lumped losses (EDFA-ML + IRLL). Accurate estimation of lumped losses has previously been shown to be crucial for QoT estimation for both partially and fully loaded networks [10, 20]. In Sec. 5 we show that the refined lumped losses also improve QoT *predictions* compared to the baseline.

In the second method, we propose to combine EDFA-ML with IR to predict *changes* relative to the current network state, the latter being characterized with high accuracy by means of IR. Since this method uses not only IR lumped losses, but also per-channel power estimates from IR, we refer to this method as EDFA-ML+IR. We explain the principle for the case of service addition. The case of service deletion is completely analogous.

We denote by $P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}}(\lambda_j)$, $P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}(\lambda_j)$ and $P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}}(\lambda_j)$ the optical output power spectrum of the i-th amplifier at wavelength λ_j predicted by IR or ML respectively, before (ba) or after addition (aa) of a service. We omit the dependence on λ_j in the following. The per-channel power prediction at the output of the i-th amplifier is then computed as follows:

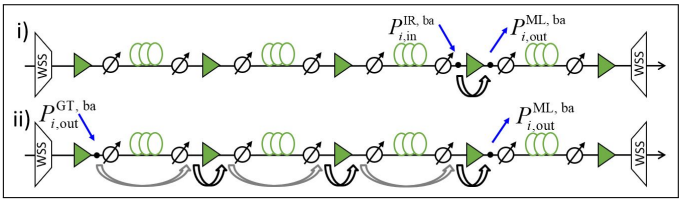

Fig. 3. Two ways of computing cascaded power predictions for amplifier i along the OMS *before* service addition. i) Using $P^{\mathrm{IR,ba}}_{i,\mathrm{in}}$ as approximation for the ground truth at amplifier input and EDFA-ML to predict output. ii) Using the ground truth at OPM level and cascade predictions along the OMS.

$$P^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} + P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}} - P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}. \tag{2}$$

Note that Eq. (2) applies to all services – existing ones and newly added services. This result can be derived as follows. Define the difference between the ground truth output power and the corresponding machine learning prediction before (respectively after) channel addition as $\Delta^{\mathrm{ba}}_{i,\mathrm{out}} = P^{\mathrm{GT,\,ba}}_{i,\mathrm{out}} - P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}$ and $\Delta^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{GT,\,aa}}_{i,\mathrm{out}} - P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}}$, respectively. If $\Delta^{\mathrm{aa}}_{i,\mathrm{out}}$ were known, the ground truth could obviously be obtained exactly as $P^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} + \Delta^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{GT,\,aa}}_{i,\mathrm{out}}$. We make two approximate assumptions: i) we assume that the correction before and after service addition is similar for the established services: $\Delta^{\mathrm{ba}}_{i,\mathrm{out}} \approx \Delta^{\mathrm{aa}}_{i,\mathrm{out}}$ and ii), since $\Delta^{\mathrm{ba}}_{i,\mathrm{out}}$ is unknown (since the ground truth is unknown), we replace the ground truth before addition by the approximate but accurate [20] IR value: $P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}} \approx P^{\mathrm{GT,\,ba}}_{i,\mathrm{out}}$. Put together, we obtain

$$\begin{aligned} P^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} + \Delta^{\mathrm{aa}}_{i,\mathrm{out}} &\approx P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} + \Delta^{\mathrm{ba}}_{i,\mathrm{out}} \\ &\approx P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} + P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}} - P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}, \end{aligned} \tag{3}$$

which is the result given above. Note that for the service(s) to be added, $\Delta^{\mathrm{ba}}_{i,\mathrm{out}}$ is zero, since both $P^{\mathrm{GT,\,ba}}_{i,\mathrm{out}}$ and $P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}$ are zero. Hence $P^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}}$ for newly added services.

Accuracy compared to the directly cascaded machine learning prediction $P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}}$ (without using IR) can be gained when the approximation errors of the two above approximations are small compared to the correction term, i.e. $\|P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}} - P^{\mathrm{GT,\,ba}}_{i,\mathrm{out}}\| \ll \|\Delta^{\mathrm{aa}}_{i,\mathrm{out}}\|$ and $\|\Delta^{\mathrm{ba}}_{i,\mathrm{out}} - \Delta^{\mathrm{aa}}_{i,\mathrm{out}}\| \ll \|\Delta^{\mathrm{aa}}_{i,\mathrm{out}}\|$. The former is directly related to the accuracy of the IR method, which we can assume to be accurate for gain profile estimation (cf., e.g., Fig. 5). We expect the latter to be fulfilled for cascaded ML predictions whose error is dominated by error amplification.

Alternatively we can interpret Eq. (2) as $P^{\mathrm{aa}}_{i,\mathrm{out}} = P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}} + d^{\mathrm{ML}}_{i,\mathrm{out}}$, where $d^{\mathrm{ML}}_{i,\mathrm{out}} = P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}} - P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}$. When changes in the network configuration due to addition or removal of a few services are small, $P^{\mathrm{IR,\,ba}}_{i,\mathrm{out}}$ estimated by IR with high accuracy as shown in Sec. 5 is the dominant term and ML provides a small correction.

Here $P^{\mathrm{ML,\,aa}}_{i,\mathrm{out}}$ must be computed by cascading, since only the input power spectrum at the first OMS is known[3]. The ML predictions $P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}$ before service addition in Eq. (2) can in principle

[3]The power of the service to be added is assumed to be provided by the operator.

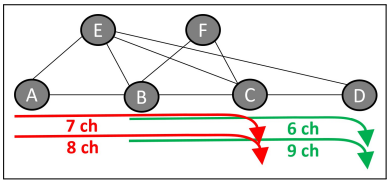

Fig. 4. Schematic of an example optical network topology. In the experimental validation, without loss of generality, we consider the four-node tandem subnetwork $\overline{ABCD}$.

be computed in three different ways. Two are schematically illustrated in Fig. 3. In Method i) the prediction at the output of amplifier i is obtained from $P_{i,\text{in}}^{\text{IR, ba}}$ estimated by IR, followed by a single EDFA-ML prediction. This is possible since the power spectrum at each EDFA is a byproduct of IR. In method ii) it is obtained by cascading the predictions starting from the GT power spectrum provided by the OPM of the first OMS. A third method is to use the output power $P_{i,\text{out}}^{\text{IR, ba}}$ provided by IR. In this case, Eq. (2) reduces to $P_{i,\text{out}}^{\text{aa}} = P_{i,\text{out}}^{\text{ML, aa}} + P_{i,\text{out}}^{\text{IR, ba}} - P_{i,\text{out}}^{\text{IR, ba}} = P_{i,\text{out}}^{\text{ML, aa}}$, i.e. the cascaded EDFA-ML prediction based on the IR lumped losses (EDFA-ML+IRLL).

While method ii) to compute $P_{i,\text{out}}^{\text{ML, ba}}$ is *per se* less accurate than i) due to cascading, it is analogous to the computation of $P_{i,\text{out}}^{\text{ML, aa}}$ *after* service addition. One may therefore expect errors to be correlated, according to our assumption i). In Sec. 5 we show that as a result, error cancellation between the ML predictions before and after service addition occurs, which greatly enhances the prediction accuracy. The resulting digital twin provides per-channel SNR and power predictions that are substantially more accurate than naively cascaded predictions.

5. EXPERIMENTAL VALIDATION

A. Experimental setup and baseline

Without loss of generality, we consider the tandem subnetwork $\overline{ABCD}$ of the network depicted in Fig. 4. Our method is applicable to generic mesh networks. Since IR can be done for each OMS independently, topological properties of the network are not used. The subnetwork consists of 3 OMS, such as the one schematically represented as in Fig. 1. The 3 OMS are composed of 5 fiber spans each (5×80 km SMF, 5×100 km PSCF, 5×100 km PSCF). Variable optical attenuators (VOAs) before and after each fiber span are used to emulate lumped losses up to 3 dB. In OMS $\overline{AB}$ all 6 EDFAs are of type A, while OMS $\overline{BC}$ and $\overline{CD}$ each contain 6 amplifiers of both types A and B.

For all experiments, we emulate various channel loading configurations of the network using an amplified spontaneous emission (ASE) source, spectrally shaped by a wavelength-selective switch (WSS). Per-channel amplifier gain spectra $G_i(\lambda)$, input and output power spectra $P_i(\lambda)$ and noise figures $F_i(\lambda)$ as well as total input and output power are measured at each amplifier. These ground truth values are used for benchmarking purposes (computation of estimation or prediction errors), or fed into our QoT tool to emulate ground truth SNRs unless they are measured. Only the power spectra at OPM level (at the output of the first and last EDFA), the total input and output power of the EDFAs, and measured or emulated ground truth SNRs are used for the input refinement.

Without IR, it is natural to assume that amplifier gain spectra are free of gain ripples and hence vary linearly with wavelength

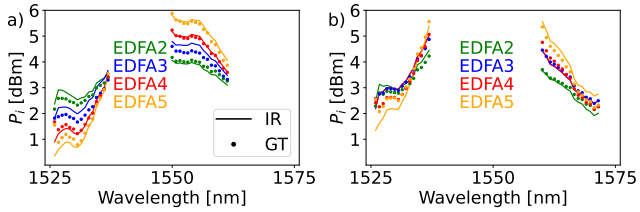

Fig. 5. Refined (IR) and ground truth (GT) power profiles for the four inner EDFAs in OMS $\overline{BC}$ for two different channel loading configurations. Dots correspond to ground truth, lines correspond to IR.

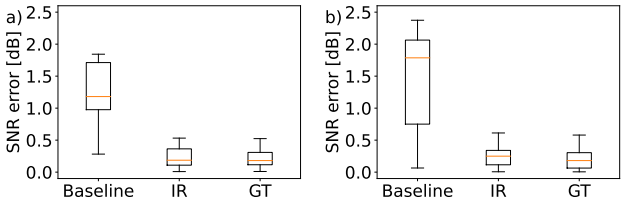

Fig. 6. Standard box plot showing the distribution of SNR estimation error based on baseline QoT parameter inputs, refined inputs (IR) and ground truth (GT) inputs, for the same two channel configurations as in Fig. 5.

(in dB) according to the nominal gain and tilt settings of each of the amplifiers ("linear tilted gains"). Lumped losses are assumed to be equally distributed between fiber input and output ($\delta_i = \delta_i'$). In the remainder of the paper, we refer to these assumptions as the baseline and use them as a benchmark for both QoT estimation and prediction. Tables 3 and 4 list the assumptions underlying the different methods.

B. Validation of QoT estimation

We first assess our digital twin for QoT estimation. We consider two network configurations with service allocation as depicted in Fig 4: Between 15 to 30 services (depending on OMS) with 75 GHz ASE-shaped channels are allocated over a 100 GHz grid in the C-band. SNR ground truth is measured by sweeping a probabilistically shaped 16-QAM channel from a commercial real-time 200 Gb/s transponder over the 30 occupied spectral positions.

Fig. 5 shows estimated output power profiles after IR (lines) vs. the measured ones (markers) of all intermediate EDFAs of OMS $\overline{BC}$ for the two block-loaded configurations. IR estimates EDFA output power profiles within 0.2 dB of the measured power profiles for both loading configurations and all EDFAs. Fig. 6 shows the corresponding distribution of SNR estimation errors relative to ground truth SNRs of all 30 channels obtained by QoT estimation for the baseline, ii) input refinement (IR) and

GT	IR	IR + baseline LL	Baseline
GT gains	IR gains	IR gains	Linear tilted gains
GT LLs	IR LLs	$\delta_i = \delta_i'$	$\delta_i = \delta_i'$

Table 3. The different cases compared in QoT estimation and their underlying assumptions.

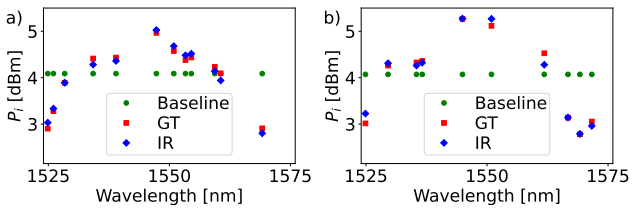

Fig. 7. Per-channel power obtained according to baseline assumptions of linearly tilted gains and lumped losses, after input refinement (IR) and measured ground truth (GT) of an intermediate EDFA in OMS $\overline{AB}$ for a sparsely loaded network configuration with 10 services randomly placed in the band.

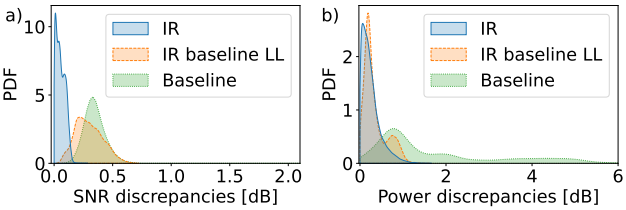

Fig. 8. a) Distribution of discrepancies of per-channel SNR estimation $|\mathrm{SNR}^{\mathrm{GT,dB}} - \mathrm{SNR}^{\mathrm{pred,dB}}|$ from IR after gain refinement and with refined or baseline distribution of lumped losses. The baseline (linearly tilted gain and equal distribution of lumped losses) is shown for comparison. Accurate estimation of the lumped loss distribution has a significant effect on QoT estimation accuracy. b) The same as in a), but for power discrepancies $|P^{\mathrm{GT,dB}} - P^{\mathrm{pred,dB}}|$.

iii) and ground truth (GT) values for gains and lumped losses (see Table 3). SNR estimation is improved by up to 2 dB after IR compared to the case of datasheet assumptions underlying the baseline. An average (maximum) error of 0.2 dB is obtained after refinement, respectively, down from more than 1 dB. The maximum error after IR is 0.6 dB. The values for IR are close to the values obtained by QoT estimation based on GT gains and lumped losses, showing that accuracy, though very high, is limited by QoT tool estimation error rather than IR.

Fig. 7 shows the measured (GT) output power profile for a random sparse loading configuration with only 10-12 services of an intermediate EDFA of OMS $\overline{AB}$ compared to the output power profile estimated using the baseline and after IR. Here up to 6 dB variation between channel powers is observed. Such power excursions can be a consequence of a "set and forget" mode of operation of an optical network, whereby the powers of existing services are left to drift as new services are added. Despite the large power variation between channels, per-EDFA output power profiles are still estimated by IR within 0.2 dB compared to the ground truth. The accuracy of the technique for power profile estimation does not depend on the specific features of the channel loading configuration (e.g., block vs. sparse loading).

In order to further validate the IR technique for lumped loss estimation, we performed a second measurement campaign using OMS $\overline{AB}$. 90 randomly generated input power spectral loads with 5 up to 40 ASE-shaped channels are generated via a WSS. Here and in the remainder of the paper we use a 75 GHz WDM grid over the 6 THz C-band, where every second channel

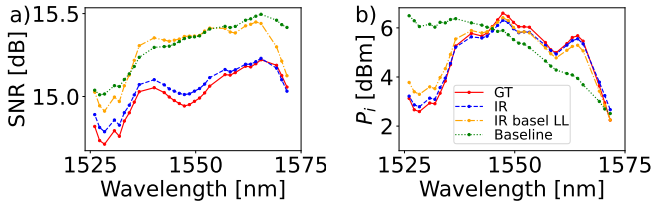

Fig. 9. Example SNR (a) and power spectra (b) measured (ground truth, GT) and estimated using baseline assumptions (Baseline), input refinement with baseline lumped losses (IR baseline LL) and IR for both gains and lumped losses. Dots show service locations of the 30 services (75% spectral occupation). Lines are a guide to the eye.

slot is blocked for noise floor measurement. For each service allocation we generate 25 distinct lumped loss configurations. For each configuration, the lumped losses δ_i, δ_i', amplifier gain spectra $G_i(\lambda)$, power spectra $P_i(\lambda)$ and noise figures $F_i(\lambda)$ are recorded and mapped into the digital twin for QoT estimation to emulate ground truth SNRs. Among the up to 40 channels, two are modulated using commercial 68 GBd 400 Gbps real-time transponders to cross-validate the emulated QoT SNR ground truth estimates.

Fig. 8a) shows the probability density function (PDF) of SNR errors for more than 40000 services. The RMSE is reduced from 0.40 dB for the baseline by more than 0.30 dB down to 0.07 dB for IR. The RMSE for IR with baseline lumped losses is reduced by only 0.08 dB compared to that of the baseline, showing that estimation of the lumped loss distribution is essential for accurate QoT estimation. Fig. 8b) shows the corresponding distribution of per-channel power estimation error at the second-to-last amplifier in the OMS (since the output power spectrum of the last amplifier is known thanks to an OPM, we do not report it). Here the large RMSE of the baseline of 2.25 dB is reduced significantly down to 0.33 dB using IR. Using IR with baseline lumped losses yields an RMSE of 0.43, i.e. the refinement of the lumped losses in addition to the amplifier gain spectrum improves power estimation accuracy by about 0.1 dB.

One can further see that the power discrepancies for IR with baseline lumped losses exhibit a bimodal distribution. The peak of the first mode of the distribution is close to the peak of the distribution using IR with refined lumped losses. The peak of the second mode is close to 1 dB. Thanks to an accurate estimation of the power profiles using IR refined gains, the majority of channel powers of the band are predicted within 0.4 dB on average, similarly as with IR. The use of baseline lumped losses however leads to inaccurate modeling of the power profile Raman tilt. This impacts mostly channels at the band edges and gives rise to a second mode in the distribution. Using refined lumped losses improves modeling of the Raman tilt and the estimation accuracy for extreme channels, which can be observed in Fig. 9b) as discussed in the following.

Fig. 9a) shows the SNR for an example load with 30 services, corresponding to 75% spectral occupation. One can clearly see that refining the amplifier gain profiles while keeping baseline lumped losses (IR + baseline LL) leads to a significant improvement of SNR estimation compared to the baseline. In particular the shape of the resulting SNR profile is better captured, however retaining an offset and tilt compared to the ground truth (GT). A corresponding improvement in power prediction accu-

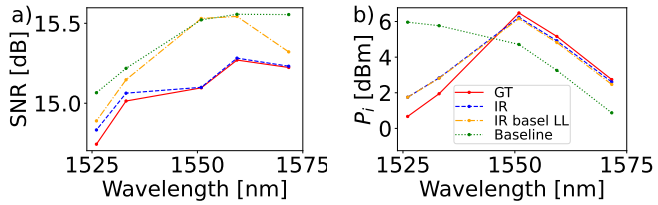

Fig. 10. Example SNR (a) and power spectra (b) measured (ground truth, GT) and estimated using baseline assumptions (Baseline), input refinement with baseline lumped losses (IR baseline LL) and with both IR gains and lumped losses for a sparse loading with only 5 services (12.5% spectral occupation) marked by dots. Lines are a guide to the eye.

racy can be seen in Fig. 9b). For both SNR and power discrepancies, estimation of the lumped loss distribution using IR reduces those errors, yielding a very close match between estimated and ground truth SNR and power.

In Fig. 10 we show SNR and power spectra for a sparse loading with only 5 services, which allows similar conclusions. One difference is that, in this particular case, total power at the input of the fiber and hence the Raman effect are reduced. Thus, changing lumped loss values (e.g. using IR baseline LL or refined IR lumped losses) has a negligible impact on the power profile estimation accuracy, as seen in Fig. 10b). The error in per-channel power estimation in these cases is dominated by the estimation error of the gain profiles, rather than lumped losses. This example illustrates that, despite the reliance of IR on nonlinear Raman and Kerr effects [10], IR performs well for power and SNR estimation in sparsely loaded scenarios with a low amount of nonlinear effects.

C. Validation of QoT prediction

In this section we validate QoT prediction of the digital twin in dynamic networks with varying service allocation. We use the same dataset as in the previous section, with 90 random loading configurations generated with 5 up to 40 ASE-shaped 75 GHz channels in every second channel slot, and with 25 different input and output lumped loss configurations each. For the assumptions underlying the different approximations considered, see Table 4.

We first predict the SNR of existing and newly established services in OMS $\overline{AB}$ using the RatioNet NN models of Sec. 3 for EDFA gain prediction to account for the load dependence of amplifier gain spectra. IR is used to estimate the lumped losses, which are load-independent (EDFA-ML+IRLL). Fig. 11a) compares distributions of SNR prediction errors for all services for the baseline and EDFA-ML+IRLL. We also consider a hybrid approach where gain profiles are predicted using EDFA-ML for configurations with up to 30 services (up to 75% band occupation). For higher occupation, gain profiles are replaced by IR gains for already existing services, while for newly added services, gain values are linearly interpolated or extrapolated from the IR gains (EDFA-ML+IRLL hybrid). This is reasonable since gain profiles tend to be rather stable when adding new services in a highly loaded scenario.

For EDFA-ML+IRLL, both RMSE and maximum error are reduced, by 0.2 and 0.7dB, respectively, compared to the baseline. We find that large values of maximum error are in general due to outliers when EDFA-ML is applied to predict amplifier gains.

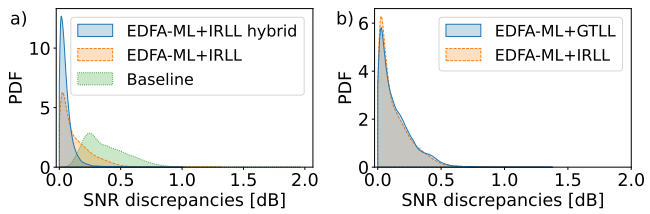

Fig. 11. a) Distribution of discrepancies of per-channel SNR predictions $|\mathrm{SNR}^{\mathrm{GT,dB}} - \mathrm{SNR}^{\mathrm{pred,dB}}|$ at the end of the OMS from EDFA-ML+IRLL and a hybrid approach where , for band occupation > 75%, gain spectra are replaced by those estimated by IR for existing services and gains are inter- or extrapolated for new services (EDFA-ML+IRLL hybrid). The baseline is shown for comparison. b) SNR discrepancies for EDFA-ML+IRLL and EDFA-ML gain prediction using ground truth lumped losses (EDFA-ML+GTLL). The distributions are virtually the same. EDFA-ML prediction accuracy is the limiting factor for EDFA-ML+IRLL, rather than lumped loss accuracy.

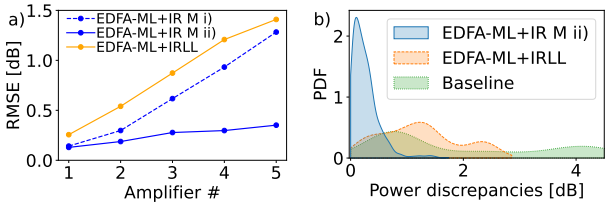

Fig. 12. a) RMSE of per-channel power predictions for EDFA-ML+IR with methods i) and ii) for the ML prediction before channel addition as illustrated in Fig. 3, compared to EDFA-ML+IRLL. b) Distribution of power discrepancies of EDFA-ML+IR based on method ii) compared to EDFA-ML+IRLL and the baseline at the last amplifier of the OMS.

In the hybrid approach the RMSE is reduced by an additional 0.1 dB while the maximum error is reduced by an additional 0.3 dB compared to EDFA-ML+IRLL. This shows that, in highly loaded scenarios (e.g. towards the network end-of-life), IR allows to improve SNR estimates by describing the gain ripples more accurately than the NN models. Note that this is based on the implicit assumption that gain spectra do not change when services are added. In other words, the NN error in some cases is larger than the change to be predicted. Fig. 11b) compares EDFA-ML+IRLL with the case where EDFA-ML is combined with ground truth values instead of IR estimated values for the lumped losses (EDFA-ML+GTLL). The distributions are almost the same, indicating that prediction accuracy is indeed limited by the EDFA-ML prediction error, rather than the error in the lumped loss distribution due to IR. These results motivate combating error amplification stemming from cascading the

EDFA-ML+IR	EDFA-ML+IRLL	Baseline
EDFA-ML gains	EDFA-ML gains	Linear tilted gains
IR δ_i, δ_i' and P_i^{IR}	IR δ_i, δ_i'	$\delta_i = \delta_i'$

Table 4. The main approximations considered in QoT prediction and their underlying assumptions.

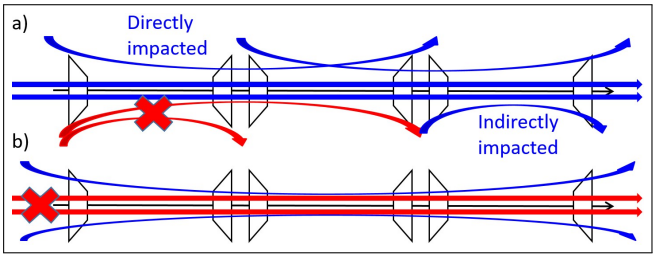

Fig. 13. a) Illustration of a typical fiber cut scenario over three OMS with services impacted directly (as they copropagate with lost services) and indirectly (as they copropagate with services that are directly impacted, but not with lost services). Lost services are marked in red and remaining services in blue. b) Worst-case scenario simulated in the experiment, where lost services are copropagating over all three OMS and all remaining services are directly impacted.

EDFA-ML predictions to further improve prediction accuracy.

To this end, we employ the EDFA-ML+IR method. We first apply it to successive addition of individual services. Here we use an experimental campaign where up to 40 ASE-shaped 75 GHz channels in every second slot of the 6 THz C-band are loaded sequentially from small to large wavelengths starting from a single service. Per-channel power for the new and previously established services is predicted by EDFA-ML+IR.

As explained in Sec. 4, the ML predictions $P^{\mathrm{ML,\,ba}}_{i,\mathrm{out}}$ in EDFA-ML+IR can be computed using different methods. Fig. 12a) compares EDFA-ML+IRLL with EDFA-ML+IR based on these methods. The RMSE of EDFA-ML+IRLL essentially increases exponentially with the number of amplifiers due to error amplification at a slope of about 0.25 dB per amplifier. EDFA-ML+IR with method i) is more accurate, but still increases almost as fast. With EDFA-ML+IR with method ii) however, the increase is significantly slower. This suggests that errors in the ML predictions before and after service addition are correlated and error cancellation occurs between them (cf. Sec. 4). Fig. 12b) shows the distribution of power discrepancies of EDFA-ML+IR based on method ii) compared to EDFA-ML+IRLL and the baseline at the last EDFA of the OMS. The assumptions underlying these approximations are summarized in Table 4. RMSE is reduced from 2.47 dB (Baseline) to 1.41 dB (EDFA-ML+IRLL) and 0.35 dB (EDFA-ML+IR M ii). This confirms significant accuracy gains of the method. We only use the more accurate method ii) in the following.

C.1. Fiber cut scenario

In this section we evaluate the EDFA-ML + IR method for a fiber cut scenario, in which multiple services are lost simultaneously. This example is motivated by the fact that changes in the network configuration are larger than for addition and deletion of individual services, which poses a more stringent test on the method. Fig. 13a) illustrates a typical fiber cut scenario in part of a meshed network where services copropagate over different routes. One can primarily distinguish directly impacted services, which copropagate with services that are lost, and indirectly impacted services, which copropagate with services that are directly impacted, but not with lost services. To simplify the test case generation, we consider in our experiment the worst-case scenario for a tandem network of up to 3 OMS shown in b), where the remaining channels copropagate with lost services over all 3 OMS and hence are directly impacted. We

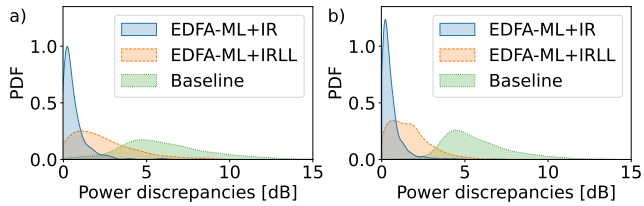

Fig. 14. Distribution of discrepancies of per-channel power predictions $|P^{\mathrm{GT,dB}}_{N,\mathrm{out}} - P^{\mathrm{ad,dB}}_{N,\mathrm{out}}|$ after service drop (ad) during a fiber cut. a) After traversing 3 OMS and $N = 18$ amplifiers and after dropping between 1–5 services. b) After traversing 2 OMS and $N = 12$ amplifiers and dropping between 1–33 services.

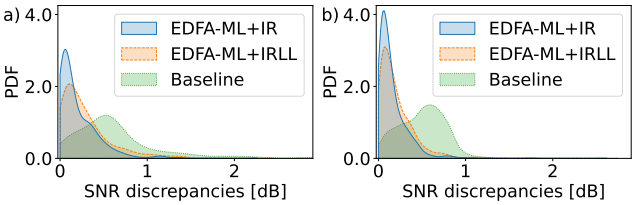

Fig. 15. Distribution of discrepancies of per-channel SNR predictions $|\mathrm{SNR}^{\mathrm{GT,dB}} - \mathrm{SNR}^{\mathrm{ad,dB}}|$ after service drop ('ad') and based on power evolution predictions. a) After traversing 3 OMS and $N = 18$ amplifiers and after dropping between 1–5 services. b) After traversing 2 OMS and $N = 12$ amplifiers and dropping between 1–33 services.

generated 94 pairs of loading cases, with the first in each pair generated randomly, having between 7 and 40 services. In the second of each pair, between 1 and 33 randomly chosen services are dropped.

The distribution of power discrepancies for this scenario are shown in Fig. 14. Here the power $P^{\mathrm{ad}}_{N,\mathrm{out}}$ predicted after service drop (ad) is computed for either EDFA-ML+IRLL, EDFA-ML+IR or the baseline. Fig. 14a) shows power discrepancies after 3 OMS (N =18 traversed amplifiers) and for fiber cuts where between 1 and 5 services of the original configuration are dropped. Fig. 14b) is for the case of 2 OMS (12 amplifiers), with up to 33 channels dropped simultaneously. In each case, IR is performed on the configuration before dropping channels to estimate the distribution of lumped losses.

The combination of EDFA-ML with IR estimated lumped losses (EDFA-ML+IRLL) greatly improves the prediction of the power of the remaining services in both scenarios. The RMSE for power prediction is reduced by more than 3 dB compared to the baseline. EDFA-ML+IR further reduces the RMSE for power prediction by more than 1dB in both cases.

Fig. 15 shows corresponding results for the SNR discrepancies $|\mathrm{SNR}^{\mathrm{GT}} - \mathrm{SNR}^{\mathrm{ad}}|$. Here ground truth SNRs are emulated using QoT estimation based on ground truth lumped losses and ground truth amplifier gain spectra. SNRs after channel deletion are computed using the digital twin with power predictions obtained by the respective method. EDFA-ML with refined lumped losses reduces global RMSE by 0.5 and 0.4dB, respectively, for the two scenarios and compared to the baseline.

Despite the large improvement in power prediction accuracy of more than 4 dB for EDFA-ML+IR compared to the baseline,

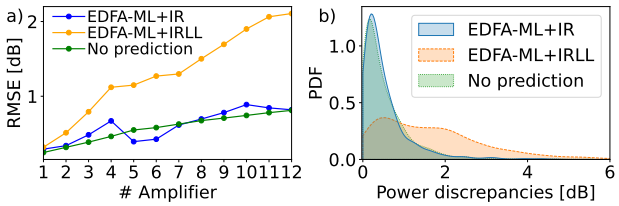

Fig. 16. a) RMSE of per-channel power predictions vs the number of cascaded amplifiers for EDFA-ML+IR, EDFA-ML+IRLL and the case of no prediction, i.e. assuming that services remain unchanged after the fiber cut. b) Power discrepancies at the last EDFA of the second OMS. The EDFA-ML+IR error distribution is similar to the one obtained without prediction.

the improvement in SNR accuracy is relatively small. This is because most services in the scenario are close to the nonlinear threshold (NLT) power, where ASE noise power is twice the nonlinear noise power. Here the SNR curve is essentially flat with respect to power, such that improvements in power accuracy do not translate into significantly higher SNR accuracy. Conversely, services for which power is far from NLT sustain a significant SNR accuracy improvement with our technique. As an example, we observed that over 3 OMS with up to 5 channels dropped, the RMSE of the SNR prediction is improved from 2.6 to 1.8 dB compared to the baseline when EDFA-ML is supplemented by IR lumped losses (EDFA-ML +IRLL) and further from 1.8 to 0.9 dB, for EDFA-ML+IR for services in the highly linear regime (power ~3 dB above NLT). This is an improvement of more than 1.5 dB compared to the baseline. The maximum SNR error reduces from 3.5 dB (baseline) to 2.4 dB (EDFA-ML+IRLL) to 1.3 dB (EDFA-ML+IR). Over 2 OMS with up to 33 channels dropped, the RMSE of SNR prediction is improved from 1.9 dB (baseline) to 1.5dB (EDFA-ML+IRLL) to 0.9 dB (EDFA-ML+IR).

In Fig. 16a) we finally compare how the RMSE of EDFA-ML+IRLL and EDFA-ML+IR increase with distance for this scenario. As before, see Fig. 12a), the error of the cascaded ML predictions of EDFA-ML+IRLL increases rapidly, much faster than that of EDFA-ML+IR. For comparison we also plot the accuracy that would be obtained using no prediction, i.e., by assuming that power of the remaining services does not change when services are dropped. One can see that the impact of a change in the network increases exponentially and the slope is comparable to that of the EDFA-ML+IR predictions. Fig. 16b) shows that the error distributions of EDFA-ML+IR and for the case of no prediction at the last amplifier are also comparable. Since the offset of the curves in Fig. 16a) is similar, the prediction error is of the order of the change to be predicted. By improving the accuracy of the individual amplifier models, the error of the cascaded predictions may also be improved.

6. CONCLUSIONS

We have presented a ML enhanced digital twin for QoT estimation and QoT prediction in optical networks. Our QoT input parameter refinement estimates the lumped loss distribution and amplifier gain spectra in the current network state. We observed significant improvements in QoT estimation accuracy compared to the baseline, with a reduction of SNR estimation error by 0.3 down to 0.07 dB.

To enable our digital twin for QoT prediction, we designed and trained novel neural network models for two types of amplifiers, which generalize over arbitrary nominal gain and tilt settings and over different physical devices of the same type. An RMSE of 0.35 dB was obtained for EDFAs of an OMS of a deployed network after simulated offline factory training.

It turns out that error amplification from the cascaded predictions is the main limitation for obtaining high accuracy for QoT prediction. We developed a method that significantly alleviates error amplification through error cancellation. Per-channel power prediction error (RMSE) after 5 cascaded amplifiers when adding a single service was reduced from 2.47 dB for the baseline down to 0.35 dB for the new method. For removal of multiple services in a simulated fiber cut, the RMSE after 12 amplifiers was reduced by as much as 4 dB compared to the baseline, leading to an improvement in SNR prediction accuracy by 0.5 dB on average and by more than 2 dB for services far from nonlinear threshold.

While being significantly more accurate than straightforward cascaded predictions, the error remained comparable to the changes to be predicted for all distances. Prediction accuracy of the cascaded predictions is expected to be improved by further reducing the error of the individual EDFA model predictions. It remains to be seen if the correlations in predictions errors that lead to their cancellation hold over links even longer than our setup. Individual prediction errors may be reduced [25] by better matching the training and test distributions, in particular by extending the training data dynamic range, better controlling for experimental uncertainties or by using alternative modeling approaches which combine machine learning methods with the device physics [26]. Transfer learning techniques may prove useful to improve model accuracy for limited training data or reduce the overhead of offline factory training.

A benefit of the modular nature of our digital twin is that amplifier models can be trained offline to update the twin when a device is replaced. We have further shown that using input refinement, the lumped losses and gain spectra for QoT estimation can be updated using conventionally monitored information in the current network state. Since parameters of fibers, filters and other passive devices of our digital twin can be inferred from datasheets, the model can be updated using input refinement without requiring online data acquisition that would otherwise impact network operation.

7. ACKNOWLEDGMENT

The authors would like to thank Dylan Le Gac for his contributions to the data acquisition for this paper.

REFERENCES

1. Y. Pointurier, "Design of low-margin optical networks," J. Opt. Commun. Netw. **9**, A9–A17 (2017).
2. Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," J. Opt. Commun. Netw. **13**, B60–B71 (2021).
3. F. da Ros, U. C. de Moura, and M. P. Yankov, "Machine learning-based EDFA Gain Model Generalizable to Multiple Physical Devices," in *2020 European Conference on Optical Communications (ECOC)*, (2020), pp. 1–4.
4. A. Mahajan, K. Christodoulopoulos, R. Martínez, S. Spadaro, and R. Muñoz, "Modeling EDFA Gain Ripple and Filter Penalties With Machine Learning for Accurate QoT Estimation," J. Light. Technol. **38**, 2616–2629 (2020).
5. Y. You, Z. Jiang, and C. Janz, "Machine Learning-Based EDFA Gain Model," in *2018 European Conference on Optical Communication (ECOC)*, (2018), pp. 1–3.

6. S. Zhu, C. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman, and D. Kilper, "Hybrid Machine Learning EDFA Model," in *2020 Optical Fiber Communications Conference and Exhibition (OFC),* (2020), pp. 1–3.
7. 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),* (2018), pp. 1–3.
8. R. Ayassi, A. Triki, M. Laye, E. L. Rouzic, N. Crespi, and R. Minerva, "Bayesian Optimization-Based Algorithm to Improve the Quality of Transmission Estimation," in *OSA Advanced Photonics Congress 2021,* (Optica Publishing Group, 2021), p. NeF2B.3.
9. 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).
10. N. Morette, I. F. de Jauregui Ruiz, and Y. Pointurier, "Leveraging ML-based QoT Tool Parameter Feeding for Accurate WDM Network Performance Prediction," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (2021), pp. 1–3.
11. G. Borraccini, A. D'Amico, S. Straullu, F. Usmani, A. Ahmad, and V. Curri, "Iterative supervised learning approach using transceiver bit-error-rate measurements for optical line system optimization," J. Opt. Commun. Netw. **15**, 111–118 (2023).
12. S. Li, D. Wang, Y. Song, Q. Fan, M. Zhang, C. Lu, and A. P. Tao Lau, "Digital twin-enabled Power Optimizer for Multi-span Transmission System Using Autoencoder," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (2021), pp. 1–3.
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 *2020 European Conference on Optical Communications (ECOC),* (2020), pp. 1–4.
14. M. P. Yankov, U. C. de Moura, and F. D. Ros, "Power Evolution Modeling and Optimization of Fiber Optic Communication Systems with EDFA repeaters," J. Light. Technol. **39**, 3154–3161 (2021).
15. 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).
16. X. Pang, S. Li, Q. Fan, M. Zhang, C. Lu, A. P. T. Lau, and D. Wang, "Digital Twin-Assisted Optical Power Allocation for Flexible and Customizable SNR Optimization," in *2022 Optical Fiber Communications Conference and Exhibition (OFC),* (2022), pp. 1–3.
17. 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).
18. S. Tariq and J. Palais, "A computer model of non-dispersion-limited stimulated Raman scattering in optical fiber multiple-channel communications," J. Light. Technol. **11**, 1914–1924 (1993).
19. Z. Wang, E. Akinrintoyo, D. Kilper, and T. Chen, "Optical Signal Spectrum Prediction Using Machine Learning and In-line Channel Monitors in a Multi-span ROADM System," in *2022 European Conference on Optical Communication (ECOC),* (2022), pp. 1–4.
20. 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 *2022 Optical Fiber Communications Conference and Exhibition (OFC),* (2022), pp. 1–3.
21. 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), pp. 1–4.
22. A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," Opt. Express **22**, 16335–16362 (2014).
23. L. Bigot, S. Guy, B. Jacquier, A. M. Jurdyc, D. Bayart, S. Blanchandin, and L. Gasca, "Homogeneous and inhomogeneous line broadening in edfa," in *Proceedings of the International Workshop on Photonics and Applications; Hanoi, Vietnam,* (2004), pp. 30–35.
24. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2016), pp. 770–778.
25. Z. J. Jiachuan Lin, Xiang Lin, "Auxiliary Neural Network Assisted Machine Learning EDFA Gain Model," in *2023 Optical Fiber Communications Conference and Exhibition (OFC),* (2023), pp. 1–3. Paper M2E.2.
26. 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),* (2021), pp. 1–3.