Experimental Comparisons between Machine Learning and Analytical Models for QoT Estimations in WDM Systems

Qirui Fan[1], Jianing Lu[2], Gai Zhou[1], Derek Zeng[1], Changjian Guo[2,5], Linyue Lu[2], Jianqiang Li[3], Chongjin Xie[4], Chao Lu[2], Faisal Nadeem Khan[1] and Alan Pak Tao Lau[1]

[1]*Photonics Research Centre, Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[2]*Photonics Research Centre, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[3]*Alibaba Group, Bellevue, WA, USA*
[4]*Alibaba Group, Sunnyvale, CA, USA*
[5]*South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China*
Email: remi.qr.fan@gmail.com

Abstract: We experimentally compare QoT estimations for WDM systems using Machine Learning(ML) and GN-based analytical models. ML estimates the side channels with better accuracy but is temporally less stable and less generalizable to different link configurations.

OCIS codes: (060.2330) Fiber Optics Communications;

Introduction

A key enabler to software defined network (SDN) is dynamic lightpath provisioning, baud rate and modulation adaptation based on quality of transmission (QoT) where the signal-to-noise ratio (SNR) of each channel is estimated. Accurate QoT estimation is under intense investigations recently to enable low-margin or margin-less networks for maximal network efficiency. In a WDM environment, cross-phase modulation (XPM) induced nonlinear distortions affect the neighboring channels and thus the SNR of one channel actually depends on the signal power of all other channels in a complicated manner, rendering QoT estimations for WDM systems a non-trivial task. In this connection, one type of proposed QoT estimation algorithms uses physics-based analytical models such as Gaussian Noise (GN) model for nonlinear distortions and linear accumulation of amplified spontaneous emission (ASE) noise [1][2]. Another line of approach uses machine learning (ML) techniques such as ANN, SVM and random forest algorithms [3]-[6]. However, physics-based analytical models suffer from parameters uncertainty in experimental settings while data-driven ML models are generally dependent on the specific link configurations and datasets. In this paper, we experimentally compare ML and analytical models for 7-channel WDM systems where both models parameters are optimized/trained from the same set of experimental data. It is shown that ML generally outperforms analytical models for the side channels while analytical models estimate the QoT of center channels with better accuracy. ML is also less stable with time and applying ML models to wider set of link configurations requires more data for re-training.

Experimental Setup

The experimental setup is shown in Fig. 1. In the data acquisition process, 2 independent 28GBaud 16-QAM digital waveforms shaped by raised cosine filter with 0.1 roll-off are generated offline and loaded into 84 GSa/s arbitrary waveform generator (AWG). The generated waveforms then drive the odd and even channels separately with up to 7 lasers. The 2 modulated tributaries are then interleaved and go into a polarization division multiplexing emulator before launching into a straight-line fiber link of length 915 km. The link consists of 12 spans of standard signal mode fibers (SSMF) with different lengths as shown Fig. 1, each is followed by a tunable attenuator-EDFA-4nm optical filter triplet to balance the launched power between each span.

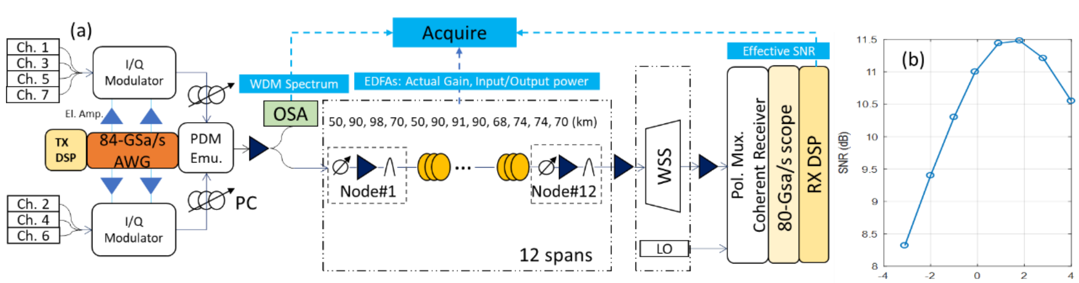

Fig. 1. (a) Experimental setup for comparing ML and analytical models for QoT estimation in WDM systems; WSS: wavelength selective switch; LO: local oscillator; PDM Emu: polarization division multiplexing emulator; (b) SNR vs. signal launched power in our study.

At the receiver, the channel of interest is filtered by a WSS and sampled by an 80 GSa/s oscilloscope. The sampled signals go through standard DSP blocks including CD compensation, constant modulus algorithm (CMA)+radiusdirected equalization (RDE) for polarization demultiplexing, 4th power algorithm for frequency offset estimation, extended Kalman filter for carrier phase estimation after which the effective SNR is calculated. The power of each laser is independently varied among the values $-4.1, -3.9, -3, -1.9, -1.4, -0.3, -0.2, 1.2, 1.3, 2.2, 3, 4.4$ dBm and we collect the mappings between effective SNR and system parameters $\boldsymbol{\theta}$, which consist of the laser launched powers P_n of the n^{th} channel, total power measurements at j^{th} amplifier outputs $\tilde{P}_{\mathrm{tot},j}$, amplifier gains G_j, noise figures NF_j and gain tilts T_j. A sample SNR vs. launched power relation for one of the channels is shown in Fig. 1(b), indicating that the data points cover the linear, quasi-linear and nonlinear regime.

Parameters Training and Optimization for ML and Analytical Models

We study feed-forward artificial neural networks (ANNs) with 1 hidden layer as the ML model where the set of parameters $\boldsymbol{\theta}$ described above are used as inputs to the ANN while the actual monitored received SNRs of various WDM channels are the ANN target vector as shown in Fig. 2 The training is performed offline in a supervised manner where we optimized various ANN *hyperparameters* such as number of hidden neurons, activation functions (ReLU or sigmoid) and learning rate for different scenarios. We consider two approaches for training and testing the ANN: (i) train and test a separate ANN model for each channel(referred to as *separate dataset*), and (ii) train and test a single ANN with multiple outputs for all the WDM channels (referred to as *combined dataset*). The training-testing ratio is 70:30 and 20% of the training data is used for validation.

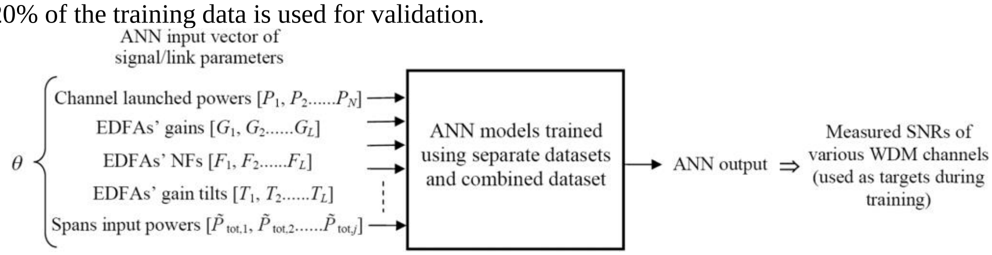

Fig. 2. ML models with different signal/link parameters as inputs and SNR estimates of various WDM channels as outputs. *N* is the total number of WDM channels while *L* is the total number of spans for a given scenario.

Next, for analytical QoT estimation models, we consider the overall SNR of a channel to be contributed by ASE noise, SPM- and XPM-induced nonlinear interference (NLI) as well as other transceiver noises σ^2_{TRx} that is analytically intractable so that the inverse of the SNR of the n^{th} channel is given by $SNR_n^{-1} = (\sigma^2_{\mathrm{TRx}} + \sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{GN}})/P_n$. We used the most recently published closed-form expression of the GN model [7] to calculate the nonlinear interference power σ^2_{GN}, which is highly flexible and can account for arbitrary launch power distributions and heterogeneous fiber spans. The ASE noise power and other undesired signal loss are included in the calculation of the noise figure NF_j of each EDFA [8] where $\sigma^2_{\mathrm{ASE}} = \sum_{j=1}^{L} NF_j(G_j - 1)h\nu B_{\mathrm{ref}}$ where L is the number of spans, h is Planck's constant, ν is the center frequency of the channel and B_{ref} is the noise reference bandwidth which is set to 0.1 nm in our study. Furthermore, we conducted back-to-back transmissions to obtain an initial estimate of σ^2_{TRx} [8]. To address the core issue of system parameters uncertainties in applying analytical models to experimental setups (e.g. uncertainties in attenuation, dispersion, and nonlinear coefficients of fibers, imprecise knowledge of actual fiber lengths; uncertainty of NF_j) and optimizing their QoT estimation performance, we performed a multi-dimensional sweep using all the recorded experimental datasets [9] and we apply them in all theoretical estimation processes. Like the ML counterpart, we consider two scenarios: (i) optimize the parameters using dataset for each channel (referred to as *separate dataset*), and (ii) obtain a single set of optimized parameters combining the datasets from all WDM channels (referred to as *combined dataset*).

Experimental Results

We study transmission of 7-WDM channels and monitor the SNR of channels 1,3,4,5 and 7. The QoT estimation results for ML and analytical models for separate dataset are shown in Fig. 3(a),(b) where we define σ_{SNR} as the standard deviation of the SNR estimation error in dB scale. For ML, the number of hidden neurons for 6- and 7-channel are optimized to 23 and 45, respectively, for training using separate dataset and is optimized to be 33 for combined dataset. For channel 1, σ_{SNR} is 0.2 and 0.13 dB for analytical and ML models, respectively. For channel 4, however, σ_{SNR} is 0.15 for ML model while analytical model gives a lower figure of 0.13 dB. It can be seen that the analytical model is better for middle channels while ML model is superior for outer channels. This can be understood by the fact that there are less neighboring channels to distort the outer channels through XPM so that the ANN output is less sensitive to the input parameter variation and thus the estimates are more accurate. In addition, most practical implementations of GN models assumes a flat NLI power spectral density (local white noise assumption) [7] which will over-estimate the nonlinear distortions for outermost channels and one can analytically show that it will lead to larger estimation errors [10]. For the center channel, the NLI estimate using the GN model is accurate while the relatively larger XPM effects degrade ML estimation accuracy, thus resulting in an overall tradeoff between analytical and ML models performance across different WDM channels. The results for combined dataset are shown in Fig. 3(c) and they are generally worse than their separate dataset counterpart as expected. In this case of combined dataset, ML generally outperforms the analytical model.

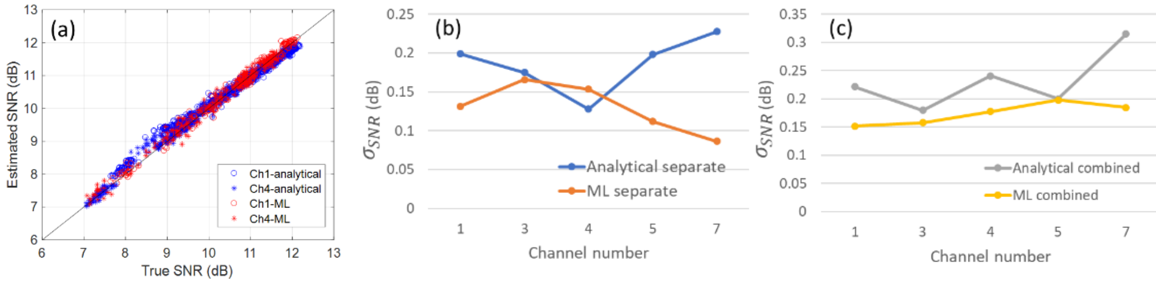

Fig. 3. (a) Estimated and True SNRs using analytical and ML models for channel 1 and 4 for 7-channel WDM transmission. The parameters are optimized using separate dataset;(b) standard deviation of SNR estimation error σ_{SNR} for separate dataset; (c) σ_{SNR} for combined dataset.

To investigate whether ML or GN-based analytical models are more generalizable to different link configurations and stable with time, we also obtained data for slightly varied link configurations including switching the positions of 4 inline EDFAs and shortening the transmission link by 1 or 2 spans. We investigated the additional training data required for the current ML and analytical models to adapt to the new configurations and the results for QoT estimation of channel 4 are shown in Fig. 4(a). Analytical models takes less than 15 data points from the new configurations to adapt while the ML model takes more than 200. Note that especially for the link shortening configuration, the analytical model virtually needs no adaption while inputting the parameters for the last span to the original ANN is not only irrelevant but also severely degrades estimation accuracy. Thus adapting the original ML model to the new configuration is often non-trivial. To further study the temporal stability of the models, we obtain 5 identical data sets for channel 4 which are taken 4 hours apart from each other. The first set is used to train both the ML and analytical models which are then tested on the other sets and the estimation errors are shown in Fig. 4(b). It can be seen that system temporal drifts due to temperature variations and other environmental factors play a bigger role in ML models.

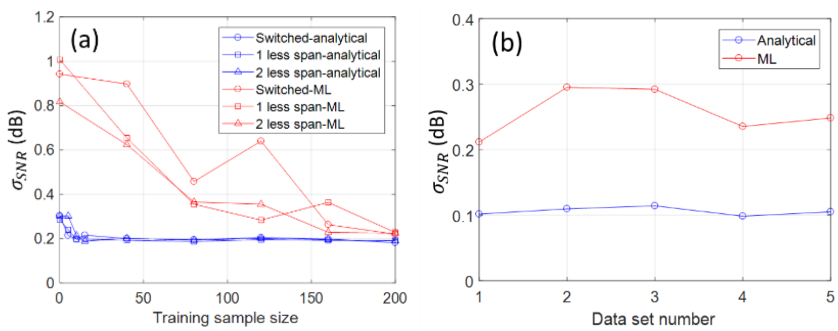

Fig. 4. (a) Adaptation of ML and analytical models to slightly different link configurations; (b) Temporal stability of ML and analytical models.

Conclusions

In this paper, we experimentally compared the performances of ML model and GN-based analytical model plus accumulated ASE noise for QoT estimation in WDM systems. Upon optimizing the model parameters with the same set of experimental data, it is shown that ML is better at estimating side channels and vice versa. However, it is more difficult for ML model to generalize to different link configurations and is less stable with time.

Acknowledgements

The authors acknowledge the support of the Alibaba Group through Alibaba Innovative Research Program, Hong Kong Government GRF PolyU 152757/16E and National Key R&D Program of China 2018YFB1801701.

References

[1] P. Poggiolini et al., *JLT*, vol. 32, no. 4, pp. 694-721, Feb 2014.
[2] M. Filer et al., *JLT*, vol. 36, no.15, pp. 3073-3082. Mar. 2018
[3] F.N. Khan et al., in *Optical Fiber Telecommunications*, 7th ed., Ch. 21, pp. 921–978, Academic Press, 2019.
[4] F.N. Khan et al., JLT, vol. 37, no. 2, pp. 493-516, Jan. 2019.
[5] R. M. Morais et al., *JOCN*, vol.10, no. 10, D84-D99, Oct. 2018.
[6] J. Yu et al. *JOCN* vol. 11, no.10, C48-C57, Aug. 2019
[7] D. Semrau et al., *JLT*, vol. 36, no. 14, pp. 3046-3055, Jul. 2018
[8] E. Torrengo et al., *OE*, vol. 19, pp. 790–798, Dec. 2011.
[9] Q. Zhuge et al., *JLT*, vol. 37, no. 13, pp. 3055-3063, Jul. 2019.
[10] J. Lu et al., in preparation for submission to *JLT*.