European Conference on Optical Communication (ECOC) 2022 © Optica Publishing Group 2022

Automated Dataset Generation for QoT Estimation in Coherent Optical Communication Systems

Caio Santos, Behnam Shariati, Robert Emmerich, Carsten Schmidt-Langhorst, Colja Schubert, and Johannes K. Fischer

Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany, (Email address of corresponding author: caio.santos@hhi.fraunhofer.de)

Abstract *We demonstrate sophisticated laboratory automation and data pipeline capable of generating large, diverse, and high-quality public datasets. The demo covers the full workflow from setup reconfiguration to data monitoring and storage, represented on a digital replica of the setup and updated in near real-time. ©2022 The Author(s)*

Overview

Current optical communication networks operate with a significant level of beginning-of-life (BoL) design margin to assure undisrupted operation during their lifetime. These margins should account for network loading as well as aging of equipment, which usually results in over-dimensioning of components. This increases the capital expenditures (CAPEX) and operating expenses (OPEX) of the operator's networks, significantly reducing their revenue. Therefore, it is of crucial importance to devise solutions to minimize margins in optical networks. Such optimization is achievable by accurately estimating the quality of transmission (QoT) of all links within the network, leading to a minimization of the difference between designed BoL margins and actual network operating values. The increase in accuracy of such estimator leads to lower margins, enabling the usage of more efficient modulation formats in the network [1], [2]. Two major requirements for improving the estimator's accuracy are an accurate QoT model and high-quality inputs [3]. The former requires efficient modeling techniques, while the latter calls for extensive high-quality dataset generation.

The ability of Machine Learning (ML) algorithms to learn system behavior from past data makes it a highly suitable technique for dealing with the complexity of optical networks, especially when dealing with QoT estimation. Previous studies have proven that ML-based QoT estimation yield reliable accuracy as reviewed and compared in [4]. Such data-driven models are even more attractive when considering the current monitoring capabilities of optical networks. However, the public availability of data is still quite scarce within the research community, meaning that most ML-based works rely on synthetic data, usually generated in-house through simulation, limited laboratory experiments or, rarely, real networks [5]. The first steps towards high-quality public datasets were taken only recently by HHI with public QoT datasets generated from data of different network simulation scenarios in [6] and by Microsoft with a published dataset from a wide-area backbone network in [7].

The challenge in obtaining these high-quality datasets stems, mainly, from the wide range of parameters present in dynamic wavelength-division-multiplexing (WDM) link setups. Altering and/or sweeping these parameters over multiple successive experiments leads to an extremely time-consuming process, which constitutes an unfeasible process for manual execution. However, through extensive laboratory automation, multi-configuration experimental setups become feasible, enabling large measurement campaigns while minimizing human intervention [2], [8], [9].

In this work, we demonstrate extensive laboratory automation applied to a digital coherent optical communication testbed with offline digital signal processing (DSP). The automation includes comprehensive system monitoring and reconfiguration, as well as an efficient data pipeline for dataset generation from the measured performance and telemetry data. The primary use-case for the dataset is QoT estimation, but the generated dataset could be useful for a variety of applications (e.g. link identification, failure prediction and localization, optical spectrum classification). A live stream of the laboratory in conjunction with a visual dashboard offer a clear picture of the automation while telemetry data is collected from the hardware and performance data is obtained from the offline DSP. We believe that our demonstration will pave the way and incentivize the future development of public, large, diverse and well-documented datasets for the optical networking community.

Innovation

We aim to showcase the full pipeline of autonomous experimental dataset generation in coherent optical communication systems. A dashboard will demonstrate all equipment information obtained at each experimental execution, while the laboratory automation changes a set of multiple parameters of the system with each iteration. In the background, all measured data is stored for dataset generation.

The key contributions of the work are:

- *System Automation and Auto-Reconfiguration*: Automation scripts interact with each other, dynamically setting different experimental characteristics and setups without human interaction. This allows automatic reconfiguration of the transmission system into various transmitter/receiver states as well as link configurations, while monitoring and storing relevant data for extensive periods. In the background, numerous alarms, fail-safes and recovery systems are ready to act in case of possible system/equipment failures, guaranteeing the safety of the automated experimental setup against unexpected crashes;
- *Visualization of Setup Parameters*: A dashboard demonstrating the state of the system in the vast space of different configurations. In this sense, a visual representation of the system is achieved containing multiple physical layer parameters, such as optical frequency and power of the channel under test (CUT), amplifier input and output powers, optical filter frequency and bandwidth, receiver input power, optical spectrum at multiple positions of the setup, distance and spans overcome by the signal. By displaying and collecting such a range of information, an in-depth knowledge of the physical layer is recorded at each measurement point. Other parameters are also exposed, such as DSP results and QoT indicators, giving a major overview of the system's performance during each iteration;
- *Dataset Generation for QoT Estimation*: All visualized data is also stored for post-processing and publishing, yielding the development of large, diverse, and well-documented datasets for the optical networking community.

Demonstration Procedure

In order to observe the described process, the dashboard depicted in Fig. 1 presents multiple characteristics, measurements and results at every iteration of the reconfiguration process of the laboratory setup in Fig. 2. It allows for a live visualization of the laboratory automation and dataset generation, showing setup auto-reconfiguration, parameter adjustments for such configurations, physical layer measurements, DSP results, QoT indicators and other relevant attributes and features of the channel under test (CUT). All the parameters and measurements are later processed in the data pipeline to create the datasets, which will also be visualized using HHI's interactive visual analytics dashboard [10].

The top part of Fig. 1 shows a digital representation of one of the experimental setups, monitoring the most relevant physical layer parameters and updating them at each iteration to enlighten the user regarding the status. The setup is a multi-channel system with an optical recirculation loop. The interfering channels result from amplified spontaneous emission (ASE) generated through amplifiers and filtered by the wavelength selective switch (WSS).

The blue boxes are each associated to an equipment and contain key measurements from each of them. The red and pink marker/lines (Fig. 1) illustrate the connections between power taps in the setup and the optical switch. Through the switch, these taps connect to the optical spectrum analyzer (OSA), to the terminal receiver or to any other monitoring/measuring device connected to the optical switch, providing flexibility within the setup for spectral measurements and different receiver configurations within the system. As an example, Fig. 1 shows the system configuration when the coherent receiver is connected to the erbium-doped fiber amplifier (EDFA) at the transmitter output (EDFA #9) and the OSA is connected to the pre-amplifier output (EDFA #11) to measure the OSNR.

Other panels within Fig.1 demonstrate more visual measurements such as the optical spectrum and constellation diagrams of the CUT. The former allows for measurement of the OSNR by extracting the waveform directly from the OSA. The latter displays a representation of the received signal through a constellation diagram of the signal after the equalizer. This exhibits the quality of the received signal in a perceptible manner.

Still within the dashboard, three boxes contain detailed information concerning the measurements. The central box shows system information (e.g., modulation format and order, symbol rate and CUT frequency/wavelength) and key QoT indicators of the setup, such as BER, error vector magnitude (EVM), OSNR and the end-to-end signal-to-noise ratio (SNR). The two other boxes exhibit physical layer measurements and DSP parameters. The former shows data obtained from components in the terminal transmitter, optical line system and terminal receiver.

In the background, the experimental data passes through a dataset generation pipeline, labeling, storing and organizing the data. All meas-

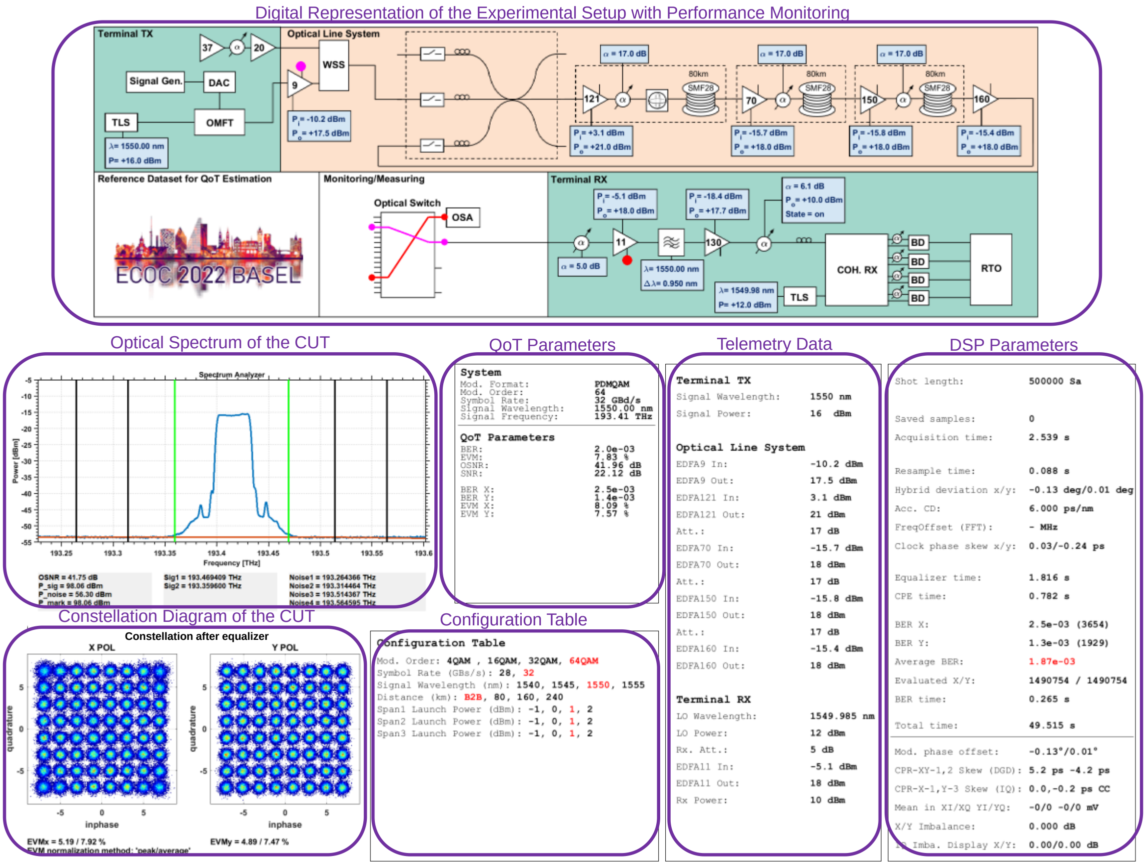

Fig. 1: A screenshot of the dashboard used for visualization of setup parameters. The configuration table shows parameters that are changing dynamically as the system reconfigures itself.

Fig. 2: Experimental setup containing the terminal transmitter (TX), optical line system, terminal receiver (RX) and monitoring/measuring for telemetry.

ured data is stored in conjunction with parameters describing component status and configuration. This indicates the exact state of every physical layer equipment during the iteration, presenting a true picture of the experimental setup. As the setup relies on offline DSP, multiple waveforms are stored in order to increase the captured statistical variation of the received data signal.

As an input to the automation, a configuration table defines a list of relevant parameters and system changes. The only other occasion that requires human intervention is in the case of a critical failure, where alarms indicate to the responsible person that an action is required. Minimal human intervention classifies this work as a powerful tool for the generation of extensive high-quality dataset.

ECOC Relevance

This work demonstrates extensive laboratory automation, facilitating diverse dataset generation through an efficient data pipeline. Such high-quality datasets allow researchers and solution providers to lower design margins, while also enabling transparent development, comparison and benchmarking of ML-based QoT estimation models through common sets of data.

Acknowledgements

The authors thank Tobias Fehenberger from ADVA Optical Networking for the fruitful discussions. This work was funded by the EU H2020 under the MSCA-ETN WON (grant no. 814276) and the German Ministry of Education and Research (BMBF) in the framework of the project AI-NET PROTECT (KIS8CEL010, FKZ 16KIS1282).

References

[1] S. J. Savory, R. J. Vincent and D. J. Ives, "Design considerations for low-margin elastic optical networks in the nonlinear regime," in Journal of Optical Communications and Networking, vol. 11, no. 10, pp. C76-C85, October 2019, DOI: 10.1364/JOCN.11.000C76

[2] J. Lu, Q. Fan, G. Zhou, L. Lu, C. Yu, A. Lau and C. Lu, "Automated training dataset collection system design for machine learning application in optical networks: an example of quality of transmission estimation," in Journal of Optical Communications and Networking, vol. 13, no. 11, pp. 289-300, November 2021, DOI: 10.1364/JOCN.431780

[3] Y. Pointurier, "Design of low-margin optical networks," in Journal of Optical Communications and Networking, vol. 9, no. 1, pp. A9-A17, Jan. 2017, DOI: 10.1364/JOCN.9.0000A9

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

[5] G. Bergk, B. Shariati, P. Safari and J. K. Fischer, "ML-assisted QoT estimation: a dataset collection and data visualization for dataset quality evaluation," in Journal of Optical Communications and Networking, vol. 14, no. 3, pp. 43-55, March 2022, DOI: 10.1364/JOCN.442733

[6] G. Bergk, et al., "QoT Dataset Collection, " [online – accessed Mar 2022] (https://www.hhi.fraunhofer.de/networkdata)

[7] Microsoft, "Wide-Area Optical Backbone Performance" [online – accessed Mar 2022] (https://www.microsoft.com/en-us/research/project/microsofts-wide-area-optical-backbone/)

[8] M. Bouda, S. Oda, O. Vassilieva, M. Miyabe, S. Yoshida, T. Katagiri, Y. Aoki, T. Hoshida and T. Ikeuchi, "Accurate prediction of quality of transmission based on a dynamically configurable optical impairment model," in Journal of Optical Communications and Networking, vol. 10, no. 1, pp. A102-A109, Jan. 2018, DOI: 10.1364/JOCN.10.00A102

[9] J. Yu, W. Mo, Y. Huang, E. Ip and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," in Journal of Optical Communications and Networking, vol. 11, no. 10, pp. C48-C57, October 2019, DOI: 10.1364/JOCN.11.000C48

[10] B. Shariati, W. Baltzer, G. Bergk, P. Safari and J. K. Fischer, "Interactive Visual Analytics Dashboard for the Paradigm of ML-assisted Autonomous Optical Networking," 2022 Optical Fiber Communications Conference and Exhibition (OFC), 2022, pp. 1-3.