

Experimental Investigation of Spectral Data Enhanced QoT Estimation

Lars E. Kruse , Sebastian Kühl , Annika Dochhan, *Member, IEEE*,
and Stephan Pachnicke , *Senior Member, IEEE*

Abstract—Obtaining telemetry data from the optical signal without demodulation at intermediate nodes of optical networks can be achieved using optical spectrum analyzers. The information gained from the spectrum analysis can be further used for different applications, which include quality of transmission estimation (QoT). Accurate QoT estimation allows to maximize network capacity and minimize margins either through reconfiguration or during the deployment. Analytical solutions for QoT estimation require exact knowledge of the parameters (e.g. fiber lengths, attenuation, non-linearity coefficients), which are not always exactly known in practice especially in multi-vendor networks. Machine learning has shown to be able to handle such parameter-agnostic scenarios. In this paper, we experimentally compare different machine learning based QoT estimators to our developed spectral data driven estimators as well as comparing it to a new approach of using automated feature extraction from the spectrum by a variational autoencoder (VAE). The VAE-based estimation approach is experimentally validated and the required optical spectrum analyzer (OSA) resolutions are investigated. The spectral data driven estimators show to be superior regarding both R^2 -score and mean absolute error. Furthermore, the automated feature extraction using the VAE is shown to be a suitable option for accurate optical performance monitoring without demodulation and QoT estimation.

Index Terms—Agnostic networks, optical performance monitoring, quality of transmission estimation, recurrent neural networks, variational autoencoder.

I. INTRODUCTION

OPTICAL networks have changed enormously to address the issue of the growing bandwidth demand. In this regard, new flexible add-drop multiplexers enable more versatile network operation as well as the implementation of flexible frequency grids. This leads to more sophisticated, configurable, and adaptable networks. Due to the increase in network complexity, monitoring and optimizing performance is of increasing importance. Today's optical networks ensure the guarantee of

service level agreements and promised capacity by including large operating margins which include unallocated system and design margins [1]. However, the disadvantage of such a conservative approach is that large margins lead to wasted capacity. In this context, accurate quality of transmission (QoT) estimation allows to maximize the capacity and may enable full self-management of the networks in the future by ensuring low-margin optical networking [2]. Multi-vendor optical networks make accurate QoT estimation a nontrivial task since exact equipment parameters are considered confidential or are in general not exactly known. Due to this, such a multi-vendor network can be considered a so-called exact component parameter agnostic network scenario. On top of the not exactly known component parameters, the parameter uncertainties and fiber nonlinearities increase the complexity of the QoT estimation task [3]. The signal quality, represented by the signal-to-noise-ratio (SNR), depends not only on the linear amplified spontaneous emission (ASE) noise from the Erbium-doped fiber amplifiers (EDFAs) in the network but also on the signal power, the power of the individual channels and the channel spacing. These influence the nonlinearities that distort the channel and lead to intersymbol and interchannel interference.

Different approaches for QoT estimation have been proposed over the past decade. The main goal has been to estimate the influence of the nonlinear impairments in a link. When using these estimation techniques, there is always a trade-off between accuracy and speed. The most accurate way of estimating the QoT is a full-fiber propagation simulation using the split-step Fourier method (SSFM) [4]. However, the SSFM involves high computational complexity and thus requires a high computation time, which renders this technique inapplicable for a real-time implementation.

Analytical QoT estimation tools, such as the Gaussian noise (GN) model [5], offer a low computation time and generally acceptable accuracy, although not being as accurate as the SSFM. The numerical integration to obtain the nonlinear interference has been most commonly used. It typically requires a computation time of a few minutes per wavelength division multiplexed (WDM) channel [3], accumulating to a few hours for full WDM systems consisting of several hundred channels. Extensions of the GN model like the incoherent GN (IGN) model or close-form approaches use approximations to reduce the computational effort with a reasonable accuracy penalty. The major downside of both simulative and analytical approaches is, however, that all link parameters have to be exactly known [6], [7], e.g. the span

Manuscript received 16 January 2023; revised 6 March 2023 and 18 April 2023; accepted 26 April 2023. Date of publication 2 May 2023; date of current version 19 September 2023. This work was supported in part by the CELTIC-NEXT Project AI-NET-PROTECT Project under Grant ID C2019/3-4 and in part by the German Federal Ministry of Education and Research under Grant 16KIS128. (*Corresponding author: Lars E. Kruse.*)

This work did not involve human subjects or animals in its research.

The authors are with the Chair of Communications, Kiel University, 24143 Kiel, Germany (e-mail: lars.kruse@tf.uni-kiel.de; sebastian.kuehl@tf.uni-kiel.de; annika.dochhan@tf.uni-kiel.de; stephan.pachnicke@tf.uni-kiel.de).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2023.3271860>.

Digital Object Identifier 10.1109/JLT.2023.3271860

lengths, fiber attenuation, chromatic dispersion, nonlinear coefficients, EDFA noise figures and non-ideal transmitter and the receiver characteristics. These parameter uncertainties generally lead to a less accurate QoT estimation [8].

A combination of both high accuracy and fast computation time is promised by the use of machine learning (ML) for QoT estimation purposes. An ML-based estimator is trained on certain input features which are correlated to the desired estimation target. This training dataset can be obtained either through full-fiber propagation simulations, i.e. SSFM, analytical models, e.g. GN-models, or experiments and field studies. However, it has to be noted that the ML-based estimator can not be more accurate than the underlying simulation, since the ML-algorithm approximates the target metric from the input features. The training of such an ML-algorithm can take between minutes up to several days depending on the training dataset size, the chosen ML-algorithm, and its dimensions. Once the algorithm is trained, however, an estimation takes only a fraction of a second making it applicable to a real-time environment.

Recently, different approaches were investigated including analytical [5], machine learning-based techniques [9], [10], and hybrid approaches [11], [12] to evaluate the performance of a certain lightpath in a system, based on different metrics. The choice of the estimator output, i.e. the performance metric, is the key enabler for responding proactively to performance degradations or potential failures in optical networks. The main QoT metric of interest to the network designer is the lightpath bit error rate (BER), which determines if the path is acceptable performance-wise or not [2]. Since forward error correction (FEC) is used in modern transmission systems the BER is usually expressed as the pre-FEC BER. However, the BER can only be obtained after the reception of the signal, which involves optical-electrical conversion and application of digital signal processing (DSP). By deploying optical spectrum analyzers (OSAs), it is possible to monitor the health of the optical system without demodulation permitting proactive maintenance and optimization of margins [13]. Measuring the optical signal-to-noise-ratio (OSNR) with the help of an OSA, however, enables network operators to validate the expected performance from the network planning stage without demodulation since the BER or Q-factor are closely connected to the OSNR. Nevertheless, measuring the OSNR of a WDM channel without interrupting the service would require dedicated hardware and algorithms. In addition to the nonlinearities from the optical link, the OSNR also includes impairments from the transceiver since the different parts of the noise term of the OSNR cannot be distinguished. This limitation leads to the definition of the generalized-OSNR (GOSNR) which is defined as the OSNR value at which the same BER is reached in the back-to-back transmission (backtrace method) [14]. Thus, the GOSNR captures only the optical impairments induced by the optical link including noise and nonlinear interference. Furthermore, the GOSNR is estimated for the destination node only due to the limitations of obtaining the OSNR within a dense WDM signal. In [15], we showed that including spectral features into an ML-based QoT estimator is beneficial for the estimation accuracy in exact component parameter agnostic networking scenarios. Furthermore, in [16] the estimator is shown to be generalizable towards feature

changes due to the heuristically distributed training features as well as being capable of properly reacting to (previously unseen) experimental data.

In this paper, we extend our work from [15] and [16] by comparing the robust ML-based QoT estimator to other ML algorithms, i.e. a traditional neural network, a support vector regressor (SVR) [17] with a radial bias function kernel, a decision tree regressor (CLF) [18], a XGradientBoost (XGB) regressor [19], and a one-dimensional convolutional neural network (CNN) [20]. On top of that, we set the topic into context and provide more information on the underlying ML algorithms as well as comparing the approaches with the new approach of using automated feature extraction from the spectrum by a variational autoencoder (VAE). We use extensive simulations based on the SSFM with heuristically varying input parameters based on realistic assumptions and margins to obtain a comprehensive data set for the training of the designed ML-algorithm. The experimental comparison of the ML-algorithms shows the superiority of spectral based estimators over to non-spectral estimators. Furthermore, the VAE shows a slightly better estimation performance than the long-short term memory (LSTM) neural network (NN) hybrid using manually selected features. The automated feature extraction and spectrum interpretation could pave the way towards fully automated optical networks and network management. In the experiments, we show that the VAE reaches a good estimation performance even on an OSA resolution of only 50 pm making it a good and cost-efficient solution for performance monitoring without demodulation of the channels.

The remainder of this paper is organized as follows: First, a brief overview of the theory of the variational autoencoder with respect to the proposed approach is given in Section II. Section III contains the description of the proposed QoT estimation approach including the simulation model and the design of the ML-based estimator. Furthermore, the estimator is compared to other frequently used ML-algorithms for QoT estimation in Section IV, and we validate the estimator trained entirely on simulation data by testing it on data obtained from experiments as well as investigating its performance depending on the resolution obtained from the OSAs. A conclusion will be drawn in Section V.

II. BACKGROUND

Traditional feed-forward neural networks (FF-NN) are only connected, as the name suggests, in the forward direction, from the input layer through each hidden layer to the output without cyclic connections. The absence of circles enables efficient learning with the backpropagation algorithm, but the fixed size of the layers is not well suited to processing data that is sequential in nature and variable in length. Sequential data can be processed with recurrent neural networks (RNNs) by inputting each symbol of the sequence individually and storing internal states between steps [21]. These networks can be trained with the backpropagation through time (BPTT) algorithm by unfolding the network, also allowing for recurrent/cyclic connections. However, this method often suffers from either vanishing or exploding gradients for long-term dependencies within the data.

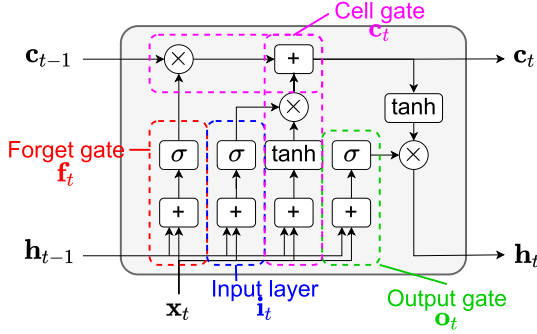


Fig. 1. Basic structure of an LSTM cell; \mathbf{x}_t : input at time t , \mathbf{h}_{t-1} : output of LSTM unit (last hidden state) at time $t-1$, \mathbf{c}_{t-1} : cell state at time $t-1$, \mathbf{c}_t : updated cell state, \mathbf{h}_t : output (new hidden state). Each gate has a certain bias value \mathbf{b} .

In this section, the theoretical concepts of the RNNs used for the QoT estimation are briefly explained. The models explained in this section both aim to overcome the vanishing gradient problem.

Feature engineering is an essential part of the preprocessing of data for the use in ML-algorithms. Especially the decision on, which features to use and the dimension reduction of the input data is of high interest. Both of these tasks can be solved using an autoencoder, which is briefly introduced in this section.

A. Long-Short Term Memory (LSTM)

Long-short term memory (LSTM) networks were first proposed by Hochreiter and Schmidhuber [22]. Since then, LSTMs are under the most popular and efficient methods for artificially understanding sequential dependencies of input features. The basic structure is depicted in Fig. 1. In an LSTM - compared to a traditional RNN - the structure of each layer is expanded to memory cells whose inputs and outputs are controlled by gates. These gates control the flow of information and preserve information from previous time steps [22]. An LSTM cell consists of input, forget, and output gates and a cell activation component [21]. The gates control the flow of information between the memory cells depending on previous inputs into the network. The different gates and their weights at the time step t are defined as [22]:

$$\mathbf{i}_t = \sigma(\mathbf{W}_I^i \mathbf{x}_t + \mathbf{W}_H^i \mathbf{h}_{t-1} + \mathbf{W}_A^i \mathbf{c}_{t-1} + \mathbf{b}^i), \quad (1)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_I^f \mathbf{x}_t + \mathbf{W}_H^f \mathbf{h}_{t-1} + \mathbf{W}_A^f \mathbf{c}_{t-1} + \mathbf{b}^f), \quad (2)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_I^o \mathbf{x}_t + \mathbf{W}_H^o \mathbf{h}_{t-1} + \mathbf{W}_A^o \mathbf{c}_{t-1} + \mathbf{b}^o), \quad (3)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_I^c \mathbf{x}_t + \mathbf{W}_H^c \mathbf{h}_{t-1} + \mathbf{b}^c), \quad (4)$$

$$\mathbf{c}_t = \mathbf{i}_t \tilde{\mathbf{c}}_t + \mathbf{f}_t \mathbf{c}_{t-1}, \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{c}_t), \quad (6)$$

where \mathbf{W}_I is the weight matrix from the input layer to the corresponding gate (i : input layer, f : forget gate, c : cell gate, o : output gate), \mathbf{W}_H denotes the weight matrix from hidden state to the corresponding gate, \mathbf{W}_A represents the weight matrix from cell activation to the corresponding gate, \mathbf{x} is the input vector,

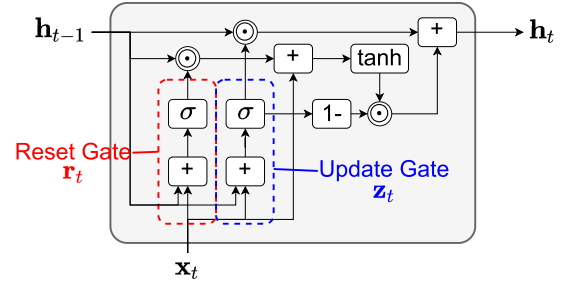


Fig. 2. Basic structure of a GRU cell; \mathbf{x}_t : input at time t , \mathbf{h}_{t-1} : output of GRU unit (last hidden state) at time $t-1$, \mathbf{h}_t : output (new hidden state) at time t . Each gate has a certain bias value \mathbf{b} .

\mathbf{h} denotes the output vector, $\tilde{\mathbf{c}}$ represents the candidate hidden state and \mathbf{b} is the bias of the corresponding gate. $\sigma(\cdot)$ is the activation function of the gate and $\tanh(\cdot)$ is the output activation function.

B. Gated Recurrent Unit (GRU)

While LSTMs tackle the problem of vanishing or exploding gradients, they require a high amount of memory due to multiple memory cells in the architecture. Similar to the LSTM unit, the GRU, which was first introduced by Cho et al. in 2014 [23], has gating units that modulate the flow of information within the unit, but without having separate memory cells. The basic structure of a GRU cell is depicted in Fig. 2. Unlike the LSTM, the GRU exposes the entire state at each time step by forming a linear sum between the existing state and the newly calculated state [21], [24]. In a similar manner to the LSTM gates equations, the updated GRU cells at each time step t are given as [24]:

$$\mathbf{z}_t = \sigma(\mathbf{W}^z \mathbf{x}_t + \mathbf{W}^z \mathbf{h}_{t-1} + \mathbf{b}^z), \quad (7)$$

$$\mathbf{r}_t = \sigma(\mathbf{W}^r \mathbf{x}_t + \mathbf{W}^r \mathbf{h}_{t-1} + \mathbf{b}^r), \quad (8)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}^h \mathbf{x}_t + \mathbf{W}^h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}^h), \quad (9)$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t, \quad (10)$$

where \mathbf{z} is the the update gate, \mathbf{r} denotes the reset gate, \mathbf{x} represents the input vector, \mathbf{h} is the output vector and \mathbf{W} and \mathbf{b} represent the weight matrix of the corresponding gate and the bias vector of the corresponding gate, respectively. As for the LSTM, $\sigma(\cdot)$ represents the activation function of the gate and $\tanh(\cdot)$ is the output activation function. In addition, the ' \odot ' denotes a element-wise product operation.

In [25], the performance of LSTMs and GRUs are compared. Several similarities and differences are presented, concluding that none of the models is inherently better than the other. Both models perform better than the other only on certain tasks. However, the GRU requires less memory than the LSTM in general.

C. Variational Autoencoder

Autoencoders (AEs) are an artificial neural network architecture consisting of an encoder $E: \mathcal{X} \rightarrow \mathcal{Z}$ and a decoder

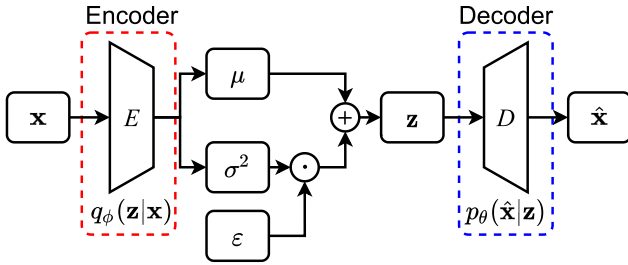


Fig. 3. Basic structure of a VAE; \mathbf{x} : input vector, \mathbf{z} : multivariate latent vector, $\hat{\mathbf{x}}$: reconstructed input vector; ε : sample of distribution.

network $D: \mathcal{Z} \rightarrow \mathcal{X}$ with $\mathcal{Z} \in \mathbb{R}^n$ jointly trained to reconstruct unlabeled data $\mathcal{X} \in \mathbb{R}^m$ distributed with an unknown probability distribution $P(\mathcal{X})$. By choosing a lower dimension $n < m$ with $n, m \in \mathbb{N}^+$ for the multivariate latent vector $\mathbf{z} = E(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, the encoder E is incentivised to learn an encoding of the input data \mathcal{X} that enables reconstruction of the data with the decoder $\hat{\mathbf{x}} = D(\mathbf{z})$, $\mathbf{z} \in \mathcal{Z}$. This concept enables applications such as dimensionality reduction of data by using \mathbf{z} instead of \mathbf{x} , denoising by using $\hat{\mathbf{x}}$ and anomaly detection by measuring the difference between \mathbf{x} and $\hat{\mathbf{x}}$.

Kingma and Welling [26] introduced Variational Autoencoders (VAEs) with a similar architecture to AEs, but with the important difference that the objective is to approximate the unknown distribution P of the data with a prior distribution p_θ parametrized with θ . In practice, the latent vector \mathbf{z} is assumed to be a multivariate Gaussian distribution, which, in addition to the applications of conventional AEs, allows the generation of new data by decoding samples from this distribution with the probabilistic decoder. Also, due to this, the VAE appears to be more generalizable than conventional AEs. Since the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$ is often intractable, it is approximated with a function $q_\phi(\mathbf{z}|\mathbf{x}) \approx p_\theta(\mathbf{z}|\mathbf{x})$ parametrized by the probabilistic encoder $E_\phi(\mathbf{x})$. The multivariate latent vector is calculated by

$$\mathbf{z} = \mu + \sigma^2 \odot \varepsilon, \quad (11)$$

where μ is the mean value, σ^2 is the standard deviation and ε is the sample of the distribution. The basic structure of an VAE can be seen in Fig. 3. The target during training is to find the optimal parameters θ and ϕ to reduce the reconstruction error at the decoder output while also maintaining the latent space probability distribution. By using the encoder of a well-trained autoencoder, the input dimension m is reduced to n ($n < m$) without losing much information. The latent space is thus a selection of reasonable features to describe the input. This can significantly reduce the effort of manual feature selection for the input of other ML-algorithms.

III. SPECTRAL DATA DRIVEN QOT ESTIMATOR

As explained in the introduction, exact component parameter agnostic network scenarios, as in a multi-vendor network, are challenging in the case of QoT estimation. This is because analytical estimators require accurate data on, for example, fiber lengths, attenuation, dispersion coefficients or EDFA noise values, etc. However, in a scenario where these or some of

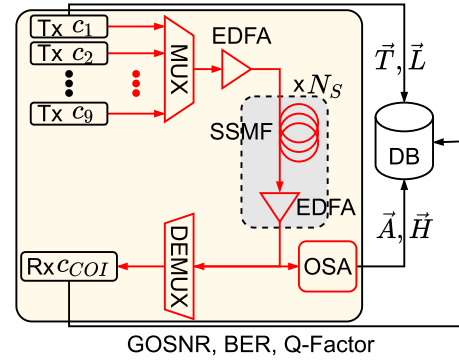


Fig. 4. Simulation setup with a central database (DB) containing the obtained feature vectors: Transmission-related feature vectors \vec{T} : transmission parameters (modulation format, launch power, channel spacing, baudrate) and \vec{L} : length vector (total link length, lengths between start, intermediate and end node). Spectrum-related feature vectors \vec{A} : Area under spectrum obtained by the OSA (i.e. total signal power) and \vec{H} : Heights of channels in spectrum (i.e. channel powers); Tx: Transmitter, EDFA: Erbium-doped fiber amplifier, SSMF: Standard single-mode fiber, OSA: Optical spectrum analyzer, Rx: Receiver, GOSNR: Generalized optical signal-to-noise ratio, BER: Bit error ratio.

these parameters are not precisely known, an analytical solution is more difficult [6]. The challenge arises from the variations in the assumed component parameters as well as the sparse monitoring data available at the intermediate nodes of a complex meshed network, where complete demodulation of the signals to obtain accurate telemetry data is not possible. For obtaining the optical spectrum with an optical spectrum analyzer (OSA), demodulation of the signal is not needed. Due to this, we focus on the usage of spectral data from the optical spectrum for the QoT estimation. Furthermore, the uncertainties of the component parameters are handled by machine learning because ML algorithms have shown to be very good at interpolation even of unseen data.

A. Simulation Setup

Data generation for neural network training is done using a simulation setup built in our Matlab based simulation tool as shown in Fig. 4. The simulation environment consists of the transmission link and a central database, where the recorded feature vectors and the obtained spectrum are stored. Up to 9 channels (c_1 to c_9) are transmitted over a coherent dual polarization (DP) WDM link with fixed channel spacing and equal launch powers per channel. The different links with up to 15 spans are analyzed for different configurations. The simulation parameters are summarized in Table I. Each span consists a standard single-mode fiber (SSMF), an EDFA, and an OSA. In an agnostic network, the exact component parameters are not precisely known, so for the simulations, the transmission parameters are calculated using a heuristic approach with a certain mean and standard deviation based on realistic assumptions and margins. Uncertainties are for example considered in the span lengths (L_S) by randomly choosing a length with a mean of 80 km and a standard deviation σ of 5 km. The varied, uncertain parameters are summarized in Table II. Therefore, the parameters are different for every span according to the

TABLE I
SIMULATION PARAMETERS

Parameter	Symbol	Value
Modulation format	MF	DP-QPSK, DP-8-QAM, DP-16-QAM
Baudrate	b	32, 64, 69 Gbaud
Channel spacing	Δf	37.5, 100 GHz
Launch power	P_L	-3 to 3 dBm
Center wavelength	λ_c	1550 nm
Number of spans	N_S	1 to 15
OSA resolution	R_{OSA}	13 pm
Channel assignment		Neighboring pairs to COI
SSFM waveplates		50 to 200 (random)
SSFM max. nonlinear rotation angle	$\varphi_{rot,max}$	0.5°

TABLE II
VARIED UNCERTAIN SIMULATION PARAMETERS

Parameter	Symbol	Mean Value	Standard Deviation
Span lengths	L_S	80 km	5 km
EDFA output power	P_E	P_L	0.5 dBm
EDFA noise figure	NF	5 dB	0.5 dB
Attenuation	α	0.2 dB/km	0.02 dB/km
Dispersion	D	17 ps/(nm·km)	0.2 ps/(nm·km)
Nonlinear coefficient	γ	$1.295 \text{ (W·km)}^{-1}$	0.05 (W·km)^{-1}

random distribution. The nonlinearities for the propagation of the signal through the fiber are calculated using the split-step Fourier method (SSFM) with a randomly chosen number of waveplates ranging between 50 and 200 per span and a PMD coefficient of $0.03 \text{ ps/km}^{1/2}$. The maximum nonlinear rotation angle in the nonlinear step is $\varphi_{rot,max} = 0.05^\circ$ and the step-size of the SSFM is chosen accordingly. The usage of the SSFM ensures exact modeling of the transmission and gives the opportunity to extract the spectrum which would not be possible with other (faster) simulation methods. To simulate more complex transmission scenarios, the number of spans is increasing with every iteration step of the simulation, i.e., one span is added for every step to a total of 15 spans. Every link can be interpreted in a different way: For example, if the link contains 4 spans, we represent a set of links with 0, 1, 2, and 3 intermediate nodes and their various possible distance variations for the distances to and from the intermediate node. This means, that there are 2^{N_i} different combinations for each link with N_i intermediate nodes considered [15]. At the receiver side, the generalized optical signal-to-noise ratio (GOSNR) is calculated by

$$GOSNR = \frac{P_{Rx}}{P_{ASE} + P_{NLI}}. \quad (12)$$

Here, the sum of the linear noise, i.e. P_{ASE} and the sum of the noise induced by nonlinearities, i.e. P_{NLI} is calculated over the deviations of the received constellation points to the ideal ones. The number of channels transmitted over the channel is varied from 1 to 9 where only neighboring pairs to the center channel are added or dropped. Furthermore, different scenarios for adding and dropping of channels at intermediate nodes are

considered within the simulation. Up to 4 neighboring pairs to the centered channel are dropped at all the intermediate nodes, while for the add scenario on the other hand these channels are added to the channel configurations in which their slots are free. Including only neighboring pair ensures worst-case scenarios, while keeping the simulation effort as low as possible.

The data extracted from the simulation is stored in a database. It can be categorized as transmission-related and spectrum-related features. In practice, it is assumed that an SDN controller has knowledge of the transmission-related features. The spectral features are extracted from the spectrum obtained by the OSAs. The overall feature structure can be seen in Fig. 5(a). The transmission-related features are composed out of the vectors \vec{T} and \vec{L} . \vec{T} includes the modulation formats, launch powers, channel spacing, symbol rates, and total link lengths, whereas \vec{L} contains the lengths of the fibers between the nodes. The spectral features are the vectors \vec{A} and \vec{H} . \vec{A} encloses the area under the envelope of the power spectral density (PSD) obtained by the OSA, i.e. the total signal power. The heights of the peaks in the PSD at the channel wavelengths, i.e. the channel powers, are included in the vector \vec{H} . Here the channel usage, the uncertainties of e.g. the EDFA gain or non-linear coefficient, and in general the influences of the non-linearities is represented. In addition, the spectra themselves are stored in the database to ensure later comparability with the variational autoencoder. Sweeping through the simulation parameters in Table I, analyzing the add and drop scenarios and including $N_i = 5$ intermediate nodes at maximum in the simulations, results in a dataset of $1.5 \cdot 10^6$ feature sets. To accommodate more uncertainties in the variation of the randomly distributed parameters from Table II, the simulation was repeated 10 times, which results in a total dataset size of $15 \cdot 10^6$ feature sets.

B. QoT Estimation Framework

Looking at the structures of the feature vectors \vec{L} and \vec{A} in Fig. 5(a), it can be seen that their size changes with the number of intermediate nodes in the considered link. The changing dimensions of the input vectors lead to another challenge in selecting a suitable ML algorithm. Most algorithms work with a fixed number of features, so we interpret each link as a set of values that can be fed into a recurrent neural network, e.g., an LSTM or GRU [15]. The generated dataset is used to train a QoT estimation framework based on LSTM and FF-NN layers. The overall structure of the framework is depicted in Fig. 5(b). First, the feature vectors are fed into the framework through an input layer. While the dimension-changing vectors are handled by LSTM-layers and a dense layer, meaning a fully-connected layer, the feature vectors, which are not changing in size due to the number of intermediate nodes, are directly handed to the concatenation layer. The concatenation layer combines the output of the RNN branch with the inputs of \vec{T} . This enables the interpretation of the output by feed-forward NN layers. After the FF-NN, the output is calculated resulting in an GOSNR value for every node in the considered link. In the case of automatic feature extraction, i.e. without manually selected features, the spectral features are replaced by the latent space of an autoencoder with

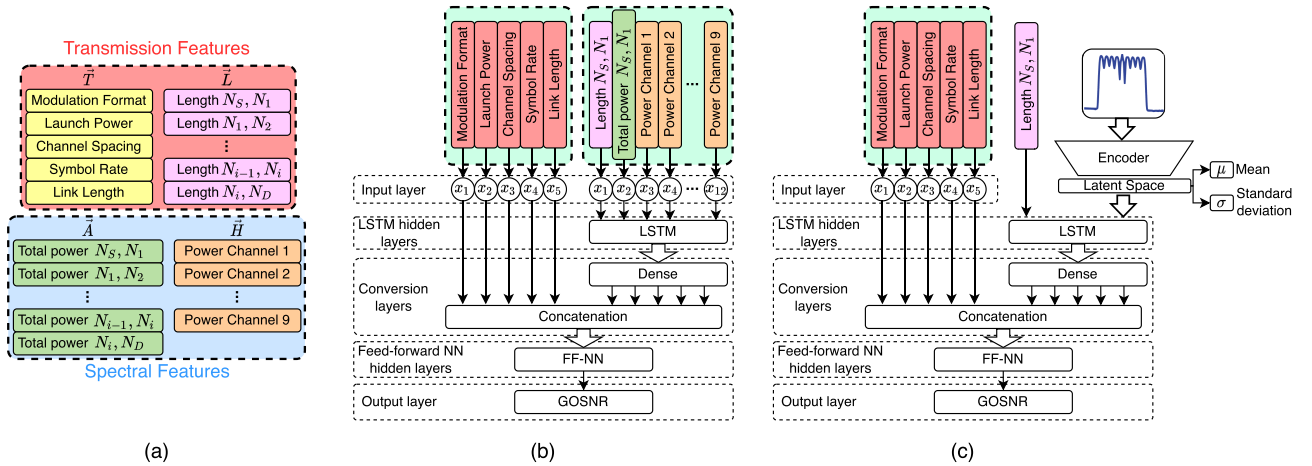


Fig. 5. Feature vectors and QoT estimation framework with used layer types; (a) Feature vector structure of \vec{T} , \vec{L} , \vec{A} and \vec{H} ; N_S : Start node, N_i : i-th node, N_D : destination node; (b) framework for manual selected features and (c) schematic view of the VAE-based feature extraction; LSTM: Long-short term memory; FF-NN: Feed-forward neural network; dense: fully-connected layer.

the same dimensions as the spectral features. So the latent space of the autoencoder is of size 12. The structure is shown in Fig. 5(c). However, the further structure of the framework is preserved to allow a comparison between manual and automatic feature selection. Before the training of the framework, the dataset is split into 60% training, 20% validation, and 20% test data. The training is performed over 800 epochs and is optimized using the Adam optimizer [27]. Choosing the optimal size of the layers is key for a high accuracy, but also a fast estimation. If the entire framework is sized too large, overfitting can occur. In addition, the required number of multiplications per neuron in the layers increases exponentially. Thus, the computational effort is much lower with a smaller size of the layers. Due to this, different sizes of the two hidden layers for two types of RNNs, i.e. GRU and LSTM, are investigated with regard to the mean absolute error (MAE). The results show, that an increase in the size of the structures does not necessarily lead to a significantly better performance with respect to the MAE. In general, the estimation accuracy of the LSTM framework is better than that of the GRU based one. Due to this, the QoT estimation framework is implemented using LSTM recurrent layers of size 24 and 12.

IV. EXPERIMENTAL INVESTIGATIONS

In general, simulations cover a wider range of complex network structures than experiments. Simulative investigations also help with choosing optimal ML algorithms and datasets. However, experimental data is essential to validate machine learning algorithms and their application in a real-world scenario to ensure a flawless usage of the trained ML algorithms in deployed networks. Furthermore, data gathered from experiments can be used to approximate real-world conditions in an optical transmission as well as enabling possible future usage of experimental data as training data for the machine learning algorithms. Here, we compare the GOSNR estimation performance on experimental data of different non-recurrent ML algorithms to the developed recurrent structures including manually and

automatically selected features from the spectrum. We further investigate the estimation performance of the VAE-based estimator with regards to the OSA resolution to make statements about the specifications of the OSAs needed. A lower resolution and low number of points for a good QoT estimation would mean that low-cost OSAs could be used at the intermediate nodes for the performance monitoring.

A. Experimental Setup

The high-level black-box model of the experimental setup is depicted in Fig. 6. The DSP is executed offline using MATLAB routines. At the transmitter side, a pseudo-random multilevel sequence (PRMS) of length $2^{17} - 1$ is generated for the channel of interest (COI) and mapped to QPSK, 8-QAM or 16-QAM symbols and the training symbols for equalization and synchronization are added. The signal is predistorted for the electrical amplifier and digital-to-analog converter (DAC) characteristics before it is up-sampled from the symbol rate of 32 GBd to the sampling rate of the DAC (88 GSa/s) and shaped using a root-raised cosine filter with a roll-off factor of 0.2, resulting in an almost rectangular spectrum. The digital-to-analog-conversion is performed by an arbitrary waveform generator (AWG) running at 88 GSa/s with an effective number of bits (ENOB) of 5.5 b. An external laser with a wavelength of 1550.004 nm in combination with a DP-IQ modulator that is driven by the DAC via 4 driver amplifiers generates the COI. The other WDM channels (loaders) are generated using a programmable wavelength-shaping filter (II-VI WS4000 A) with an amplified spontaneous emission (ASE) noise source as input. This results in shaped ASE noise which represents the WDM channels in the vicinity of the COI. The waveshaper has a periodically repeating filter bandwidth corresponding to the considered channel spacing and is configured to level all channels at the output. The advantage of using noise-loading over generating traditional channels arises from the lower complexity of the transmitter side by only using one modulator, one

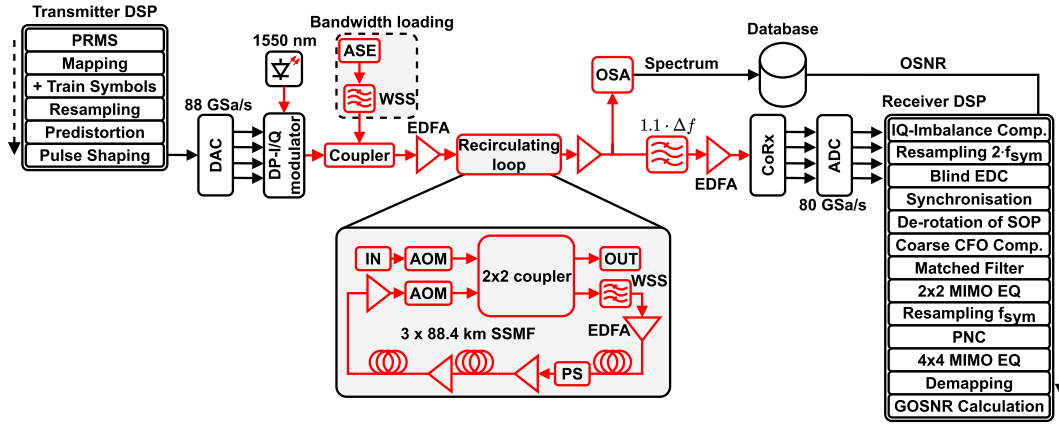


Fig. 6. Experimental transmission setup with a recirculating loop. PRMS: pseudo-random multilevel sequence, DAC: digital-to-analogue converter, ASE: amplified-spontaneous emission, WSS: wavelength selective switch, EDFA: Erbium-doped fiber amplifier, PS: polarization scrambler, EDC: electrical dispersion compensation, SOP: state of polarization, CFO: carrier frequency offset, PNC: phase-noise compensation.

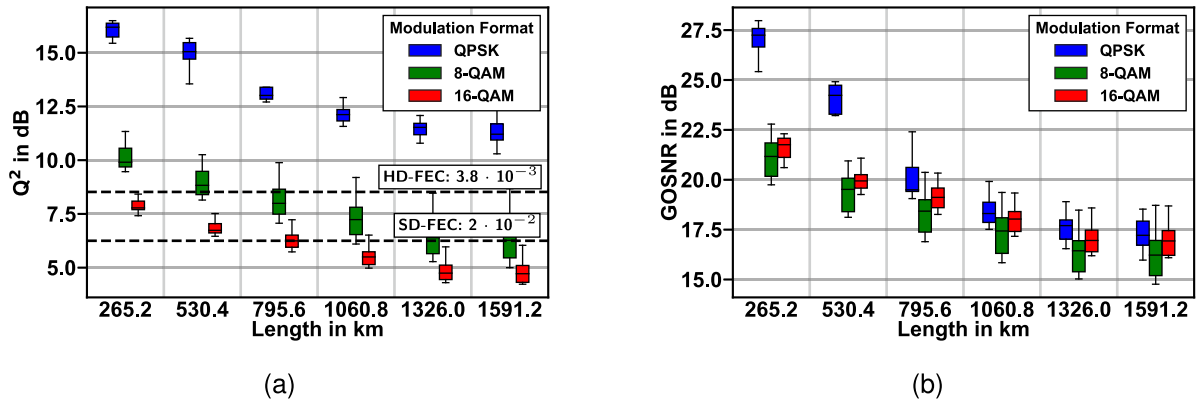


Fig. 7. Measurements of (a) Q-factor distribution over length; SD-FEC limit for 15% overhead according to OpenROADM specifications [28]; (b) GOSNR distribution over length for QPSK, 8-QAM and 16-QAM; HD-FEC: hard decision forward-error-correction; SD-FEC: soft decision forward-error-correction.

laser, and one DAC. The characteristics of a noise-loaded signal compared to a traditional WDM signal are very similar [29]. The COI and the loaders are combined using a 3 dB-coupler before being amplified using an EDFA. The EDFA output is fed into the recirculating loop. The loop is composed of another wave-shaper (Finisar WS4000S) being used as a gain-flattening filter followed by three spans and a polarization scrambler (Fig. 6). Each span consists of an EDFA running at a constant output power of 10.5 dBm, a VOA after the EDFA to get the desired launch power for the following 88.4 km SSMF. After the first span, the polarization scrambler is localized to randomize the polarization shift effects from the fibers. At the receiver side, the signal is first amplified using another EDFA before the COI is filtered. The COI is then detected using a coherent receiver. The analog-to-digital conversion (ADC) is performed by an oscilloscope with 80 GSa/s. The received signal is impaired by several disturbances which can be either uncompensated (mostly noise and nonlinearities) or compensated, i.e. IQ-skews and IQ-imbalance from the transmitter and receiver, laser phase noise from the transmitter and receiver, chromatic dispersion, polarization mode dispersion (PMD), rotation of the state of polarization (SOP), carrier frequency offset, and laser phase

TABLE III
EXPERIMENTAL SYSTEM PARAMETERS

Parameter	Symbol	Value
Modulation format	MF	DP-QPSK, DP-8-QAM, DP-16-QAM
Baudrate	b	32 Gbaud
Channel spacing	Δf	37.5 GHz
Number of channels	N_{Ch}	1, 3, 5
Launch power	P_L	-3, -2, -1, 0 dBm
Center wavelength	λ_c	1550.004 nm
Loop length	L_{Loop}	265.2 km
Loop iterations	N_{Loop}	1 to 6
OSA Resolution	R_{OSA}	10, 20, 50, 100, 200, 500 pm
OSA Points	N_{OSA}	501

noise from the receiver. The receiver DSP is done offline using standard DSP algorithms for coherent DP WDM systems [30]. At the end of the DSP chain, the GOSNR is calculated using pre-measured look-up tables of relations of OSNR and Q-factor for the considered configurations using the back-trace method. The spectrum is obtained using an OSA (Advantest Q8384)

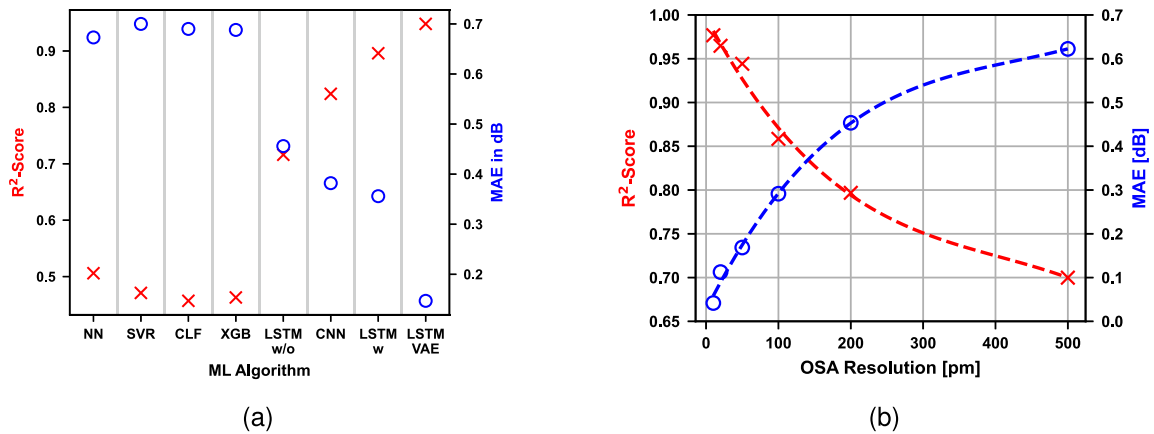


Fig. 8. (a) Estimation performance of the QoT estimators regarding R^2 -score and MAE; OSA resolution is 10 pm (b) R^2 -score and MAE over OSA resolution; MAE: Mean absolute error.

with a flexible optical resolution. The experimental parameters for the investigations are summarized in Table III.

B. Experimental Results

1) *Experimental Dataset*: The performance of the experimental setup is summarized in Fig. 7. The plot contains the minimum and maximum of the metrics for each modulation format for a certain length indicated by the whiskers. The values in the range between the first and third quartile are indicated by the boxes. The median is indicated by the horizontal line within the boxes. Fig. 7(a) shows the Q-factor distribution over length in the obtained dataset, which contains different modulation formats, number of channels and launch powers per channel. As expected, the lower-order modulation formats have higher Q-factors than higher-order modulation formats. Furthermore, a decrease of the Q-factor is visible for higher lengths. It has to be noted, that the hard decision forward-error-correction (HD-FEC) limit is always surpassed for QPSK for any length. For 8-QAM, only the median for 265.2 km and 530.4 km are above the HD-FEC limit, whereas for 16-QAM it is not reached for any length. This is mainly caused by non-optimal precompensation of the electrical amplifiers and amplifier noise after the AWG as well as shot noise and thermal noise of the coherent receiver. However, in an optical long-haul transmission system strong FEC algorithms are used anyways. So, the SD-FEC (soft decision FEC) limit [28] (15% overhead) is added to the graph. It can be seen that for 8-QAM the maximum reach is increased to be over 1326.0 km while for 16-QAM over 795.6 km transmission reach can be achieved. The experimentally obtained Q-factors are used for the back-trace to a Q-factor graph for a B2B configuration to obtain the GOSNR. The GOSNR distribution over the lengths is depicted in Fig. 7(b). In general, it can be observed that the curves characterized by the medians approach a certain GOSNR value, as expected. This is because the curves of the Q-factor OSNR become steep below a certain low OSNR value. A similar behavior could also be observed at high OSNR values since the Q-factor OSNR curve is flat there. Furthermore, it is noticeable that the modulation formats QPSK and 16-QAM have smaller boxes than 8-QAM and generally the medians of the 8-QAM

values are below those of 16-QAM. This is due to the fact that 8-QAM is more prone to nonlinearities from multi-channel transmissions reasoned by the non-equal symbol distances.

2) *ML-Algorithm Comparison*: All of the considered ML-algorithms are trained on simulation data and tested on the obtained experimental dataset. A comparison is made with different ML algorithms. For the algorithms, which do not include any spectral information, an FF-NN with two hidden layers and 40 neurons per layer, a support vector regressor (SVR) [17] with a radial bias function kernel, a decision tree regressor (CLF) [18], a XGradientBoost (XGB) regressor [19], and the above mentioned LSTM-based framework without the spectral features are compared. When comparing the algorithms with spectral features, a one-dimensional convolutional neural network (CNN) [20], the LSTM-based framework, and the proposed LSTM-based framework with feature extraction by the VAE are considered. All algorithms are trained with the simulation data and tested on the network topology. The hyperparameters of the non-spectral estimators are optimized using a grid search with 250 different configurations. The CNN is chosen to be one-dimensional since the size of the spectral data inputs changes according to the number of considered intermediate nodes. This enables to use a single CNN for all node configurations rather than implementing one CNN for each number of intermediate nodes. The LSTM structures are built as described in Section III-B. Furthermore, the OSA resolution is assumed to be 13 pm for the simulations. The results are summarized in Fig. 8(a). At first glance, it can be seen that the R^2 -scores for the non-spectral algorithms are low (below 0.5). On a closer look, it can be seen that the ML solutions based on neural networks achieve the best performance of the non-spectral estimators with over 0.5. Thus, the tree structures CLF and XGB perform worst followed by the support vector regressor (SVR). The simple FF-NN achieves the highest R^2 -score of the non-recursive ML algorithms with 0.51137. Due to the recursive structures and the inclusion of the individual lengths between the nodes, the framework without spectral features achieves a higher R^2 -score with 0.71563. This is only surpassed by the algorithms which use spectral features. The CNN shows an R^2 -score of 0.8238 and the presented LSTM framework achieves an R^2 -score of 0.8964 for the manually

selected features and 0.94826 for the features extracted by the VAE. An equal performance order is observed regarding the mean absolute error. These differences show that the usage of spectral information for the QoT estimation is beneficial. The LSTM framework without spectral feature inputs just learns the dependency between the lengths and the nonlinearities, whereas the spectral features obtained from the OSAs in the network reduce the error for the GOSNR estimation.

3) *OSA Resolution Investigation*: The developed QoT estimator using the spectral features extracted by the variational autoencoder is trained with simulation data and is then tested on the obtained experimental datasets with the different OSA resolutions. The performance of the estimator regarding the R^2 -score and the mean absolute error (MAE) is depicted in Fig. 8(b). First, it can be stated, that the estimator can reliably estimate GOSNR values from experimental data even though being trained on simulation data only. Second, a lower resolution results in a better estimation performance due to the estimator being trained on simulation spectra obtained with an OSA resolution of 13 pm. It can be seen, that the estimation performance is accurate with low errors up to a resolution of 50 pm.

V. CONCLUSION

In this paper, we compared the performance of different ML-algorithms for QoT estimation purposes when they are trained on simulation data and tested on experimental data. Furthermore, we investigated the influence of spectral data and recursive ML structures on the estimator's performance. The considered ML-algorithms include feed-forward neural network, support vector regressor, tree structures, such as XGradientBoost, a one-dimensional convolutional neural network, and long-short term memory networks. The LSTMs are either trained without or with spectral data while this spectral data can be either manually selected features or automatically extracted features by a variational autoencoder. These approaches were compared on experimental data acquired with a recirculating loop for 32 GBaud DP-QPSK, DP-8-QAM, and DP-16-QAM with up to 5 channels with 37.5 GHz spacing. The results show, that the algorithms leveraging spectral features perform very well on experimental data while being trained on simulation data surpassing R^2 -scores of over 0.9. On top of that, an estimation with a mean absolute error below 0.2 dB can be achieved with VAE-based spectral feature selection with only 50 pm of OSA resolution. We showed that heuristically distributed input features for the representation of not exactly known component parameters together with spectral features obtained from OSAs increase the QoT estimation accuracy. This enables reliable QoT estimation in e.g. multi-vendor networks on the road to fully-disaggregated networks without the need for confidential component data sharing.

REFERENCES

- [1] Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. A9–A17, 2017.
- [2] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, 2021.
- [3] J. Lu et al., "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B35–B44, 2021.
- [4] J. Shao, X. Liang, and S. Kumar, "Comparison of split-step fourier schemes for simulating fiber optic communication systems," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–15, Aug. 2014.
- [5] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.
- [6] V. Curri, "GNPy model of the physical layer for open and disaggregated optical networking," *J. Opt. Commun. Netw.*, vol. 14, no. 6, pp. C92–C104, 2022.
- [7] A. D'Amico et al., "Experimental validation of GNPy in a multi-vendor flex-grid flex-rate WDM optical transport scenario," *J. Opt. Commun. Netw.*, vol. 14, no. 3, pp. 79–88, 2022.
- [8] Q. Zhuge et al., "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," *J. Lightw. Technol.*, vol. 37, no. 13, pp. 3055–3063, Jul. 2019.
- [9] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A286–A297, 2018.
- [10] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of transmission estimation and short-term performance forecast of lightpaths," *J. Lightw. Technol.*, vol. 38, no. 10, pp. 2807–2814, May 2020.
- [11] I. Sartzetakis, K. K. Christodouloupoulos, and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," *J. Opt. Commun. Netw.*, vol. 11, no. 3, pp. 140–150, 2019.
- [12] 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.*, vol. 13, no. 6, pp. C21–C30, 2021.
- [13] I. Khan, M. Bilal, M. U. Masood, A. D'Amico, and V. Curri, "Lightpath QoT computation in optical networks assisted by transfer learning," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B72–B82, 2021.
- [14] H. J. Cho et al., "Constellation-based identification of linear and nonlinear OSNR using machine learning: A study of link-agnostic performance," *Opt. Exp.*, vol. 30, no. 2, pp. 2693–2710, 2022.
- [15] L. E. Kruse, S. Kühn, and S. Pachnicke, "Exact component parameter agnostic QoT estimation using spectral data-driven LSTM in optical networks," in *Proc. Opt. Fiber Commun. Conf.*, 2022, Paper Th1C–1.
- [16] L. E. Kruse, S. Kühn, and S. Pachnicke, "Generalizable QoT estimation based on spectral data driven LSTM in exact component parameter agnostic networks," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.
- [17] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
- [18] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees. belmont, ca: Wadsworth," *Int. Group*, vol. 432, no. 151-166, 1984, Art. no. 9.
- [19] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
- [20] H. Lun et al., "Soft failure identification for long-haul optical communication systems based on one-dimensional convolutional neural network," *J. Lightw. Technol.*, vol. 38, no. 11, pp. 2992–2999, Jun. 2020.
- [21] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*.
- [22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [23] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, Doha, Qatar, 2014, pp. 1724–1734.
- [24] K. Abdelli, H. Griebler, and S. Pachnicke, "A machine learning-based framework for predictive maintenance of semiconductor laser for optical communication," *J. Lightw. Technol.*, vol. 40, no. 14, pp. 4698–4708, Jul. 2022.
- [25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop Deep Learn.*, 2014.
- [26] D. P. Kingma et al., "An introduction to variational autoencoders," *Foundations Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019.
- [27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.
- [28] M. A. Sluysk, "Open ROADM MSA 3.01 W-port digital specification (200G-400G)," 2019. Accessed: Mar. 02, 2023. [Online]. Available: <http://openroadm.org/download.html>

- [29] T. Richter, J. Pan, and S. Tibuleac, "Comparison of WDM bandwidth loading using individual transponders, shaped, and flat ASE noise," in *Proc. IEEE Opt. Fiber Commun. Conf. Expo.*, 2018, pp. 1–3.
- [30] S. Ohlendorf, T. Wettlin, S. Pachnicke, and W. Rosenkranz, "Optimized flexible mappings with multidimensional modulation for coherent optical transport," in *Proc. IEEE 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.

Lars E. Kruse received the M.Sc. degree in electrical and information engineering in 2020 from Kiel University, Kiel, Germany, where he is currently working toward the Ph.D. degree with the Faculty of Engineering. Since 2020, he has been a Research and Teaching Assistant with the Chair of Communications, Kiel University. His research interests include predictive maintenance and optimization of optical networks with the help of machine learning and machine learning based equalization.

Sebastian Kühn received the M.Sc. degree in computer science from Bremen University, Bremen, Germany, in 2020, and is currently working toward the Ph.D. degree with the Chair of Communications, Kiel University, Kiel, Germany. The primary focus of his work is the development and evaluation of photonic reservoir computing architectures for signal equalization in fiber optical transmission systems. Past research was centered on the utilization of reinforcement learning for capacity optimization in elastic optical networks.

Annika Dochhan (Member, IEEE) received the Dipl. Ing. and Dr. Ing. degrees in electrical engineering and information technology from Kiel University, Kiel, Germany, in 2004 and 2013, respectively. From 2005 to 2012, she was a Research Assistant with the Chair of Communications, Kiel University. From 2012 to 2020, she has been with the Advanced Technology Group of ADVA Optical Networking SE, Meiningen, Germany, as a Principal Engineer responsible for the High-Speed Transmission Lab. From 2020 to 2022 she was with Dataport AÄR, an IT Service Provider and Network Operator for Public Administration in Altenholz, Germany. Since 2022, she has been a Research Staff Member with the Chair of Communications, Kiel University, where she is responsible for the Optical Communications Lab. She is involved in many numerical and experimental research activities covering data center networks and WDM systems.

Stephan Pachnicke (Senior Member, IEEE) received the M.Sc. degree in information engineering from City University, London, U.K., in 2001, the Dr.-Ing. degree in electrical engineering from TU Dortmund, Dortmund, Germany, in 2005, and the Dipl.-Wirt.-Ing. degree in business administration from Fern-Universität, Hagen, Germany, in 2005. From 2011 to 2015, he was with ADVA Optical Networking SE with the Advanced Technology Group (CTO Office). Since 2016, he has been a Full Professor and heading the Chair of Communications, Kiel University, Kiel, Germany. He is the author or co-author of more than 240 scientific publications, author of a book on *Fiber-Optic Transmission Networks* (Springer, 2011), and holds several patents. Dr. Pachnicke is currently serving on the technical program committees of the European Conference on Optical Communications (ECOC), the Optical Fiber Communication Conference (OFC), the Signal Processing in Optical Communications Conference (SPPCom), the Asia Communications and Photonics Conference (ACP) and the International Conference on Photonics in Switching and Computing (PSC). He is a Member of the European Academy of Sciences and Arts.