# QUALITY OF TRANSMISSION MODELING AND ESTIMATION FOR LIVE COMMERCIAL OPTICAL FIBER NETWORKS

**YAN HE**

**PhD**

**The Hong Kong Polytechnic University**

**2026**

The Hong Kong Polytechnic University

Department of Electrical and Electronic Engineering

# Quality of Transmission Modeling and Estimation for Live Commercial Optical Fiber Networks

Yan He

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

August 2025

# Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

______________________ (Signed)

______Yan He______ (Name of student)

# Quality of Transmission Modeling and Estimation for Live Commercial Optical Fiber Networks


## Abstract

Optical fiber networks form the critical backbone infrastructure for high-speed, high-bandwidth transmission supporting modern AI and Internet applications. Recently, open and disaggregated optical networks have gained popularity, transitioning from legacy single-vendor closed systems to multi-vendor disaggregated systems. This architecture allows network operators to choose best-in-class devices, enabling customized network functionality and management planes. However, its multi-vendor characteristics make physical layer modeling much harder due to the accuracies of system parameter characterizations and the implementation principle of the devices can vary with different vendors. To design and manage disaggregated networks efficiently, accurate knowledge of physical layer impairments (PLI), or Quality of Transmission (QoT), on each lightpath is essential. Given that inaccurate input parameters still degrade the accuracy of the QoT estimation, in this regard, this thesis proposed novel methods to improve the accuracy of QoT estimation for different types of commercial networks with disaggregated architecture.

This thesis first studied a large-scale long-haul optical network and proposed to refine the signal power measurement and amplifier gain profile used in the QoT model. This method was further extended by adding simultaneous learning of EDFA NFs, additional biases for different OMS and frequencies. The monitored data has a period of 8 months and represents the largest-scale disaggregated network study for QoT estimation with 1,478,354 data samples. The std of estimation error using the proposed techniques can be reduced from 0.6337 to 0.2377dB. In addition to long-haul, this thesis also studied metro data center interconnect (DCI) networks where the data-rate of TRX varies from 200Gbps to 800Gbps and are more disaggregated with multiple vendors. We first trained an ANN model to learn inherent transponder effects for accurate QoT estimations. Besides, this thesis also proposed input refinement method for analytical SNR model with a specific focus on one type of 400Gb/s TRx. The refinement method was based on experimentally characterizing the input-dependent TRx SNR penalty and input-dependent NFs of multi-vendor multi-type EDFAs. The RMSE of QoT estimation using our method can be reduced from 0.662dB to 0.287dB. This thesis also proposed accurate polynomial-based amplifier gain profile characterization model for multi-vendor multi-type commercial EDFAs. The accuracy of greenfield QoT estimation enabled by the power prediction and EDFA gain profile model is comparable to that of the brownfield network, demonstrating that even at the pre-deployment stage of the network, reliable QoT estimation can be achieved. Finally, this thesis

provided an open-source dataset extracted from a commercial live network where the optical power, bit error ratio (BER) and etc., are made publicly available on GitHub, providing a benchmark for comparing different QoT models.

In summary, this thesis presented improved brownfield and greenfield QoT estimations that enable accurate and fast assessment of signal impairments applicable in different commercial network architectures. These models and methods facilitate automated network design, planning and operation, enabling operators to optimize capacity utilization, reduce unnecessary overprovisioning and avoid manual interventions or rework.

# Publications


[1] **He, Yan***, Zhiqun Zhai, Liang Dou, Lingling Wang, Yaxi Yan, Chongjin Xie, Chao Lu, and Alan Pak Tao Lau. "Improved QoT estimations through refined signal power measurements and data-driven parameter optimizations in a disaggregated and partially loaded live production network." *Journal of Optical Communications and Networking* 15, no. 9 (2023): 638-648.

[2] Zhai, Zhiqun, Liang Dou, **Yan He**, Alan Pak Tao Lau, and Chongjin Xie. "Open-source data for QoT estimation in optical networks from Alibaba." *Journal of Optical Communications and Networking* 16, no. 1 (2023): 1-3.

[3] **He, Yan***, Sai Chen, Liang Dou, Zhiqun Zhai, Huan Zhang, Yuanchao Su, and Alan Pak Tao Lau. "Accelerating Hollow-Core Fiber Deployment: An Efficient One-Shot Field Measurement Technique to Characterize $CO_2$ Absorption-induced OSNR Penalty." In *Optical Fiber Communication Conference*, Optica Publishing Group, 2026. (Accepted)

[4] **He, Yan***, Zhiqun Zhai, Sai Chen, Huan Zhang, Chuang Xu, Liang Dou, Yuanchao Su, Chao Lu, and Alan Pak Tao Lau. "Transceiver Penalty and Amplifier Noise Figure Characterization for Accurate QoT Estimation in Hyperscale Disaggregated DCI Networks." In *Optical Fiber Communication Conference*, pp. M1J-6. Optica Publishing Group, 2025.

[5] **He, Yan***, Kausthubh Chandramouli, Zhiqun Zhai, Sai Chen, Liang Dou, Chongjin Xie, Chao Lu, and Alan Pak Tao Lau. "QoT estimation for large-scale mixed-rate disaggregated metro DCI networks by artificial neural networks." In *Optical Fiber Communication Conference*, pp. W3G-2. Optica Publishing Group, 2024.

[6] **He, Yan***, Zhiqun Zhai, Lingling Wang, Yaxi Yan, Liang Dou, Chongjin Xie, Chao Lu, and Alan Pak Tao Lau. "Improved QoT estimations through refined signal power measurements in a disaggregated and partially-loaded live production network." In *Optical Fiber Communication Conference*, pp. Tu2F-5. Optica Publishing Group, 2023.

[7] Chen, Sai, **Yan He**, Liang Dou, Alan Pak Tao Lau, Huan Zhang, Zhiqun Zhai, and Yuanchao Su. "Characteristics and Impacts of $CO_2$ Absorption Effects in Hollow Core Fiber (HCF) Transmission Systems," 2025 European Conference on Optical Communications (ECOC), Copenhagen, Denmark, 2025, pp. 1-4.

[8] Zhai, Zhiqun, Liang Dou, Sai Chen, Huan Zhang, **Yan He**, Alan Pak Tao Lau, and Yuanchao Su. "Improving QoT Estimation Accuracy in Production Networks: A Data-Driven Approach to Address OLS Imperfections." In *Optical Fiber Communication Conference*, pp. Tu3I-4. Optica Publishing Group, 2025.

[9] Yan, Yaxi, Liwang Lu, Xiong Wu, Jingchuan Wang, **Yan He**, Daru Chen, Chao Lu, and Alan Pak Tao Lau. "Simultaneous communications and vibration sensing over a single 100-km deployed fiber link by fiber interferometry." In *Optical Fiber Communication Conference*, pp. W1J-4. Optica Publishing Group, 2023.

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Alan Pak Tao Lau, for his invaluable guidance, professionalism, support, and kindness throughout my entire PhD journey. I am sincerely grateful for the opportunity to join his group and for his trust in allowing me to collaborate with the outstanding colleagues from both our Photonics Research Institute and Alibaba Cloud.

My heartfelt thanks go to Dr. Liang Dou, Dr. Sai Chen, Ms. Zhiqun Zhai, Mr. Huan Zhang, and Dr. Chongjin Xie from Alibaba Cloud. I truly appreciate their patience, encouragement, and mentorship during my research internship. Over the past three years, they have consistently supported me, greatly enhancing my understanding of the industry and the nature of practical, impactful research.

Special thanks to Dr. Qirui Fan, Dr. Chuang Xu, Dr. Zhuili Huang and Dr. Xiong Wu for their ongoing support, patience, and kindness—especially when I first joined the group four years ago. Their warm welcome and generous teaching played a crucial role in my growth as a researcher.

I would also like to thank Dr. Yaxi Yan, Jingchuan Wang, Liwang Lu, and all my friends at the Photonics Research Institute for their camaraderie and encouragement.

I am deeply grateful to my family—my mother, father, grandparents, and paternal aunt—for their unwavering love and support throughout my life. Your love has been my constant source of strength.

Last but not least, I would like to acknowledge myself for nurturing self-love, perseverance, and resilience over the past 27 years and for continuing to do so in the years ahead.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **ADC** | Analog-to-digital converter/conversion |
| **AGC** | Automatic Gain Control |
| **AI** | Artificial Intelligence |
| **AL** | Active learning |
| **ANN** | Artificial neural network |
| **APC** | Automatic Power Control |
| **APS** | Automatic protection switch |
| **ASE** | Amplified spontaneous emission |
| **AWGN** | Additive white Gaussian noise |
| **B2B** | Back-to-back |
| **BA** | Booster amplifier |
| **BER** | Bit error ratio |
| **BoL** | Beginning of life |
| **CapEx** | Capital expenditure |
| **CD** | Chromatic dispersion |
| **CM** | Center of mass |
| **CNN** | Convolutional neural network |
| **COI** | Channel of interest |
| **CUT** | Channel under test |
| **D-margin** | Design margin |
| **DAC** | Digital-to-analog converter/conversion |
| **DB** | Database |
| **DC** | Direct current |
| **DCI** | Data center interconnection |
| **DCN** | Data center network |
| **DGD** | Differential group delay |
| **DGT** | Dynamic gain tilt |
| **DP** | Dual-polarization |
| **DSP** | Digital signal processing |
| **EDF** | Erbium-doped fiber |
| **EDFA** | Erbium-doped fiber amplifier |
| **EoL** | End of life |
| **FEC** | Forward error correction |

| | |
|---|---|
| **flex-TRX** | Rate-flexible transponder |
| **FWM** | Four-wave mixing |
| **GFF** | Gain flattening filter |
| **GN** | Gaussian noise |
| **GP** | Gaussian process |
| **GSNR** | Generalized signal-to-noise ratio |
| **GVD** | Group velocity dispersion |
| **HCF** | Hollow core fiber |
| **IL** | Insertion loss |
| **ISI** | Inter-symbol interference |
| **ISRS** | Inter-channel stimulated Raman scattering |
| **LO** | Local oscillator |
| **LPN** | Laser phase noise |
| **ML** | Machine learning |
| **MLP** | Multi-layer perception |
| **MSE** | Mean squared error |
| **NF** | Noise figure |
| **NLI** | Nonlinear interference |
| **NLSE** | Nonlinear Schrödinger equation |
| **NN** | Neural network |
| **OCH** | Optical channel |
| **OCM** | Optical channel monitoring |
| **OLA** | Optical line amplifier |
| **OLE** | Optical line equalizer |
| **OLP** | Optical line protection |
| **OLR** | Optical line router |
| **OMS** | Optical multiplex section |
| **OMSP** | Optical multiplex section protection |
| **OSNR** | Optical signal to noise ratio |
| **OTDR** | Optical time-domain reflectometer |
| **OTN** | Optical transport network |
| **OTS** | Optical transport section |
| **OXC** | Optical cross-connect |
| **PBC** | Polarization beam combiner |
| **PD** | Photodetector |
| **PDL** | Polarization-dependent loss |
| **PLI** | Physical layer impairment |
| **PM** | Performance monitoring |
| **PMD** | Polarization-mode dispersion |
| **PS** | Probabilistically-shaped |

| | |
|---|---|
| **PSD** | Power spectral density |
| | |
| **QAM** | Quadrature amplitude modulation |
| **QoS** | Quality of service |
| **QoT** | Quality of transmission |
| **QPSK** | Quadrature phase shift keying |
| | |
| **RMSA** | Routing, Modulation, and Spectrum Assignment |
| **RMSE** | Root mean squared error |
| **ROADM** | Reconfigurable optical add-drop multiplexer |
| **Rx** | Receiver |
| | |
| **S-margin** | System margin |
| **SCI** | Self-channel interference |
| **SDN** | Software-Defined Networking |
| **SE** | Spectral efficiency |
| **SHB** | Spectral hole burning |
| **SLA** | Service-level agreement |
| **SMF** | Single-mode fiber |
| **SNR** | Signal to noise ratio |
| **SPM** | Self-phase modulation |
| **SRS** | Stimulated Raman scattering |
| **SSFM** | Split step Fourier method |
| **std** | Standard deviation |
| **SVD** | Singular value decomposition |
| | |
| **TDM** | Time-division multiplexing |
| **TL** | Transfer learning |
| **TPD** | Transponder |
| **TRX** | Transceiver |
| **Tx** | Transmitter |
| | |
| **VOA** | Variable optical attenuator |
| | |
| **WDM** | Wavelength-division multiplexing |
| **WSS** | Wavelength selective switch |
| | |
| **XCI** | Cross-channel interference |
| **XPM** | Cross-phase modulation |

# 1

# Introduction

## 1.1 Background

The global fiber-optic infrastructure serves as the cornerstone of today's digital communications and Artificial Intelligence (AI) society, which is fundamentally anchored in more than half a century of sustained efforts in optical physics and engineering. In the 1970s, optical carriers only provided 64 kb/s point-to-point transport delivering data from voice, FAX and modem. Time-division multiplexing (TDM) technique was applied at that time, which utilized electronics to aggregate data to high data rates. The early stage of optical fiber transmission witnessed steady development with a moderate rate of increase in bit rate of information delivered in a single optical carrier. Drastic growth in the capacity of optical fiber network has begun in the late 1990s, due to the introduction of wavelength-division multiplexing (WDM) enabled by erbium-doped fiber amplifier (EDFA) to achieve broadband amplification[1]. The capacity growth was enabled by multiplexing many wavelengths with the bit rate of information carried on each independent wavelength into a single fiber. Following the fixed-grid WDM techniques, the scalability, flexibility and restoration of optical fiber networks then become reality due to the commercialization of tunable lasers, reconfigurable optical add-drop multiplexer (ROADM) and optical cross-connect (OXC) which enable flexible grid (flex-grid) WDM transmission with finer granularity (e.g. 6.25 GHz). Together with the automated network management and control plane, current optical fiber networks are reconfigurable with a wavelength granularity that supports automated optical connection provisioning and flexible bandwidth allocation to address rapid changes in capacity demand between source-destination pairs[1], which plays a significant role in supporting explosive data traffic in the AI era.

In parallel development, the research community and industry also focus on increasing the bit rate per wavelength to achieve higher spectral efficiency (SE) in WDM optical fiber networks. Advanced modulation format such as higher order quadrature amplitude modulation (QAM) and/or probabilistically-shaped (PS), coherent reception incorporating advanced digital signal processing (DSP), and polarization multiplexing techniques have been extensively explored. Coherent receiver operates on the beating optical signals by a local laser oscillator, which are then detected by the photodetector (PD) with the phase of the carrier preserved. Hence, the continuous phase or frequency modulation signals can be processed in the electronic domain[9]. DSP, enabled by ultra-high sampling rate analog-to-digital converter/conversions (ADCs), can achieve polarization de-multiplexing, clock and carrier phase recovery, as well as combat the linear and nonlinear impairments due to chromatic dispersion (CD), polarization-mode dispersion (PMD) and other effects. From terrestrial to submarine, today's optical fiber network is capable of transmitting tens of terabits per second on a single fiber over a distance that spans the Pacific Ocean[1], thanks to cutting-edge WDM and advanced coherent optical transponders. The evolution of the optical fiber network architecture and its key enabler techniques is illustrated in **Figure 1.1**[1].

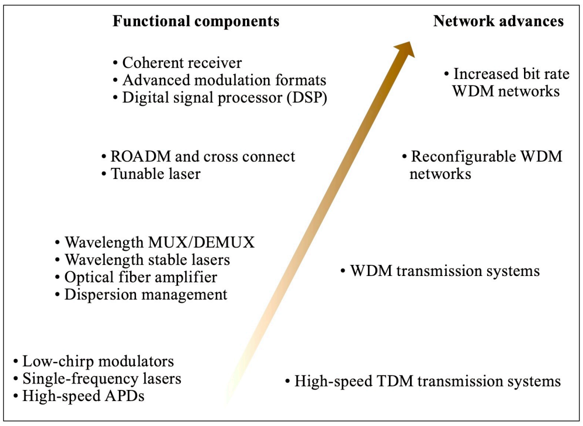



**Figure 1.1:** Evolution of the optical fiber network functionality and architecture and its underlying enabler techniques (from[1]).

Modern optical fiber networks can be divided into long-haul, metro, access, and data center network (DCN) according to the covering area and distance, as shown in **Figure 1.2**[2]. Long-haul networks, including both terrestrial and submarine, are the backbone/core of global network supporting over 10,000 kilometers high-speed high-bandwidth optical transmission. Metro networks support shorter distance compared to the long-haul counterpart, from tens of kilometers to hundreds of kilometers covering a smaller geographic areas such as a city. However, long-haul networks tend to be more stable in topology, unlike their metro counterparts. Optical access network is built to cover distances ranging from hundreds of meters to 20 km. DCN typically refers to the intra-connect of data centers, which is designed to support the large data workloads exchanged between the parallel server machines. In terms of data center inter-connect, typically referred to as DCI, can be regarded as a type of metro networks according to its typical transmission distance of less than 200 km but much longer than that of DCN. Coherent transponder technology was first introduced in long-haul transmission and subsequently penetrated other areas such as metro and DCI networks as well. In general, delivered bandwidth and capacity should be considered first for long-haul backbone networks or metro DCI networks while for other shorter links, the performance and SE of coherent solutions are not always required where the cost and power consumption of this technique should be taken into account. In this thesis, two types of live commercial optical networks are the main research focus, i.e., metro DCI networks and long-haul backbone networks, where commercial coherent transponders and WDM techniques are widely adopted.

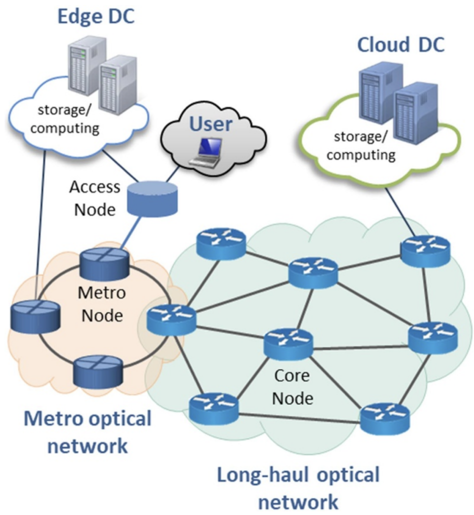



**Figure 1.2:** Long-haul (comprising backbone/core nodes), metropolitan (comprising metro nodes), and access (comprising access nodes) optical network segments supporting cloud and edge infrastructures (data centers) from [2].

The fundamental architecture, definition, and data model of optical transport network (OTN) are essential for the management of a commercial optical network. An OTN is composed of a set of optical network elements connected by fiber spans, capable of providing the functionality of transmission, routing and multiplexing of optical channels delivering client signals, according to the requirements given in ITU-T G.872 [17]. As the data communication layer is moving towards Ethernet, OTN has a strong role to play in supporting 100 Gbps and beyond Ethernet and providing the per-channel monitoring capabilities [1]. The optical layer of OTN was defined by the standard ITU-T G.872, consisting of three layers: optical channel (OCH) layer, optical multiplex section (OMS) and optical transport section (OTS) [17]. As illustrated in **Figure 1.3** [3], OCH fundamentally forms the basis of the optical connection between regeneration sites or client access points, while OMS and OTS manage the multiplexing and transport of these channels, with overhead information added at each level for monitoring and man-

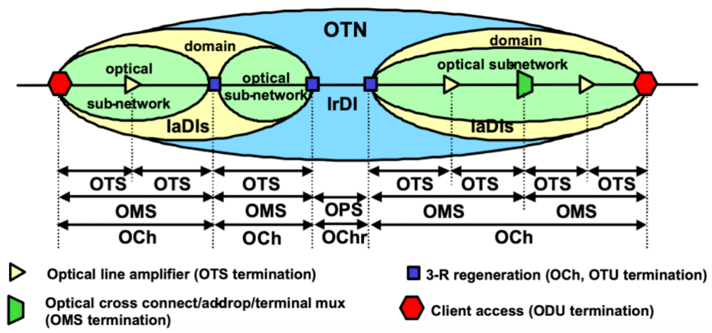



**Figure 1.3:** OTN Network Layers[3]. ODU: Optical Data Unit, OTU: Optical Transport Unit, OCH: Optical Channel, OMS: Optical Multiplex Section, OTS: Optical Transport Section, 3R: reshaping, retiming and reamplification of the signal.

agement purposes. The OCH layer network provides for the transport of digital optical channel transport unit (OTU) signals through an OCH trail between access points[3]. The characteristic information of an OCH layer involves an optical signal defined by a set of parameters such as central frequency, required bandwidth and signal to noise ratio (SNR) associated with the network media channel are of particular interest[3]. The OMS is mostly terminated in network elements such as the ROADMs to realize the functionalities of channel add/drop and multiplexing/demultiplexing. The OTS is a single unidirectional fiber between points of two network elements, e.g., between amplifiers or between an amplifier and the point where the OMS is terminated. The definition of optical layer in OTN will be useful in the thesis since the data we used are extracted from the network controller of commercial optical fiber networks which comply with these standardized definitions.

The ever-growing demand to access cloud-based computational services such as machine learning (ML) applications running in data-center infrastructure, continues to drive the demands of transmission bandwidth and capacity on DCI and longhaul backbone networks even further. Given the state of current commercial systems, this suggests the need for 1 Pb/s per-fiber systems in the not-too-distant future[1]. To meet the ever-growing demands, optical transmission networks have been constantly evolving with a series of advanced techniques such as modulation with probabilistically-shaped (PS), bandwidth-variable transceivers and hollow core fiber (HCF). However, the **hy-**

**perscalers** (e.g. Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP) or Alibaba Cloud) need not only to scale up in transported bits/fiber but also to reduce the cost of network design, deployment, and operation for a full network lifecycle at an accelerated pace to sustain this growth. This necessitates a transition towards fully autonomous, or "self-driving", optical networks that are predicated on the ability to automate the entire lightpath provisioning and operation life-cycle[18]. The modeling of physical layer impairment (PLI), or quality of transmission (QoT), will be the critical enabler that provides the necessary physical layer awareness for a wide range of automated functions, from intelligent routing and dynamic provisioning to proactive maintenance and efficient capacity management.

## 1.2 Fundamentals of Impairments in Optical Transmission System

In this section, we revisit the fundamental principles and models of impairments originating from transponders, amplifiers, and fibers, since they are indispensable components to establish a lightpath in the network and have an inevitable impact on the deterioration of signal quality.

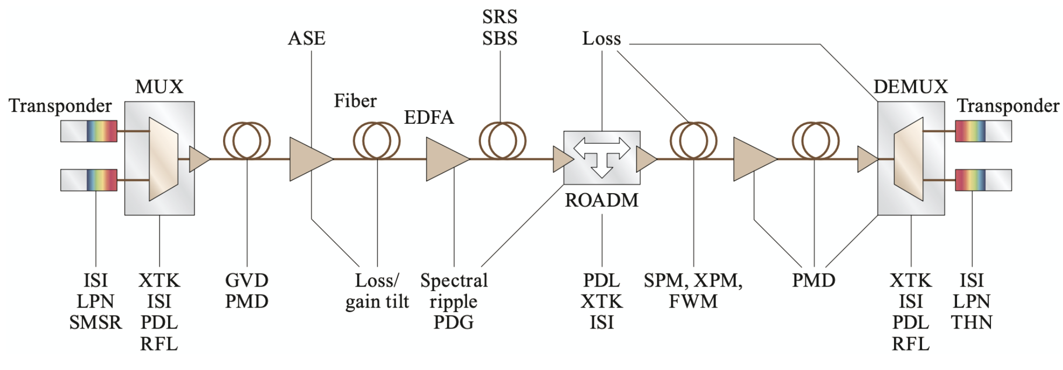



**Figure 1.4:** A typical WDM line system and related transmission impairments, extracted from[1]. ISI: Intersymbol interference; LPN: Laser phase noise; SMSR: Side-mode suppression ratio; RFL: Reflections; XTK: Cross-talk; THN: Thermal noise; PDL/PDG: Polarisation-dependent loss/gain[1].

An illustrative optical fiber transmission link is shown in **Figure 1.4** (extracted from[1]), which shows the indispensable devices/components for establishing a lightpath and the corresponding physical impairments that a lightpath can encounter along the transmission link. Transponders centered at different wavelengths are multiplexed onto one fiber by a WDM multiplexer, then amplified by a booster amplifier (BA) and sent over a fiber link. The typical length of a fiber span is 80 km in terrestrial systems. During fiber propagation, impairments such as fiber loss, fiber nonlinearity, polarization-dependent loss (PDL), group velocity dispersion (GVD) and polarization-mode dispersion (PMD) limit the performance of the signal. Fiber nonlinearities including stimulated Raman scattering (SRS) leading to power transfer and Kerr effect leading to intra-channel and inter-channel nonlinearities, impose an upper limit on the launch power of optical channels. EDFAs are used to compensate for the power attenuation after each fiber span, but it introduced additional amplified spontaneous emission (ASE) noise and non-ideal spectral loss/gain tilt and ripples. Before the receiver (Rx), the WDM signal is pre-amplified and demultiplexed. The signals with different wavelengths are then received by the transponders whose performance can be limited by PD thermal noise (THN), laser phase noise (LPN), and ADC quantization noise and etc[1]. If adding/dropping WDM channels is necessary at intermediate sites, then reconfigurable optical add-drop multiplexers (ROADMs) are placed between the transmission links.

### 1.2.1 Transmitter

A typical structure of transmitter is illustrated in **Figure 1.5**. It is usually composed of a tunable laser, dual-polarization I/Q modulator, DAC, linear drivers and polarization beam combiner (PBC)[1]. DAC generates the analog voltages and the linear RF amplifiers between the DACs and the I/Q modulator are used to amplify the DAC output voltage[1]. The linearity is an important aspect for these drivers, as any departure from linearity would result in distortions that are passed on to the modulated optical signal[1]. To achieve polarization multiplexing, dual-polarization I/Q modulators split optical carrier and modulate two independent complex streams via two I/Q modulators. The two modulated signals are combined via a PBC[1]. The footprint, bandwidth, $V_\pi$ and insertion loss (IL) of the modulators are important figures of merit.

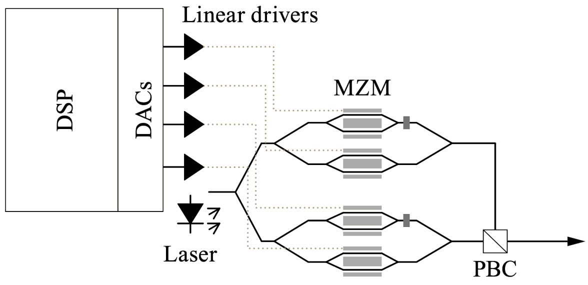



**Figure 1.5:** Typical architecture of an optical transmitter.[1]

## 1.2.2 Fiber

The main impairments in a single fiber span that affect the complex envelope of the guided electric field $E(z,t)$, is a function of fiber propagation distance $z$ and time $t$[16], and obeys the nonlinear Schrödinger equation (NLSE)[19]:

$$\begin{aligned}\frac{\partial E(z,T)}{\partial z} &= \underbrace{-\frac{\alpha}{2}E(z,T)}_{\text{Loss}} \underbrace{- j\frac{\beta_2}{2}\frac{\partial^2 E(z,T)}{\partial T^2} + \frac{\beta_3}{6}\frac{\partial^3 E(z,T)}{\partial T^3}}_{\text{Dispersion}} \\ &\quad + \underbrace{j\gamma|E(z,T)|^2 E(z,T)}_{\text{Kerr}} \underbrace{- j\gamma T_r \left\{\frac{\partial}{\partial T}|E(z,T)|^2\right\} E(z,T)}_{\text{Raman}}\end{aligned} \tag{1.1}$$

which demonstrates that fiber loss, fiber dispersion, Kerr nonlinearity and stimulated Raman scattering (SRS) are the main impairments. Accordingly, NLSE in frequency domain can be obtained by taking the Fourier transform of **Equation (1.1)** as[16],

$$\begin{aligned}\frac{\partial E(z,f)}{\partial z} &= \underbrace{-\frac{\alpha}{2}E(z,f)}_{\text{Loss}} + \underbrace{\left(j2\pi^2\beta_2 f^2 + j\frac{4}{3}\pi^3\beta_3 f^3\right)E(z,f)}_{\text{Dispersion}} \\ &\quad + j\gamma\left(\underbrace{E(z,f) * \overline{E(z,-f)} * E(z,f)}_{\text{Kerr}} \underbrace{- jT_r(2\pi f E(z,f) * \overline{E(z,-f)} * E(z,f))}_{\text{Raman}}\right)\end{aligned} \tag{1.2}$$

where $\overline{E(\cdot)}$ denotes the complex conjugate of $E(\cdot)$ and $*$ denotes the convolution operation. Next, we will elaborate on these four main impairment model and analyze their causes and influence on the signal transmission.

## I Fiber Loss

For the term of fiber loss, we can rewrite **Equation (1.2)** as

$$\frac{\partial E(z,f)}{\partial z} = -\frac{\alpha E(z,f)}{2} \tag{1.3}$$

by negelecting other effects and then obtain the solution of the electrical field as

$$E(z,f) = E(0,f)e^{-\frac{\alpha(f)}{2}z} \tag{1.4}$$

where $\alpha(f)$ represents the wavelength-dependent fiber attenuation coeffficient. Taking the optical power $P(z,f) = |E(z,f)|^2$ and **Equation (1.4)** can be changed into

$$P(z,f) = P(0,f)e^{-\alpha(f)z} \tag{1.5}$$

The optical power loss in dB units due to propagation in a fiber of length $L$ is defined as

$$\text{loss}\,[\text{dB}] = -10log_{10}\frac{P(z,f)}{P(0,f)} \approx 4.343\,\alpha L \tag{1.6}$$

Obviously, fiber loss is a linear effect. The signal power is attenuated as it propagates along the fiber mainly due to Rayleigh scattering and material absorption. A typical fiber loss profile[2] for a low-water peak single-mode fiber (SMF) is shown in **Figure 1.6**, showing that C-band (1528–1565 nm) and L-band (1570–1610 nm) have comparatively low attenuation and are commonly used communication bands in large-scale networks.

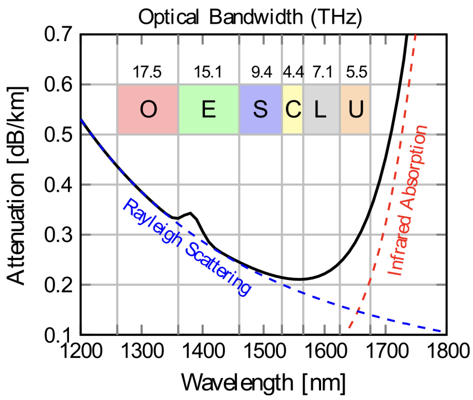



**Figure 1.6:** Typical fiber loss spectra for a low-water peak single-mode fiber (SMF)[2].

## II Fiber Dispersion

The group velocity dispersion (GVD), also referred to as chromatic dispersion (CD), is the dependence of the refractive index of the medium upon the propagating optical frequency, which is the result of different wavelengths in a pulse propagating at slightly different speeds thus leading to the broadening of an optical pulse propagating through the fiber. It is modeled by the Taylor series expansion of the mode-propagation constant, $\beta$, with respect to the central frequency of the pulse. In **Equation (1.2)**, $\beta_2$ is defined as the second derivative of $\beta$ with respect to the optical frequency computed in the pulse central frequency, and is the parameter that describes the pulse broadening; $\beta_3$ is the third-order dispersion coefficient or dispersion slope, $T = t - z/v_g$ is the time of the reference time window moving at group velocity $v_g$[16]. By neglecting the fiber loss, Kerr nonlinearity and Raman effects, we can rewrite **Equation (1.2)** and obtain its solution as

$$E(z, f) = E(0, f)e^{j(2\pi^2\beta_2 f^2 + \frac{4}{3}\pi^3\beta_3 f^3)z} \tag{1.7}$$

from which we can see that $E(0, f)$ experiences a frequency-dependent phase change refereed as CD, while its amplitude remains constant[16]. The term of fiber dispersion in **Equation (1.2)** is also a linear effect.

In addition, the phenomenon of polarization-mode dispersion (PMD) exists where two different polarizations of light in a waveguide travel at different speeds due to random imperfections and asymmetries[1]. PMD is related to the differential group delay (DGD) caused by birefringence in optical fibers. Both CD and PMD cause a time spreading of the optical pulses, which can translate into an inter-symbol interference (ISI) effect. CD and PMD compensation becomes a critical function in DSP to counteract signal distortion[16].

## III Fiber Kerr Nonlinearity

Similarly, by neglecting the terms of fiber loss, fiber dispersion and Raman effect in **Equation (1.1)**, we can rewrite **Equation (1.1)** as:

$$\frac{\partial E(z,T)}{\partial z} = j\gamma|E(z,T)|^2 E(z,T) \tag{1.8}$$

and the corresponding solution can be written as

$$E(z,T) = E(0,T)e^{j\gamma|E(0,T)|^2 z} \tag{1.9}$$

where $\gamma$ is the fiber nonlinear coefficient that quantifies the strength of self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) effects. It is defined in terms of optical power as:

$$\gamma(\lambda) = \frac{2\pi}{\lambda}\frac{n_2}{A_{eff}} \tag{1.10}$$

where $n_2$ is the nonlinear Kerr parameter and $A_{eff}$ is the effective mode area[4]. For WDM optical transmission delivering $n$ optical channels with $E = \sum_{i=1}^{n} E_i$, **Equation (1.8)** can be simplified by only taking the terms of channel of interest (COI) $E_1$,

$$\frac{\partial E_1}{\partial z} = j\gamma|\sum_{i=1}^{n} E_i|^2 E_1 \tag{1.11}$$

The fiber Kerr nonlinear impairments can be mainly categorized into: self-channel interference (SCI) and cross-channel interference (XCI). More specifically, the intra-channel nonlinearities are self-phase modulation (SPM)-induced nonlinear phase noise, modulation instability, isolated pulse SPM, intrachannel cross-phase modulation (IXPM), and intrachannel four-wave mixing (IFWM)[1]. The inter-channel nonlinearities can be mainly categorized into: interchannel cross-phase modulation (XPM) and four-wave mixing (FWM)[1].

It can be observed from **Equation (1.9)** that Kerr nonlinearity is also weighed by a factor $|E(0,T)|^2$ in the time domain, hence proportional to the propagating signal power. Thus, it is expected that Kerr effect becomes relevant only at high power[1]. Hence, we observe that an optimal power exists in the **Figure 1.7**, corresponding to the maximum achievable SNR. **Figure 1.7** also shows an ASE-limited

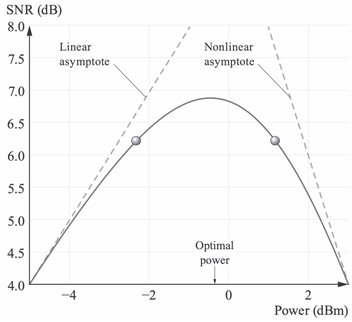



**Figure 1.7:** Example of behavior SNR versus power after a fiber-optic link[1].

region and a nonlinearity-limited region, where two asymptotic behaviors are recognized at small/high powers. Beyond the optimal power the SNR drops, with a slope which is actually steeper than the slope of increase at small power[1].

## IV Raman Effect

Another nonlinear effect described by **Equation (1.2)** is the stimulated Raman scattering (SRS). Raman time constant is related to the slope of the Raman gain $T_r = \frac{1}{2\pi\gamma f_r}\frac{\partial}{\partial f} g_r(f)\Big|_{f=0}$ with the fractional contribution of the delayed nonlinear response[16]. The Raman gain coefficient $g_i(|\Delta f|)$ governs the power coupling between two frequency components with a frequency shift of $\Delta f = f_p - f_s$, where $p$ and $s$ denote the carrier at higher (pump) and lower (Stokes wave) frequencies, respectively[4]. This coefficient depends on several features of the fiber and the propagating channel modes[4]. An exmaple Raman gain coefficient from[4] is shown in **Figure 1.8**.

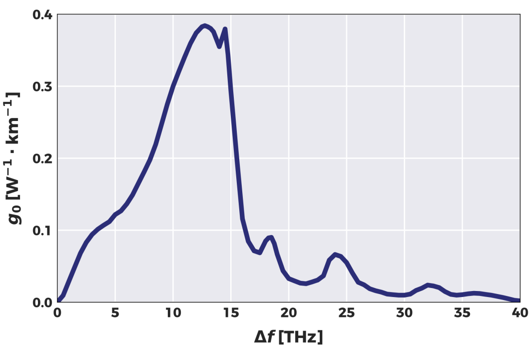



**Figure 1.8:** An experimental Raman gain coefficient curve from[4].

The following coupled Raman equations of the $i^{th}$ COI can be derived from **Equation (1.2)**, which describes its power evolution along with the fiber:

$$\frac{\partial P_i}{\partial z} = \underbrace{\sum_{k=1}^{i-1} g_i(|\Delta f|)P_k P_i}_{\text{ISRS gain}} - \underbrace{\sum_{k=i+1}^{n} \frac{f_k}{f_i} g_i(|\Delta f|)P_k P_i}_{\text{ISRS loss}} \overbrace{-\alpha_i P_i}^{\text{fiber loss}} \tag{1.12}$$

where $P_i$, $f_i$ are the power and frequency of the $i^{th}$ COI, $P_k$, $f_k$ are the power and frequency of the remaining WDM channels. The amount of power that is transferred from one channel to another follows the Raman gain spectrum $g_i(|\Delta f|)$. The coupled Raman equations show that SRS causes the transfer of power from lower wavelength signals to higher wavelength signals in a WDM system[1], resulting in SNR degradation of the lower wavelength signals and introducing crosstalk in the higher wavelength signals. In particular, SRS-induced power transfer incurs a tilt of the WDM spectrum during fiber propagation and thus affects the power-dependent, Kerr effect-induced NLI impairment accumulation[4].

### 1.2.3 Amplifier

The aspects of modeling gain and noise behaviour of erbium-doped fiber amplifier (EDFA) are introduced in this section. EDFA is an indispensable device to compensate for fiber power attenuation underpinning C and/or C+L band optical networks. The typical structure[5] of single-stage EDFA is shown in **Figure 1.9**, consisting of EDF, couplers, isolators, forward/backward pump, and gain flattening filter (GFF) which is placed at the output of EDF to flatten the output power spectra.

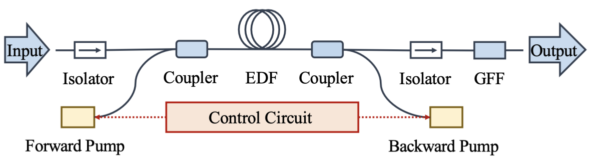



**Figure 1.9:** Structure of a typical single-stage EDFA, extracted from[5].

## I Population Inversion

In principle, the gain and noise response of single-stage EDF are usually modeled using the propagation and rate equations of a homogeneous, two-level inversion model[20,21]. The third-level ion population is negligible and decays rapidly in the stationary regime, justifing the inversion model that accounts only for excited ($N_2$) and ground ($N_1$) states[1].

Consider an EDFA with the EDF length of $L$, with a density of active atoms $\rho$ within an active volume of cross-sectional area $A$[20]. An arbitrary number $N$ of beams of wavelength $\lambda_k$ and power $P_k(z,t)$ travel through the amplifier in a direction of $\mu_k$[20]. The rate equation of the excited state $N_2(z,t)$ is given by

$$\frac{\partial N_2(z,t)}{\partial t} = -\frac{N_2(z,t)}{\tau} - \frac{1}{\rho A}\sum_{j=1}^{N}\mu_j \frac{\partial P_j(z,t)}{\partial z} \tag{1.13}$$

where $\tau$ is the spontaneous lifetime of the excited state. The fractional population of the lower state $N_1(z,t)$ can be obatined by[20]

$$N_1(z,t) + N_2(z,t) = 1 \tag{1.14}$$

The light in the fiber is subject to absorption or stimulated emission at rates that depend on wavelength[20]. The cross sections for stimulated emission and absorption at wavelength $\lambda_k$ are $\sigma_k^e$ and $\sigma_k^a$, respectively. A sample of emission and absorption cross-sections for Al/P-silica fiber are shown in **Figure 1.10**. More cross sections such

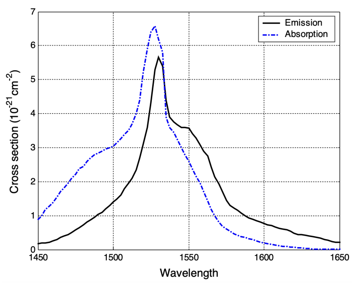



**Figure 1.10:** The example emission and absorption cross-sections for Al/P-silica fiber from[6].

as Giles'[7] for Ge/silicate and Al/Ge/silicate fibers are shown in **Figure 1.11**.

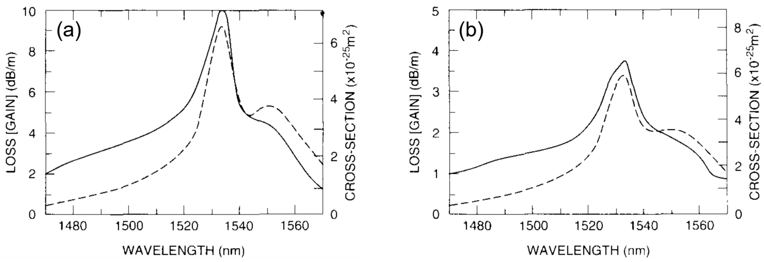



**Figure 1.11:** The gain (– – –) and absorption (——) spectra for (a) Ge/silicate and (b) Al/Ge/silicate fibers. Cross section calculated from Ladenburg-Fuchbauer equation is shown on right axis[7].

The change of power in the $k^{th}$ beam is described by[20]

$$\frac{\partial P_k(z,t)}{\partial z} = \rho\mu_k\Gamma_k\left[(\sigma_k^e + \sigma_k^a)N_2(z,t) - \sigma_k^a\right] - P_k(z,t) \tag{1.15}$$

where $\Gamma_k$ is the confinement factor of the fiber amplifier at wavelength $\lambda_k$[20]. By recognizing the unique role played by average inversion[1,22]

$$\overline{N}_2 = \frac{1}{L}\int_0^L N_2(z)\mathrm{d}z \tag{1.16}$$

one can gain a further insight into EDFA response prediction. By defining the average gain,

$$\overline{g}_n = \frac{1}{L}\ln\frac{P_{k,out}}{P_{k,in}} \tag{1.17}$$

and combining the above equations, it is possible to obtain a unique relation between the input power and the average population inversion at excited state $\overline{N}_2$[1,22]:

$$\overline{N}_2 = -\frac{\tau_0}{LA\rho}\sum_{k=1}^{N} P_{k,in}(e^{\left[(\sigma_k^e+\sigma_k^a)\overline{N_2}-\sigma_k^a\right]L} - 1) \tag{1.18}$$

its solution not only allows for complete EDFA analysis but also provides important insights that aid the amplifier design process[1]. With the cross sections $\sigma_k^e$ and $\sigma_k^a$, one can calculate the average inversion level across the entire spectral range in any EDFA operating regime in principle[1].

## II Gain Response

The signal gain can be expressed in terms of the average inversion of the erbium ions along the fiber[23]:

$$G_k = \exp\left[(\overline{N}_2\sigma_k^e - \overline{N}_1\sigma_k^a)\Gamma_k L\right] \tag{1.19}$$

where $\Gamma_k$ is the overlap factors between the light-field modes and the erbium distribution. This shows that the signal gain is only dependent on the average inversion and does not depend on the details of the shape of the inversion as a function of position along the fiber length[23]. On the other hand, ASE will depend on the detailed shape of the inversion, due to the local nature of ASE generation[23]. Therefore, noise figure (NF) of EDFA will depend on the profile of the inversion, and not just the average inversion[23]. The average inversion relationship for the gain, **Equation (1.19)**, is a result of the assumption of homogeneous broadening for the erbium ion transition at 1550nm[23]. Notably, a more explicit interpretation of the above equation is the singular value decomposition (SVD). Specially, **Equation (1.19)** has been shown to be modeled by a SVD-based black-box gain characterization model[24]:

$$G_{[\mathrm{dB}]} \coloneqq L \cdot \begin{bmatrix} \overline{\mathbf{N}}_2 & -\overline{\mathbf{N}}_1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\sigma}^e \\ \boldsymbol{\sigma}^a \end{bmatrix} \tag{1.20}$$

## III Noise Figure

The spontaneous emission factor $n_{sp}$ characterizes the noise performance of the amplifier[25]:

$$n_{\mathrm{sp}} \equiv \frac{\eta N_2}{\eta N_2 - N_1}, \qquad \eta \triangleq \frac{\sigma_e}{\sigma_a} \tag{1.21}$$

where $\sigma_e$ and $\sigma_a$ are cross-sections for the spontaneous emission and absorption processes[25]. Due to $N_2 > N_1$, the stronger the population inversion, the more efficient the stimulated emission of signal photons becomes. Under complete population inversion, $n_{sp}$ reaches the minimum value $n_{\mathrm{sp}} = 1$ and EDFA behaves ideally with the lowest ASE generation[25]. Depending on different EDFA internal designs, $n_{sp}$ lies in the range $1 < n_{\mathrm{sp}} < 4.0$. Typical values for a low-noise EDFA are $1.2 < n_{\mathrm{sp}} < 1.5$. The in-band uniform ASE power spectral density $\rho_{\mathrm{ASE}}$ at the output of the optical amplifier is given as[25]:

$$\rho_{\mathrm{ASE}} = hv(G-1)n_{\mathrm{sp}} \qquad (\mathrm{W/Hz}) \tag{1.22}$$

where $G$ is the total gain, $v$ is the center frequency and $h$ is Planck's constant. The noise figure is commonly used to characterize the noise added by an amplifier and is defined as the ratio of the input SNR to the output SNR[8],

$$F_n = \frac{SNR_{\text{in}}}{SNR_{\text{out}}} \tag{1.23}$$

which is illustrated in **Figure 1.12** where the SNR is measured in the electrical domain using PD at the input and output of the EDFA, and measuring the electrical signal and noise powers[8].

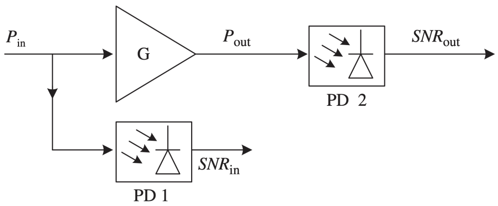



**Figure 1.12:** Measurement of the amplifier noise figure[8].

Let us consider ideal photodetectors with 100% quantum efficiency without thermal noise and dark current, and the input power is $P_{\text{in}}$, the photocurrent is

$$I_{\text{in}} = RP_{\text{in}}, \qquad R = \frac{q}{hv} \tag{1.24}$$

where $R$ is the responsivity. The electrical signal power $S_{\text{in}} = I_{\text{in}}^2 R_L$ where $R_L$ is the load resistor. Assume that there is no noise before the amplifier, the noise power at the output of PD1 is only due to the shot noise $N_{shot} = 2qI_{\text{in}}R_L B_e$ where $B_e$ is the electrical bandwidth. Then, we have

$$SNR_{\text{in}} = \frac{S_{\text{in}}}{N_{shot}} = \frac{RP_{\text{in}}}{2qB_e} = \frac{P_{\text{in}}}{2hvB_e} \tag{1.25}$$

Let us now consider $SNR_{\text{out}}$. Assume the output optical power is given by $P_{\text{out}} = GP_{\text{in}}$, the corresponding photocurrent and electrical signal power are

$$I_{\text{out}} = RGP_{\text{in}}, \qquad S_{\text{out}} = (RGP_{\text{in}})^2 R_L \tag{1.26}$$

respectively. The noise power delivered to a resistor $R_L$ consists mainly of two components: shot noise and signal–ASE beating noise without considering the ASE–ASE beating noise[8]. Assume that the optical filter is absent, the total noise power can be obtained by[8]

$$N_{\text{out}} = N_{shot} + N_{s-sp} = 2qRP_{\text{out}}R_LB_e + 4R^2\rho_{\text{ASE}}P_{\text{out}}R_LB_e \tag{1.27}$$

The $SNR_{\text{out}}$ at PD2 can be written as

$$SNR_{\text{out}} = \frac{S_{\text{out}}}{N_{\text{out}}} = \frac{RGP_{\text{in}}}{(q+2R\rho_{\text{ASE}})2B_e} \tag{1.28}$$

Thus, combining **Equation (1.23)** we can write

$$F_n = \frac{q+2R\rho_{\text{ASE}}}{qG} \tag{1.29}$$

Using $R = q/hv$ then the power spectral density of ASE is $\rho_{\text{ASE}} = (GF_n - 1)hv/2$. Consider $\rho_{\text{ASE}}$ can also be expressed in **Equation (1.22)**. Equating the right-hand sides of the two equations, the amplifier noise figure and the spontaneous emission factor $n_{\text{sp}}$ can be related[8]:

$$F_n = \frac{2n_{\text{sp}}(G-1)}{G} + \frac{1}{G} \tag{1.30}$$

since $G \gg 1$, we have

$$F_n \cong 2n_{\text{sp}} \tag{1.31}$$

Since the minimum value of $n_{\text{sp}}$ is 1, the lowest achievable noise figure is 2. If it is expressed in dB units, we have $F_n$ [dB] $= 10\log_{10}F_n$. When $n_{\text{sp}} = 1$, $F_n = 3\,\text{dB}$, corresponding to an ideal amplifier with the lowest ASE noise[8].

Notably, the gain and noise characterizations described above are based on single-stage EDF. However, commercially available EDFAs are usually constructed by **dual-stage** EDFs accompanied with gain flattening filter (GFF) and VOA inside to realize Automatic Gain Control (AGC) or Automatic Power Control (APC) and optimize amplified spontaneous emission (ASE) noise level. There are many spaces for vendors to design dual-stage EDFA with different EDF section lengths and materials, and various pump methods such as co-pumping and counter-pumping. The modeling of multi-vendor EDFAs becomes more challenging for network operators in such a black-box background. In addition, as shown in **Figure 1.10** and **Figure 1.11**, different materials of doped fiber exhibit distinct frequency responses, which directly influence the wavelength dependence of gain and noise figure. Although GFFs are designed to counteract the non-flat gain response but still exhibit residual wavelength-dependent gain ripples to be further investigated. In summary, the impairments introduced by EDFAs involove 1) non-ideal gain ripples leading to non-uniform amplification; 2) input-dependent ASE noise profile.

### 1.2.4 Receiver

**Figure 1.13** shows a schematic diagram of a digital optical coherent receiver. Coherent receivers are widely used in commercial high-speed optical transponders, which utilize a local oscillator (LO) $C_{\mathrm{LO}} = Re^{j\omega_{\mathrm{LO}}t}$ to mix with the received optical field $C_{\mathrm{SIG}} = Ae^{j\varphi}e^{j\omega_c t}$ [1]. The mixed light is detected by PD,

$$E = \left| \frac{C_{\mathrm{SIG}} + C_{\mathrm{LO}}}{\sqrt{2}} \right|^2 = \frac{R^2}{2} + \mathbf{Re}\left\{RAe^{j\varphi}e^{j\Delta\omega t}\right\} + \frac{A^2}{2} \tag{1.32}$$

where $\mathbf{Re}\{\cdot\}$ means the real part of a complex number, $\Delta\omega = \omega_{\mathrm{c}} - \omega_{\mathrm{LO}}$ is the angular frequency difference of the signal carrier and the LO, $1/\sqrt{2}$ is induced by a 3 dB power attenuation from coupler[1]. In **Equation (1.32)**, the first term is a DC term and the second term is the signal of interest[1]. If the LO power is much larger than the signal, the second term will dominate over the third term. Only the lower-frequency term, which falls within the absorption range of the PD, is converted into the electronic current preserving both the phase and amplitude of the modulated signals[9]. Two important noise sources are added in this process: (i) thermal noise and (ii) shot noise.

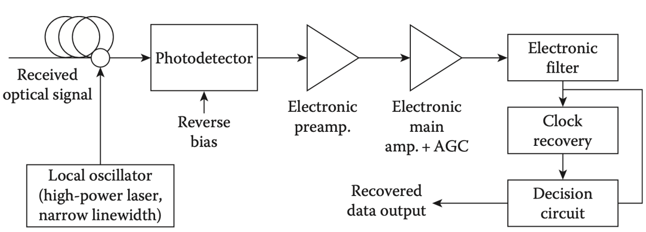



**Figure 1.13:** Schematic diagram of a digital optical coherent receiver from[9].

### I Thermal Noise

With the temperature increases, electrons move faster and the resulting current is noisy and is called "thermal noise"[8]. If $B_e$ is the effective bandwidth of the receiver, its noise variance can be calculated as[8]

$$\sigma^2_{\text{thermal}} = \frac{4k_B T B_e}{R_L} \tag{1.33}$$

where $k_B = 1.3806 \times 10^{-23} J/K$ is Boltzmann's constant, $R_L$ is the resistor value. Note that **Equation (1.33)** does not consider the noise sources in the amplifier circuit, but it can be modified to account for the noise sources within the amplifier as[8]

$$\sigma^2_{\text{thermal}} = \frac{4k_B T B_e}{R_L} F_n \tag{1.34}$$

where $F_n$ is the amplifier noise figure.

### II Shot Noise

For a non-ideal PD with quantum efficiency $\eta < 1$, the probability that a photon incident on the PD generates an electron–hole pair that contributes to the photocurrent is $\eta$[8]. The fluctuation of the photocurrent is defined as the shot noise process. The shot noise is a white noise process, and its power spectral density is constant. The shot noise process generated by the dark current $I_D$ is independent of the signal optical power and its power spectral density can be given by

$$S_D = 2qI_D \tag{1.35}$$

Compared with the thermal noise power, the dark current shot noise does not represent any limitation on receiver performance[25].

The photodetection process of the incident optical field produces a corresponding electron flux whose fluctuations are identified with the *signal shot noise*[25]. The linear relationship between this shot noise power and the input optical power is fundamental in understanding the input signal-dependent behavior of this noise contribution[25]. The signal shot noise process is stationary and its unilateral power spectral density is uniform which is given by

$$S_{shot} = 2qI_{\text{in}} \tag{1.36}$$

Since the average photocurrent $I_{\text{in}}$ can be expressed as $I_{\text{in}} = RP_{\text{in}}$ where $R$ is the responsivity of PD, then

$$S_{shot} = 2qRP_{\text{in}} \tag{1.37}$$

thus the input-dependent shot noise power can be calculated as $\sigma^2_{\text{shot}} = 2qRP_{\text{in}}B$ through an effective bandwidth of $B$.

## 1.3 Fundamentals of Quality of Transmission and Low-margin Optical Network

This section introduces the fundamentals of the physical layer impairment (PLI) and quality of transmission (QoT) modeling, low-margin operation, network control and management in commercial optical networks.

For commercial optical networks, network operators must guarantee a given end-to-end quality of transmission (QoT) to their customers to meet the service-level agreements (SLAs). The measure of QoT is usually given by the pre-forward error correction (FEC)-BER, which can be reported in real time from the coherent transponder of a given optical connection, or estimated by combining a B2B OSNR-BER model with a physical layer impairment (PLI) model utilizing other optical layer data such as optical power, amplifier gain and noise figures, etc. Before service deployment where there is no BER monitored, PLIs must be evaluated before establishing an optical connection to ensure good signal quality and avoid potential degradation of quality of transmission (QoT). The modeling of PLIs can be divided into analytical models and machine learning (ML)-based models. We will discuss the uncertainties of input parameters in analytical models and the data pre-processing procedures for ML-based models, and the possible methods to improve the accuracy of QoT/PLI estimation. Then we elaborate on the definition of network margin and how the accurate estimation of lighpath's

QoT can facilitate optical network margin design and management, translating to a maximized utilization of network capacity and bandwidth. Finally, the importance of QoT models/estimations in realizing the automated network control and management plane is discussed.

### 1.3.1 QoT Metrics

#### I BER, Q-factor, SNR and OSNR

In optical networks, the most important criterion to evaluate QoT is the pre-forward error correction (FEC) BER, which reflects the raw signal quality and the likelihood of errors occurring without any error correction. BER is usually translated to the Quality(Q)-factor by[1]:

$$Q\,[\mathrm{dB}] = 20\log_{10}\left[\sqrt{2}erfc^{-1}(\,2\,\mathrm{BER}\,)\right] \tag{1.38}$$

where $erfc^{-1}(\cdot)$ is the inverse of the complementary error function. The higher the value of the Q-factor, the better the BER. In some cases, Q-factor in dB scale is more frequently used because it is almost linearly related to signal to noise ratio (SNR) with slope 1 in a dB/dB scale for several modulation formats, thus being helpful for empirical rule design[1]. With the development of high-speed coherent transponders for commercial optical networks, optical signal to noise ratio (OSNR) and signal to noise ratio (SNR) have become the widely adopted QoT metrics since optical channels are mostly degraded by additive white Gaussian noise (AWGN). SNR is defined as the ratio of the electrical signal power $P_{\mathrm{Rx}}$ at the receiver side to the noise power $\sigma^2$ accumulated along the fiber links by

$$\mathrm{SNR} = \frac{P_{\mathrm{Rx}}}{\sigma^2} \tag{1.39}$$

The relation between BER and SNR depends on the modulation format of the transmitted signal. For the AWGN channel with square M-QAM, their relation is given by[26]

$$\mathrm{BER} = \frac{2}{\log_2 M(1-\frac{1}{\sqrt{M}})}\,erfc\left(\sqrt{\frac{3}{2(M-1)}\mathrm{SNR}}\right) \tag{1.40}$$

where $erfc(\cdot)$ is the complementary error function.

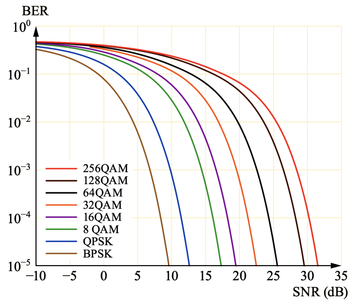



**Figure 1.14:** The relation between SNR and BER for the most popular modulation formats from[1].

**Figure 1.14** shows the relation between SNR and BER for the most popular modulation formats[1]. Since current performance of coherent receivers is almost ideal, there is a direct one-to-one mapping between the optical domain to electrical domain[14]. SNR is often related with the optical signal to noise ratio (OSNR), which is calculated with respect to a specific noise bandwidth $R_n$, typically 0.1nm or 12.5GHz. However, the SNR by default refers to the signal bandwidth $R_s$ (symbol rate) of the matched filter used for demodulation. Thus, the relationship between SNR and OSNR is given by[14]:

$$\mathrm{OSNR} = \mathrm{SNR}\frac{R_s}{12.5\mathrm{GHz}} \tag{1.41}$$

OSNR is also an important QoT metric for estimating optical layer impairments since in the optical domain BER cannot be measured.

## II Generalized SNR

As discussed in the previous section, the most significant physical layer impairments are ASE noise and NLI for dispersion uncompensated coherent optical transmission links. In dispersion-uncompensated coherent links, accumulated CD makes the Kerr-induced nonlinear distortion decorrelate and behave like Gaussian noise (GN). The well-known GN model has verified that NLI noise can be regarded as approximately Gaussian and additive noise, similar to ASE noise[10]. This property leads to the definition of the generalized or effective SNR (GSNR)[10,14], a QoT metric that accounts for both ASE and NLI noise given by:

$$\mathrm{GSNR} = \frac{P_{ch}}{P_{ASE} + P_{NLI}} \tag{1.42}$$

where $P_{ch}$ denotes the launched optical chanel power, $P_{ASE}$ is the ASE noise power introduced by EDFA and $P_{NLI}$ is a suitably-calculated power of NLI noise[10], as illustrated in **Figure 1.15**. **Equation (1.42)** is additive in noise power across spans so that it can be calculated via inverse-sum, enabling accurate span-by-span planning and SLA budgeting.

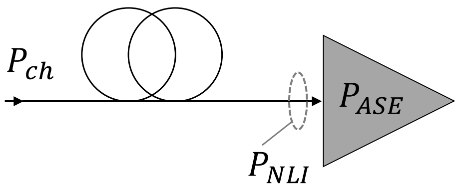



**Figure 1.15:** Assumed span layout with the span-relevant quantities.[10]

## 1.3.2 Physical Layer Impariment Models

In recent years, the types of PLI-based QoT estimation models can be divided into (i) analytical models and (ii) machine learning (ML) based models.

### I Analytical Models

- **SNR$_{\text{ASE}}$**

  Traditionally, a classical formula to estimate the linear noise induced $SNR_{\text{ASE}}$ of a signal with symbol rate of $R_s$ and traveling over $N$ homogeneous fiber spans with identical EDFA working at gain of $G$ and noise figure of $NF$ is given as[1]:

  $$SNR_{\text{ASE}}\,[dB] = 58 + P\,[\text{dBm}] - NF\,[\text{dB}] - G\,[\text{dB}] - 10\log_{10} N - 10\log\left(\frac{R_s}{12.5\,\text{GHz}}\right) \tag{1.43}$$

  where 58 is a constant related to Plank's constant[27], $P$ is the EDFA output power and OSNR can also be obtained by neglecting the last term of **Equation (1.43)**[1]. However, this formula only gives a rough estimate of ASE-induced SNR degradation that ignores NLI noise and does not apply to heterogeneous fiber spans and channel loadings either. Since heterogeneous fiber spans are more common to see in real networks, a more accurate estimation $SNR_{\mathrm{ASE}}$ is to estimate the ASE noise individually for a signal centered at $f_c$ with a bandwidth of $B$ passing through the $i^{th}$ EDFA using the standard EDFA noise power model:

$$N_{\mathrm{ASE},i} = hf_cB(G_i - 1)NF_i \tag{1.44}$$

and denote the signal power at the output of the $i^{th}$ EDFA as $P_i$ we have

$$SNR_{\mathrm{ASE},i} = \frac{P_i}{N_{\mathrm{ASE},i}} = \frac{P_i}{hf_cB(G_i - 1)NF_i} \tag{1.45}$$

and lump them together to obtain the overall $SNR_{\mathrm{ASE}}$ after a total of $i$ EDFAs:

$$SNR_{\mathrm{ASE}} = \frac{1}{\sum_i SNR^{-1}_{\mathrm{ASE},i}} \tag{1.46}$$

- **SNR$_{\mathrm{NLI}}$**

On the other hand, NLI noise is defined as:

$$P_{\mathrm{NLI}} = \eta_i(f_j)P_j^3 \tag{1.47}$$

where $\eta_i(f_j)$ is the nonlinear coefficient evaluated at the end of the $i^{th}$ fiber span and $P_j$ is launch power of the $j^{th}$ COI centered at $f_j$. $\eta_i$ can be approximated by[28]

$$\eta_i(f_j) \approx \frac{B}{P_j^3}G(f_j) \tag{1.48}$$

where the $G(\cdot)$ is the power spectral density (PSD) of the nonlinear interference. A series of Gaussian noise (GN) models[10,28–30] have aimed to find an analytical expression for $G(\cdot)$ and their proposed models have been shown to accurately estimate NLI noise accumulation over fiber spans by treating NLI as additive white Gaussian noise (AWGN). It is based on the assumptions[10] that 1) NLI noise is the perturbation of the signals, 2) the optical signal can be regarded as a stationary GN statistically and 3) the interference in the fiber is an additive

GN[31]. The NLI power spectral density (PSD) can be expressed as[28]

$$G_{NLI}(f) = \frac{16}{27}\gamma^2 \iint G_{\mathrm{Tx}}(f_1)G_{\mathrm{Tx}}(f_2)G_{\mathrm{Tx}}(f_1+f_2-f) \left| \int_0^L \frac{\rho(\zeta,f_1)\rho(\zeta,f_2)\rho(\zeta,f_1+f_2-f)}{\rho(\zeta,f)} e^{j\phi(f_1,f_2,f,\zeta)} d\zeta \right|^2 df_1 df_2 \tag{1.49}$$

where $G_{\mathrm{Tx}}(f)$ is the PSD of the signal on the transmitter (Tx) side, $\gamma$ is the nonlinear coefficient, $\beta_2$ is the group velocity dispersion (GVD), $\beta_3$ is the slope of the GVD, $\phi(f_1, f_2, f, \zeta) = -4\pi^2(f_1-f)(f_2-f)[\beta_2+\pi\beta_3(f_1+f_2)]\zeta$ and $\rho(z, f) = \frac{P(z)}{P(0)}$ is the normalized signal power profile[16]. This model assumes that, for the induced nonlinear perturbation of COI at frequency $f$, every frequency component in the triplet $(f,\ f_1,\ f_2)$ attenuates as $f$ during the nonlinear mixing during the FWM process[28]. To correctly account for arbitrary modulation formats, the extend GN (EGN) model was proposed in[32]. However, either the EGN model or the GN model **Equation (1.49)** is given as an integral format which requires extensive calculations and its complexity scales with the number of channels[16], which is not cost effective for QoT estimation. In this connection, the closed-form of GN model with the presence of ISRS was then proposed in[33,34], which calculate the nonlinear coefficient by:

$$\eta_i(f_j) = \eta_{\mathrm{GN},i}(f_j) + \eta_{\mathrm{corr},i}(f_j) \tag{1.50}$$

where $\eta_{\mathrm{GN},i}(f_j)$ can be approximated by[33]:

$$\eta_{\mathrm{GN},i}(f_j) \approx \sum_{i=1}^{n} \left(\frac{P_{i,j}}{P_j}\right)^2 \cdot \left[\eta_{\mathrm{SPM}_i}(f_j) n^{\epsilon} + \eta_{\mathrm{XPM}_i}(f_j)\right] \tag{1.51}$$

where the SPM and XPM contributions to NLI are denoted as $\eta_{\mathrm{SPM}_i}(f_j)$ and $\eta_{\mathrm{XPM}_i}(f_j)$ respectively, and $\epsilon$ is the coherent factor to account for the NLI accumulation with the number of spans $i$[16]. The closed-form expression of $\eta_{\mathrm{SPM}_i}(f_j)$ was given by[33]:

$$\eta_{\mathrm{SPM}_i}(f_j) = \frac{4}{9}\frac{\gamma_j^2}{B_j^2}\frac{\pi}{\phi_j \alpha'(2\alpha+\alpha')} \left[\frac{T_j' - \alpha^2}{\alpha}\operatorname{asinh}\left(\frac{\phi_j B_j^2}{\pi\alpha}\right) + \frac{A^2 - T_j'}{A}\operatorname{asinh}\left(\frac{\phi_j B_j^2}{\pi A}\right)\right] \tag{1.52}$$

where $\alpha'$ is an optimization parameter to characterize fiber loss and different from

**Table 1.1:** The value of $\Phi$ for typical modulation formats[16].

| Modulation format | Excess kurtosis $\Phi$ |
|---|---|
| QPSK | -1 |
| 16QAM | -0.6800 |
| 64QAM | -0.6190 |
| Gaussian distribution | 0 |

$\alpha$, $\phi_j = \frac{3}{2}\pi^2(\beta_2 + 2\pi\beta_3 f_j)$, $A = \alpha + \alpha'$ and $T_j' = (A - P_{tot}C_r f_j)^2$ where $C_r$ is the slope of Raman gain spectrum and $P_{tot}$ is the total Tx optical power[16]. And the closed-form of total XPM contribution was given by[33]:

$$\eta_{\mathrm{XPM}}(f_j) = \frac{32}{27}\sum_{k=1,k\neq j}^{N_{ch}}\left(\frac{P_k}{P_j}\right)^2 \frac{\gamma_j^2}{B_k\phi_{j,k}\alpha'(2\alpha+\alpha')} \left[\frac{T_k' - \alpha^2}{\alpha}\mathrm{atan}\left(\frac{\phi_{j,k}B_j}{\alpha}\right) + \frac{A^2 - T_k'}{A}\mathrm{atan}\left(\frac{\phi_{j,k}B_j}{A}\right)\right] \tag{1.53}$$

where $\phi_{j,k} = 2\pi^2(f_k - f_j)[\beta_2 + \pi\beta_3(f_j + f_k)]$[33]. Now, the closed-form expression of $\eta_{\mathrm{GN},i}(f_j)$ in **Equation (1.50)** has been obtained, which accounts for the non-linear contribution of Gaussian modulated symbols. Another term $\eta_{\mathrm{corr},i}(f_j)$ in **Equation (1.50)** works as a correction term to account for the NLI dependence on the modulation format, whose closed-form expression was given by[34]:

$$\eta_{\mathrm{corr},i}(f_j) = \frac{80}{81}\Phi\sum_{k=1,k\neq j}^{N_{ch}}\left(\frac{P_k}{P_j}\right)^2 \frac{\gamma_j^2}{B_k}\left\{\frac{1}{\phi_{j,k}\alpha'(2\alpha+\alpha')}\left[\frac{T_k' - \alpha^2}{\alpha}\mathrm{atan}\left(\frac{\phi_{j,k}B_j}{\alpha}\right) + \frac{A^2 - T_k'}{A}\mathrm{atan}\left(\frac{\phi_{j,k}B_j}{A}\right)\right] + \frac{2\pi\tilde{n}T_k'}{|\phi|B_k^2\alpha^2A^2}\left[(2|\Delta f| - B_k)\log\left(\frac{2|\Delta f| - B_k}{2|\Delta f| + B_k}\right) + 2B_k\right]\right\} \tag{1.54}$$

where $\Delta f = f_k - f_j$, $\phi = -4\pi^2[\beta_2 + \pi\beta_3(f_j + f_k)]L$ and $\tilde{n} = 0$ for a single span, otherwise $\tilde{n} = i$[16]. And $\Phi = \mathbb{E}[X^4]/\mathbb{E}^2[X^2] - 2$ is the excess kurtosis which takes different modulation format into account, whose typical values are given in the **Table 1.1**. Combining the above equations, the complete closed-form expression of **Equation (1.50)** can be obtained and thus the NLI noise can be also efficiently calculated using **Equation (1.47)**.

Considering that a lightpath transmitting through $N_s$ fiber spans, the analytical models introduced above can be used to estimate $P_{\text{ASE}}$ and $P_{\text{NLI}}$ so that the total GSNR can be obtained by:

$$GSNR = \left[\sum_{i=1}^{N_s}\left(\frac{1}{SNR_{\text{ASE},i}} + \frac{1}{SNR_{\text{NLI},i}}\right)\right]^{-1} \tag{1.55}$$

## II Machine Learning (ML)-based Models

A fundamentally different paradigm for QoT estimation has emerged with the rise of machine learning (ML) techniques. Instead of solving the complex physical equations discussed above, ML-based models can be trained to learn more intricate mapping between the physical layer features and the actual QoT values, which may not be possible when applying analytical models only. For example, the GN models to estimate the noise of NLI were obtained with specific assumptions and approximations that lead to several known inaccuracies, such as overestimating NLI in the initial span of a link. Meanwhile, the uncertainty of physical layer parameters monitored in the network degrades the accuracy of analytical model. To overcome those limitations, a wide variety of ML-based models or ML-enhanced models have been proposed over the past few years. In this section, the typical architectures and algorithms of the ML model are introduced.

Multi-layer perception (MLP) is a typical supervised ML architecture used to approximate any continuous function[11], which is essentially a feedforward neural network consisting of fully connected neurons with nonlinear activation functions. MLPs are composed of neurons called *perceptrons*. The general structure of a perceptron is illustrated in **Figure 1.16** where there are $n$ input features $\{x_1, x_2, ..., x_n\}$, each of which is associated to a weight $\{w_1, w_2, ..., w_n\}$. The function of $u(x) = \sum_i^n w_i x_i + b_i$ computes the weighted sum of input features, and also adds a bias term, then passes through an activation function $f$ to introduce nonlinearity, finally the output of a perceptron $y = f(u)$ can be obtained.

Activation functions $f$ are mathematical transformations applied to the output of

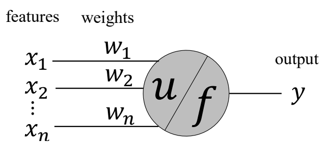



**Figure 1.16:** A perceptron with $n$ input features[11].

each perceptron and determine whether the perceptron should be activated. Common activation functions include (i) Sigmoid: squashes input values into a range between 0 and 1, useful for probabilistic outputs; (ii) Rectified Linear Unit (ReLU): outputs zero for negative inputs and passes positive inputs unchanged, promoting sparse activations and efficient training; (iii) Tanh: similar to sigmoid but outputs in the range -1 to 1. The choice of activation function impacts the learning capacity and convergence speed of the network.

The structure of MLP contains at least three layers: input layer, hidden layer and output layer, as is shown in **Figure 1.17**[11]. The input layer receives the initial features representing the physical layer information, such as optical signal power and center frequency, number of channels, fiber length, fiber loss and any other link characteristics. The hidden layer contains the perceptrons and is fully connected and placed between the input and output layers, which captures intricate patterns and relationships in the data that are not linearly separable. The choices of various depths (number of layers) and widths (perceptron per layer) influence the model's ability to represent complexity. Finally, the output layer provides the final prediction or estimate, such as a QoT metric like SNR or OSNR.

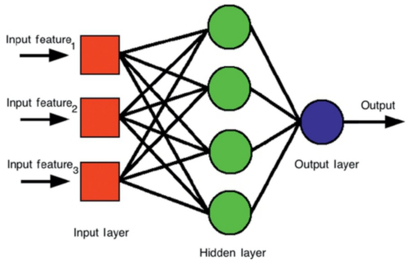



**Figure 1.17:** A typical three-layer MLP from[11].

Learning in MLPs involves optimizing its weights and biases to minimize the errors between predicted and actual outputs on the training dataset. This optimization is commonly achieved using the gradient descent algorithm, which iteratively adjusts the parameters by moving them in the direction that reduces the error. Variants like stochastic gradient descent (SGD) use mini-batches of data for efficiency. To build a reliable QoT estimation model, the dataset is typically divided into: 1)training set which is used to train the model by adjusting parameters based on input-output pairs, 2)testing set which is used to evaluate the model's performance on unseen data to ensure generalization and avoid overfitting. A well-designed split helps in assessing how the model will perform on real network scenarios. Meanwhile, hyperparameters are configuration settings external to the model, such as learning rate, number of hidden layers, perceptrons per layer, and activation function choices. Hyperparameter tuning involves systematically searching for the set of hyperparameters that yield the best performing model, often via techniques like grid search, random search, or Bayesian optimization.

Neural network (NN)-based ML methods like MLPs have demonstrated powerful approximation capabilities in various domains, including QoT estimation for optical networks[35,36]. However, QoT estimation or any other modeling tasks for optical networks still need to address some limitations, especially ones that are based on numerous physical layer data and still need to obey physical principles and contexts. Hence, one of the most significant limitations is the interpretability. NNs models are often viewed as "black-box" systems. Their decision process lacks transparency, making it challenging to gain physical insight into how inputs affect QoT metrics. Another significant limitation is overfitting, which means, if the NN architecture is excessively complex relative to the dataset, the model may overfit, capturing noise rather than actual physical principles and contexts. This impairs generalizability to unseen data. The limitations of interpretability and generalizability for NN-based models have thus motivated researchers to choose other ML methods like linear regression and curve fitting to be implemented as an enhancement of analytical model such as QoT estimation and

EDFA gain modeling. Linear regression is a fundamental approach for modeling the relationship between an independent variable $x$ and a dependent variable $y$. A simplest linear regression model is formulated as

$$y = mx + b \tag{1.56}$$

where $m$ represents the slope, and $b$ is the intercept. The goal is to find $m$ and $b$ that minimize the discrepancy between the predicted values $y$ and the actual observations/measurements. This discrepancy is commonly measured by the mean squared error (MSE):

$$J(m, b) = \frac{1}{n} \sum_{i=1}^{n} [(mx_i + b) - y_i]^2 \tag{1.57}$$

The least squares method is a popular strategy for identifying the best-fit curve by minimizing the MSE. Given data points $(x_1, y_1), ..., (x_n, y_n)$, the optimal slope and intercept are computed as:

$$\begin{aligned} m &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ b &= \frac{\sum y_i - m \sum x_i}{n} \end{aligned} \tag{1.58}$$

For more complex relationships, polynomial or nonlinear functions can be fit using least squares optimization by defining an appropriate model $f(x; \theta)$ and minimizing:

$$\min_{\theta} \sum_{i=1}^{n} [f(x_i; \theta) - y_i]^2 \tag{1.59}$$

The simple least-squares fitting method is foundational in physical layer modeling[37,38] when seeking simple, interpretable models with well-understood mathematical properties that complement or outperform neural networks in scenarios constrained by interpretability, computational resources, or limited data. For large or high-dimensional datasets, direct formulaic approaches can be computationally expensive. Instead, the gradient descent algorithm is employed. For example, for the cost function of **Equation (1.57)**, we can first compute the gradient:

$$\begin{aligned} \frac{\partial J}{\partial m} &= -\frac{2}{n} \sum_{i=1}^{n} x_i \left[(mx_i + b) - y_i\right] \\ \frac{\partial J}{\partial b} &= -\frac{2}{n} \sum_{i=1}^{n} \left[(mx_i + b) - y_i\right] \end{aligned} \tag{1.60}$$

Then the update rule in the $k^{th}$ iteration is:

$$
\begin{aligned}
m^{(k)} &= m^{(k-1)} - \alpha \cdot \frac{\partial J}{\partial m} \\
b^{(k)} &= b^{(k-1)} - \alpha \cdot \frac{\partial J}{\partial b}
\end{aligned}
\tag{1.61}
$$

where $\alpha$ is the learning rate controlling the step size of updates.

### 1.3.3 Data Aspects

In addition to the aforementioned NN structure and algorithm for training robust ML models, accurate and high-quality input data is essential to yield accurate ML model output, which is also true for obtaining accurate QoT estimates with analytical models[39]. However, raw data of physical layer are usually from performance monitoring in optical network or from datasheet specification provided by device/fiber vendors, which are not always accurate and may be the worst-case value, and even noisy, incomplete, contains outliers. In this connection, analyzing the uncertainty of input data and its impact on QoT results will also be a crucial step, before feeding them into the model, either in ML or in analytical. Possible solutions include applying some pre-processing or input parameter refinement methods[40,41]. In this section, we review the data source and parameter uncertainty in the optical fiber network, and introduce some possible solutions to deal with the data aspects for QoT estimation.

#### I Data Source in Commercial Optical Networks

In commercial optical networks, there are mainly two sources of data that can be used for QoT modeling and estimation: performance monitoring (PM) and device specification sheet. The performance of the network devices and the transmission status have to be periodically monitored to comply with SLAs. In a real network, the network controller provides control and management capability of the hardware equipment and the overall network, including configuration, performance monitoring (PM), and alarm management[15]. The standardized data model, YANG[42], can be used to provide a standard abstraction of equipment capabilities and functions[15]. The management and configurations of devices can be done by NETCONF[43] (Network Configuration Protocol), including Optical Transport Network (OTN) devices[15]. PM can be done through NETCONF[43] or streaming telemetry.

As shown in **Figure 1.3**, OCH layer data contain the optical connection's center frequency and baud rate that can be configured in optical terminals/transponders, which provide the PM data involving pre-FEC-BER, Q-factor, CD, differential group delay (DGD), polarization-dependent loss (PDL) and etc., with instant, average, maximum and minimum values during a period of 15 minutes. In the OMS layer, key PM data contain the actual gain and tilt, total input power, total output power from EDFAs, and channel power at signal bandwidth or power at 12.5GHz grid from OCMs, also with instant, average, maximum and minimum values during a period of 15 minutes.

Although dynamic network PM data constantly change with time, other static data from specification datasheet used to characterize the capabilities of the equipment may remain stable for extended periods, but may still degrade or age over time. These static data contain the noise figures (NFs) of EDFAs, fiber length and loss, back-to-back (B2B) curve of BER vs. OSNR, and etc. The amplifier NF is usually given in product specification provided by the vendor with a typical or worst-case NF value as a function of gain target only. The length of the optical fiber is generally provided by the optical cable supplier, but can be corrected using optical time-domain reflectometer (OTDR). The quality of optical fibers can degrade or even suffer interruptions due to various factors such as external pressure and bending. This often necessitates fiber re-splicing, which may alter the fiber length and introduce additional fiber loss. B2B curves of BER versus OSNR are typically measured under laboratory conditions, often through repeated testing of TRX modules sharing identical specifications from the same vendor. However, such test results may still be difficult to adequately characterize all TRX modules in non-ideal networks, given the substantial scale of deployed TRX modules and inherent manufacturing variations.

It should be noted that the calculation methods of the PM data reported by devices from different vendors may vary. For example, the calculation of the "actual gain" values of EDFAs provided by some vendors does not include the ASE noise power introduced by EDFA itself, while the actual gain values reported by EDFAs from other vendors do not exclude the ASE noise power added at the output power. On the other hand, some vendors of optical transponders do not provide complete four types of data: instant, average, maximum, and minimum. To achieve a more precise monitoring of time, it is necessary to rely on streaming telemetry to poll the device with a period as small as 1 second. This further introduces the challenge of how to preprocess data from commercial multi-vendor disaggregated optical network so as to construct a universal PLI model, which is the main research objective of this thesis and will be discussed in the following sections.

### II Input Parameter Uncertainty and Refinement Methods

Analytical QoT models achieve high accuracy with the assumption of accurate knowledge of model's input parameters[44]. In **Figure 1.18**, various physical parameters are illustrated including EDFA gain $G$, tilt $T$, noise figure $NF(G)$ as a function of gain $G$, total input power $P_{\mathrm{in}}^{tot}$ and total output power $P_{\mathrm{out}}^{tot}$; channel output power $P_{\mathrm{out}}^{i}$ from OCM; lumped loss $\alpha$ including VOA, insertion loss and connector loss; fiber loss $A$ and length $L$, fiber NLI coefficient $\eta_{\mathrm{NLI}}$; and the received optical power $P_{\mathrm{Rx}}$ at the receiver (Rx) side. Among these parameters, the optical power data may have different levels of precision. For example, the total power $P_{\mathrm{in}}^{tot}$ and $P_{\mathrm{out}}^{tot}$ are usually more accurate than the OCM power $P_{\mathrm{out}}^{i}$ that has a measurement error of ±0.8dB[45,46]. The lumped loss $\alpha$ is unknown and its associated uncertainties is hard to quantify in a live production network, but it can greatly affect the launched power into the fiber and thus the NLI noise and SRS effect. A common practice is to use a typical value of 0.5dB∼1.5dB in the QoT model.

One of the critical parameter uncertainties comes from the EDFA gain profile. Typ-

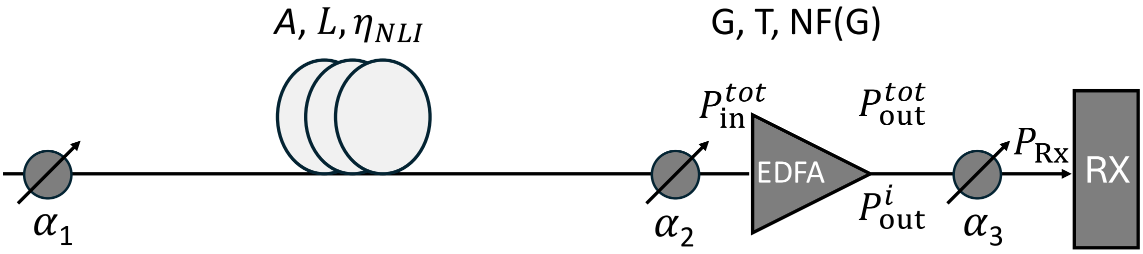



**Figure 1.18:** An optical fiber span with amplifier, fiber, and input/output lumped loss, and their corresponding physical parameters.

ically EDFAs are operated in Automatic Gain Control (AGC) mode for the terrestrial transmission network, each with a gain target $G$ to compensate for fiber attenuation and a tilt target $T$ to counteract the SRS power transfer effect. An ideal implementation of $G$ and $T$ of EDFA will result in an ideal gain profile without gain ripples. However, this is not true in practical cases where the input signal is not uniformly amplified by EDFA. Thus, the gain ripple $\mathbf{g_r}(f)$ always exists and is defined as the difference between target gain profile $G_{\text{target}}(f)$ and actual gain profile $G_{\text{actual}}(f)$. As presented in the hollow scattered points of **Figure 1.19**[12], gain profiles can be approximated by a first-order linear approximation using the target parameters of $G$ and $T$, while the residual gain ripples $\mathbf{g_r}(f)$ are difficult to obtain when only using two parameters: $G$ and $T$. In fact, the residual gain ripple $\mathbf{g_r}(f)$ is a result of gain flattening filter (GFF) and cross sections for stimulated emission $\sigma_k^e$ and absorption $\sigma_k^a$ at wavelength $\lambda_k$. In **Section I**, we have discussed the cross sections are related to the different materials of the doped fiber.

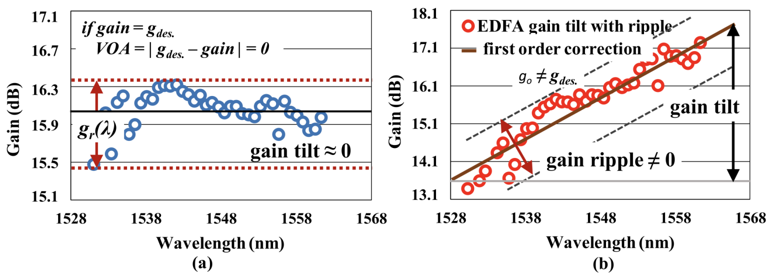



**Figure 1.19:** The residual gain ripples at different gain tilt, from[12].

To maintain a constant total gain, commercial EDFAs use a dual-stage structure to implement the gain control. The gain ripple has usually been considered zero with acceptable results in some cases[47]. Although the total gain after two stages EDF is controlled to be the target value, the actual response of gain and ASE still changes with variations in input signal power, which can induce a change in the gain ripples. Any inaccuracy in the input parameter of QoT estimation would result in a higher cost of network management and control. To obtain an accurate estimation of gain ripple, semi-analytical models[5,48,49] have been proposed based on pre-measured ripple profile under operating conditions. ML models have also been actively explored[41,50]. The related gain profile modeling approaches are summarized in **Table 1.2**.

Another uncertain parameter of EDFA is the noise figure (NF), generally assumed to be flat and only dependent on $G$. However, it is actually frequency dependent as a result of intrinsic characteristics of the doped fiber[51]. It also depends on different input power and tilt targets, as these situations mean that the amplifier has a different gain response (i.e., the population inversion of erbium ions) within the two-stage EDF[51]. To characterize a more exact NF, an empirical semi-analytical model has been proposed to characterize it as a function of $\{G, T, f\}$ in[52]. In addition, a high-order polynomial regression model was proposed to approximate the NF measurements as a function of $\{G, T, f, P_{\text{in}}^{tot}\}$[53]. The coefficients of those characterization models are obtained from the experimentally measured NF dataset. ML model can be also leveraged to learn a better NF from services already deployed, as validated in[54].

The lumped loss $\alpha$ is another type of uncertainty that significantly affects the estimation of NLI noise and signal power evolution along the fiber. Lumped loss is hard to be modeled exactly because of the lack of optical power monitor at the beginning and end of each span. In[55], connector loss at the input of each fiber span was evaluated with **GNPy**[56] as the QoT tool and the distribution of GSNR computation error was analyzed based on different assumptions on the value of connector loss. In[40,57], the lumped loss was estimated by minimizing the errors of power/SNR estimation using gradient descent algorithm. The distribution of insertion loss was investigated for minimizing the power estimation error in C+L scenario[58]. The uncertainties in fiber span lengths has been discussed which influence the accuracy of QoT estimation using machine learning[59].

**Table 1.2:** Uncertain input parameters for QoT modeling and the corresponding refinement methods proposed in recent years.

| **Uncertain parameters** | **Proposed refinement methods** | **Testing Data** |
|---|---|---|
| Gain profile | Semi-analytical model of $\mathbf{g}(G, T, f)$[49] | Experimental measurements |
| | Semi-analytical model of $\mathbf{g}(G, P_{\text{in}}(f), f)$[5] | Experimental measurements |
| | Analytical model of $\mathbf{g}(G, P_{\text{in}}(f), f)$ with a lookup table[48] | Experimental measurements |
| | ML models of $\mathbf{g}(G, T, P_{\text{in}}(f), f)$[12,41,50] | Experimental measurements |
| | Hybrid model integrating ML with analytical model[60] | Field trial |
| | ML model trained with current network state[12] | Experiments Simulations |
| Noise figure | Semi-analytical model of $\mathbf{NF}(G, T, f)$[52] | Experimental measurements |
| | Polynomial fitting model of $\mathbf{NF}(G, T, P_{\text{in}}^{tot}, f)$[53] | Experimental measurements |
| | ML model to refine the **NF** from datasheet[54] | Simulation |
| Lumped loss | Pre-measured connector loss in the lab[55] | Live network |
| | ML models to refine lumped loss by minimizing the end-to-end power/SNR estimation errors[40,57] | Experimental measurements |
| | Dichotomous algorithm to refine lumped loss by minimizing power estimation errors[58] | Simulation |
| OCM power | ML models to refine OCM power by minimizing the end-to-end power/SNR estimation errors[40,57] | Experimental measurements |
| Filter penalty | ML model trained with current network state[12] | Experiments Simulations |
| NLI coefficient | ML model trained from precomputed per-channel SCI values[61] | Simulation |
| $P_{\text{Rx}}$ induced SNR penalty | Analytical model with linear least squares fitting[62] | Experimental measurements |
| | Analytical model with linear least squares fitting[38] | Experimental measurements |

We summarize different types of uncertain parameters and the refinement methods proposed in recent years in **Table 1.2**. As we can see, most of the proposed parameter refinement methods utilize data-driven approaches or models built from pre-measured datasets. Typically, the input parameter refinement process depends heavily on data-driven ML techniques[12,40,41,54,57,63] or pre-measured datasets obtained from controlled laboratory environments[5,49,52,53,55]. Data-driven methods such as ML leverage actual network measurements collected during operation, enabling models to adapt dynamically to real-world variations and degradations. This approach ensures the input parameters such as signal power, amplifier gain, NF and lumped loss, reflect the current state of the physical network, supporting more accurate QoT predictions. Alternatively, laboratory pre-measured data play a crucial role in establishing baseline performance characteristics under well-defined conditions. Models[5,49,52] built or calibrated using such data benefit from a solid foundational understanding of underlying physical phenomena. Certainly, this pre-measured data and the corresponding characterization model may still involve some degree of error because even if the laboratory equipment is of the same specification as that currently deployed in the network, subtle differences may still exist. Combining data-driven refinement method with models built from laboratory measurements often yields robust and reliable QoT estimation. The dual approach balances adaptability to actual network conditions with the physical insights derived from controlled settings.

However, only a few related works have demonstrated the feasibility of their methods in commercial live networks, so the applicability of the proposed approaches in **Table 1.2** remains an open question. This gap highlights the challenges of translating laboratory or simulation-based models into dynamic networks, where factors such as channel add/drop, component aging, and environmental variations introduce complexities not fully captured by existing datasets. Furthermore, the scalability and real-time adaptability of these refinement techniques are critical considerations for their deployment in live networks. Many data-driven methods require extensive measurement campaigns or retraining to maintain accuracy, which can be resource-intensive and impractical for large-scale systems. Consequently, there is an increasing interest in hybrid approaches that integrate physical modeling with data-driven updates, aiming to balance model precision and operational efficiency. In summary, while significant progress has been made in refining input parameters for QoT estimation, further research is needed to improve the robustness, scalability, and real-world validation of these methods, ensuring they meet the stringent demands of commercial optical network management.

## III Data Preprocessing and Feature Selection for ML training

In the previous section, we discuss the input parameter uncertainties and refinement methods for analytical model. For ML-based QoT models, data preprocessing steps such as cleaning, balancing, and normalization become more critical to achieving a reliable model with better accuracy and generalization capabilities. First, data cleaning involves identifying and removing outliers, inconsistencies, or missing values in the dataset to ensure the training data accurately reflect the real-world conditions[39]. This step is essential because noisy or incomplete data can mislead the learning algorithm, resulting in poor model performance. Data balancing addresses the issue of uneven representation of different classes or parameter ranges within the dataset. For instance, if certain QoT degradation scenarios are underrepresented, the model may become biased towards more common cases, reducing its effectiveness in predicting rare but impactful events. Techniques like oversampling, undersampling, or synthetic data generation can be employed to mitigate this imbalance[64]. Normalization transforms the input features into a consistent scale, preventing variables with larger numeric ranges from dominating the learning process. This adjustment helps the optimization algorithms converge more efficiently and avoids skewed model behavior[39]. These data preprocessing approaches enable ML models to better capture the complex relationships underlying QoT and improve their predictive robustness across dynamic networks.

The selection of input features is another critical aspect for obtaining a reliable ML-based QoT model[65]. Feature selection involves identifying the most relevant types of parameters that significantly affect QoT in the available dataset. The features used for QoT estimation are typically drawn from two categories, reflecting the hierarchical nature of a lightpath[66]: 1) Span-level features: describing the properties of the individual fiber span. For each fiber span, this could include its length, fiber type, loss coefficient, and dispersion parameter. For each amplifier, it could include its type (e.g., EDFA), gain and tilt setting, and noise figure. 2) Channel-level features: describing the OCH itself and its spectral occupation, including the modulation format, baud rate, launch power, and information about adjacent channels (e.g., their power and spectral distance from the CUT).

Choosing appropriate features is essential to reduce model complexity, improve training speed, and enhance predictive performance by minimizing the inclusion of redundant features. In **Table 1.3**, we summarized some related works on ML-based QoT model with a focus on their selections of input features and sources of dataset. It can be seen that these works have significant variation in the selection of input features due to different network operational scenarios they are considering. In the physics-based analytical QoT model, channel-level features such as signal power and span-level features such as amplifier gain, are extremely crucial parameters and their values directly determine the magnitude of GSNR. However, these parameters are not included in the input features in some ML models based on simulated datasets such as[67,68]. Although it might be useful when none of the channel and span information is available, the reliability and generalization ability of their models are still questionable. In particular, the authors in[35] chose the same input features as the input of the analytical model. Their results show that ML is better at estimating side channels upon optimizing the model parameters with the same set of experimental data, but it is more difficult for ML models to generalize to different link configurations and is less stable with time[35]. Recent research has explored the use of Explainable AI (XAI) techniques, such as Shapley Additive Explanations (SHAP), for managing the trade-off between the richness of the feature set and the practical costs associated with it[65,69]. SHAP can quantify the contribution of each individual feature to the model's predictions, allowing network engineers to identify and discard less informative features. This process can lead to simpler, more efficient models that speed up inference and even improve predictive accuracy. Consequently, effective input feature selection is a key step toward robust and scalable ML-based QoT estimation frameworks in commercial optical fiber networks.

**Table 1.3:** Details of papers utilizing regression task to estimate the QoT of lightpath

| Ref | Model | Input Features | Output | Data Source | Performance |
|---|---|---|---|---|---|
| 35 | ANN | Per channel launch power, per amplifier [gains, NFs, gain tilts], per span input power | SNR | Experimental | Error STD $< 0.14$dB |
| 70 | ANN | Launch power, laser bias, per amplifier [input power, output power] | Q-factor | Experimental | RMSE $< 0.02$dB |
| 71 | ANN | CUT [symbol rate, number of neighbor channels, transmit power, distance to channel], span length, number of WDM channels/spans, total used bandwidth, average power level | SNR | Trained on synthetic and applied real data | MAX error $< 0.5$dB |
| 72 | CNN, ANN | Per channel [Power, ASE, NLI, number of spans, total length] | GSNR | Simulation | MAX error: 0.37dB |
| 73 | TL | average power of background channels, position of the testing channel, number of active background channels | BER | Simulation | $> 95\%$ accuracy |
| 67 | GP, AL | number of links, lightpath length, max link length, traffic volume and modulation format | BER | Simulation | Reducing 75% training data |
| 68 | AL | total length, the length of longest link, number of links and spans traversed, transmit power | GSNR | Simulation | Best: 0.01 MSE |
| 74 | CNN | Span features, amplifier features, channel features | GSNR | Experimental | Error STD $< 0.25$dB |
| 75 | Composable ANN | Span features, amplifier features, channel features | Average GSNR | Simulation, experimental | 0.83 STD, 1.5 MSE |

### 1.3.4 Low-margin Optical Network

Since each impairment in the physical layer results in deterioration of the QoT of the received signal, network operators should provide enough QoT **margin** to guarantee the normal performance of optical channels throughout the lifetime of the network. In this section, we introduce the concept and considerations of low-margin optical network and how it is associated with the QoT modeling and estimation.

The margin of a light path is defined as the difference between the actual QoT metric (e.g., SNR, OSNR or BER) of a light path, and the BER threshold above which the signal is deemed recoverable "error-free" (i.e., the forward error correction (FEC) limit). In the greenfield scenario, optical networks are usually designed with a high OSNR margin to account for the inaccuracy of QoT estimator of a lightpath to be deployed.

For a brownfield network, the actual margin of a deployed lightpath still remains high in order to account for potential OSNR degradation because of time-varying impairments, e.g., fiber loss degradation, fiber and amplifier aging, increasing channel loadings and etc. The evolution of the margin over time is shown in **Figure 1.20**[13] where we can see that SNR margin is overly high at the network planning phase, but decreases with time during network operation due to aging, fiber bends/cuts, increasing network loadings and thus NLI noise. Even at the end of life (EoL), SNR margin is still much higher than the FEC limit so that the SLA is met during the full life cycle of the network.

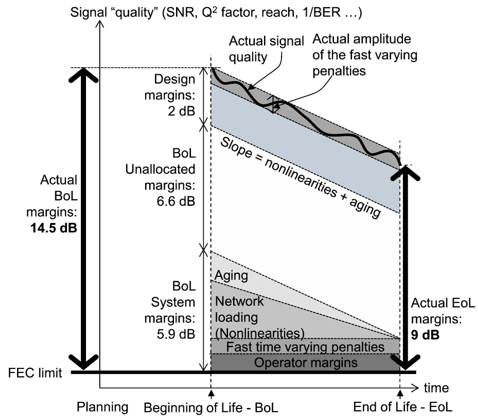



**Figure 1.20:** Types of network margins and their evolution with time, from[13].

Nonetheless, as in any industry, margins come with a cost, as margin represents that network capacity and resources have not been fully utilized, which is not desirable or cost-effective for network providers[13]. On the one hand, an excessively high network margin indicates that a significant amount of capacity and bandwidth have been wasted. In other words, the over-allocated OSNR margins can be translated into additional optical channels which could have been established/served for more connection demands, or higher modulation-format/data-rate of existing channels which could have been upgraded. On the other hand, the delivered OSNR and OSNR margin of a lightpath cannot be enhanced without limitations. It is physically limited by the signal power constraint, which is subject to the total output power constraint of EDFAs and fiber nonlinearity. Therefore, those trade-offs between network margins and capacity lead to network design and optimization/re-optimization problems possibly with various objectives such as maximized capacity or performance. Network margin can be categorized into two types: (i) design margin and (ii) system margin, based on their distinct characteristics and network scenarios. In this section, a comprehensive explanation of their definitions, network application scenarios, and optimization methods of lowering these margins will be presented.

## I Design Margin

In[13], design margins (D-margins) are defined as the difference between the planned beginning of life (BoL) QoT value (e.g., SNR) and the real value of the QoT monitored after network deployment. This difference is the result of inaccurate QoT estimators used during the phase of network planning, stemming from (i) the inaccuracy of input parameters of QoT estimation model and (ii) the inaccuracy of the QoT estimation model itself. Unallocated margins are defined as the capacity margins as a result of the mismatch between the capacity demand and the TPDs under consideration, but it can be minimized by utilizing rate-flexible transponders (flex-TRXs) with adaptive data rates and modulation formats in the network.

To reduce D-margins and improve the utilization rate of available bandwidth and capacity resources, an accurate QoT estimation/prediction is the key to characterizing QoT degradation over fiber spans. Whether in an analytical or black-box manner, the model of QoT should have the capability of abstracting the optical signal power variation and its accumulated impairments along the traversed optical fiber and devices, caused by a variety of linear and nonlinear phenomena. However, a significant uncertainty comes from the parametric characterization of fibers/equipments that is hard to establish before network deployment, which will be the input paramters of QoT estimation models. Even if accurate input parameters can be obtained in a laboratory environment, devices of the same specification deployed in the live production network, such as EDFAs, may still exhibit variations in noise figures. Moreover, differences in fiber quality, such as those caused by bending and Rayleigh scattering, further contribute to performance variability and thus estimation inaccuracies. Therefore, taking laboratory-derived parameters as a baseline, it is essential to learn and refine input parameters in QoT tool with real-time monitoring parameters from live production network. Before deployment, the initial network demand set $\{D_0\}$ is only designed with a possibly high design margin $m_0$ to incorporate the uncertainty $\epsilon_0$ of network equipment characterizations $E_0$[13]. After network/service deployment, one can benefit from the advanced Software-Defined Networking (SDN) controller to access the real-time monitoring parameters such as pre-FEC BER from coherent TPDs and optical channel power from optical channel monitorings (OCMs), which means the ground-truth D-margins is now available. Then, a closed-loop parameter training/learning process is enabled for reducing D-margins[13]. On the $k^{th}$ upgrade or service provisioning, i.e., a new demand set of $\{D_k\}$ arrives, a "training" process is firstly launched based on existing monitoring dataset to refine the physical layer parameters and reduce their uncertainties $\epsilon_k^{'} < \epsilon_k$; on the next step of "prediction", network operators can use a lower D-margin $m_{k+1} = f(\epsilon_k^{'}) < m_k$ to upgrade service or establish a new service. These two procedures, in turn, enable the reduction of uncertainty in the characterizations of the equipment $(E_{k+1} + \epsilon_{k+1})$[13]. The closed-loop process then repeats for each arrival of new demand and finally minimizes the D-margins of optical network for the entire life cycle[13].

## II System Margin

The system margins (S-margins) are usually specified by network operators to account for time-varying impairments such as aging, increasing fiber losses due to splices to repair fiber cuts, channel add/drop operations, polarization effects and etc., which are frequently seen but not predictable in real optical networks[13]. These impairments might lead to quality of service (QoS) issues and even fail to meet the SLA for customers. In this connection, S-margins define the minimum QoT value of an optical signal to be met at BoL.

S-margins are only available after network/service deployment, that is, in a brownfield scenario. Since it accounts for time-varying impairments, considerable power margins should be provided to enhance QoT margins for guaranteed physical-layer performance. This requires an optimized power allocation and adjustment strategy/algorithm, which can be integrated in the centralized SDN controller and respond promptly to immediate impairments. However, as mentioned before, most of commercial EDFAs has a saturated output power around 23dBm and power enhancement can lead to higher NLI noise. If the power margin is used up or reaches its limit, one may consider decreasing the order of modulation format or bit rate in TPDs to guarantee the QoT margins. The prompt solution of an RMSPA (routing, modulation format, spectrum, power allocation) optimization problem is expected to lower the S-margins in a dynamic live production network[13]. Therefore, instead of increasing system margins, one may think of introducing efficient, dynamic and automated, impairment-aware cross-layer control to network management system as an important alternative solution to reduce S-margins[76]. This also necessitates an accurate QoT evaluation tool to decrease the outage probability, and in turn decrease the cost of the transport network infrastructure.

### 1.3.5 Intelligent and Automatic Network Control and Management

A critical aspect in low-margin optical network operation is to build an intelligent and automatic network control and management plane, introduing greater operational dynamism and efficiency. Guided by the principles of Software-Defined Networking (SDN), the ability of network controller to make intelligent decisions regarding resource allocation and network management is paramount. The modules for routing, capacity planning, monitoring, (re)optimization and failure management are essential components in the centralized network controller, their decisions depends entirely on the reliability and fidelity of the physical-layer performance evaluations it receives[58]. As shown in **Figure 1.21**, physical layer modeling serves as a vital central functionality, providing the physical layer awareness necessary to transform the control plane from a reactive, rule-based system into a proactive, predictive, and truly autonomous entity[77]. The QoT estimator does not operate in isolation; rather, it serves as a foundational service for other critical control plane modules, creating a dependency hierarchy where its performance dictates the effectiveness of higher-level functions.

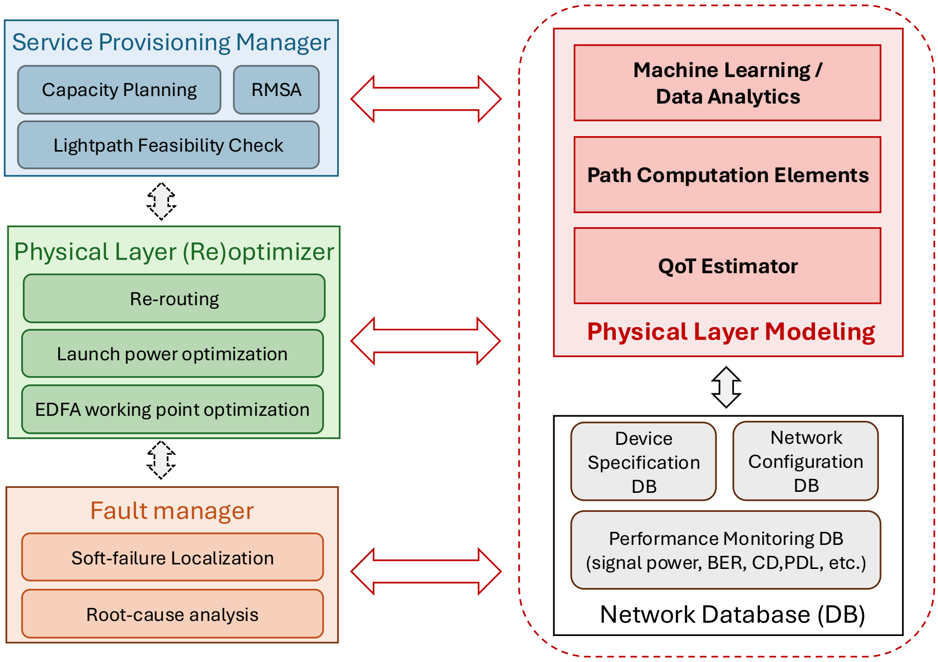



**Figure 1.21:** A fundamental structure of automatic network control and management plane, which comprises modular functions and their interactive relationships, with physical layer modeling serving as the most critical central module[14].

The module of service provisioning manager[14] is responsible for capacity planning, Routing, Modulation, and Spectrum Assignment (RMSA), and feasibility check. Capacity planning undertakes the evaluation of the available transmission capacity between a target source-destination pair. This process involves assessing the current utilization, the filling ratio of WDM channels, and the fragmentation ratio along the targeted fiber links. RMSA computes and allocates the routing and spectral resources for a new service request. However, a computed path is only a candidate until its physical feasibility is validated. The physical layer modeling with QoT estimator and data analytics acts as the gatekeeper in this process. If the predicted QoT (e.g., SNR or BER) meets the required margin threshold, the path is validated, and resources are provisioned. If not, this module must consider an alternative path[77] or enable the physical layer (re)optimizer module to achieve a higher QoT margin.

The fault management reacts proactively and analyzes the alarms generated in the PM DB, and provides rapid intervention to maintain service quality. In this process, the identification and localization of the root cause can be achieved by calling the path computation elements and QoT estimator as powerful diagnostic tools. A significant and persistent discrepancy between the QoT predicted by a trusted model and the actual QoT can provide a crucial clue for the fault management module to pinpoint the source of the degradation.

The physical layer (re)optimizer is also an indispensable functionality to enable automated network control and management. In dynamic networks, time-varying fiber loss necessitates dynamic gain re-adjustment in subsequent EDFAs. Fluctuations in the number of transmitted OCHs require adaptive tilt optimization in these amplifiers to counteract the SRS effect. However, excessive signal power amplification risks introducing significant NLI impairments, which necessitates optimization of launch power profile. This operational paradigm considering many trade-offs therefore demands iterative optimization of network configuration parameters combined with continuous QoT estimator invocation to identify latent optimal solutions for performance enhancement.

Ultimately, the goal is a form of closed-loop, autonomous network management where the network can self-optimize: it plans new connections with accurate QoT prediction, verifies, and monitors them in real time, and adapts to any changes or degradations with minimal human intervention. Achieving this requires the synergy of all the modules described and, fundamentally, a highly reliable physical layer modeling mechanism at its heart. Physical layer modeling should be not only accurate and intelligent, but also computationally efficient enough to operate within the stringent time constraints of dynamic, real-time applications.[78].

## 1.4 New Challenges of QoT Modeling and Estimation in Optical Fiber Networks

This chapter delineates recent evolutionary trajectory of commercially deployed large-scale optical networks. The discussion further elucidates the accompanying technical challenges in the modeling of the physical layer precipitated by these developments, which constitutes the primary research motivation for this study.

### 1.4.1 The Challenge in Open and Disaggregated Optical Network

The recent trend of large-scale commercial optical networks is openness and disaggregation. Different levels of optical network disaggregation are shown in **Figure 1.22**. Traditionally, the optical fiber network is composed of "closed transmission systems" where the system is proprietary and from a single vendor, including ROADMs, EDFAs, coherent TRXs, and associated control and management software[15]. The shifting of the optical network architecture from closed and proprietary to openness and disaggregation is mainly because of[15]: 1) advanced coherent TRXs which simplifies the design of transmission system and enables the decoupling of "optical line system" from TRXs; 2) the development of SDN which promises network operators to develop their own network management softwares; and 3) the great demand of DCI networks that grow much faster than telecom networks, require much larger bandwidths and have a much shorter technological iteration cycle[15].

Network disaggregation decouples traditional closed single-vendor transmission systems into individual functional blocks, enabling network operators to choose the best-in-class components from different vendors and thus breaking vendor lock-in and possibly reducing capital expenditure (CapEx)[14]. This "open module system" architecture is also called a white-box or **fully-disaggregated** network architecture[79]. Between

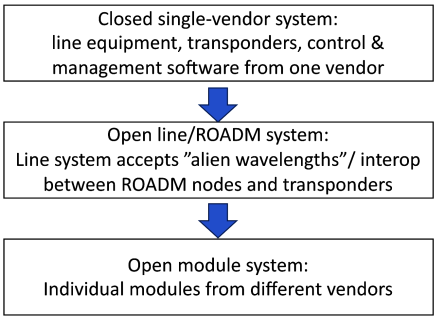



**Figure 1.22:** Different disaggregation levels of optical transmission systems, from [15].

these two extremes is the "open line system" architecture, which means that a line system (e.g., EDFAs) belongs to a single vendor but accepts "foreign" or "alien" wavelengths [15]. The network with multiple open line systems (each from different vendors) is also called as a **partially-disaggregated** optical network, which is the most widely deployed network architecture at present.

The multi-vendor disaggregation context promises greater flexibility and reduced costs, however, imposes a lot of burdens in the modeling of physical layer and the overall network behavior. The reason is that, conventionally, any variation of network performance can be well understood and fully managed by the vendor who designs and manufactures these devices and owns the network management and controller. However, for open and disaggregated network, network components work as a black-box and the behind physical principles and logic are blind to the network operators, further adding the uncertainties of physical modeling and parameter characterizations in such heterogeneous and disaggregated network. For instance, transponders from different suppliers will use proprietary DSP algorithms and third-party components, leading to subtle but significant variations in performance even for the same nominal configuration [14]. Similarly, EDFAs and ROADMs from different vendors will have unique performance characteristics, i.e., signal power measurements, gain flatness, NF profile, and filtering penalty. This diversity makes it extremely difficult to create a single, accurate physical layer characterization model. Inaccurate performance characterization and the difficulty of identifying precise parameters for every component in a multi-vendor network are primary sources of QoT estimation errors [80]. In particular, this operational reality requires the development of vendor-specific PLI models by network operators, thereby enabling granular performance awareness and control of the network. Such an implementation demands efforts such as extensive and rigorous experimental validations, possibly integrated with data-driven reverse-engineering methodologies to establish high-fidelity physical layer abstractions.

### 1.4.2 The Challenge of Green-field QoT Estimation

In live production optical fiber networks, QoT estimation heavily relies on precise knowledge of channel parameters, among which the optical power per channel is crucial. Typically, this parameter is monitored in real time by optical channel monitorings (OCMs), enabling accurate evaluation of signal impairments and adjustments in network management systems. However, this approach becomes unrealistic when applied to **green-field** network deployments or newly provisioned lightpaths where OCM readings are either not available or not feasible. In a green-field scenario, absence of historical data and live measurement feedback creates a void in crucial monitoring information. The lack of direct per-channel power measurement means that any QoT estimation results definitely have significant uncertainties, which is undesirable for network operators. This uncertainty complicates nonlinear impairment estimation, OSNR prediction, and margin calculation, all of which are essential for optimizing transmission performance and maintaining robust level of service.

One conventional approach to mitigate this challenge involves assuming ideal or flat gain profiles in optical amplifiers, but such over-simplifications fail to capture the dynamic behaviors encountered in real networks. The variation in amplifier gain across the wavelength spectrum, influenced by factors including input channel power and amplifier settings, introduces additional uncertainty into the QoT estimation process. In particular, EDFAs, which are ubiquitous in long-haul and metro networks, exhibit gain spectra sensitivity to total input power and channel power, making simplistic models insufficient. On the other hand, as discussed in **Section 1.2.3**, there are ample flexibility for vendors to design EDFA with different EDF section lengths and partitioning, various pump methods and etc. The modeling of multi-vendor EDFAs becomes more chanllenging for open and disaggregated network in which the EDFA design information is not available to network operators.

To bridge this gap, an effective solution involves developing an accurate, adaptive and vendor-agnostic EDFA gain profile model that can infer channel power distribution in the absence of direct OCM measurements. By leveraging physical principles governing amplifier dynamics combined with network configuration data, such a model can estimate the gain spectrum of EDFAs deployed along the transmission path. This inferred gain profile serves as a foundation for estimating channel powers post-amplification, thus enabling more precise QoT predictions even in green-field deployments.

### 1.4.3 The Data Dependency Challenge

The greatest impediment to the widespread validation and adoption of ML-based QoT estimators is the scarcity of large, public, and well-structured datasets from real-world deployed networks[35]. While the ML paradigm is built on the premise of learning from data, the field of optical networking has historically lacked the data-sharing culture and infrastructure necessary to support it. This "data scarcity" problem may lead to:

- A crisis of reproducibility and comparability: individual research groups are forced to generate their own data because of the absence of common benchmark datasets, typically through proprietary and time-consuming SSFM simulations or limited-scope lab experiments. This practice makes it extremely difficult to reproduce published results and nearly impossible to perform fair, head-to-head comparisons of different ML models proposed in the literature[81]. A model that performs well on one custom dataset may not perform well on another, making it difficult to assess true progress in the field.

- Barriers to broader ML community engagement: the lack of accessible, high-quality data prevents the wider machine learning community, i.e., experts in algorithm development who may not have access to optical network testbeds, from contributing to solving the QoT estimation problem. Public datasets have been instrumental in driving rapid innovation in fields like computer vision and natural language processing, a catalyst that is currently missing in optical networking[81].

- Hindrance to operator trust and adoption: for a network operator, deploying an ML model in a mission-critical control plane is a high-stakes decision. The inability to independently validate a model's performance against a standardized, trusted dataset extracted from real network creates a significant barrier to adoption. Operators are understandably hesitant to rely on black-box solutions whose performance claims cannot be objectively verified, contributing to the continued reliance on established legacy models[82].

## 1.5 Research Objectives and Organization of the Thesis

Motivated by these problems, the research objectives of this thesis are to investigate the input refinement methods for accurate QoT modeling and estimation in commercial large-scale long-haul optical network and data center interconnection (DCI) network, the parameter characterization methods for accurate QoT modeling and estimation in green-field network, and the publicly available dataset extracted from commercial network to benchmark the performance of different QoT models. In particular, the following aspects will be investigated:

**A)** To improve the accuracy of QoT estimation for commercial long-haul network: we studied GN model-based QoT estimation in a large-scale disaggregated and partially loaded live production long-haul network with monitored physical layer data spanning across 8 months. Vendor-agnostic input parameter refinement methods of signal power measurements, gain and noise power profiles of each inline amplifier were proposed to improve QoT estimation accuracy. A data-driven parameter optimizations to incorporate intractable location-specific and spectral effects in the network were also investigated using simple gradient descent algorithm.

**B)** To improve the accuracy of QoT estimation for commercial DCI network: we studied different models of QoT estimation in a large-scale disaggregated and partially loaded live production DCI network with monitored physical layer data spanning across 3 months. We first studied an artificial neural network (ANN) model to incorporate inherent transponder effects for accurate QoT estimations. The effectiveness of such ANN-based QoT model has been demonstrated in optical line protection (OLP) switching scenarios that are common in metro DCI network. In addition to ANN-based model, we also investigated the vendor-agnostic input parameter refinement methods for improving the accuracy of GN model-based QoT estimation. These vendor-agnostic methods include conducting extensive experimental B2B measurements of 400Gb/s coherent transceivers to model inherent transponder effects at different optical signal powers and to obtain the input power dependent OSNR-SNR model, and extensive experimental measurements to characterize NF as a function of gain, tilt, frequency and input power for EDFAs from different vendors.

**C)** To extend the applicability of QoT estimation model to green-field scenario: the biggest obstacle to applying the aforementioned improved QoT models in green-field scenario is the lack of OCMs so that realistic per-channel signal power is not available. Extensive experiments were conducted to measure the gain profiles under different gain and tilt target, input power for 9 types of EDFAs from 2 vendors, which has been in operation in the considered DCI network. We proposed a vendor-agnostic polynomial model to characterize the gain profile under different gain and tilt target, input power, with minimized RMSE and validation for improved QoT estimation in live production DCI network.

**D)** To benchmark the performance of QoT estimation with different models on real network dataset: we provide open-source dataset from *Alibaba* production optical network on GitHub, which consists of 4 OCH groups and 25 OCHs. All OCH groups have data available for 162 hours, with one data sample per hour. More than 10,000 G-OSNR can be calculated using our open-source data, thus benchmarking the performance of QoT estimation with different models and contributing to the data-sharing culture in our optical network community.

The organization of this thesis is as follows:

In **Chapter 1**, we first briefly introduce the technical background and development of the optical network and the motivation of the thesis in **Chapter 1.1**. The fundamentals of transmission impairments, amplifier physics and receiver noise in fiber-optic communication systems are elaborated in **Chapter 1.2**. Then, the fundamentals of QoT models and its data aspects, and low-margin optical network control, automation, and management are detailed in **Chapter 1.3**. Next, the main challenges of QoT modeling in an open and disaggregated live optical network and a green field network, and the challenge of data dependency and scarcity are given in **Chapter 1.4**. Finally, the objectives and organization of this thesis are presented in **Chapter 1.5**.

In **Chapter 2**, effective methods of improving the accuracy of analytical model based QoT estimation in a long-haul large-scale live production network are proposed and validated. An overview is firstly provided in **Chapter 2.1** with the background and literature review. Then, the details of the link model and network database understudy are given in **Chapter 2.2**. The proposed methodology is introduced in **Chapter 2.3**, including refinements of the input parameters described in **Chapter 2.3.1**, and the data-driven optimizations described in **Chapter 2.3.2**. The results of margin reduction enabled by improved QoT estimation are discussed in **Chapter 2.4**. A summary in **Chapter 2.5** concludes this Chapter.

In **Chapter 3**, an overview is firstly provided in **Chapter 3.1** with the background and research question of QoT modeling for large-scale metro DCI networks. ML-based

QoT modeling and validation in OMSP scenario for disaggregated DCI networks are given in **Chapter 3.2**. The improved QoT estimations based on analytical model and experimentally characterized transponder penalty and amplifier noise figure for disaggregated DCI networks are discussed in **Chapter 3.3**. A summary in **Chapter 3.4** concludes this Chapter.

In **Chapter 4**, an overview is firstly provided in **Chapter 4.1** with the background and research question of greenfield QoT modeling and EDFA gain profile modeling. Then, the details of experimental setup and multi-vendor EDFAs is provided in **Chapter 4.2**. The methodology and results of building an input pre-emhpasis aware gain model are provided in **Chapter 4.3**. The validation of greenfield QoT estimation using the proposed gain models is presented in **Chapter 4.4**. A summary in **Chapter 4.5** concludes this Chapter.

In **Chapter 5**, an open-source network dataset containing real-time performance data, network configurations and topology from several segments of a commercial large-scale long-haul network is provided and described.

In **Chapter 6**, a conclusion is given for the thesis and all the proposed QoT models and methods, and a prospect for further study on enhanced physical layer modeling for commercial optical fiber networks is discussed.

# 2

# Improved QoT Estimation for Commercial Long-haul Optical Network

## 2.1 Overview

In this Chapter, we proposed efficient methods to improve the accuracy of QoT estimator applied in a large-scale long-haul live network. Quality of transmission (QoT) estimation is a common tool that can be applied to evaluate the performance of in-service lightpaths in optical transport network. The accuracy of QoT estimation hinges on both accurate models and accurate input parameters. In general, QoT estimation models being studied can be divided into analytical models[10,32,33,56] and machine learning (ML)-based models[36,61,83,84]. From a practical point of view, network service providers prefer to use analytical models instead of ML-based models because analytical models originate from physical modelling of the optical layer including the linear and nonlinear impairments[85] and are hence more interpretable and scalable[86]. It is known that standard erbium-doped fiber amplifier (EDFA) noise models and Gaussian noise (GN) model[10] and its variants[33] can well predict amplified spontaneous emission (ASE) noise and nonlinear interference (NLI) to accurately calculate the generalized signal-to-noise ratio (GSNR)[56]. Consequently, the accuracy of QoT estimation using analytical models is mainly limited by the accuracy of the input parameters i.e. signal power measurements, fiber loss, amplifier noise figures, etc. This is particularly problematic for deployed links where network equipment aging, link maintenance and other unaccountable factors in practical settings add to the difficulty in obtaining accurate physical parameters[87].

In a parallel development, techno-economic developments over the past decade in optical communications technology has led to the evolution from closed and single-vendor optical transport networks to disaggregated networks composed of multi-vendor open line systems[15,56]. Network disaggregation has the advantages of operational flexibility and cost reduction but also has additional complexity[86]. One example is the complexity of system parameters characterizations such as signal power measurements of the amplifiers, optical channel monitors (OCM) and other link components come from different vendors with varying levels of inaccuracies and even for the same model of the same vendor, there exist statistical fluctuations of their performance and parameter measurement accuracies. Large-scale dataset can be used for system calibrations and improving QoT estimation accuracy, but they are usually proprietary to optical hardware vendors[63] and not network operators. Another issue is the incomplete knowledge of parameters such as amplifier gain spectra, actual fiber length, etc., in a live network due to the limited ability of actual monitors and practical limitations in estimating these parameters before deployment. To address the aforementioned challenges, researchers have proposed ML-based or GNPy-based input parameter uncertainties reduction techniques especially for power level and noise figure of amplifiers[54], span length[59], and connector loss[55]. Other studies have specifically proposed vendor-agnostic QoT estimators for disaggredgated networks with the utilization of ML[83,88] or GNPy[89,90]. However, the open-source GNPy project which interfacing GN model to practical system has assumed fully-loaded conditions until recently[91]. Before that, applying it in partially loaded networks results in conservative QoT estimates, introduce additional system margins and unnecessarily reduce network capacity[86]. On the other hand, although ML can be used to address this issue as long as good dataset is available, a common practical concern is the specificity of the developed ML algorithm to the dataset and the validity/scope of applicability with network growth and loading/configuration changes. In addition, the performance of any ML-based QoT estimation hinges on the accuracy of input physical parameter accuracies[92]. Therefore, on a more fundamental level, it is imperative to consolidate all the real-time monitored data and various component specification information to minimize parameter uncertainties in analytical model and improve QoT estimation in disaggregated live production network settings[93]. Such uncertainty minimizations can include but are not limited to refining the measured physical parameters, using the measured parameters to estimate other unknown parameters through well-understood analytical models and/or additional data-driven optimizations to incorporate intractable impairments into the model, which is particularly relevant in the deployed network settings.

There have been a lot of research works over the past few years in QoT estimations. The study on parameters uncertainty can be traced back to[54] where ML algorithms were proposed to learn power levels and noise figures of amplifiers to reduce QoT prediction error to 0.02dB and their method was validated in simulations with the assumption of fully-loaded/on-demand wavelength allocations. P. Ramantanis *et al.*[94] numerically emulated SNR uncertainty as a function of the uncertainties of physical parameters, e.g., fiber attenuation, noise figure, span input power and etc., which also has guiding significance for QoT estimation accuracy improvements and network design in presence of physical parameters uncertainties. Also, the impact of fiber length uncertainty on QoT using ML was investigated in[59] assuming fully-loaded artificial data base. Another parameter, lumped loss, was studied in[57] and estimated along with amplifier gain profiles using ML on fully-loaded experimental data and typical SNR estimation error less than 0.1 dB was obtained. The method was then further extended and validated in a partially-loaded experimental scenario[40]. In addition to lumped loss, the impact of connector loss that is hard to measure in practice on QoT estimation using GNPy was studied in[55] on a multi-vendor fully-loaded test-bed. Results show that assuming a connector loss of 0.25dB on live production network, 90% of the GSNR estimation errors will be within ±1dB.

For QoT estimation in multi-vendor experiments, M. Filer *et al.* proposed a vendor-agnostic QoT estimator based on GN model and experimentally validated its accuracy within ±0.75dB[89]. Since then, there have been many QoT papers developing the well-known open-source project GNPy such as[56,95] together with multi-vendor experimental validation of GNPy[90] or the combination of ML and GNPy[83]. These multi-vendor experiments fill ASE channels to C-band and are thus fully-loaded settings. Furthermore, for live production networks, J. Muller *et al.*[71] proposed machine learning-based QoT model trained on synthetic data and tested on two OMSes of their live network with a maximum SNR deviation less than 0.5dB. H. Chouman *et al.*[84] also proposed machine learning-based models trained with field data from two optical networks for short-term QoT forecasting. In[96], the authors assumed optical layer data is not available as network infrastructure information cannot always be shared with end customers and they studied QoT estimation in live network with a channel-probing method and slice-able coherent transceivers to assess the GSNR. It is worth noting that the authors in[63] presented a good QoT estimation result for their partially loaded live network based on large-scale calibration data. However, such specialized calibration data deep in their devices are often only available for equipment vendors internally and not network operators.

Consequently, most of related work studied parameter uncertainty reductions in numerical simulations, fully-loaded disaggregated lab experiments, or non-disaggregated live production networks. A thorough QoT study on disaggragated and partially loaded live production network is lacking and much needed for practical realization of dynamic mesh networks in the future. In this connection, we studied physical model-based QoT estimation of a large-scale live production network in which the signal power in each span can be measured either by EDFA total input/output power or OCM for each individual channel (**Figure 2.1**). We proposed to refine the signal power used in the model by calculating an offset between EDFA total power and the sum of OCM powers for each EDFA and multiplying the offset to each OCM power. Such an approach reduces the mean and standard deviation (std) of GSNR estimation error compared to those using either EDFA total input/output power or OCM power only. Another critical parameter to be refined is the amplifier gain profile which is not explicitly available but can be estimated from the monitored power profiles and their evolution along each span of fiber with attenuation and stimulated Raman scattering (SRS). The estimated gain profiles in turn further improve the ASE noise power profiles and thus GSNR estimations across different channels of a wavelength-division multiplexing (WDM) network.

We hereby further extend our previous contribution by adding simultaneous learning of EDFA noise figure (NF), additional biases for different optical multiplex section (OMS) and frequency channels to incorporate location-specific disturbances and residual unaccounted frequency effects across different components of the network respectively. The data-driven approach involves the optimization of 230 variables and simple gradient descent algorithm is built on the JAX[97] platform with auto-differentiation and parallel computation for fast and effective optimizations. It is found the biases learning can further improve QoT estimation accuracy while optimizing NF values from lab measurements offer marginal improvements. We also investigated different ways to divide the datasets for regular re-training/update of biases values, e.g. uniform division according to time or non-uniform division according to variations of channel loadings and their impacts on the QoT estimation performance. The dataset is also extended from 4 months[45] to 8 months and represent the largest-scale *disaggregated* and *partially-loaded* live production network study for QoT estimation with 1478354 data samples, 76 optical multiplex sections with a total length of 34939.44 kilometers and channel loading in each OMS varying from 2 to 48. The final standard deviation of QoT estimation errors using all the aforementioned techniques can be reduced from 0.6337dB to 0.2377dB. It should be noted that the work presented herein is not meant to argue against the use of ML. Rather, we hope to illustrate what disaggregated network operators can do to "clean up" the physical layer measurements regardless of whether ML will be used subsequently.

## 2.2 Link Model and Network Database

The long-haul live production network under study is a *partially disaggregated* network where the amplified lines are managed as aggregated subsystems[56], which means that EDFAs, OCMs and other link components in a given OMS are from the same vendor as illustrated in **Figure 2.1**. The optical channel (OCH) signal[98] is defined as a frequency channel $\lambda$ (spectral slot) allocated for a service to transmit information between a particular network node (site "A") and another node (site "Z") of the network. Hence, two different lightpaths, one from "A" to "Z" and the other from "Z" to "A", share the same OCH ID, center frequency and set of OMS, but they go through separate EDFAs and OCMs that are separately calculated in our study. The number of frequency channels in each OMS varies from 2 to 48 across the whole C band depending on the actual services at the moment and hence it is also a *partially-loaded* network.

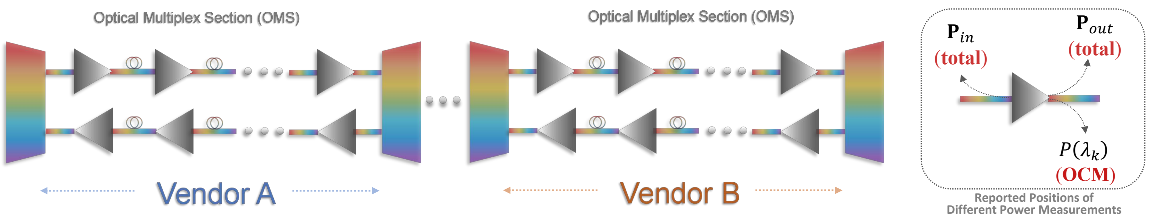



**Figure 2.1:** Disaggregated link model for the live production network under study and various signal power measurements around each EDFA.

The network monitoring and management system is developed by **Alibaba Cloud**[15]. The whole network database can be divided into electrical layer data and optical layer data. The electrical layer data mainly includes information from transceivers such as data rate $R_b$, Baud rate $R_B$, center frequency $f_c$, pre-forward error correction bit error ratio (pre-FEC BER), Q-factor, received optical power etc. The optical layer data consists of fiber link information such as span length $L$, typical values of insertion loss, configured value of variable optical attenuator (VOA), EDFA vendor and type. For each EDFA, multiple optical power measurements including EDFA total input power $\mathbf{P}_{in}$, EDFA total output power $\mathbf{P}_{out}$ and OCM output power $P(\lambda_k)$ for each channel $\lambda_k$

(as shown on the right side of **Figure 2.1** are being monitored every hour and reported to the network database. EDFA nominal gain $G_{ij}$ and total signal loss of each span are computed from real-time reported power data. **Table 2.1** shows the details of the live production network under study. The network consists of 76 OMSes, 694 EDFAs (2 vendors and 6 types), and about 140∼200 bi-directional OCH services at a given time. The time range lasts 243 days and results in a total data size of 1478354 after anomalous data were removed.

**Table 2.1:** Details of the Live Production Network Database

| | |
|---|---|
| Net Data Rate [Gbps] | 200 |
| Baud Rate [GHz] | 69 |
| Total Number of EDFA | 694 |
| Total Number of OMS | 76 |
| Total Length of OMS [km] | 34939.44 |
| Number of OCHs per Hour | 140∼200 |
| Time Range [days] | 243 |
| Total Number of Data Samples | 1478354 |

## 2.3 Methodology

The QoT estimator performance metric for an OCH data is defined by the GSNR estimation error $\epsilon$

$$\epsilon = GSNR_{\mathbf{E}} - GSNR_{\mathbf{M}} \tag{2.1}$$

where $GSNR_{\mathbf{M}}$ is the "measured" GSNR given by the mapping function from pre-FEC BER to GSNR provided by the vendor, and $GSNR_{\mathbf{E}}$ is the "estimated" GSNR using the real-time measured power levels and is given by

$$GSNR_{\mathbf{E}} = \frac{1}{\sum_i \left(\frac{1}{GSNR_i}\right)} \tag{2.2}$$

where

$$\frac{1}{GSNR_i} = \frac{1}{SNR_{i,\,\mathrm{ASE}}} + \frac{1}{SNR_{i,\,\mathrm{NLI}}} \tag{2.3}$$

accounts for an OCH signal passing through the $i^{th}$ OMS. We used the closed-form intra-channel Stimulated Raman Scattering GN (ISRSGN) model in[33] to calculate the nonlinear interference coefficient $\eta(\lambda_k)$ and hence $SNR_{i,\,\mathrm{NLI}}$. In addition, the contribution by the $j^{th}$ EDFA in the $i^{th}$ OMS to $\frac{1}{SNR_{i,\,\mathrm{ASE}}}$ is given by

$$\frac{1}{SNR_{i,\,\mathrm{ASE}}} = \sum_j \frac{hf_c B_{ch}[G_{ij(\lambda_k)} - 1]NF_{ij}(G_{ij})}{\widetilde{P}_{ij}(\lambda_k)} \tag{2.4}$$

where $h$, $f_c$, $B_{ch}$, $\widetilde{P}_{ij}$, $G_{ij}$, $NF_{ij}$ respectively denotes the Planck's constant, signal center frequency, bandwidth, and single-channel output signal power, gain, noise figure of the $j^{th}$ EDFA. Since the non-flat gain profile $G_{ij}(\lambda_k)$ of EDFA is not available in the optical layer data, we initially assume a flat EDFA gain $G_{ij}$

$$G_{ij} = \frac{\mathbf{P}_{out,ij}}{\mathbf{P}_{in,ij}} \tag{2.5}$$

in the calculation. For each type of EDFA, the noise figures $NF_{ij}(G_{ij})$ as a function of gain $G_{ij}$ were obtained from lab measured values with a polynomial interpolation of degree 4,

$$NF_{ij}(G_{ij}) = c_{t4}G_{ij}^4 + c_{t3}G_{ij}^3 + c_{t2}G_{ij}^2 + c_{t1}G_{ij} + c_{t0} \tag{2.6}$$

where $\{c_{t4}, c_{t3}, c_{t2}, c_{t1}, c_{t0}\}$ are the polynomial coefficients and $t \in \{1, 2, 3, 4, 5, 6\}$

denotes the type of EDFA.

### 2.3.1 Input Parameters Refinement

In this section, we will demonstrate how to refine the input physical parameters step by step to improve the accuracy of QoT estimation.

#### I Optical Power

An accurate output power measurement of optical amplifier is critical in GSNR estimation because it will directly affect the calculation of $\widetilde{P}_{ij}(\lambda_k)$ in $SNR_{i,\,\mathrm{ASE}}$ in **Equation (2.4)** and launch power in $NLI$ and hence ultimately GSNR. According to the amplifier model in **Figure 2.1**, we have two monitored power measurements, total power $\mathbf{P}_{out,ij}$ and single-channel power $P_{ij}(\lambda_k)$ at the output of each amplifier. Denote $K$ as the total number of active channels in an OMS. In theory, the sum of OCM power $\sum_{k=1}^{K} P_{ij}(\lambda_k)$ should equal to total output power $\mathbf{P}_{out,ij}$ but it is not always the case in practice. In **Figure 2.2(a)**, we show the distribution of power offsets

$$\Delta P_{ij} = \frac{\mathbf{P}_{out,ij}}{\sum_{k=1}^{K} P_{ij}(\lambda_k)} \tag{2.7}$$

between total output power and sum of OCM power for all of the EDFAs in the network at a given time. Those power offsets motivated us to further investigate which power measurement can provide us with better QoT estimations.

We firstly compared the statistical results of GSNR estimation errors using (1) the average per-channel power ($\mathbf{P}_{out,ij}/K$), and (2) OCM power $P_{ij}(\lambda_k)$ as $\widetilde{P}_{ij}(\lambda_k)$ in **Equation (2.4)**. The results are shown in **Figure 2.2(b)** and **Figure 2.2(c)** in which (mean, std) of GSNR estimation errors are (-0.1043, 0.6037) dB for using average power and (-0.7875, 0.6337) dB for using OCM power. It is obvious that using the average per-channel power leads to a smaller mean error than using OCM power. Note that the results only represent the statistical results of a large dataset, different from the accu-

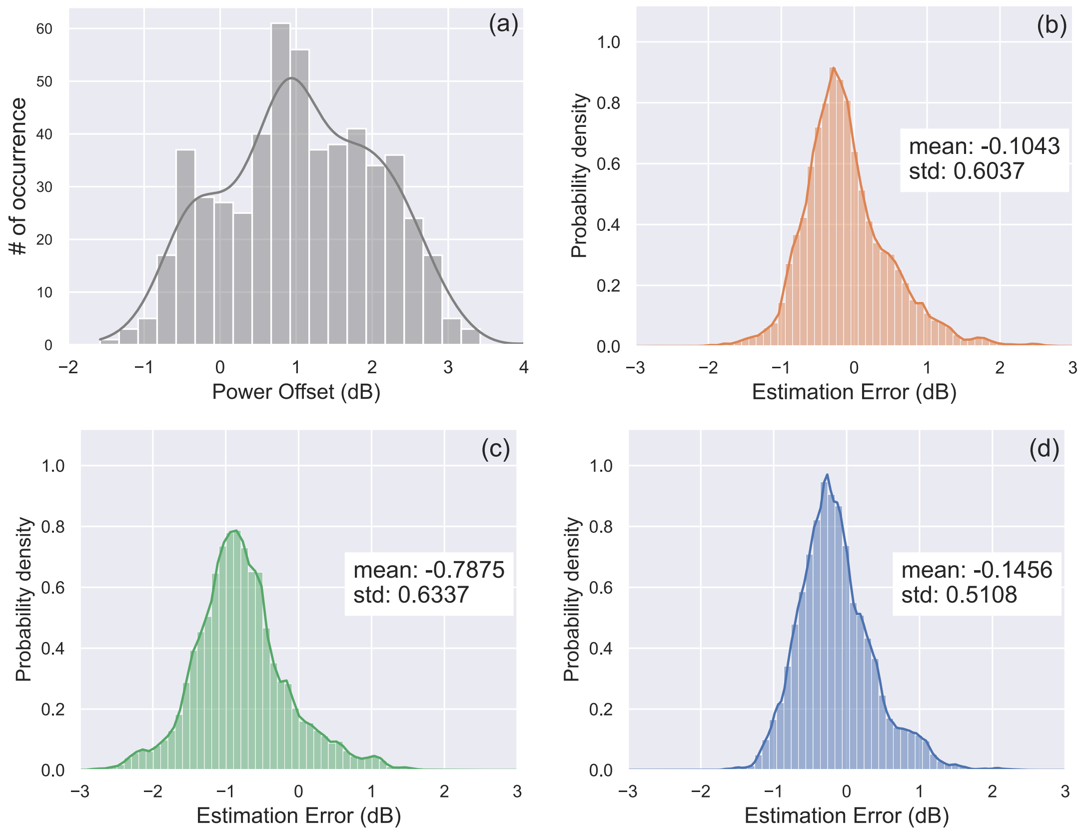



**Figure 2.2:** (a) Distribution of power offsets $\mathbf{P}_{out,ij}/\sum_{k=1}^{K} P_{ij}(\lambda_k)$ in decibels for all EDFAs across the network at a given measurement time instant; Probability distribution of GSNR estimation errors of the 8-month dataset when using (b) average power ($\mathbf{P}_{out,ij}/K$), (c) OCM power $P_{ij}(\lambda_k)$ as $\widetilde{P}_{ij}(\lambda_k)$ in Eq. (2.4) and (d) the proposed power offset method.

racy of a single OCH. Hence, a possible explanation is that OCM power is measured after variable optical attenuators (VOA), fiber interface unit (FIU) and other internal components following the output of amplifier in some EDFA models. Hence, there is always additional insertion loss before OCM in those EDFAs. Such insertion loss is hard to estimate accurately due to different internal EDFA constructions by different vendors in a disaggregated network setting. The typical values of insertion loss are used in the calculation for those EDFA, which possibly result in less accurate estimation with an obvious mean deviation when using OCM power. Also, the standard deviation of the estimation errors using OCM power is also a bit larger than that using the average power, which can be explained by the fact that OCM power measurements in the deployed networks are prone to an error up to 0.8 dB according to operational experience and previous communications with vendors.

Although it seems to be more accurate to use the average power to estimate the GSNR with a relatively low mean estimation error, the assumed flat power profile of ($\mathbf{P}_{out,ij}/K$) is not realistic and cannot capture the variation of $P_{ij}(\lambda_k)$ across different frequency channels in practice. The EDFA total output power $\mathbf{P}_{out,ij}$ will also inevitably include ASE noise from empty channels which degrades its accuracy. Therefore, we propose to first remove the ASE noise power from $\mathbf{P}_{out,ij}$ especially for the case of partial loading where number of OCH in an OMS is less than 19. Then we calculate a power offset for each EDFA and multiply this offset $\Delta P_{ij}$ to all the OCM powers $P_{ij}(\lambda_k)$ of that EDFA so as to obtain a more accurate total power level while preserving the power variations across different channels. Note that in this case, $\widetilde{P}_{ij}(\lambda_k)$ in $SNR_{i,\mathrm{ASE}}$ will be changed to

$$\widetilde{P}_{ij}(\lambda_k) = \Delta P_{ij} \cdot P_{ij}(\lambda_k) \tag{2.8}$$

and also launch power in $SNR_{i,\mathrm{NLI}}$ calculations will be changed. **Figure 2.2(d)** shows the distribution of estimation error with this signal power refinement method. It can be seen that the mean estimation error are kept at a similar level compared to that using average power while the std is reduced to 0.5108 dB, thus indicating the power refinement method helps improve the QoT estimation accuracy.

## II EDFA Gain Profile

In the above QoT-E model we used, the value of gain in **Equation (2.4)** is the same for each frequency channel $\lambda_k$, which is also another approximation that is not realistic. Note that the $j^{th}$ EDFA gain profile $G_{ij}(\lambda_k)$ of the $i^{th}$ OMS can be computed by

$$G_{ij}(\lambda_k) = \frac{\widetilde{P}_{ij}(\lambda_k)}{P_{in,ij}(\lambda_k)} \tag{2.9}$$

where $\widetilde{P}_{ij}(\lambda_k)$ is the refined single-channel output power in **Equation (2.8)** indeed. However, the OCM in our live production network can only measure and report single-channel output power but not input power $P_{in,ij}(\lambda_k)$ and thus we cannot calculate the frequency-dependent gain profile $G_{ij}(\lambda_k)$ in principle. To circumvent this problem, we assume a simple linear power attenuation model to estimate the $j^{th}$ EDFA input power profile $P_{in,ij}(\lambda_k)$ from the output power profile of the $(j-1)^{th}$ EDFA through

$$P_{in,ij}(\lambda_k) = \frac{\widetilde{P}_{i(j-1)}(\lambda_k)}{\alpha_{ij}} \tag{2.10}$$

where $\alpha_{ij}$ is the total loss of the fiber span following the output of the $(j-1)^{th}$ EDFA. Then combining with the refined $j_{\text{th}}$ EDFA output power profile, we can simply use **Equation (2.9)** to estimate the $j^{th}$ EDFA gain profile $G_{ij}(\lambda_k)$. Thus we can estimate the gain profile of every EDFA in an OMS except the first amplifier (also called boost amplifier, BA). With this method, the gain and hence noise power of each frequency channel will be different in **Equation (2.4)** and the resulting (mean, std) of QoT estimation errors is improved to (-0.1211, 0.4952) dB in **Figure 2.3**(b) compared to that of **Figure 2.2**(d).

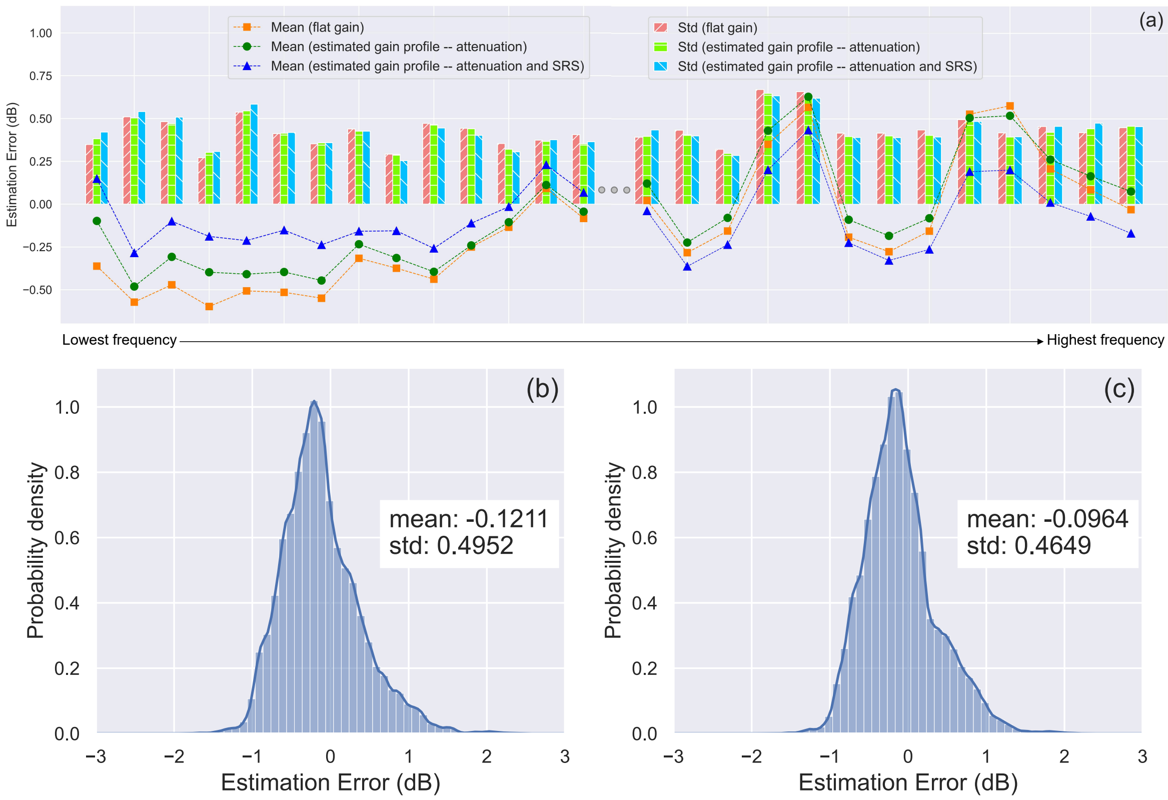



**Figure 2.3:** (a) Mean and std of GSNR estimation errors vs. channel frequency using different estimated EDFA gain profiles with and without SRS. Probability distribution of GSNR estimation errors using the estimated gain profile (b) without SRS and (c) with SRS taken into account.

We also compared the mean and std of QoT estimation error for each frequency channel ($\lambda_k$) illustrated in **Figure 2.3**(a). The orange dot-line and bar are the mean and std result of using the power refinement but flat gain in the estimation model, while the green dot-line and bar are the mean and std result of replacing the flat gain with estimated gain spectra. There is a clear reduction of the mean error on the part of lower frequencies in the green dot-line of **Figure 2.3**(a) compared to orange dot-line (flat-gain). However, note in **Figure 2.3**(a) that the mean estimation error for lower frequencies tends to be positive while the mean error for higher frequencies tends to be negative. This phenomenon suggests that a frequency-dependent optical power attenuation over fiber may be at play here. Therefore, we further incorporate the SRS effect in the fiber attenuation calculation and use a numerical method that solve a discretized version of differential equation combining fiber attenuation and inter-channel Raman gain[99]. In this paper, the standard Raman gain spectrum of G.652 optical fiber is used[100]. The mean errors at different frequencies are shown in the blue dot line in **Figure 2.3**(a) and its overall distribution of QoT estimation error is shown in **Figure 2.3**(c) with (mean, std) improved to (-0.0964, 0.4649) dB. It is clear that incorporating SRS effect can further improve the mean and std and the overall estimation accuracy.

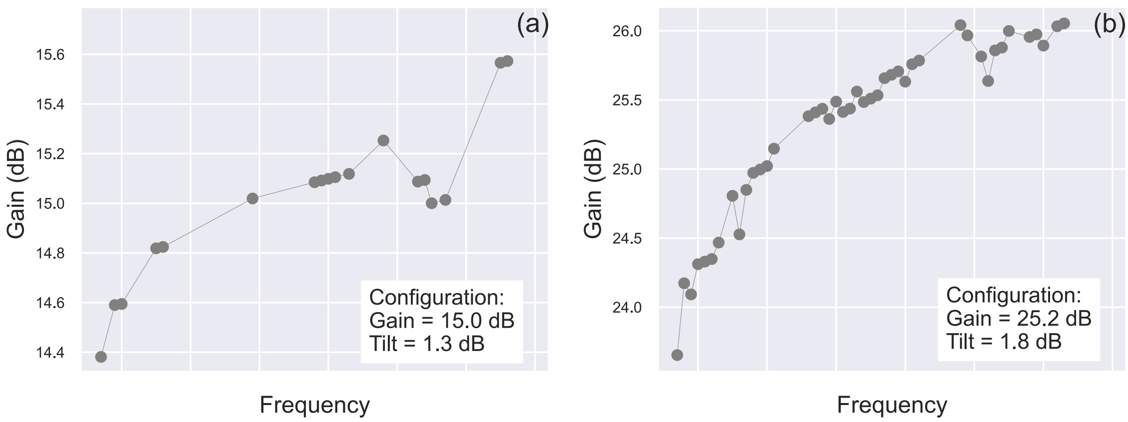



**Figure 2.4:** Example estimated gain profiles for two EDFA with (a) partially-loading and (b) fully-loading in the live network.

Two examples of estimated EDFA gain profiles carrying partial and full channel loadings are shown in **Figure 2.4**(a) and **Figure 2.4**(b) respectively. In **Figure 2.4**(a), the estimated gain profile correspond to EDFA gain $G$ and tilt $T$ configuration of $G = 15.0$ dB, $T = 1.3$ dB with 18 OCHs while in **Figure 2.4**(b), the estimated gain profile correspond to EDFA gain $G$ and tilt $T$ configuration of $G = 25.2$ dB, $T = 1.8$ dB with full loading of 43 OCHs. The gain profiles estimated by our model have average gain and tilt values that are roughly consistent with those from network configuration settings even in different channel loading conditions. Therefore, such information can also serve as an additional quasi real-time monitoring technique for each individual deployed EDFA.

### 2.3.2 Data-Driven optimizations

In addition to refining the signal and noise power values, we will now show how additional data-driven optimizations help further improve the QoT estimation error in a disaggregated live production network from the operator's perspective.

#### I Data-driven optimization parameters

In a production network whose OMS may come from different vendors, it is reasonable to expect that the underlying optical fiber/devices have different characteristics which leads to variations of physical parameters such as power measurement, gain, fiber loss, passband narrowing effects as well as other intractable interactions specific to the OMS concerned. Although the bi-directional components of an OMS belong to the same vendor, the characteristics of individual components can still differ from each other. In this connection, we introduce additional bias $\beta_{\mathrm{oms}_i}^{\mathrm{z}\rightarrow\mathrm{a}}$ and $\beta_{\mathrm{oms}_i}^{\mathrm{a}\rightarrow\mathrm{z}}$ for the $z \rightarrow a$ and $a \rightarrow z$ direction respectively of the $i^{th}$ OMS in the GSNR model to be learnt from data. Another bias we proposed to add is the frequency bias $\beta_{\lambda_k}$ which aims to capture the residual frequency-dependent effects across the network as evident in **Figure 2.3**(a) after the input parameter refinements in Section 2.3.1 due to model imperfections, components inaccuracies and other intractable interactions. Furthermore, it is not uncommon in practice that reported values of EDFA noise figures (from datasheets or lab measurements) differ from those in deployed live production networks. In our case, we have lab-measured and interpolated $NF(G)$ for 6 types of EDFA as shown in the grey dots of **Figure 2.5** that can serve as initialized values for learning the polynomial interpolation coefficients $\{c_0^t, c_1^t, c_2^t, c_3^t, c_4^t\}$ for the $t^{th}$ type of EDFA.

$$\begin{aligned}
GSNR_{\mathbf{E}} &= \frac{1}{\sum_i \frac{1}{\left(\frac{1}{1/SNR_{i,\,\mathrm{ASE}}+1/SNR_{i,\,\mathrm{NLI}}}\right)\cdot \beta^d_{\mathrm{oms}_i}}} \cdot \beta_{\lambda_k} \\
&= \frac{1}{\sum_i \frac{1/SNR_{i,\,\mathrm{ASE}}+1/SNR_{i,\,\mathrm{NLI}}}{\beta^d_{\mathrm{oms}_i}}} \cdot \beta_{\lambda_k} \\
&= \frac{1}{\sum_i \frac{\sum_j \left[h f_c B_{ch}(G_{ij}-1) NF_{ij}/\widetilde{P}_{ij}\right]+1/SNR_{i,\,\mathrm{NLI}}}{\beta^d_{\mathrm{oms}_i}}} \cdot \beta_{\lambda_k} \\
&= \frac{1}{\sum_i \frac{\sum_j \left\{h f_c B_{ch}(G_{ij}-1)\left[c_0^t G_{ij}^4 + c_1^t G_{ij}^3 + c_2^t G_{ij}^2 + c_3^t G_{ij} + c_4^t\right]/\widetilde{P}_{ij}\right\}+1/SNR_{i,\,\mathrm{NLI}}}{\beta^d_{\mathrm{oms}_i}}} \cdot \beta_{\lambda_k}
\end{aligned} \tag{2.11}$$

Incorporating all the above parameters, we have 230 variables in total to be learnt from data. The cost function is the mean squared error $\overline{\epsilon^2}$ over the selected OCH dataset in which the estimated GSNR is given by **Equation (2.11)** where $d \in \{z \to a, a \to z\}$ denotes the direction of the $i^{th}$ OMS that an OCH signal with wavelength $\lambda_k$ passes through. All the OMS and frequency biases are initialized at zero and the coefficients of $NF(G)$ are obtained by fourth order polynomial interpolation of the measured points. Simple gradient descent algorithm is used to update the bias and polynomial coefficients. Note that the $i$, $j$ and $d$ can be totally different for an OCH at different time as they have dynamic transmission paths and **Equation (2.11)** and hence the parameters update equations need to be amended accordingly. Thanks to JAX's auto-differentiation and auto-vectorization[97], computing partial derivatives of **Equation (2.11)** and applying the mini batch-based training to accelerate convergence is straightforward. Although the modeling of dynamic traffic may introduce additional overheads in the execution of some state-of-the-art optimizations, a mini-batch size of 300 data samples is found to achieve 20 times speedup (with Intel i7-11700 CPU) without compromising final performance.

We first determine the relative importance of learning the OMS, frequency bias and

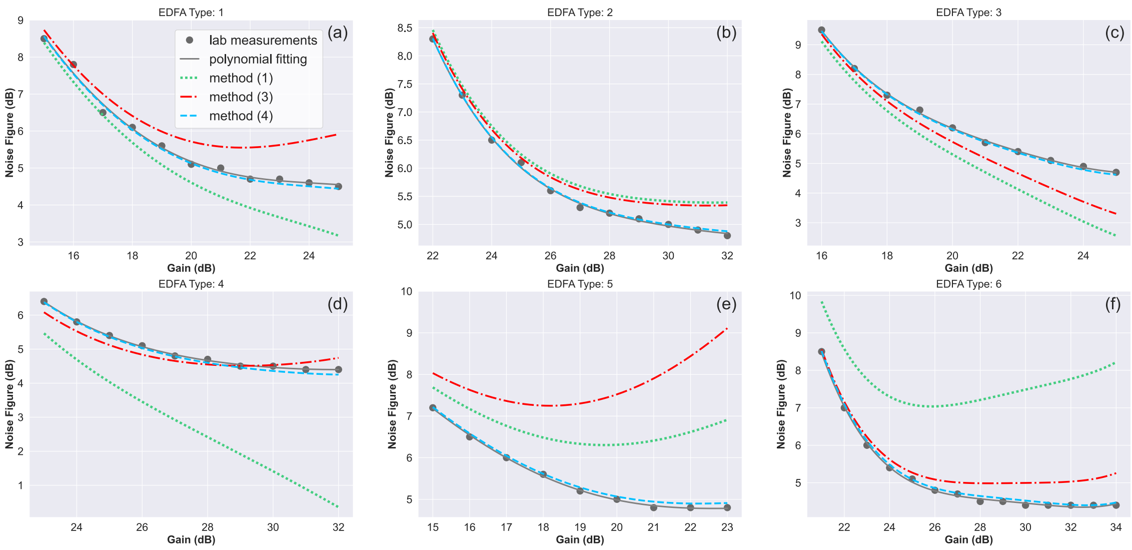



**Figure 2.5:** Learned Noise Figures vs. Gain relations for different parameter learning strategies and their comparison with the original $NF(G)$ from lab measurements and curve fitting for each type of EDFA.

the NF polynomial interpolation coefficients $\{c_0^t, c_1^t, c_2^t, c_3^t, c_4^t\}$ from a 4-day dataset: training dataset with 3 days for training and 1 day for testing. Four learning strategies are studied: (1) learn the biases and NF coefficients simultaneously, (2) only learn the biases, (3) only learn the NF coefficients, (4) first learn the biases and then learn the NF coefficients. Prior to the learning, the (mean, std) of QoT estimation errors of the test dataset is $(-0.0544, 0.4355)$dB. With the learnt OMS and frequency biases and NF coefficients, the (mean, std) of QoT estimation errors are improved to $(0.0084, 0.2082)$dB, $(0.01, 0.2109)$dB, $(0.0071, 0.417)$dB and $(0.0119, 0.2109)$dB for method (1), (2), (3) and (4) respectively. The estimation performances are nearly the same for method (1), (2) and (4) but not for (3), thus suggesting that the NF learning alone is not as good as OMS and frequency biases learning. The learnt $NF(G)$ curves for 6 EDFA types of method (1), (3) and (4) are shown in **Figure 2.5**. It can easily be seen that some of the learnt $NF(G)$ curves clearly go against our experience and knowledge of amplifiers e.g. method (1) for EDFA type 4 has close-to-zero NF at high $G$ value, or $NF(G)$ by method (3) for EDFA type 5 increase with $G$ for high $G$ values. Together with the fact that the (mean,std) for method (1), (2) and (4) are similar, one can conclude that learning the biases alone already constitute most of the performance gain and additional learning of the $NF(G)$ coefficients (from the lab-measured values) bring marginal benefits. This suggests that maybe the lab-measured values are already quite good and the strange $NF(G)$ curves represents pure numerical artifacts that marginally improve the cost function while the parameter values significantly deviate from expected values. Therefore, we chose method (2) as our training method hereafter.

### II Training and Testing Dataset Selection

We have 8 months of data from the live production network and due to system drifting over time, it is typical to divide the data into multiple periods within which the network state is assumed to be stable and the data-driven parameters are re-trained/updated in every period. One can simply divide the data uniformly across time or according to the network dynamics such as the dates of major channel add/drop in the network. An example of major channel add/drop for 6 randomly selected OMS over four month is shown in **Figure 2.6(c)** where we can see certain same variations of add/drop. In our study, we divided the 4-month data non-uniformly into 8 periods accordingly to dates of major channel add/drop as shown in **Figure 2.6(b)** and compare with uniform division by time (half-month) into 8 periods in **Figure 2.6(a)**. We applied the same process on the extended 4-month data so that the total number of periods is 16. The training dataset are the first two days of data of each period while the remaining data of that period are the corresponding testing dataset.

The QoT estimation performance using the data-driven approaches are further improved from those in **Section 2.3.1**. The statistical results of estimation errors for each period of both methods are shown in **Table 2.2** and **Table 2.3**. The overall (mean, std) combining data from all the periods is (0.0046, 0.2377) dB and (0.0047, 0.2444) dB respectively for the uniform and non-uniform division respectively. The results are similar and thus indicate that defining specific periods of apparent network stability for data-driven parameter updates do not give significantly more benefits. We also perform training/testing for the case of uniformly dividing the data into 8 periods

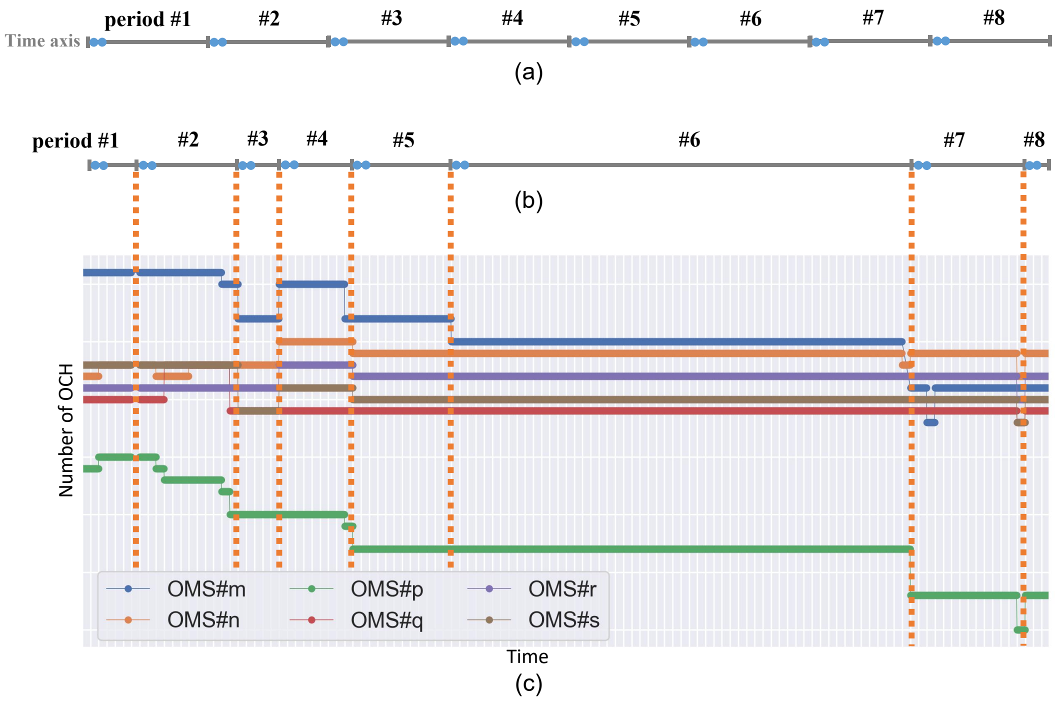



**Figure 2.6:** Division of dataset (a) uniformly over time and (b) non-uniformly according to channel loading variations; (c) Examples of channel loading variations in six sample OMS over four months.

by time (one-month) and the overall (mean,std) is slightly worsened to (0.0138, 0.2605) dB. We further consider a worst case scenario where the biases are learnt from the first two days of data only and then are applied to all the subsequent 8-month of data and the results are much degraded to (-0.0184, 0.4394) dB. This is in agreement with expectations as longer duration for each period experience more system drifting and sample learnt OMS and frequency biases and their updates along different training periods are shown in **Figure 2.7** to better illustrate such effects. Therefore, it is important to update the data-driven parameters regularly to capture the intractable system drifts and maintain good QoT estimation performance in a live production network. From the results, we determined that re-training the biases every half a month are good trade-off between reducing the estimation errors while keeping the training procedure simple and effective.

Finally, after the proposed input parameter refinements technique in the last section and the data-driven parameter optimization, the final mean and std of GSNR estimation errors is significantly reduced to (0.0046, 0.2377) dB as shown in the blue

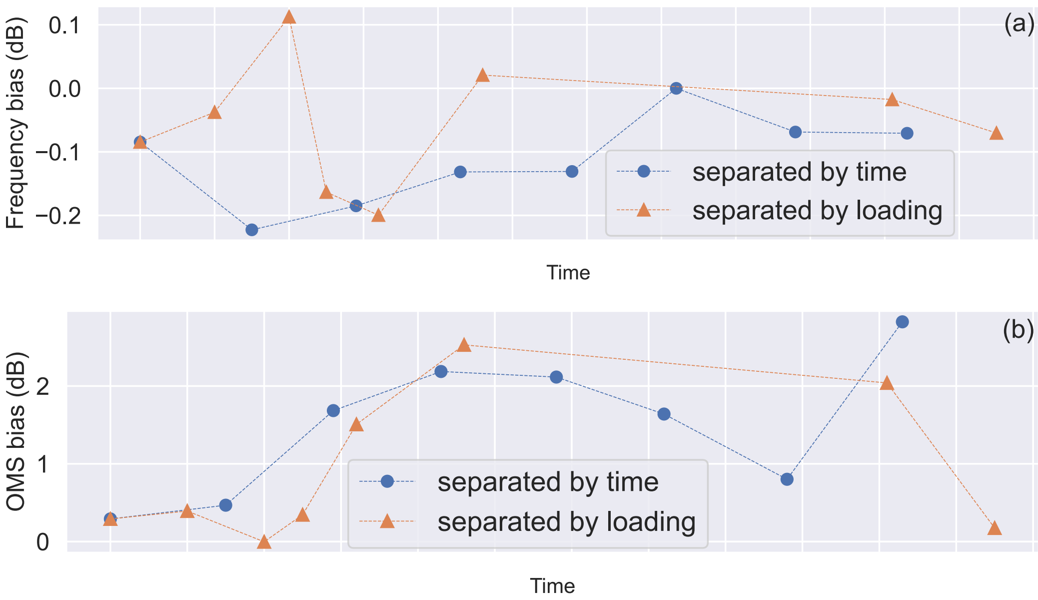



**Figure 2.7:** Drifting of sample (a) frequency bias and (b) OMS bias over time using different dataset division methods.

**Table 2.2:** Statistics of GSNR estimation errors for dataset division uniformly over time.

| **Period** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mean**[dB] | 0.03 | 0.00 | 0.02 | 0.02 | 0.01 | -0.01 | -0.03 | 0.03 | 0.01 | -0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.01 | -0.02 |
| **Std**[dB] | 0.29 | 0.29 | 0.23 | 0.23 | 0.18 | 0.30 | 0.21 | 0.24 | 0.20 | 0.27 | 0.24 | 0.21 | 0.22 | 0.22 | 0.24 | 0.23 |
| **Max**[dB] | 2.18 | 2.36 | 2.53 | 1.14 | 0.77 | 1.15 | 1.55 | 1.26 | 2.15 | 2.39 | 2.28 | 1.85 | 2.91 | 2.85 | 1.61 | 2.73 |
| **Min**[dB] | -1.85 | -2.40 | -2.66 | -2.06 | -1.04 | -2.95 | -1.62 | -2.37 | -1.33 | -2.99 | -3.25 | -2.96 | -1.09 | -2.93 | -1.51 | -2.69 |

histogram of **Figure 2.8** compared to those using average of total EDFA output power and OCM power measurements. The final GSNR estimation performance is much more accurate without much additional computational efforts, and each parameter refinement/optimization step is based on physical insights from the underlying transmission system which add to the intepretability and scalability of the proposed methods and are vital for network operators.

**Table 2.3:** Statistics of GSNR estimation errors for dataset division non-uniformly according to channel loading variations.

| **Period** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mean**[dB] | 0.03 | -0.01 | -0.03 | 0.01 | -0.01 | 0.02 | 0.01 | 0.00 | 0.01 | -0.01 | 0.01 | 0.00 | 0.00 | 0.02 | -0.01 | 0.01 |
| **Std**[dB] | 0.27 | 0.33 | 0.26 | 0.21 | 0.23 | 0.26 | 0.26 | 0.21 | 0.20 | 0.27 | 0.24 | 0.22 | 0.21 | 0.22 | 0.23 | 0.19 |
| **Max**[dB] | 2.28 | 1.85 | 2.91 | 2.85 | 1.61 | 2.73 | 1.22 | 1.11 | 2.16 | 2.41 | 2.18 | 1.88 | 2.91 | 2.88 | 2.94 | 2.10 |
| **Min**[dB] | -2.99 | -2.96 | -1.09 | -2.93 | -1.51 | -2.69 | -2.23 | -0.97 | -1.15 | -3.00 | -3.20 | -2.98 | -1.09 | -2.65 | -3.41 | -2.69 |

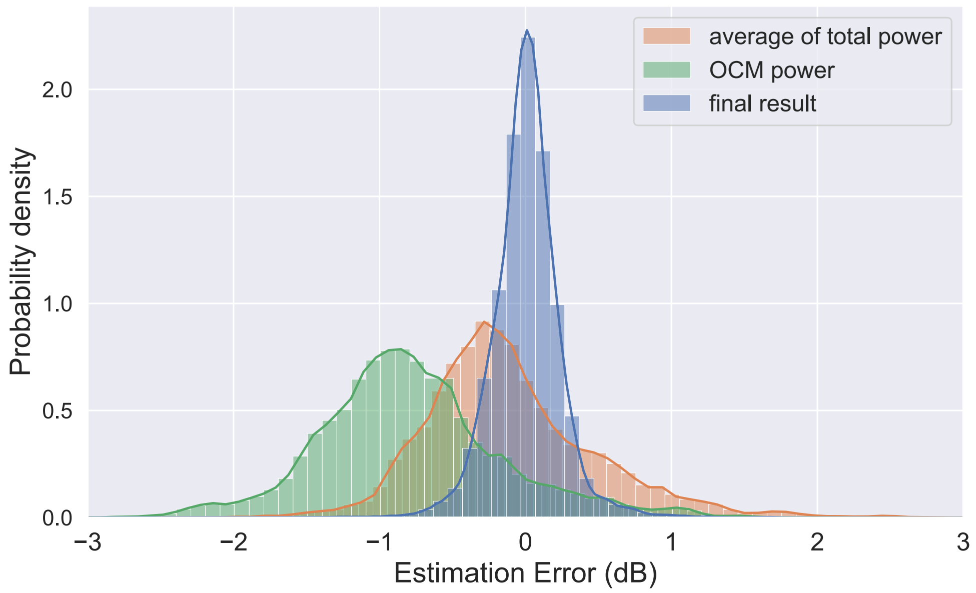



**Figure 2.8:** Probability density of GSNR estimation errors employing all the parameter refinement and data-driven optimization techniques compared with those using OCM or average of total EDFA output power measurements only.

## 2.4 Results and Discussion

The improvements in QoT estimation accuracy also translates into margin reductions. In **Figure 2.9**, we show the complementary cumulative distribution function (CCDF) plots of underestimation errors ($\epsilon < 0$) and overestimation errors ($\epsilon > 0$) for the case of using OCM power only, average total power only, after power and gain profile refinement and after data-driven parameter optimization. For an outage probability of $10^{-3}$, the average of total power method (pink line) gives an (under-estimation, over-estimation) errors of $(2.03, 3.62)$ dB, and the OCM power method (red line) gives an (under-estimation, over-estimation) errors of $(2.65, 1.99)$ dB. After power and gain profile refinements, the under-estimation error is reduced to 1.41 dB while the over-estimation of 2 dB is inferior to that of the OCM power method. This is because the OCM method has a significant mean under-estimation error of 0.7875 dB as shown in **Figure 2.2(c)** which leads to a disproportionately small over-estimation error. Nevertheless, adding data-driven parameter optimization further reduce the (under-estimation, over-estimation) errors to $(1.41, 1.34)$ dB. If one assumes a margin design for an outage probability of $10^{-3}$, a margin reduction of from 3.62 dB to 1.34 dB can by achieved by our proposed methods.

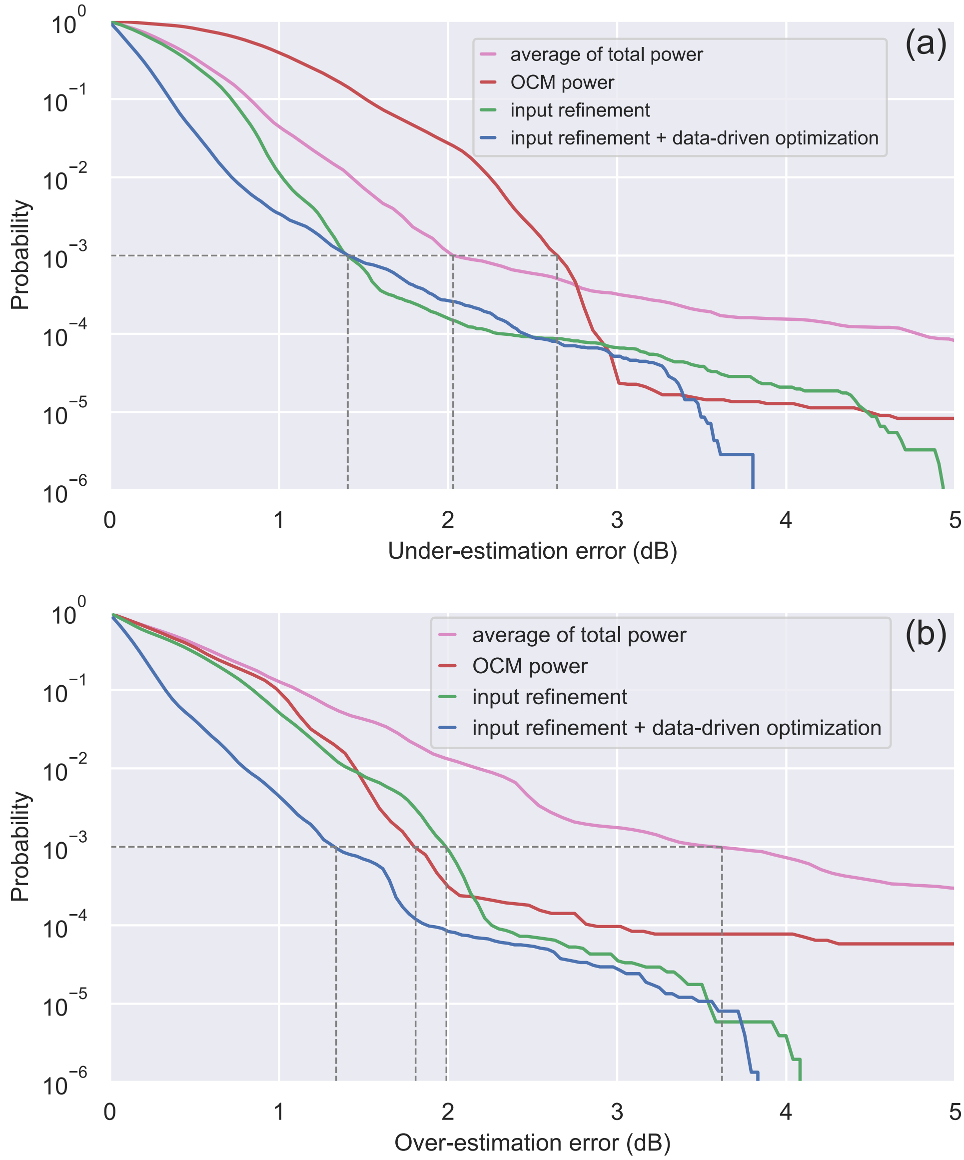



**Figure 2.9:** CCDF of GSNR (a) under-estimation and (b) over-estimation error.

## 2.5 Summary

In this Chapter, we presented an analytical model-based QoT estimation study based on the largest-scale disaggregated and partially-loaded long-haul live production network with 8 months of data. We demonstrated step by step how to reduce the QoT estimation error by refining the inline amplifier and OCM signal power measurements and estimating gain spectra of each inline amplifier which led to more accurate amplifier noise power hence QoT estimations. We then further introduce OMS and frequency bias to incorporate intractable location-specific and spectral effects in the network and proposed data-driven parameter optimizations to learn the biases as well as EDFA noise figures. Combining all the aforementioned methods, the standard deviation of the QoT estimation error can be reduced from 0.6337dB to 0.2377dB. The work demonstrated practical steps to improve QoT estimation accuracy, reduce network margins and potentially increase network capacities without compromising reliability in realistic current and future mesh networks. Further investigations on network automation and active hard/soft failures identifications/management with QoT estimation for disaggregated production networks will be future topics of interests.

# 3

# Improved QoT Estimation for Commercial Metro DCI Networks

## 3.1 Overview

In this Chapter, different QoT estimation models designed for large-scale metro data center interconnection (DCI) networks are proposed. Data Center Interconnect (DCI) optical networks are critical infrastructures that facilitate high-speed, low-latency communication between geographically distributed data centers, which is of great significance for the Hyperscalers. Architecturally, DCI networks feature high-speed short-reach coherent communications, a more varying traffic pattern, and has physical layer protection, which make it different from the long-haul optical network. Optical multiplex section protection (OMSP) with primary/secondary path switching is commonly adopted in metro DCI networks to provide high-quality transmission services without interruption. In this situation, there is always one backup path, either primary or secondary, that does not have the actual BER/QoT monitoring data because it is not connected to the receiver. Therefore, accurate QoT estimation/prediction become significantly important for DCI networks with OMSP, which can improve reliability in operations such as optical line protection (OLP) switching[101].

On the other hand, metro DCI networks have much shorter transmission distance which makes the transponder effects become a dominant impairment, instead of fiber nonlinearity and ASE noise in the long-haul network. This is because the higher modulation format ($\geq$ 16QAM) is applied with a lower noise tolerance $SNR_{\mathrm{TRX}} \simeq \Phi^{-1}(\mathrm{BER}_{\mathrm{floor}})$. However, metro DCI networks are more disaggregated for the adopted coherent TRX. Data rate for metro DCI networks covers a wide range from 200Gbps to 400Gbps, 600Gbps and 800Gbps with potentially different coherent transponder structure and DSP design, and they are more disaggregated with multiple coherent transponder vendors, which adds to the difficulty in accurate QoT.

Given these challenges, we proposed two different QoT models with acceptable accuracies which can be applied in the commercial metro DCI networks. Two models are proposed and discussed in the following sections: (i) ML-based models which incorporates different TRX features as input and (ii) analytical model with experimentally characterized input-dependent B2B OSNR-BER relationship.

## 3.2 Machine Learning based QoT Estimation

### 3.2.1 Overview

QoT estimation for optical transport network has been actively investigated in recent years. Researchers have proposed various algorithms including analytical models, data-driven parameter optimizations as well as machine learning (ML) to improve generalized signal-to-noise ratio (GSNR) estimation accuracy[36,45,46,102]. However, relatively few research on QoT estimation for large-scale metro data center interconnection (DCI) networks are reported. As multi-vendor mixed-rate coherent modules are becoming more common in DCI scenario, a more accurate QoT tool is needed for network performance monitoring and resource allocation[15]. However, metro DCI networks have much shorter transmission distance with almost 90% of lightpaths less than 100 km and generally have higher SNR values than their long-haul counter parts. This nullify the advantages of accurate QoT estimation based on analytical models (such as the GN model) as amplifier noise and fiber non-linearity effects are no longer dominant impairments compared to transponder effects[37] that are difficult to model explicitly. Also, data rate for metro DCI networks covers a wide range from 200Gbps to 400Gbps, 600Gbps and 800Gbps with potentially different coherent transponder structure and DSP design and are more disaggregated with multiple coherent transponder vendors, which adds to the difficulty in accurate QoT. Even for the same model from a given vendor, our own experimental characterizations on a few transponders show that SNR performance exhibits a larger variation from one transponder to another for higher SNR values, making it difficult for accurate QoT in practical large-scale deployed networks. There are some related works for metro network field trial[103] and experimental test-bed[104], but a QoT study over large-scale metro production networks has yet to be investigated.

In this connection, we study large-scale disaggregated metro DCI networks and propose an artificial neural network (ANN) model to incorporate inherent transponder effects for accurate QoT estimations. Results show that the ANN model can outperform analytical models which incorporate amplifier noise, fiber nonlinearity distortions and transponder distortions with data-driven parameter estimation, thus suggesting the unique advantage of ML techniques in such scenarios. The effectiveness of such ANN-based QoT model is demonstrated in OLP switching scenarios that are common in metro DCI networks.

### 3.2.2 Methodology

We started our QoT study by applying the same optimization methods i.e., refined physical parameters and bias leaning, proposed in[46] to disaggregated metro DCI case where have around 10,000 lightpaths across 300 OMSes at a given time. We mainly study 8 types of transponders from 4 vendors as shown in **Table 3.1**. We randomly selected 200,000 data samples of each transponder type from June to August 2023 from metro DCI networks. Each data sample contains EDFA total input/output power, gain and noise figure, single-channel power from OCM placed at the output of each EDFA, fiber length and loss along the links that an OCH has passed though, which will be used to estimate GSNR. At the receiver, pre-FEC BER will be mapped into ground-truth GSNR from pre-measured B2B curves. The training and testing datasets are randomly chosen from these 200,000 data samples with the ratio of 7:3 for bias optimizations and ANN-based method.

$$\begin{aligned}\frac{1}{GSNR} &= \sum_i \left(\frac{1}{SNR_{i,\,\mathrm{ASE}}} + \frac{1}{SNR_{i,\,\mathrm{NLI}}}\right) \\ &= \left(\sum_i \frac{1/SNR_{i,\,\mathrm{ASE}} + 1/SNR_{i,\,\mathrm{NLI}}}{\beta^d_{\mathrm{oms}_i}} + \frac{1}{SNR_{\mathrm{TRX}}}\right) \cdot \beta_{\lambda_k} \cdot \beta_{\mathrm{TRX}_{type}}\end{aligned} \tag{3.1}$$

Analytical model with refined physical parameters by combining optical power measurements from EDFA input/output and optical channel monitors and EDFA gain

Table 3.1: Statistics of GSNR estimation error (test dataset) for different rate and vendor of transponders

| Vendor | Rate [Gbps] | Analytical Model with Refined Parameters | | with Refined Parameters and Bias Learning | | ANN | |
|---|---|---|---|---|---|---|---|
| | | Mean | Std | Mean | Std | Mean | Std |
| A | 400 | 0.063 | **1.828** | 0.007 | **0.521** | 0.024 | **0.164** |
| | 600 | −1.424 | **0.888** | −0.008 | **0.184** | 0.008 | **0.054** |
| B | 400 | −0.899 | **2.119** | 0.013 | **1.173** | 0.007 | **0.29** |
| | 800 | −1.243 | **1.774** | 0.006 | **0.584** | −0.019 | **0.185** |
| C | 200 | 2.363 | **2.301** | 0.005 | **0.62** | 0.002 | **0.151** |
| | 400 | −0.407 | **2.512** | 0.011 | **0.685** | 0.008 | **0.156** |
| D | 200 | 0.118 | **2.458** | 0.000 | **0.505** | 0.005 | **0.171** |
| | 400 | −0.133 | **2.021** | −0.006 | **0.504** | 0.017 | **0.173** |

profile estimation[46] was first used to estimate GSNR and compared to true GSNR (converted from the pre-FEC BER) to obtain the estimation error. The expression as shown in Eq. (1) where $SNR_{\mathrm{ASE}}$, $SNR_{\mathrm{NLI}}$ and $SNR_{\mathrm{TRx}}$ denotes the GSNR if only ASE noise, nonlinearity or transponder effects are present. The mean and standard deviation (std) of GSNR estimation errors of this method are shown in the third and fourth columns of **Table 3.1**. It can be seen that the mean and std (0.8∼2.5dB) obtained for each type of transponders in metro DCI networks are larger compared to around 0.4dB std in long-haul production network[46]. This is possibly because the refined physical parameters mainly improve the accuracy of $SNR_{\mathrm{ASE}}$ and $SNR_{\mathrm{NLI}}$ but these impairments are negligible in short DCI transmission distance as experimentally validated[37]. As $SNR_{\mathrm{TRX}}$ can play a more important role than $SNR_{\mathrm{ASE}}$ and $SNR_{\mathrm{NLI}}$, we measured OSNR-BER characteristics of different transponders in our experimental test-bed and the results for 800Gbps transponders (vendor B) are shown in **Figure 3.1(a)** for different received optical power (`opr`). We can see the minimum OSNR is largely affected by `opr` outside -8∼2 dBm. We can also deduce $SNR_{\mathrm{TRX}}$ from testing results and then put them in **Equation (3.1)** with adding OMS/frequency/transponder type bias learning based on gradient descent algorithm. Additional OMS bias $\beta^d_{\text{oms}_i}$ where $d \in (a \rightarrow z, z \rightarrow a)$ denotes the direction of the $i^{th}$ OMS, frequency bias $\beta_{\lambda_k}$ to capture frequency-dependent effects and $\beta_{\text{TRX}_{type}}$ to compensate for different types of transponder's inherent effects are added in the GSNR model to be learnt from data. The results can be seen from the fifth and sixth columns of **Table 3.1** in which the mean and std are reduced to an acceptable level for most types of transponders except for the 400G transponder (vendor B) which still has estimation errors std of around 1 dB.

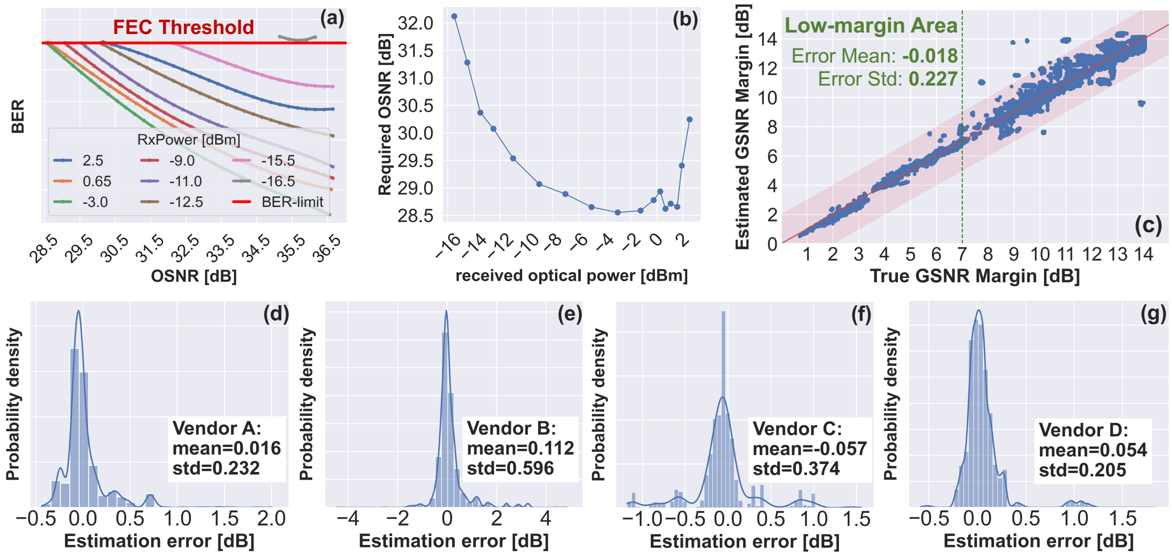



**Figure 3.1:** (a) Measured BER vs. OSNR curves for B2B at different received optical power and (b) received optical power vs. required minimum OSNR of experimental results for 800G transponder(vendor B); (c) all data samples which have switched working paths in September and GSNR is estimated by the corresponding ANN model, and statistical results of estimation errors for (d) vendor A, (e) vendor B, (f) vendor C and (g) vendor D.

### 3.2.3 Results

These results indicates that even with an accurate analytical model, input parameters with refined accuracy, $SNR_{\text{TRX}}$ characterization from a pair of transponders in lab test-bed as well as data-driven linear bias parameter learning, such methodologies still falls short of accurately characterizing the GSNR of a great number of line cards/transponders in large-scale metro DCI networks. Hence, a completely data-driven neural network model may be needed to learn such characteristics unknown to network operators to realize more accurate QoT estimation in such disaggregated metro scenarios. A multi-layer perception (MLP) neural network with 5 hidden layers and number of nodes (4096, 2048, 1024, 128, 8) was built to train QoT estimation model. The activation function is ReLU and input features include EDFA total input/output power, gain, noise figure, single-channel output power from OCM, fiber length and loss, and center frequency of an OCH. Test results are shown in the seventh and eighth columns in **Table 3.1** with std of estimation errors below 0.3dB for all types of transponders.

### 3.2.4 Validation on Optical Line Protection (OLP) Switching

1+1 OMSP which prevents OCH services disrupting from fiber cut/bend and device/component failure, is preferred in metro DCI scenarios[101]. There are always two different physical paths in an OMS (i.e., primary path and secondary path) with the same source and destination and duplicated signals. The switching between primary and secondary path is performed through an automatic protection switch (APS) at the receiver side, which depends on the optical power difference at the OMS destination and the selected path is named as working path while another is backup path[101]. Statistically, the number of OLP switching that occur in a month can be around 200 in the metro DCI networks under study. As the received signal and its pre-FEC BER are only available for the working path but not the backup path, continuous QoT estimations of backup paths on top of monitored optical powers of the backup path are important to ensure enough operating margin in advance of next OLP switching event.

Since the BER/ground-truth GSNR can be reported from backup path only after the switching happened, we specially selected those lightpaths (around 700 lightpaths) that have switched its working paths to backup paths in September to validate the proposed ANN-model and the results are shown in **Figure 3.1**(c-g). Due to mixed-rate multi-vendor transponders have different received SNR ranges and data disclosure confidentiality reasons, we present GSNR margin instead of true GSNR values, which is computed by measured/estimated GSNR minus the minimum SNR required for the specific modulation format and type of transponder and the results are shown in **Figure 3.1**(c) where the red-shadow covers estimation errors within 2dB. As the QoT estimation accuracy in lower-margin areas should be more important and guaranteed than that in higher-margin area, **Figure 3.1**(c) shows that the switched paths working in the lower-margin area ($< 7$dB) has a small std is only 0.227dB which is a good guarantee of QoT estimation accuracy for most of the switching events. The distributions of GSNR estimation errors are shown in **Figure 3.1**(d-g) for different vendors. It should be noted that since the OLP switching is mostly caused by the decreased optical powers of the original working path and most of these OCHs are highly likely at the high-margin region after the switching with higher QoT estimation errors as shown in **Figure 3.1**(c). This in turn lead to a larger std in **Figure 3.1**(d-g) than those in **Table 3.1** as the data from **Table 3.1** do not differentiate between switched and unswitched paths. Overall, these results show that the trained ANN models can be used to accurately estimate QoT of backup paths in metro DCI networks.

### 3.2.5 Summary

In this section, we proposed an ANN-based QoT estimator for large-scale mixed-rate disaggregated metro DCI networks with 4 transponder vendors and multiple transmission rates of 200, 400, 600 and 800Gb/s. A QoT estimation error standard deviation of 0.3 dB is achieved, outperforming analytical-based methods with transponder SNR characterizations and data-driven parameter optimizations. QoT estimations for protection switching scenarios are demonstrated its value for future disaggregated and mixed-rate metro DCI networks.

## 3.3 Analytical Model based QoT Estimation

### 3.3.1 Overview

Optical network disaggregation brings interoperability and flexibility to transmission system and enables operators to select best-in-class network equipment for individual network function[105]. Nonetheless, it also brings challenges in accurate modeling of multi-vendor devices either physically or in black-box models for optimizing network services design, management and control. Signal-to-noise ratio (SNR) margin monitoring and estimation are vital for network operators to proactively manage and protect their services in disaggregated deployed networks. It has been experimentally validated that transceiver (TRx) noise and EDFA amplified spontaneous emission (ASE) noise are the dominant impairment sources in coherent transmission system with short distance[38,106], frequently seen in metro networks such as data center interconnection (DCI). TRx-constrained SNR indicates an upper limit to the available SNR and is generally measured in back-to-back (B2B) configuration[106]. Recently, Mano *et al.* proposed a received optical power penalty term to estimate the input power dependency of SNR-OSNR assuming such power penalty stems from the shot noise of the photodetectors (PD)[38]. Thus, modeling SNR penalty as a function of received optical power can better characterize the limitations of optical front-end and digital signal processing (DSP)[38,60,107]. In addition to input-dependent SNR penalty in TRx, ASE noise introduced by EDFA is also a dominant impairment source. Most of current research focus on EDFA gain profile estimation to improve the accuracy of ASE noise estimation but assume flat noise figure (NF)[41,46]. However, modeling non-flat amplifier NF in non-ideal scenarios is also crucial for accurate OSNR estimation which has been shown in simulations[51], experiments[53] but its significance in a live production network with frequent non-ideal operation conditions such as extremely low total input power caused by partially-loaded links and fiber loss degradation are yet to be demonstrated.

In this work, we conducted extensive experimental measurements to characterize NF as a function of gain, tilt, frequency and input power for EDFAs from different vendors, followed by least-square fitting the experimental B2B measurements of 400Gb/s coherent transceivers for different optical signal powers to obtain the input power dependent OSNR-SNR model. We applied such OSNR-SNR and NF models to the real-time measurements in our hyperscale disaggregated DCI networks with over 10,000 lightpaths across 300 OMSes from July to September in 2024, constituting the largest deployed network study of this topic to date. Compared with simply using NF as a function of gain from datasheets and measured B2B OSNR-SNR model at beginning of life (BoL), we show that the root mean squared error (RMSE) of SNR estimation can be reduced from 0.662 dB to 0.287 dB.

### 3.3.2 Methodology

Amplifier noise figure provided by datasheet is simply a function of gain as shown in the blue dotted lines of **Figure 3.2**(a) and (b). However, NF can be greatly affected by other factors such as tilt, frequency and input power[51,53]. For more accurate noise power modeling that can account for different operation conditions in deployed networks, we measured NF under different target gain $G_{target}$, target gain tilt $T_{target}$, frequency $f_i$ and total input power $P_{in}^{tot}$ using EDFA mode of EXFO OSA5240S after total input/output power calibration. 8 types of EDFAs (4 types from each vendor and 2 vendors in total) that have been applied in the deployed network were measured with fully-loaded 48 channels under test (CUT) over C band. The total input power ranges from -20 dBm to the largest value corresponding to saturated output. NF values at 194.1THz under saturated output are shown in the solid lines of **Figure 3.2**(a-b) where NF increases when $G_{target}$ decreases or $T_{target}$ increases. This trend of NF variation with gain and tilt is the same for all types of EDFAs in our network. It can be seen that NF provided by datasheet is actually a worst-case value resulting in more conservative OSNR estimation. NF variations with $P_{in}^{tot}$ under different $G_{target}$

and $T_{target} = 0$ are shown in **Figure 3.2**(c) and (d) for EDFA type#1 (vendor A) and EDFA type#2 (vendor B) respectively. The trend of NF with $P_{in}^{tot}$ has shown different vendor-specific characteristics, i.e., NF increases with $P_{in}^{tot}$ drops for vendor A (**Figure 3.2**(c)) while NF increases with $P_{in}^{tot}$ increases for vendor B (**Figure 3.2**(d)). On the other hand, the NF spread across 48 CUTs widens when $P_{in}^{tot}$ drops for vendor A while the spread shrinks for vendor B.

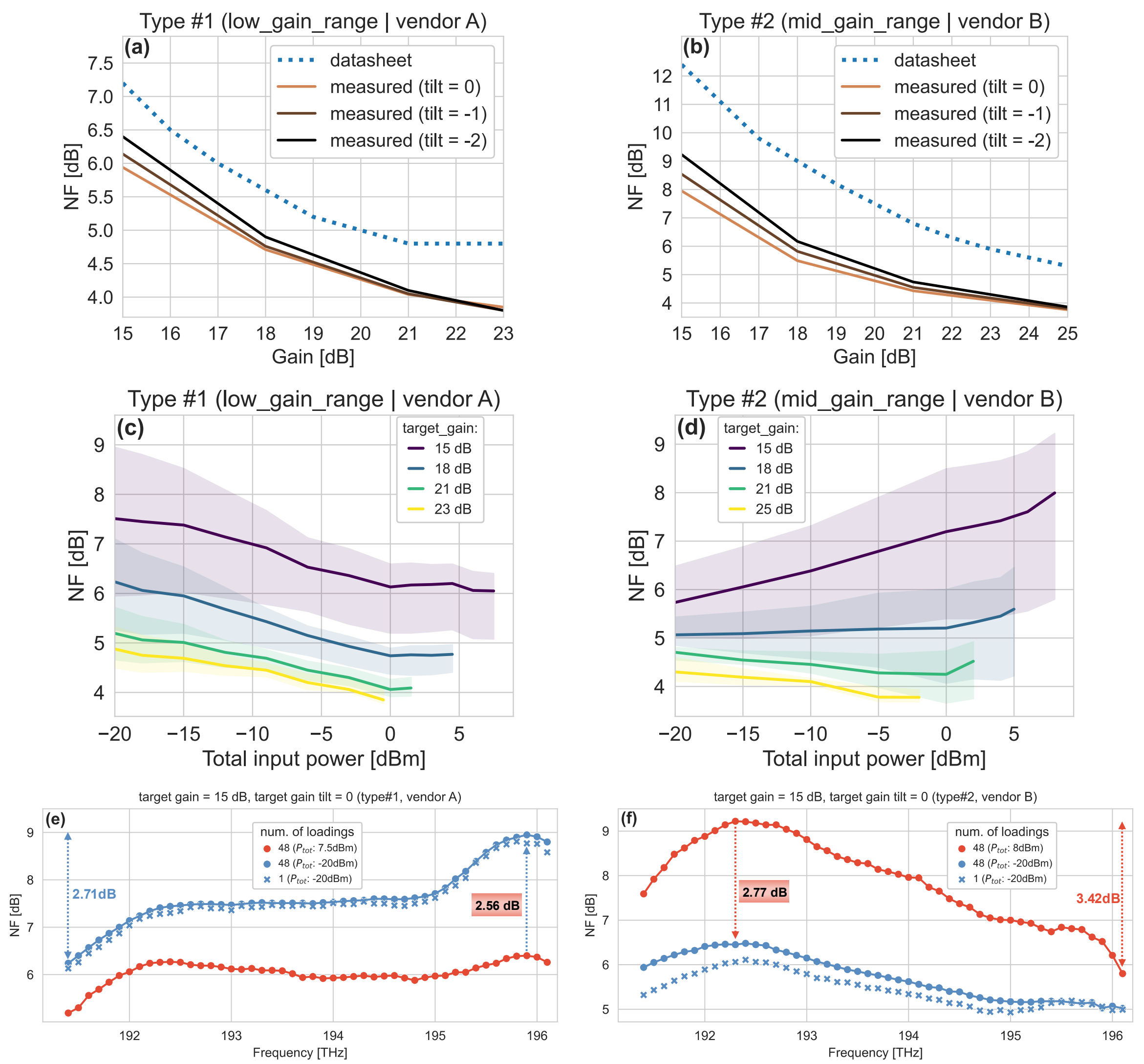



**Figure 3.2:** NF vs. gain for (a) EDFA type#1 and (b) EDFA type#2; measured NF vs. total input power under different gain settings and target tilt $T_{target} = 0$ for EDFA (c) type#1 and (d) type#2. Solid lines denote the median value of NF among the 48 CUTs and the upper/lower edge of color-bar corresponds to the maximum and minimum NF across frequency; measured NF vs. frequency at the largest/smallest total input power with 48/1 channel loading and target tilt $T_{target} = 0$ for EDFA (e) type#1 and (f) type#2.

The dependence of NF vs. frequency is also vendor specific as shown in **Figure 3.2**(e-f). For EDFA from vendor A, the NF profile generally decrease with increasing total input power but EDFA from vendor B exhibits the opposite trend. The blue cross of **Figure 3.2**(e-f) denotes the NF when the system only has 1 existing channel with the same $P_{in}^{tot}$. This shows that different loading conditions with the same total input power has certain impact on NF values especially at lower frequencies for vendor B. However, such effect is still negligible compared to that of total input power and so that we will ignore the impact of loading conditions in our study. Overall, EDFAs NF from different vendors are clearly different as observed in **Figure 3.2**(c-f) especially when modeling NF as a function of frequency and total input power. It should be noted that vendor-specific NF dependence on input power/frequency can be found in all four types of EDFAs of the same vendor but are not shown here due to space limitations.

The input power dependent transceiver SNR penalty model is also crucial for modeling SNR with different intractable distortions from practical networks due to device/fiber/connector failures. The input power-dependent TRx OSNR-SNR performance was measured in a B2B configuration. The $j^{th}$ measurement data point contains a set $\{\mathrm{SNR_j}, \mathrm{OSNR_j}, \mathrm{opr_j}\}$ where $\mathrm{opr_j}$ denotes received optical power and $\mathrm{SNR_j}$ can be calculated from $\mathrm{BER_j}$ by Eq.(18) in[26]. We measured three line-cards of the same type/vendor TRx that are used in the deployed network and operated at 400Gbps/69GBaud/16QAM and the measurements are shown in **Figure 3.3**(a) for different received optical powers. The yellow dashed line $\Psi_{BoL}(OSNR)$ in **Figure 3.3**(a) represents the baseline OSNR-SNR curve measured at BoL and optimal received power of -4dBm. $\Psi_{BoL}(OSNR)$ is obviously different from the yellow scatters measured recently at opr=-4dBm thus calibration of OSNR-SNR curve after the TRx aging is a necessity. On the other hand, it can also be seen that the received optical power induced SNR degradation can be as large as 1 dB if opr decreases from -4dBm to -15dBm at high OSNR region as shown in **Figure 3.3**(a).

Then, by least-square (LS) fitting the measurements under different received optical powers utilizing the model in[38], we can obtain a set of parameter estimates such as the SNR degradation caused by received optical power degradation $\mathrm{SNR_p^{-1}} \approx$

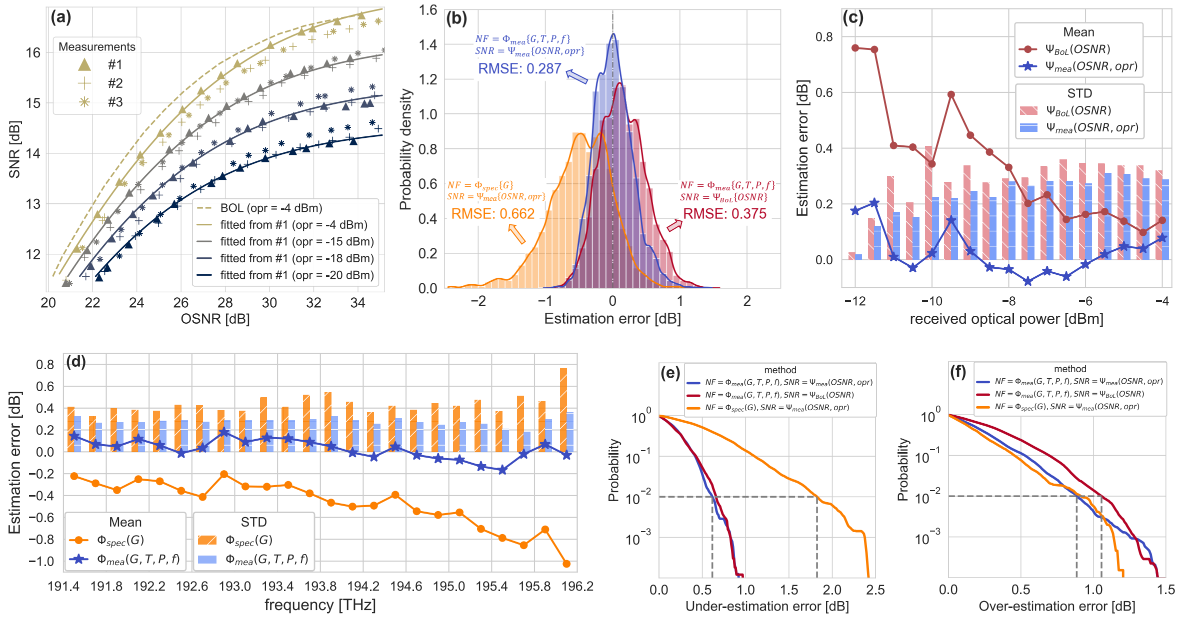



**Figure 3.3:** (a) B2B measurements and the fitting model from linecard#1 measurements; (b) Distribution of SNR estimation errors using different methods for all OCH data in the network; (c) comparison of mean and std of SNR estimation errors using BoL and experimentally measured OSNR-SNR models at different received optical power; (d) comparison of mean and std of SNR estimation error using NF model from specifications vs. experimentally measured for OCHs at different frequency; outage probability of (e) SNR under-estimation and (f) SNR over-estimation using different NF and OSNR-SNR models.

−38dBm, TRx penalty excluding the noise induced by received optical power degradation $\mathrm{SNR}_{\mathrm{TRx}'} \approx 17$dB. These parameters can be used to fully characterize the B2B measurements by $\mathrm{SNR}^{-1} = \mathrm{SNR}^{-1}_{\mathrm{ASE}} + \mathrm{SNR}^{-1}_{\mathrm{NLI}} + \mathrm{SNR}^{-1}_{\mathrm{TRx}'} + \mathrm{SNR}^{-1}_{\mathrm{p}}\mathrm{opr}^{-1}$ where $\mathrm{SNR}^{-1}_{\mathrm{ASE}}$ and $\mathrm{SNR}^{-1}_{\mathrm{NLI}}$ model the contribution to SNR by amplifier noise and fiber nonlinearity respectively. The solid lines of **Figure 3.3**(a) are obtained by least-square fitting measurements of line-card#1 and the RMSE of the obtained fitting model $\Psi_{mea}(OSNR, opr)$ of #1 is just 0.05dB. However, the #1 model $\Psi_{mea}(OSNR, opr)$ for the measurements of line-card#2 and #3 at the optimal received power of -4dBm can result in RMSE of 0.23dB and 0.19dB respectively. Consequently, the RMSE of SNR estimation in the deployed networks will be lower bounded by such measurement variations across different transceivers of the same vendor and same model.

### 3.3.3 Results and Discussion

The adopted QoT estimation method is primarily based on the analytical model and power/gain profile refinement in our previous work[46] and the details of our hyperscale DCI networks with around 10,000 lightpaths across 300 OMSes at a given time can be found in[107]. First, we compare the GSNR estimation performance using different NF profiles: $NF = \Phi_{spec}(G)$ where NF is interpolated by gain from datasheet specifications, or $NF = \Phi_{mea}(G, T, P, f)$ where NF is interpolated by gain, tilt, input power and frequency from our experimental measurements. The different NF values are applied in the **Equation (2.4)** to obtain the linear noise estimation. Then we compare different B2B OSNR-SNR model to obtain the SNR estimation: $SNR = \Psi_{BoL}(OSNR)$ where SNR is directly interpolated by OSNR from measurements at BoL, or $SNR = \Psi_{mea}(OSNR, opr)$ where SNR is calculated by the fitting model from linecard#1 measurements.

The QoT estimation result using $NF = \Phi_{mea}(G, T, P, f)$ and $SNR = \Psi_{mea}(OSNR, opr)$ is shown in the blue histogram in **Figure 3.3(b)** with RMSE of 0.287dB, which is much better than the orange histogram with RMSE=0.662dB using $NF = \Phi_{spec}(G)$ and slightly better than the red histogram with RMSE=0.375dB using $SNR = \Psi_{BoL}(OSNR)$. **Figure 3.3(c)** shows our final optimized method (blue line and bar) performs better at all the received optical powers in our dataset than those using only optimized NF $NF = \Phi_{mea}(G, T, P, f)$ but with $SNR = \Psi_{BoL}(OSNR)$ (red line and bar). On the other hand, **Figure 3.3(d)** shows our final optimized method (blue line and bar) effectively eliminate the frequency dependent errors which is better than than those using only $SNR = \Psi_{mea}(OSNR, opr)$ but with $NF = \Phi_{spec}(G)$ (orange line and bar). For an outage probability of $10^{-2}$, our final optimized method (blue line in **Figure 3.3(e) and (f)**) gives an (underestimation, overestimation) errors of (0.62, 0.88) dB, better than the one uses $NF = \Phi_{spec}(G)$ (orange line) that has underestimation error of 1.82 dB and the one uses $SNR = \Psi_{BoL}(OSNR)$ (red line) that has overestimation error of

1.06 dB.

In this section, we demonstrated the importance to measure vendor-specific EDFA NF and its dependence on frequency, loading and signal power in disaggregated deployed networks. Modeling received power-dependent OSNR-SNR can also account for intractable distortions and play an important role in the QoT estimation accuracy. We show that utilizing the above refined models can reduce the RMSE of SNR estimation from 0.662dB to 0.287dB in hyperscale metro DCI networks.

## 3.4 Summary

In this Chapter, we proposed two applicable methods for improving the accuracy of QoT estimation for large-scale disaggregated DCI networks where transponder effects and ASE noise become dominant impairments rather than fiber nonlinearity in the long-haul. Using a real PM dataset from large-scale DCI networks, we trained ML models to learn the SNR incorporating various input-dependent transponder effects in 8 types of coherent transponders from 4 different vendors. The output of the trained ML model has been demonstrated to be able to estimate SNR for different types of TRX with good accuracy. Then, focusing specifically on the 400Gb/s coherent transponder that has been widely used in the DCI network understudy, we conducted extensive experiments on that type of transponder and obtained an input-dependent OSNR-SNR model by least-squares fitting the measurements. We also conducted extensive experiments to characterize the noise figures of 8 types of EDFAs from 2 different vendors, which is modeled as a function of gain, tilt, input power and frequency. Applying these two experimentally obtained models in the real PM dataset from the DCI networks, we showed that the RMSE of analytical model based SNR estimation from 0.662dB to 0.287dB which validated its effectiveness.

# 4

# Green-field QoT Estimation and Optimization Enabled by EDFA Gain Profile Modeling

## 4.1 Overview

In live production optical fiber networks (**brown-field**), QoT estimation relies heavily on the measurements of signal parameters such as optical signal power obtained from Optical Channel Monitors (OCMs) deployed at the output of each EDFA. OCMs provide real-time channel power and spectral shape that can be fed directly into QoT models, enabling an effective impairment assessment and connection feasibility evaluation. The refined OCM signal power using EDFA total output power in the live network has been shown to be helpful in obtaining an accurate QoT estimation in both long-haul network (**Chapter 2**) and metro DCI network (**Chapter 3**)[46,107–109].

However, in **green-field** scenarios such as fiber link before deployment, newly provisioned or reconfigured lightpaths in evolving network segments, the absence of OCM data poses a fundamental challenge to the aforementioned QoT estimation approaches. Without direct optical power readings per channel, the accuracy of QoT models degrades, leading to higher uncertainty margins or conservative overprovisioning that reduces network utilization. This is because practical EDFAs do not amplify wavelength division multiplexing (WDM) channels uniformly, i.e., gain ripples, and also introduce varying levels of amplified spontaneous emission (ASE) noise across different channels, i.e., noise figure ripples[49,51,53,108,110]. During the design phase in green-field network, the EDFA response directly influences configuration strategies for amplifier gain and tilt[111,112], as well as channel power allocation[60,113]. These strategies often target multiple objectives such as capacity maximization, power equalization, or uniform signal to noise ratio (SNR) at the receiver. Achieving these goals with sufficient SNR margin requires a realistic and accurate EDFA model[93]. In brown-field networks, maintaining SNR margins across services demands reliable estimation of signal and noise power profiles . This too relies on accurate EDFA models, as the amplifier's response is input-dependent and affected by variations in power levels due to fiber aging/bend, channel add/drop, or topology changes. Central to this effort is the need to model the EDFA gain profile accurately under varying operational conditions, including different target gains, tilts, and input power profiles.

Physics-based EDFA gain models utilize a two-level inversion model of stationary state which characterize fundamental principles with propagation and rate equations governing the population dynamics of erbium ions and the evolution of signal, backward and forward ASE[1,7,21,114]. Most of these models contain coupled differential equations that must be solved numerically and require detailed knowledge of the internal erbium-doped fiber (EDF) which is hard to measure[22]. The homogeneous gain model of Saleh *et al.*[20] has simplified the coupled differential equations to a single transcendental equation by setting the balances between input fluxes and output fluxes[48]. More recently, the gain dependence on signal and pump power of single-stage EDF is studied and fitted by singular value decomposition (SVD) with the indenpendence assumption between mean population vectors and emission and absorption cross section vectors[24]. The gray-box model of Liu *et al.*[5] demonstrated that a transcendental equation can model the scenario of multi-stage EDFA by assuming the EDFs of each stage are of the same specification and modeling the gain flattening filter (GFF) response as a constant insertion loss spectra. However, vendors retain considerable design flexibility in practical C- or L-band EDFAs. Variations in EDF length, pump wavelength and power, VOA/GFF configurations and internal control logic can be tailored to meet different performance objectives, such as gain flattening or noise figure optimization. As a result, knowledge of physical parameters of EDF alone may not be sufficient to fully capture the amplifier's behavior across diverse operating conditions. The advent of open and disaggregated optical networks introduces additional complexity[15,115–117], as these environments typically incorporate multi-vendor multi-type EDFAs with different internal components and constructions and those EDF related parameters are often vendor locked-in. Also, these gain models are not suitable for the aforementioned network design and optimization problems, compared to continuous and differentiable models such as neural networks (NNs)[41,50,60,118–121]. But NN-based approaches often require extensive datasets to capture the full range of operational scenarios and are susceptible to issues of overfitting[60] and lack of explainability/transparency.

Given these challenges, a gain profile characterization model that is partially analytical and partially data-driven, which makes it independent of internal component knowledge, is preferred. This can provide greater transparency and usability for network operators to adopt. In addition, a gain characterization model itself consisting of continuous and differentiable functions with low complexity is more desirable, which can act as an effective digital twin of EDFA and necessarily accelerate network optimization. Focusing on commercial EDFAs that work in Automatic Gain Control (AGC) mode and terrestrial transmission systems, gain profile modeling depends on the specific device implementation and different operating states governed by signal and pump power[114,122]. The dynamic gain tilt (DGT) curve[114,123] has been validated to represent the amplifier gain profile at different states, which has also been adopted in the well-known open-source software GNPy[56,124] and can be obtained by the normalized difference with respect to a specific frequency of two different gain profiles. Recently, Lan *et al.*[125] approximated the change in gain profile at different operating states by adding or subtracting a fraction of the EDF's DGT curve to the gain profile. A semi-analytical model proposed by Borraccini *et al.*[49] characterized the gain ripple as a function of target gain and tilt with flat input WDM channels, which can be considered as a variant of implementing the DGT model. However, the model input of Borraccini *et al.*[49] did not take input total power into account, but varying fiber loss in the real network can be one of the uncertainties leading to a variety of input total power and thus operating states. In this paper, we will first show that this semi-analytical model can be changed into twice continuously differentiable function leveraging simple cubic spline interpolation[126]. Then, the continuous model will be further enhanced by linearly fitting two surfaces that characterize the gain-dependent and power-dependent changes using simple cubic polynomials. We will also show how this simple surface fitting method aligns with fundamental principle of EDF and is explainable using SVD. Thus, a continuous gain profile model as a function of target gain, tilt, total input power with flat WDM channels can be obtained.

Current network operators typically fill the unused channels with ASE power to maintain the fully-loaded condition which is commonly adopted to maintain consistent performance. A more practical scenario is that the input of EDFA has fully-loaded WDM channels but with non-flat channel power because of the imperfection of each network element the signal has passed through. We refer to this non-flat WDM input power profile as *pre-emphasis* same as the case in[48]. In this practical case, our continuous gain model built from measurements with flat WDM inputs at different total power can be seen as a fundamental dataset of gain profiles at all the possible average population inversion states. The model of Meseguer *et al.*[48] has shown that based on a fundamental gain profile dataset with flat WDM input, one can easily obtain the gain profiles with input pre-emphasis by assuming the total input powers of a non-flat spectra and a flat spectra can yield the same population inversion. It has been validated in[48] under a constant pump current that is frequently seen in the subsea transmission system with an Automatic Current Control mode and extended to the scenario of arbitrary pump current in[127]. It was also proved in[60] for EDFA in Automatic Power Control (APC) mode. In terms of AGC mode, the center of mass (CM) model[128] is a simple analytical model characterizing gain variation given the changes of input channel loading but with compromised accuracy which limits its utility in real networks[118,129]. In this paper, focusing on AGC mode, we will show that the gain profile with pre-emphasis can be approximated by simply refining the input parameters of our continuous gain function with flat WDM input, while the property of quadratic continuous differentiability of cubic polynomials is still preserved.

Therefore, our research objective is to build a continuous and differentiable cubic polynomial function that can reproduce the gain profiles under various target gain and tilt configurations and input power with fully-loaded channels, which will be beneficial for network operators to fully model the gain behaviors of EDFAs. The obtained gain model can be used to estimate the signal power evolution which is an indispensable parameter for an accurate QoT estimation in green-field scenario where OCM is not available, and can also be used to optimize the design of network architecture and service provisioning. We first extended the semi-analytical model of Borraccini *et al.*[49] to a continuous and differentiable model using curve fitting, which was further enhanced by linearly fitting a surface that characterizes the gain-dependent and frequency-dependent changes at a fixed low output power using simple polynomials with a maximum degree of 3. The same surface fitting method was then applied again to fit the power-dependent and frequency-dependent changes incorporating full ranges of all possible total input powers until the EDFA saturates. Only 3-rd order polynomial regression model is utilized to fully characterize the parameter space of gain profile with significant accuracy of root mean squared error (RMSE) well below 0.083 dB achieved in 9 types of EDFAs from 2 vendors. Then, we study WDM inputs with pre-emphasis and show that introducing input gain and tilt parameter refinements to the cubic polynomial-based gain profile characterization model can further reduce the gain prediction errors. Finally, we show this pre-emphasis aware gain profile model can help obtain accurate QoT estimation results in live production network without the real-time OCM data as input. Such reliable greenfield QoT estimation enables network operators to avoid unnecessary re-work during the design, planning, and construction phases, thereby maximizing the utilization of network capacity and resources.

## 4.2 Experimental Setup and Details of EDFA Under Study

The details of EDFAs understudy are shown in **Table 4.1** for two vendors, A and B, respectively. There are 8 types of C-band EDFAs and 1 type of L-band EDFA investigated. **Figure 4.1** shows the experimental setup for measuring the gain profile of a device under test (DUT) EDFA. A broadband ASE light source was shaped into 48×25GHz WDM channels with 100GHz spacing and flat channel power across the C or L band. The DUT EDFA works in AGC mode, which is the same as the configuration in our terrestrial production network. An optical spectrum analyzer (OSA, EXFO FTBx-5245) was used to sweep the input/output traces which are calibrated according to the total input/output power reported from the built-in photo-detectors (PDs) in the DUT. The sweep and configuration were performed automatically using a Python script that executes Command Line Interface (CLI) commands to remotely control each device and read the measurements. Gain measurement of a channel with frequency centered at $f$ was calculated by

$$g_{\mathrm{dB}}(f) = 10 \log_{10} \frac{P_{out}(f) - N_{out}(f)}{P_{in}(f) - N_{in}(f)} \tag{4.1}$$

where $P_{in}(f)$ and $P_{out}(f)$ denote the measured channel total power at the input and output of DUT EDFA, respectively. $N_{in}(f)$ and $N_{out}(f)$ denote the measured channel noise power at the input and output of the DUT EDFA, which are obtained by interpolating the noise floor.

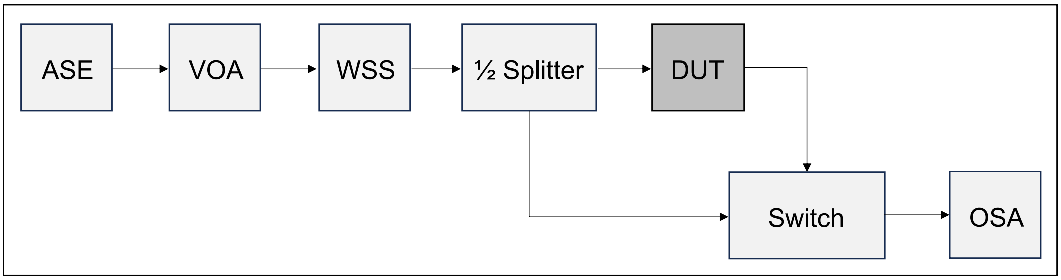



**Figure 4.1:** Experimental setup for measuring EDFA gain profile.

In this paper, independent variables such as frequency, gain, tilt or power are written in *italics*. Any function of those independent variables which does not have an explicit mathematical expression is also written in *italics*. In contrast, continuous and differentiable functions with explicit mathematical expressions will be written in **bold**. Unless otherwise specified, all variables are expressed in dB or dBm.

## 4.3 Cubic Polynomial based Gain Profile Modeling

### 4.3.1 Cubic polynomial fitting model

#### I Cubic spline based curve fitting model

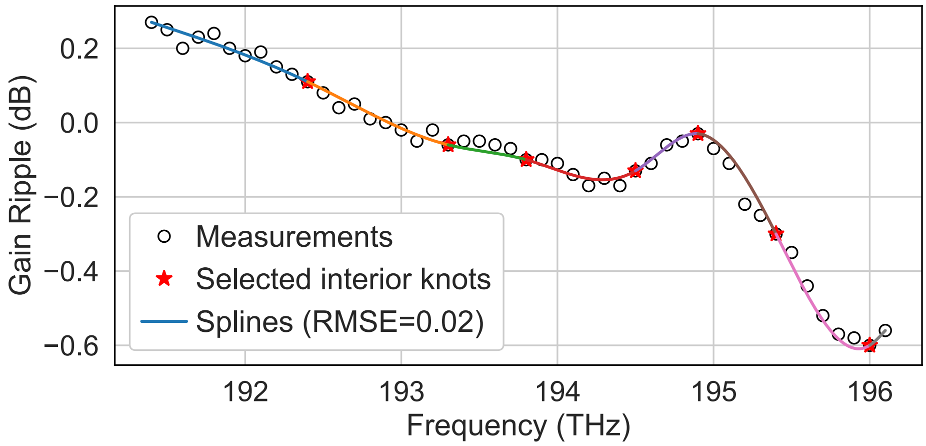
To establish a continuous and fully differentiable function, we first introduce a curve-fitting method to obtain the function of any given spectral profile, e.g., the gain ripples

**Table 4.1:** The details of EDFAs understudy.

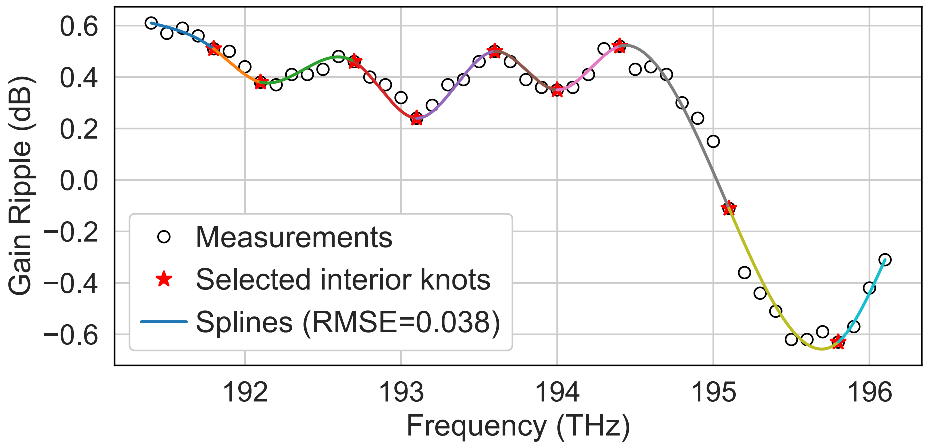
| Vendor | Type | Band | Target Gain | | Target Tilt | | Total Output power | |
|---|---|---|---|---|---|---|---|---|
| | | | min | max | min | max | min | max |
| A | 1 | L | 15 | 24 | −4 | 0 | 5 | 23.5 |
| | 2 | C | 15 | 24 | −4 | 0 | 5 | 23.5 |
| | 3 | C | 15 | 25 | −2 | 0 | 5 | 23.5 |
| | 4 | C | 15 | 25 | −2 | 0 | 5 | 23.5 |
| | 5 | C | 22 | 32 | −2 | 0 | 11 | 23.5 |
| | 6 | C | 22 | 32 | −2 | 0 | 11 | 23.5 |
| B | 7 | C | 15 | 21 | −3 | 0 | 3 | 22.5 |
| | 8 | C | 16 | 21 | −3 | 0 | 6 | 23.5 |
| | 9 | C | 23 | 28 | −3 | 0 | 11 | 23.5 |

w.r.t. frequency. Linear regression such as polynomials is the most common model; however, its accuracy may be guaranteed by increasing the degree of polynomials when fitting a variety of shapes[53]. Cubic spline interpolation was found to be more robust than high-degree linear regression due to large oscillation at the edge of intervals in linear regression, which was also known as "Runge's Phenomenon"[130]. Thus, we adopted cubic spline interpolation as a basic method to fit a curve of any given discrete measurements such as gain ripple profile and obtain its explicit mathematical expression.

A set of interior knots $\{f_j\}$ should be determined to obtain the cubic spline function $\zeta(f)$ from a set of discrete measurements, as shown in **Figure 4.2**. The interior knots can be selected uniformly or non-uniformly according to different shapes of spectra as long as the RMSE between the fitted curve and the measured data samples is minimized. **Figure 4.2** shows two examples of the non-uniformly selected interior knots for gain ripples of two different types of EDFA, whose corresponding cubic spline model have minimized the RMSE to 0.02dB for type 3 (**Figure 4.2**(a)) and 0.038dB for type 9 (**Figure 4.2**(b)) while also keep the number of model coefficients as small as possible. For example, 7 interior knots are selected in **Figure 4.2**(a) leading to 8 piecewise cubic polynomials in total so that $4 \times (7 + 1) = 32$ coefficients should be




(a)




(b)

**Figure 4.2:** Examples of the cubic spline curve model fitted from different gain ripple measurements for EDFA (a) type 3 and (b) type 9. The selected interior knots in red stars divide the fitted curve into a set of continuous piecewise cubic polynomials such that the lines are plotted in different colors.

---

estimated to fully characterize the gain ripple, while 9 interior knots are selected in **Figure 4.2**(b) so that $4 \times (9 + 1) = 40$ coefficients will be estimated.

The method of obtaining the model coefficients $\{a_{i,j}\}$ is described as follows. Assume there are $n$ selected data points in total, which includes the selected interior knots with the lowest and highest frequency points as boundaries, leading to $(n - 1)$ cubic spline pieces. Denote each cubic piece between two adjunct knots as $\mathbf{S}_j(f)$,

$$\mathbf{S}_j(f) = a_{0,j} + a_{1,j}(f - f_j) + a_{2,j}(f - f_j)^2 + a_{3,j}(f - f_j)^3 \tag{4.2}$$

where $j = 1, ..., n-1$. The gain ripple measurements at interior knots and boundary are denoted as $y_j$. Since cubic spline interpolation requires the values of two adjunct cubic spline functions at the knot to be continuous, as well as the values of first derivative and second derivative, we have,

$$\mathbf{S}_j(f_j) = y_j, \quad j = 1, ..., n-1 \tag{4.3}$$

$$\mathbf{S}_j(f_{j+1}) = y_{j+1}, \quad j = 1, ..., n-1 \tag{4.4}$$

$$\mathbf{S}_j^{'}(f_{j+1}) = \mathbf{S}_{j+1}^{'}(f_{j+1}), \quad j = 1, ..., n-2 \tag{4.5}$$

$$\mathbf{S}_j^{''}(f_{j+1}) = \mathbf{S}_{j+1}^{''}(f_{j+1}), \quad j = 1, ..., n-2 \tag{4.6}$$

Additionally, we adopt "natural" end condition as two more constrains which means the second derivative at the start and end of the curve are zero,

$$\mathbf{S}_1^{''}(f_1) = 0 \tag{4.7}$$

$$\mathbf{S}_{n-1}^{''}(f_n) = 0 \tag{4.8}$$

Combining **Equation (4.3)**,**(4.4)**,**(4.5)**,**(4.6)** with assuming two boundary conditions at the two end points **Equation (4.7)**,**(4.8)**, the coefficients can be obtained by writing out these constraints explicitly as a system of linear equations with $4(n-1)$ unknowns and solving the system in matrix form[126]. Then, the obtained cubic spline function can be written as

$$\boldsymbol{\zeta}(f) = \mathbf{S}_j(f), \quad f \in [f_j, f_{j+1}] \text{ and } j = 1, ..., n-1 \tag{4.9}$$

One of the advantages of the curve fitting model **Equation (4.9)** is that it is continuous over the full usable parameter space and twice continuously differentiable. It can be seen from **Equation (4.2)**, **(4.9)** and the lines plotted in different colors of **Figure 4.2** that the cubic spline model is essentially a set of piecewise cubic polynomials. Compared to higher degree polynomials, it avoids Runge's phenomenon while still having a smooth curve offering better stability.

### II Bivariate surface fitting model

Frequency dependence is the first-order effect in modeling EDFA gain ripple variation. It can also be related to the EDFA pump power or input power, which we call the "operating state". Its ideal characterization necessitates multi-dimentional fitting whose complexity scales up exponentially with dimensions. For example, EDFA gain ripple profile $\mathbf{\Gamma}(f, z)$, as a function of frequency $f$ and gain or input power state $z$, can be approximated by polynomials with orders $p$ and $q$ to control the accuracy,

$$\mathbf{\Gamma}(f, z) = \sum_{i=0}^{p} \sum_{j=0}^{q} a_{ij} f^i z^j \tag{4.10}$$

In fact, by our measurements of commercial EDFA for WDM transmission, the $z$-dependance is found to be weak. Either being fundamental or engineered, this property is essential to the operation of WDM system on gain control while stabilizing signal transmission. Using this justification for the mild dependence on $z$ leads to an efficient approximation of **Equation (4.10)**,

$$\mathbf{\Gamma}(f, z) = \mathbf{s}(z)\boldsymbol{\zeta}(f) + \mathbf{t}(z) \tag{4.11}$$

where $\mathbf{s}(z)$ is the scaling factor and $\mathbf{t}(z)$ is the offset. And this model is devised based on our observation that the bounded and continuous surface $\mathbf{\Gamma}(f, z)$ can be represented by scaling and shifting a $z$-trace of $\boldsymbol{\zeta}(f)$. In principle, $\mathbf{s}(z)$ and $\mathbf{t}(z)$ can be arbitrary functions, provided that the errors between the fitted model and the measurements are minimized. In addition, due to mild $z$-dependence, we found that $\mathbf{s}(z)$ and $\mathbf{t}(z)$ in the form of quadratic or cubic polynomials are sufficient for overall accuracy, going beyond that may cause overfitting.

On the other hand, this observation-based approximation model of gain ripple profile in **Equation (4.11)** is successfully linked to the physical picture of how EDFA functions. In particular, as reported in[24], the EDF's gain profile is a product of two independent terms: 1) The emission and absorption cross section vectors represent wavelength-dependent internal EDF characteristics but is relatively independent of signal and pump power; 2) The mean populations of excited and non-excited ions in EDF, which largely relate to different operating states $z$[24]. Therefore, $\mathbf{s}(z)$ can be roughly deemed as a quantity that governs how the gain response of two-stage EDF varies from a complete/suppressed population inversion to a suppressed/complete population inversion, i.e. the variation of $\overline{N}_2$. In short, the independence assumption when converting **Equation (4.10)** to **Equation (4.11)** has been validated to conform with

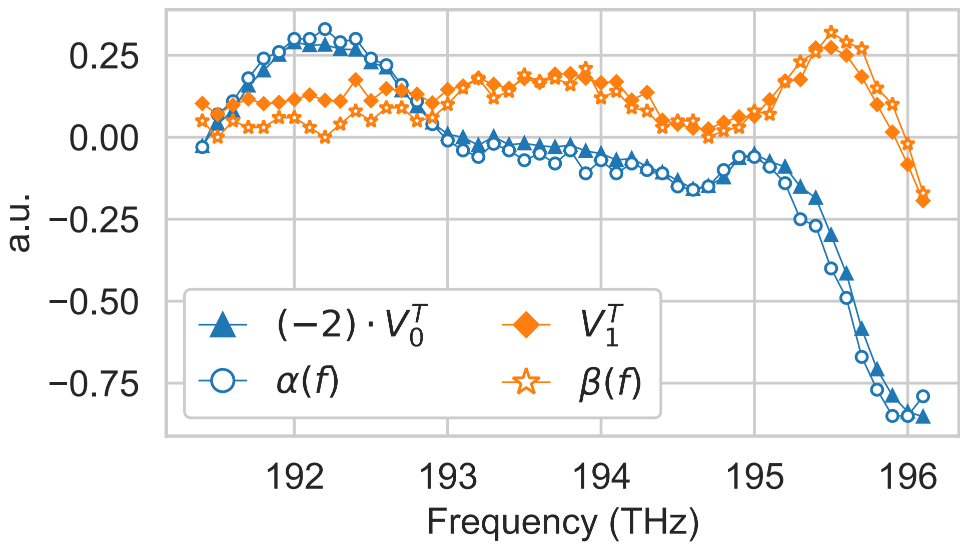



**Figure 4.3:** An example of similarity between components of SVD decomposition and $\boldsymbol{\alpha}(f)$ and $\boldsymbol{\beta}(f)$ of $r_0(f, G, P_{\min})$ which is the gain ripples at different gain targets and a fixed low total output power for EDFA type 2.

the physical intuition.

One useful heuristic developed in implementing Equation (4.11) is to approximate the changes part $[\boldsymbol{\Gamma}(f,z)-\boldsymbol{\Gamma}(f,z_{\min})]$ which has a smoother surface compared to $\boldsymbol{\Gamma}(f,z)$. It turns out to be more effective to re-condition Equation (4.11) to

$$\begin{aligned}\boldsymbol{\Gamma}(f,z) &= \boldsymbol{\Gamma}(f,z_{\min}) + [\boldsymbol{\Gamma}(f,z) - \boldsymbol{\Gamma}(f,z_{\min})] \\ &= \boldsymbol{\alpha}(f) + [\mathbf{s}(\Delta z)\boldsymbol{\beta}(f) + \mathbf{t}(\Delta z)]\end{aligned} \tag{4.12}$$

where $\boldsymbol{\alpha}(f)$ denotes the continuous cubic spline function of $\boldsymbol{\Gamma}(f,z_{\min})$ and $\boldsymbol{\beta}(f)$ is the continuous cubic spline function of $[\boldsymbol{\Gamma}(f,z_{\max})-\boldsymbol{\Gamma}(f,z_{\min})]$ since $z$ is a bounded variable and $\boldsymbol{\Gamma}$ is a bounded surface by definition. An additional constraint naturally exists, that is, $[\mathbf{s}(\Delta z=0)\boldsymbol{\beta}(f) + \mathbf{t}(\Delta z=0)] = 0$. Since it holds for all of $f$, the only way it can be true if both $\mathbf{s}(\Delta z=0)=0$ and $\mathbf{t}(\Delta z=0)=0$, which means this constraint eliminates the constant term of either a quadratic or cubic polynomial used to mathematically express $\mathbf{s}(\Delta z)$ and $\mathbf{t}(\Delta z)$.

Indeed, Equation (4.11) is similar to SVD of $\boldsymbol{\Gamma}(f,z)$ to the first order but utilizing $\mathbf{t}(z)$ to lump the sum of higher-order SVD components. Due to the surface shape of ripple profile $\mathbf{\Gamma}(f,z)$ is dominated by $\boldsymbol{\zeta}(f)$, the $\mathbf{s}(z)$ is found to be rather constant without much need to fit. As for the first and second order, the SVD of $\mathbf{\Gamma}(f,z)$ can be expressed as,

$$\begin{aligned} SVD\{\mathbf{\Gamma}(f,z)\} &= \Sigma_{i=0}^{1}\mathbf{U}_i S_i \mathbf{V}_i^T + \epsilon \\ &= (S_0\mathbf{U}_0)\mathbf{V}_0^T + (S_1\mathbf{U}_1)\mathbf{V}_1^T + \epsilon \end{aligned} \tag{4.13}$$

where $\epsilon$ means the lump of higher order components. As validated and presented in the example of **Figure 4.3**, $\boldsymbol{\alpha}(f)$ in **Equation (4.12)** is very close to $\mathbf{V}_0^T$ multiplied by a scaling factor and it is also confirmed that $\mathbf{U}_0$ remains almost constant. Meanwhile, $\boldsymbol{\beta}(f)$ in **Equation (4.12)** is also close to the second-order governing term $\mathbf{V}_1^T$. In this connection, we infer that

- $\boldsymbol{\alpha}(f)$ in **Equation (4.12)** that dominates the curvature of $\mathbf{\Gamma}(f,z)$ corresponds to the first order SVD component $(S_0\mathbf{U}_0)\mathbf{V}_0^T$,
- $\mathbf{s}(\Delta z)\boldsymbol{\beta}(f)$ in **Equation (4.12)** captures the second order SVD component $(S_1\mathbf{U}_1)\mathbf{V}_1^T$,
- $\mathbf{t}(\Delta z)$ in **Equation (4.12)** applies bias and lumps the SVD higher order residuals $\epsilon$ seen as perturbations.

Therefore, the proposed model of **Equation (4.12)** serves as an intuitive interpretation of SVD conditioned by the property of EDFA gain surface governed by theoretical model[24]. On the other hand, **Equation (4.12)** dictates that only 2 traces $\mathbf{\Gamma}(f,z_{\min})$ and $\mathbf{\Gamma}(f,z_{\max})$ are needed to effectively characterize the $\mathbf{\Gamma}$, leading to more efficient tests and extended specifications of existing EDFA device.

### 4.3.2 Gain profile model with flat WDM input

#### I A baseline model with continuous function

The semi-analytical gain model $g(f,G,T)$ proposed in[49] was adopted as a baseline method to compare. With target gain set as $G$ and target tilt set as $T$, the model can be summarized as

$$g(f, G, T) = G + \frac{T}{B}(f - f_c) + r_0(f) + T\,K(f) \tag{4.14}$$

where $f$ denotes center frequency of a optical channel, $f_c$ is the rotation pivot point of the tilt and $B$ is EDFA bandwidth. $r_0(f)$ denotes the gain ripple when tilt target is set to zero $T_0 = 0$, whose value can be calculated from the gain profile measurement $g_m(f, G, T_0)$ at a certain gain target $G$,

$$r_0(f) = g_m(f, G, T_0) - G \tag{4.15}$$

Another parameter $K(f)$ can be derived by

$$K(f) = \frac{g_m(f, G, T) - g_m(f, G, T_0)}{T} - \frac{f - f_c}{B} \tag{4.16}$$

which represents a tilt-invariant curve associated with the type of EDFA that quantifies the variation in gain ripple in proportion to the set tilt target $T \neq 0$[49]. While tilt target $T$ can be set to any non-zero value to obtain $K(f)$, it is preferred to set a higher tilt which can lead to lower measurement error according to the suggestion in[49].

Given that $g_m(f, G, T)$ is essentially a set of gain profile measurements under different gain and tilt targets, $r_0(f)$ and $K(f)$ derived from **Equation (4.15)** and **Equation (4.16)** are also discrete if curve fitting is not applied. We applied the curve fitting method described in **Section I** for making them continuous, i.e.,

$$\mathbf{r_0}(f) = \sum_{i=0}^{3} a_{i,j}(f - f_j)^i, \quad f \in [f_j, f_{j+1}] \tag{4.17}$$

and

$$\mathbf{K}(f) = \sum_{i=0}^{3} b_{i,j}(f - f_j)^i, \quad f \in [f_j, f_{j+1}] \tag{4.18}$$

for a given center frequency $f$ falling within an interval $f \in [f_j, f_{j+1}]$ and $f_j$ denotes the *knot* of cubic spline model. $\{a_{i,j}\}$ and $\{b_{i,j}\}$ denote the cubic polynomial coefficients to be estimated. Combining the semi-analytical model **Equation (4.14)** and curve fitting models **Equation (4.17)** and **Equation (4.18)**, the continuous function of "baseline" model $\mathbf{g_0}$ can be reconstructed by

$$\mathbf{g_0}(f, G, T) = G + \frac{T}{B}(f - f_c) + \mathbf{r_0}(f) + T\,\mathbf{K}(f) \tag{4.19}$$

The baseline model has been validated to be able to characterize the gain ripple changes with the tilt target. Hereafter, we will no longer model the ripple dependence on tilt variation since the baseline model is supposed to be enough to characterize it.

Nonetheless, the baseline model itself did not claim the model should be learned at which operating state, i.e., "total input/output power" or "target gain". Following the same setting of relatively low total output power to avoid saturation[49], our "baseline" model coefficients were estimated at the minimum total output power and minimum gain target given in **Table 4.1**. Then, the obtained model was tested on all the measurements tested by the step size of 1dB and within the range specified in **Table 4.1**. The error statistics of our "baseline" model **Equation (4.19)** for different types of EDFAs are given in **Table 4.2**, demonstrating this model can result in a large maximum error beyond 1dB, as well as RMSE beyond 0.2dB.

### II Gain and power dependent model

To improve the accuracy of the baseline model, the consistency of the parameters $r_0(f)$ and $K(f)$ at the minimum total output power was firstly analyzed. It can be observed in **Figure 4.4** that $r_0(f)$ and $K(f)$ show obvious variations with different gain target $G$, indicting that these two tilt-independent parameters are probably dependent on gain target $G$ at a fixed total output power. This is quite reasonable as this situation is essentially different total input powers that are amplified by different gain targets (pump) to maintain the same total output power, corresponding to different operating states as is discussed in **Section II**. Total output power instead of total input power was focused here since the maximum saturated output power is a fixed value given by specification, which makes the input parameter space $\{f, G, T, P\}$ of our proposed model more consistent and easier to establish. Therefore, in this subsection, the aim is to obtain continuous functions to characterize the unknown dependency on gain targets and total output power.

The first step is to obtain the function of two new parameters $\mathbf{r_0}(f, G, P_{\min})$ and

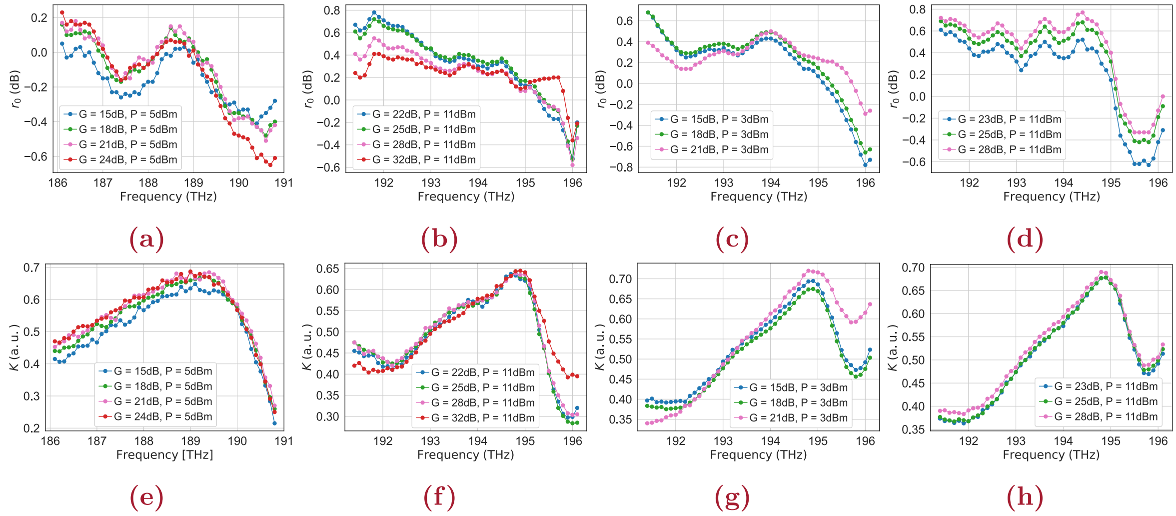



**Figure 4.4:** The measured profile of $r_0(f)$ at a minimum total output power with different target gain of different EDFA types: type 1 (a), type 5 (b), type 7 (c) and type 9 (d); and the measured profile of $K(f)$ at a fixed output power with different target gain of different EDFA types: type 1 (e), type 5 (f), type 7 (g) and type 9 (h).

$\mathbf{K}(f, G, P_{\text{min}})$ at a fixed minimum total output power $P_{\text{min}}$. The minimum output power parameter $P_{\text{min}}$ will be omitted in parentheses of these two parameters for simplicity hereafter as it simply implies that the model coefficients are estimated from all the measurements under the setting of $P_{\text{min}}$, which does not impact the modeling approach itself. A reference cubic spline curve $\boldsymbol{\alpha}(f)$ can be easily fitted from ripple measurements $r_0(f, G_{\text{min}})$ at a reference gain target $G_{\text{min}}$ which is given by **Table 4.1**. Then, according to the model described in **Section II** we have

$$\Delta r_0(f, G - G_{\text{min}}) = r_0(f, G) - \boldsymbol{\alpha}(f) \tag{4.20}$$

we denote $G - G_{\text{min}}$ as $\Delta G$ for simplicity hereafter, which satisfies $G_{\text{min}} \leq G \leq G_{\text{max}}$. Then, since $\boldsymbol{\beta}(f)$ can be fitted from the measurements at the maximum gain target $\Delta r_0(f, \Delta G_{\text{max}})$, the continuous function of $\boldsymbol{\Delta}\mathbf{r_0}(f, \Delta G)$ will be obtained by fitting

$$\boldsymbol{\Delta}\mathbf{r_0}(f, \Delta G) = \mathbf{s}(\Delta G)\boldsymbol{\beta}(f) + \mathbf{t}(\Delta G) \tag{4.21}$$

which has been explained in Section 4.3.1-II and the coefficients of $\mathbf{s}(\Delta G)$ and $\mathbf{t}(\Delta G)$ are estimated by simple least-squares. Now, the improved version of $\mathbf{r_0}$ can be summarized as

$$\mathbf{r_0}(f, G) = \boldsymbol{\alpha}(f) + [\mathbf{s}(\Delta G)\boldsymbol{\beta}(f) + \mathbf{t}(\Delta G)] \tag{4.22}$$

where $\boldsymbol{\alpha}(f)$ and $\boldsymbol{\beta}(f)$ are cubic spline curves, $\mathbf{s}(\Delta G)$ and $\mathbf{t}(\Delta G)$ are either quadratic or cubic polynomials without constant term. In **Figure 4.5**, we can see the obtained model (surface) agrees with the real measurements (dots) for different types of EDFA. Due to space limitation, only 4 of 9 types are shown here. The same method can also be applied to obtain the function of $\mathbf{K}(f;G)$. Then, our improved model can be expressed as

$$\mathbf{g_1}(f,G,T) = G + \frac{T}{B}(f - f_c) + \mathbf{r_0}(f,G) + T\,\mathbf{K}(f,G) \tag{4.23}$$

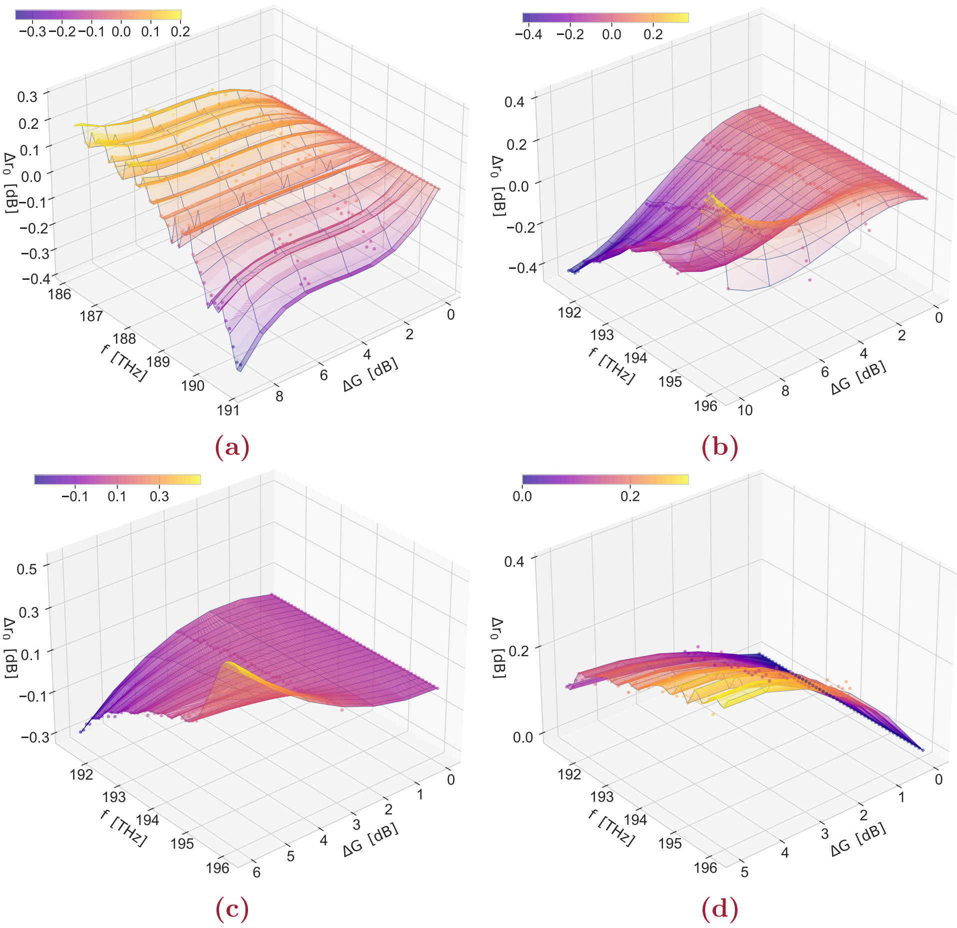



**Figure 4.5:** The obtained model (surface) and real measurements (dots) of $\boldsymbol{\Delta}\mathbf{r_0}(f;\Delta G)$ for EDFA types: type 1 (a), type 5 (b), type 7 (c) and type 9 (d).

Due to the fact that the residual errors of the proposed model $\mathbf{g_1}(f,G,T)$ are highly related to the increase in the output power until it saturates, our final improved model is to build a continuous function of $\mathbf{g_2}(f,G,T,P)$. We define a residual modeling error term $\mathbf{u}(f,P)$ dependent on the output power based on the previous improved model, which can be calculated by

$$\mathrm{u}(f, P) = \mathrm{g}_2(f, G, T, P) - \mathbf{g_1}(f, G, T) \tag{4.24}$$

For a given type of EDFA, the value of its saturated output power $P_{\max}$ is determined by the specification/data-sheet. Also, since Equation (4.23) was obtained at $P_{\min}$, we have $\mathrm{u}(f, P_{\min})$ approaching zero asymptotically so that $\boldsymbol{\alpha}(f) = 0$. Then, denote the cubic spline function of $u(f, P_{\max})$ as $\boldsymbol{\beta}(f)$, similarly we have

$$\mathbf{u}(f, \Delta P) = \mathbf{s}(\Delta P)\boldsymbol{\beta}(f) + \mathbf{t}(\Delta P) \tag{4.25}$$

where $\Delta P = P - P_{\min}$ and $P_{\min} \leq P \leq P_{\max}$. The obtained models of $\mathbf{u}(f, \Delta P)$ are shown in the surfaces of Figure 4.6 and the real data samples are shown in the dots of Figure 4.6. It can be seen that the trend of residual errors vary with increasing output power stays nearly the same for all types of EDFA, especially at the higher frequencies and when total output power is approaching its saturation point. However, the amplitudes of residual errors for different types are obviously different suggesting an type-specific characteristic. Thus, our final optimized model $\mathbf{g_2}(f, G, T, P)$ can be written as

$$\mathbf{g_2}(f, G, T, P) = G + \frac{T}{B}(f - f_c) + \mathbf{r_0}(f, G) + T\,\mathbf{K}(f, G) + \mathbf{u}(f, P) \tag{4.26}$$

From Table 4.2, we can see that the RMSE of our optimized model $\mathbf{g_2}(f, G, T, P)$ has been greatly reduced to less than 0.083 dB for all types of EDFA understudy and with the minimum RMSE of 0.038 dB for EDFA type 6. The mean error has also centered around 0.0dB and $3\sigma$ ranges from 0.113 dB to 0.249 dB for all types of EDFA. All of these results are acceptable for mirror modeling EDFA gain profile simply using cubic polynomials, validating the feasibility of the proposed model.

### 4.3.3 Input pre-emphasis aware gain model

Considering the cascading of EDFAs is common in real networks, the accurate prediction of gain profiles with different pre-emphasis WDM inputs will be beneficial for

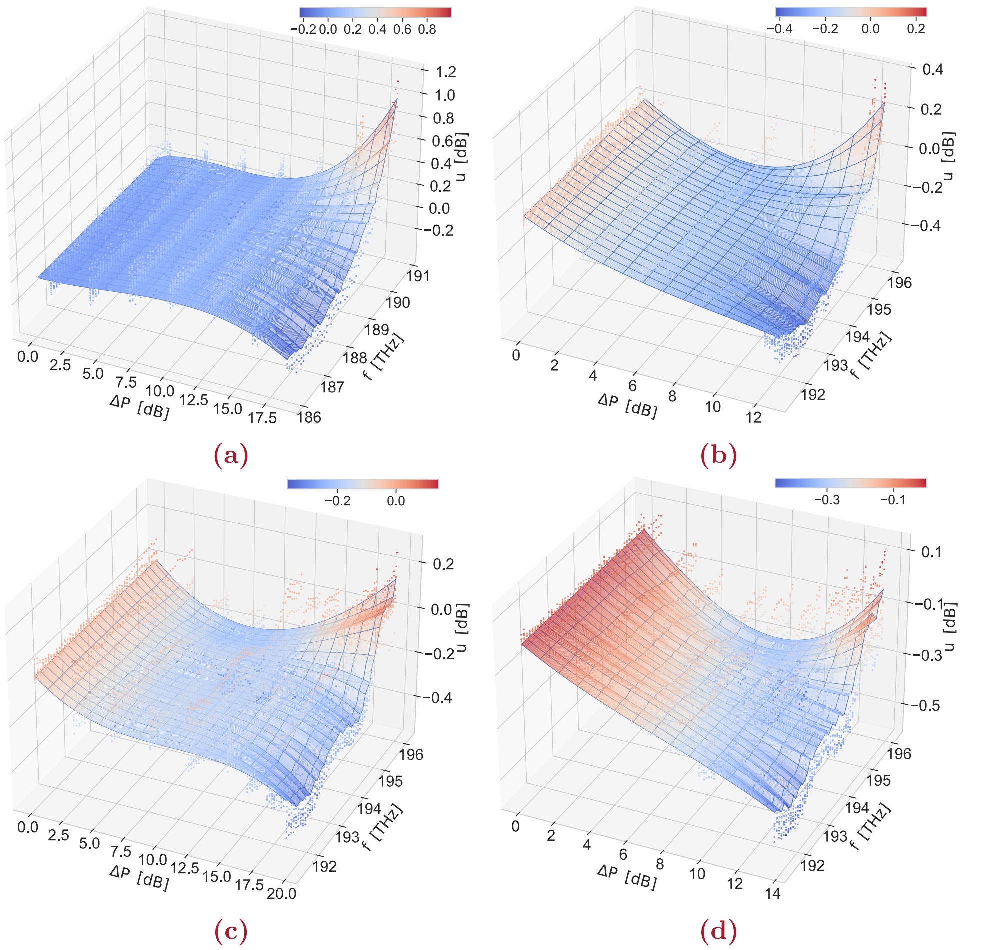



**Figure 4.6:** The obtained model (surface) and real measurements (dots) of $\mathbf{u}(f;\Delta P)$ for EDFA types: type 1 (a), type 5 (b), type 7 (c) and type 9 (d).

building the network digital twin[131]. In this section, we will demonstrate the application of $\mathbf{g_2}(f,G,T,P)$ to a pre-emphasis aware gain profile model based on two observations and simple refinement of the input parameters of $\mathbf{g_2}(f,G,T,P)$.

## I Dataset Generation

We measured the gain profile of EDFA type 1 and type 2 for comparing the different gain behaviours in L-Band and C-Band, with 45 shape patterns of pre-emphasis WDM inputs. Since each pattern of input corresponds to a fixed total input power, we will set different gain targets until the output saturated and also under each gain target the tilt target was set from 0 to -4. The step size is set to 1dB when tuning the target gain or tilt. This allows for a total number of 88320 of data samples understudy for each type. The amplitude of pre-emphasis ranges from a minimum of 6dB to a maximum value of 15dB since we care more about a worst case modeling performance. Most shapes of input pre-emphasis is set to some fixed shapes, for example, the shape of slash "/" denotes this shape has higher power at higher frequency while lower power at lower frequencies, with some randomly generated fluctuation imposed on that standard shape of "/", as is shown in **Figure 4.7**. Conversely, the shape of backslash "\" denotes higher power appears at lower frequencies while lower power appears at higher frequencies. And the shape of "W" denotes the higher powers are around the maximum and minimum and the center of the whole frequency range, so on and so forth. The same 45 input power profiles of **Figure 4.7** were applied when measuring EDFA type 2 but the x-axis changed to the frequency range of C band.

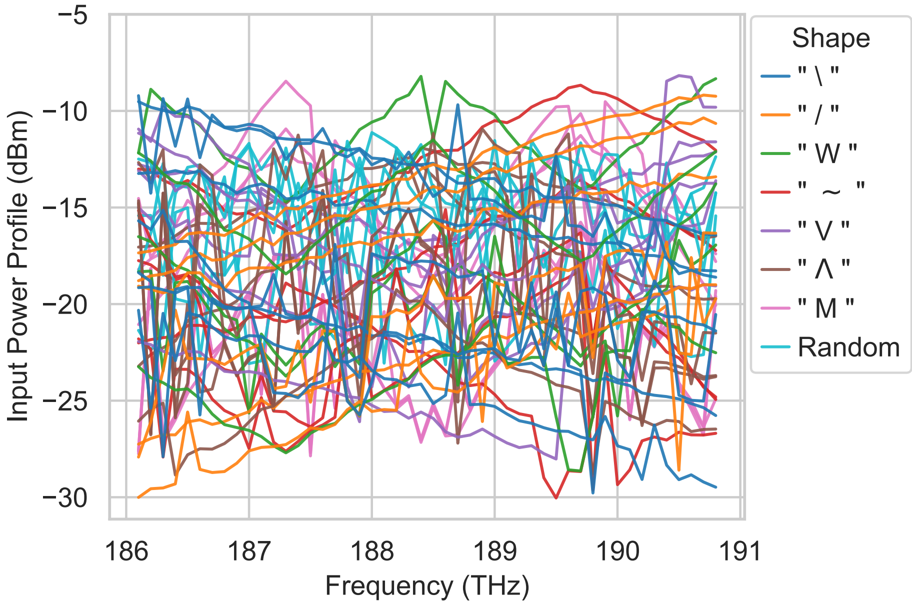



**Figure 4.7:** 45 pre-emphasis input power profiles for measuring gain profile of EDFA type 1 (L-Band).

## II Methodology

Consider EDFAs working in AGC mode and with pre-emphasis WDM inputs, our pre-emphasis aware (PA) gain profile model is based on the two observations directly from the dataset:

1. EDFAs adjust the pump power dynamically to maintain a contant gain by monitoring the total input and output power;
2. the gain spectrum of an EDFA tends to exhibit stronger gain effects in frequency ranges with lower input power, while experiencing gain suppression in regions with higher input power levels.

The function of $\mathbf{g_2}(f, G, T, P)$ acts as an "equivalent flat (EF)" gain function, which actually incorporates the information of most of the combinations of various inputs and pumps, which is built as a look-up table in[48] and a neural networks in[60]. However, these two references did not consider EDFA working in the AGC mode, making it hard to be utilized. Therefore, base on the $1^{st}$ observation, we first assume a feedback control existing when EDFA has pre-emphasis input. Denote a random input power profile $p_{\text{in}}(f)$ and its corresponding total input power as $P_{\text{in}}$. When the target gain is $G$, it will result in a target total output power of

$$P_{\text{out}}^{\text{Target}} = P_{\text{in}} + G \tag{4.27}$$

and its corresponding equivalent flat gain profile $\mathbf{g}^{\text{EF}}(f)$ is calculated by

$$\mathbf{g}^{\text{EF}}(f) = \mathbf{g_2}\left(f,\, G,\, T,\, P_{\text{out}}^{\text{Target}}\right) \tag{4.28}$$

Then the assumed equivalent flat output profile $p_{\text{out}}^{\text{EF}}(f)$ is supposed to be

$$p_{\text{out}}^{\text{EF}}(f) = p_{\text{in}}(f) + \mathbf{g}^{\text{EF}}(f) \tag{4.29}$$

And denote the corresponding total power of the power profile $p_{\text{out}}^{\text{EF}}(f)$ as $P_{\text{out}}^{\text{EF}}$. When the pre-emphasis input profile deviates from the flat input, it will result in a power offset between the target $P_{\text{out}}^{\text{Target}}$ and the equivalent $P_{\text{out}}^{\text{EF}}$. Define this power offset $\delta P$

as

$$\delta P = P_{\text{out}}^{\text{EF}} - P_{\text{out}}^{\text{Target}} \tag{4.30}$$

Recall that EDFA adjusts pump power dynamically to maintain the constant gain, this power offset $\delta P$ will be deducted in the input parameter, target gain $G$ to mimic this procedure. The input parameter $G$ of $\mathbf{g_2}$ will be replaced by $(G - \delta P)$ to become

$$\mathbf{g_2}\left(f,\, G - \delta P,\, T,\, P_{\text{out}}^{\text{Target}}\right) \tag{4.31}$$

leading to a new refined gain profile estimation that has been obtained by utilizing an input pre-emphasis aware (PA) term $\delta P$.

As for the $2^{nd}$ observation, this power offset $\delta P$ can also be utilized to refine the input parameter of tilt $T$ in the EF gain function $\mathbf{g_2}$ by replacing $T$ with $(T + \delta P)$ such that our proposed PA gain profile model $\mathbf{g}_2^{\text{PA}}$ will be

$$\mathbf{g}_2^{\text{PA}} = \mathbf{g_2}\left(f,\, G - \delta P,\, T + \delta P,\, P_{\text{out}}^{\text{Target}}\right) \tag{4.32}$$

This tilt modification can be supported by the following arguments. First, the target tilt ranges from 0 to -4 for EDFA type 1 and 2 (Table 4.1) so that most gain profile measurements exhibit a higher gain in a higher frequency range when $T < 0$ or a relatively flat gain when $T = 0$. For simplicity, we assume the center frequency of the EDFA bandwidth to be 0 and express the gain profile as $af_k$ where $a > 0$ models tilt with frequency as x-axis. Denote the flat WDM input channel power as $p_0$. If the pre-emphasis input profile is $(p_0 - bf_k)$ where $b > 0$ corresponds to positive power slope (e.g., $p(f_{min}) - p(f_{max}) \gg 0$), $\delta P$ will be negative because

$$\sum_k \left[(af_k)(p_0 - bf_k) - (af_k)p_0\right] = \sum_k \left(-abf_k^2\right) < 0 \tag{4.33}$$

Recall the $2^{nd}$ observation aforementioned, the pre-emphasis input profile with a large positive power slope, $p(f_{min}) - p(f_{max}) \gg 0$, will suppress gain at the lower frequency range while enhance gain at the higher frequency range. Consequently, the input parameter $T$ moves in a negative direction in Equation (4.32) justifying the replacement of $(T + \delta P)$. Analogously, we conclude $\delta P$ will be a positive value if the input pre-emphasis profile exhibits a large negative power slope (e.g. $p(f_{min}) - p(f_{max}) \ll 0$).

On the other hand, according to **Equation (4.33)**, the absolute value of $\delta P$ is proportional to the slope amplitude of the input power profile and also proportional to the slope amplitude of gain profile $T$. From this perspective, the overall slope amplitude of the pre-emphasis input can be reflected in the amplitude of $|\delta P|$ which controls the level of final gain slope by using **Equation (4.32)**.

### 4.3.4 Results and Discussions

#### I EF Gain Model

**Table 4.2:** Error statistics between baseline model ($g_0$) and optimizied model ($g_2$).

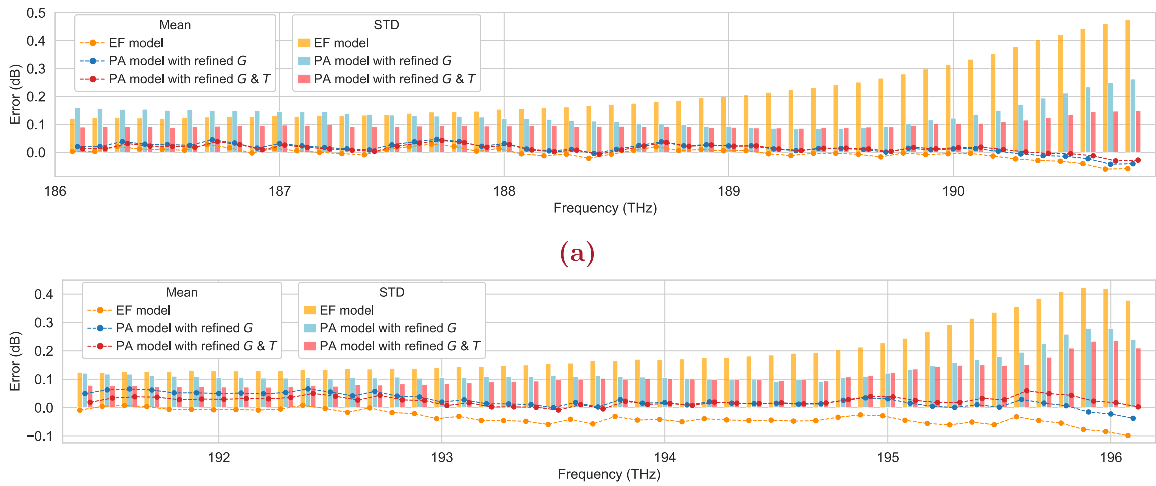
| Vendor | Type | Model | Statistics of modeling error | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **Min** | **Max** | **MAE** | **Mean** | **3$\sigma$** | **RMSE** |
| A | 1 | $g_0$ | -0.51 | 1.14 | 0.106 | 0.047 | 0.415 | 0.146 |
| | | $g_2$ | **-0.398** | **0.325** | **0.066** | **0.003** | **0.249** | **0.083** |
| | 2 | $g_0$ | -0.488 | 0.827 | 0.089 | 0.022 | 0.364 | 0.123 |
| | | $g_2$ | **-0.444** | **0.339** | **0.064** | **0.001** | **0.245** | **0.082** |
| | 3 | $g_0$ | -0.515 | 1.005 | 0.174 | -0.018 | 0.696 | 0.233 |
| | | $g_2$ | **-0.264** | **0.256** | **0.053** | **-0.0** | **0.215** | **0.072** |
| | 4 | $g_0$ | -0.49 | 1.07 | 0.15 | -0.067 | 0.564 | 0.2 |
| | | $g_2$ | **-0.324** | **0.456** | **0.046** | **0.001** | **0.204** | **0.068** |
| | 5 | $g_0$ | -0.73 | 0.39 | 0.247 | -0.226 | 0.555 | 0.292 |
| | | $g_2$ | **-0.249** | **0.266** | **0.039** | **0.002** | **0.156** | **0.052** |
| | 6 | $g_0$ | -0.29 | 0.77 | 0.107 | 0.032 | 0.455 | 0.155 |
| | | $g_2$ | **-0.243** | **0.125** | **0.029** | **-0.0** | **0.113** | **0.038** |
| B | 7 | $g_0$ | -0.59 | 0.56 | 0.148 | -0.11 | 0.416 | 0.177 |
| | | $g_2$ | **-0.341** | **0.25** | **0.057** | **0.0** | **0.228** | **0.076** |
| | 8 | $g_0$ | -0.39 | 0.71 | 0.111 | 0.045 | 0.436 | 0.152 |
| | | $g_2$ | **-0.373** | **0.241** | **0.057** | **0.0** | **0.228** | **0.076** |
| | 9 | $g_0$ | -0.46 | 0.333 | 0.154 | -0.117 | 0.412 | 0.181 |
| | | $g_2$ | **-0.254** | **0.219** | **0.063** | **0.002** | **0.229** | **0.076** |

It should be noted that the flat gain range is limited for some types of EDFAs because of a limited operational range of VOA inside the two stages of EDF, whose setting controls the realistic gain tilt implementation. Especially at the maximum gain target, we found that the tilt target $T = 0$ is no longer guaranteed so that its actual gain profile with actual tilt possibly much higher than 0, as the VOA inside the EDFA module is approaching its limit. This will cause a failure of the baseline model **Equation (4.14)** because the model itself assumes the gain ripple varies linearly with the first-order approximation of the target tilt. For instance, the normal tunable gain range given by the product specification is from 15dB to 25dB for EDFA type 1 and type 2 but we only model the measurements with gain target ranges from 15dB to 24dB as shown in **Table 4.1**. The reason is that the tilt target $T = 0$ at 25dB gain target is not guaranteed, which means the actual gain profiles appear the same regardless of whether the tilt target is set to 0 or -1. The same phenomenon has been observed with EDFA type 7/8/9 so that their target gain ranges shown in **Table 4.1** are not the full tunable ranges given by the product specification. Focusing on the flat gain range, our model is based on the assumption that **Equation (4.14)** is accurate enough to characterize the variations of the gain ripple w.r.t. the tilt target. However, we believe that a more complex regression model that can characterize the tilt dependence of the gain ripple beyond the flat gain range can further improve the usability of the proposed model and this will be one of our future research directions.

In summary, the optimized gain model fitted by the aforementioned methods can be utilized to fully characterize the gain profile variations within the normal operational gain, tilt and power range with flat WDM inputs. It facilitates better mirror modeling of the gain behaviors under different input and pump settings of EDFA and also simplifying network design and optimization problems and contribute to building better network digital twins[132,133].

## II PA Gain Model

Table 4.3 shows the error statistics of three models: 1) EF model in Equation (4.28), 2) PA model with refined gain parameter in Equation (4.31) and 3) PA model with refined gain and tilt in Equation (4.32), testing on a total of 88320 data samples each for EDFA type 1 (L band) and type 2 (C band). It can be seen from Table 4.3 that our EF model obtained with flat input becomes invalid and leads to a large maximum and minimum error of -1.803 dB and 1.65 dB for EDFA type 1 when the input has pre-emphasis. The proposed PA model with refined gain and tilt parameters obviously reduces the maximum and minimum error and also the RMSE by almost half to around 0.112 dB for type 1 and 0.12 dB for type 2, which validated the proposed model.




(a)

(b)

**Figure 4.8:** The error statistics at different frequency of the equivalent flat (EF) model and the pre-emphasis aware (PA) models with different refined parameters for (a) EDFA type 1 and (b) type 2.

The frequency-dependent error statistics are illustrated in Figure 4.8(a) and (b) for EDFA type 1 and type 2, respectively, to compare the performance of EF model, PA model only with refined gain and the PA model with both refined gain and tilt. Obvious frequency-dependent errors can be seen in the orange dotted lines and bars for both types with the EF model Equation (4.28). After refining the input parameter of gain, the frequency-dependent errors are significantly reduced as shown in the blue

**Table 4.3:** Error statistics of EF and PA models testing on the dataset measured under pre-emphasis WDM inputs.

| Type | Model | Statistics of modeling error | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Min** | **Max** | **MAE** | **Mean** | **3$\sigma$** | **RMSE** |
| 1 | EF | -1.803 | 1.65 | 0.172 | -0.003 | 0.775 | 0.259 |
| | PA w/ refined $G$ | -0.924 | 0.802 | 0.118 | 0.018 | 0.466 | 0.156 |
| | PA w/ refined $G\&T$ | **-0.578** | **0.471** | **0.088** | **0.017** | **0.331** | **0.112** |
| 2 | EF | -1.382 | 1.286 | 0.154 | -0.034 | 0.657 | 0.222 |
| | PA w/ refined $G$ | -0.888 | 0.923 | 0.108 | 0.025 | 0.42 | 0.142 |
| | PA w/ refined $G\&T$ | **-0.66** | **0.976** | **0.09** | **0.023** | **0.354** | **0.12** |

dotted lines and bars in **Figure 4.8**(a) and (b). Then, after refining both gain and tilt parameter applied in the PA model, the frequency-dependent errors can be further reduced as shown in the red dotted lines and bars in **Figure 4.8**(a). However, the frequency-dependent errors at the higher frequency range around 195.8 THz for EDFA type 2 in **Figure 4.8**(b) is not as good as that of type 1 in **Figure 4.8**(a). It is possibly due to severe spectral hole burning (SHB) phenomenon appearing around the frequency of 195.8 THz where there has saturating signal power. It was confirmed by subtracting the measured gain spectra at 195.8 THz from a reference gain spectra at 191.8 THz far from the frequency of interest. And the measurement method of SHB is the same as the spectral subtraction technique described in [134,135]. The resulting gain difference is shown in **Figure 4.9** which was obtained at different gain targets. The results in **Figure 4.9** confirm SHB effect at higher frequencies around 195.8THz is dominant than other frequencies, which in turn explains the result in **Figure 4.8**(b) since we do not consider a specific modeling of SHB effect in this work. It will be our future work direction, for example, learning SHB mechanism and pattern from our dataset of QoT estimation, e.g., at what situation (input power shape) it becomes non-trivial.

The inhomogeneous gain saturation, generally manifests as SHB which was only observed at C-band EDFA type 2 not L-band EDFA type 1, which agree with the gain behaviour presented and discussed in [136,137] where the authors showed that L band gain saturation manifests as a broad-band gain depression. This gain depression or

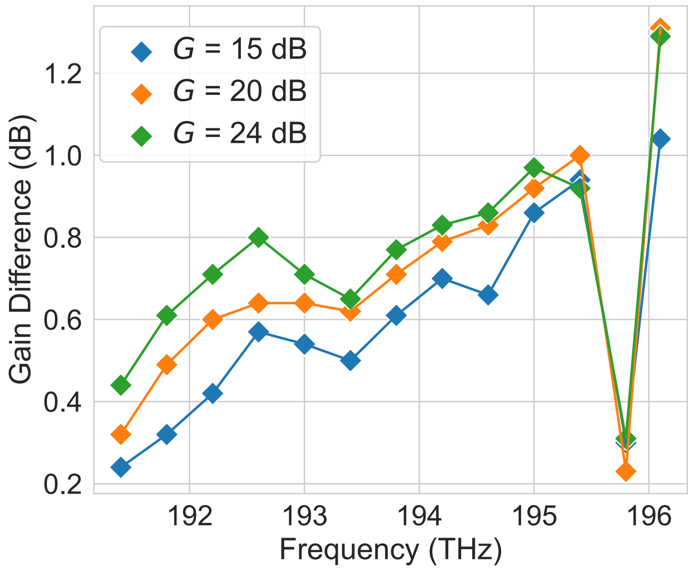
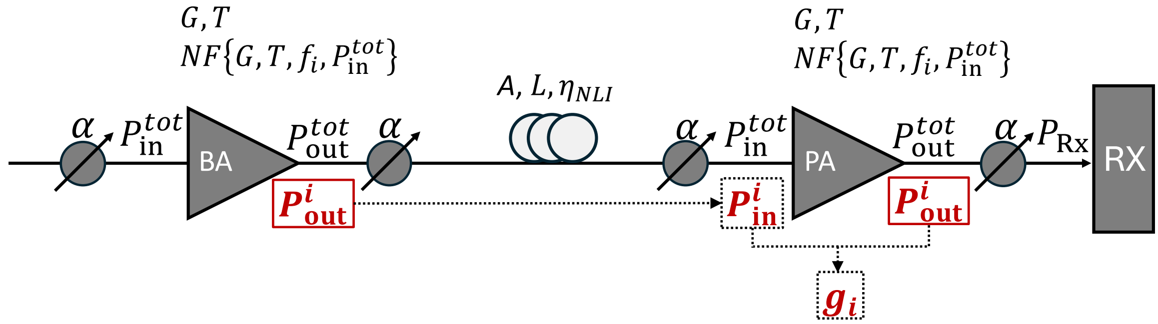



**Figure 4.9:** The gain difference calculated by subtracting the gain spectra at 195.8THz from a reference gain spectra at 191.8THz, measured at different gain targets with tilt target set to 0.

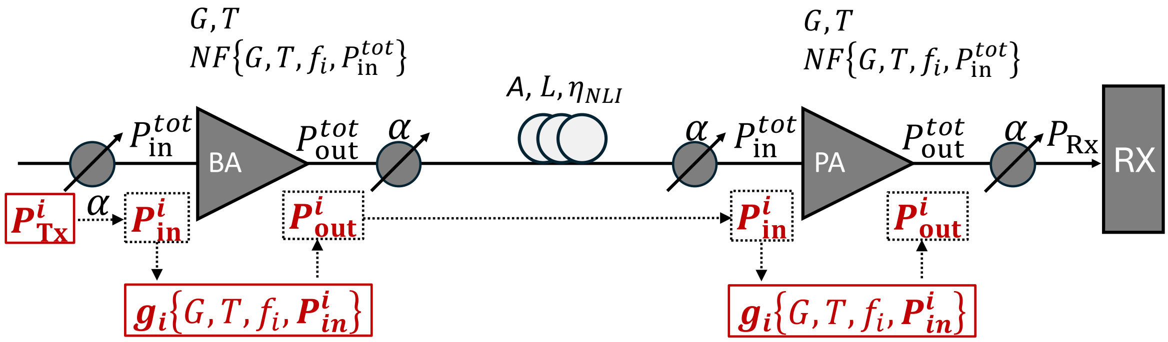
suppression of L-band EDFA exhibit different amplitude of suppression over lower and higher frequencies, as it can be seen in the red dotted lines and bars in **Figure 4.8**(a) where there are still slight higher error at higher frequencies. This may lead to a data-driven approach, which means, we can refine our input parameter tilt to be $(T + c \cdot \delta P)$ in the PA model of **Equation (4.32)**, where $c$ is a correction factor that can be learned directly from the measurements. This will also be one of our future research works to make the model more accurate.

## 4.4 Validation of Greenfield QoT Estimation with the proposed Gain Profile Models

To bridge the gap between brownfield QoT estimation and greenfield QoT estimation, we adopted the proposed gain profile modeling to estimate the signal power of each span and applied the analytical model of GSNR described in **Chapter 1.3.2 I** and **Chapter 2.3**, with the focus on the metro DCI networks. To validate the feasibility of the greenfield QoT estimation, we compared the resulting statistics of GSNR estimation errors with those using our previous brownfield methods with OCM power as one of the input parameters of GSNR model.




**(a) Brownfield**: real-time PM data including the OCM.




**(b) Greenfield**: OCM is not available so that gain profile models are utilized.

**Figure 4.10:** A comparison of (a) brownfield QoT estimation where the OCM is available so that $P^i_{\text{out}}$ can be readily obtained at the output of each EDFA, which is enclosed with a solid-line box. Meanwhile, the input power profile $P^i_{\text{in}}$ is enclosed with a dashed-line box to show it is estimated rather than monitored by actual physical devices; and (b) greenfield QoT estimation where the OCM power is not available but can be estimated from our gain model with the designed parameters of Tx power profile and WSS attenuation, and EDFA gain and tilt targets.

The comparison of two different network scenarios is shown in **Figure 4.10**. The difference lies in the OCM power, which determines the signal power and the approach of gain profile estimations. In brownfield QoT model, gain profiles are obtained by estimating the input power profile (enclosed with a dashed-line box) of the next EDFA and then subtracting the OCM power profile (enclosed with a solid-line box) of the next EDFA, which has been described in **Chapter 2.3**. It has been validated that this approach of gain profile estimation facilitates a better estimation of the ASE noise power and $SNR_{\text{ASE}}$ in **Chapter 2.3**. However, in greenfield QoT model where the

OCM is not available, we directly use our polynomial-based gain model to obtain the gain profile of each EDFA, from which the input power profile and output power profile of each EDFA can also be estimated (enclosed with a dashed-line box). In this case, we assume the signal power per channel after Tx and WSS is a known parameter $P_{\text{in}}^{i}$ and can be designed/configured by the attenuation of WSS. Other parameters such as EDFA gain target $G$, tilt target $T$, total input power $P_{\text{in}}^{tot}$ and total output power $P_{\text{out}}^{tot}$ are assumed to be fixed the same as the settings in the brownfield since they are also easily designed/configured. Other parameters such as lumped loss are also assumed to be fixed the same as the settings in the brownfield.

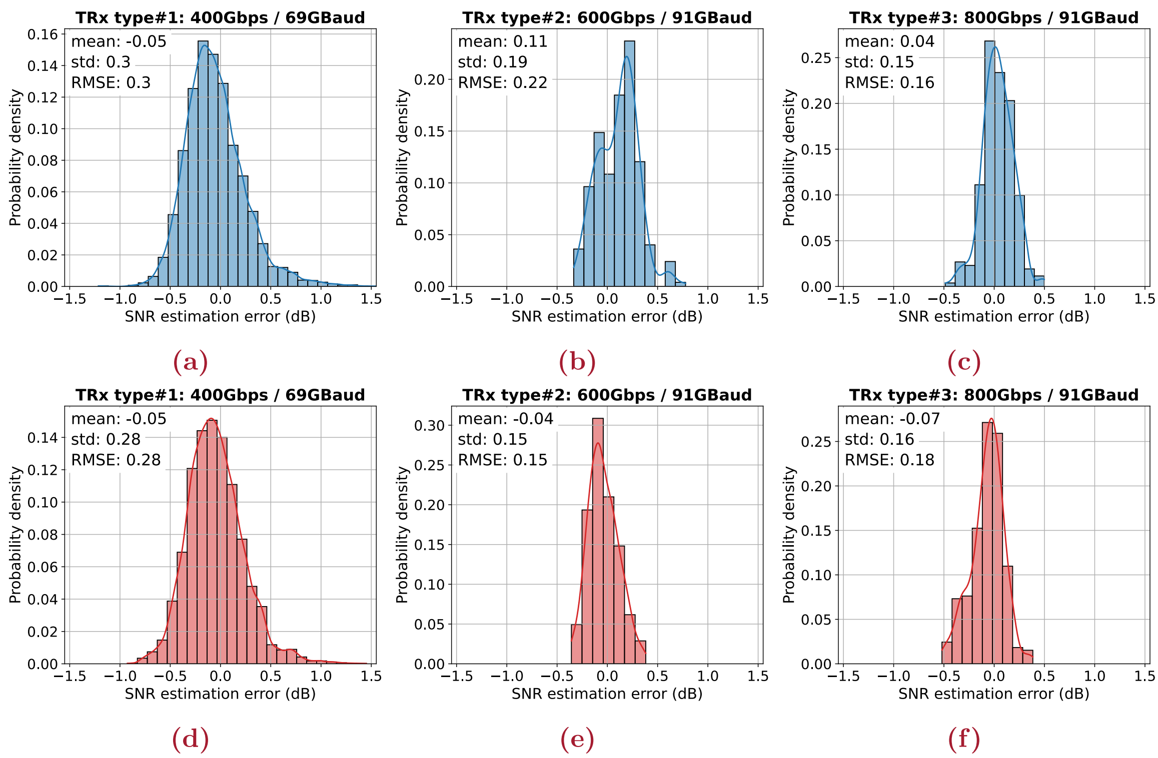



**Figure 4.11:** Error statistics of **brownfield** QoT estimation for (a)TRx type 1 (400G/69GBaud) and (b)TRx type 2 (600G/91GBaud) and (c)TRx type 3 (800G/91GBaud), and **greenfield** QoT estimation for (d)TRx type 1 (400G/69GBaud) and (e)TRx type 2 (600G/91GBaud) and (f)TRx type 3 (800G/91GBaud).

By substituting the input parameters of OCM power profile and estimated gain profile in brownfield QoT estimation with the gain model built from measurements and estimated signal power profile, we can obtain the SNR estimations for greenfield network scenario. These SNR estimations are then compared with the ground-truth SNR converted from pre-FEC BER using B2B models described in **Chapter 3.3** to obtain the estimation errors. The results of error statistics are shown and compared in **Figure 4.11**. The error distributions shown in blue are brownfield QoT estimations using OCM power from PM database, and the error distributions shown in red are greenfield QoT estimations using the gain profile model proposed in this section. These results show that there is no obvious performance discrepancy between brownfield QoT and greenfield QoT evaluated at different types of TRXs, and thus validate the effectiveness of using our gain profile model to estimate the signal power profile evolution and ASE noise power profile with using accurate NF measurements.

## 4.5 Summary

In this Chapter, we proposed a low-complexity polynomial fitting model to characterize gain profiles with WDM input under the normal range of gain and tilt targets and total input power. The model was built based on cubic polynomials and with curve fitting and surface fitting method, and the coefficients of this model are obtained by simple least-squares fitting from the measurements. The feasibility of our proposed model was validated on 9 types of EDFAs including both C and L band from two different vendors. The RMSE of proposed model on 9 types of EDFA is well below 0.083 dB and can reach to a minimum RMSE of 0.038dB. The applicability of the proposed gain model on EDFA gain spectra with pre-emphasis WDM inputs has also been investigated. With simple refinements to the input parameters of our proposed polynomial-based gain model, an acceptable RMSE of the model was also achieved with no more than 0.12 dB for two C and L band EDFAs. This input pre-emphasis aware gain profile model can be applied in the real network where there are multiple EDFAs cascading, whose property of quadratic continuous differentiability can also accelerate the network optimization.

The accuracy of greenfield QoT estimation enabled by our gain profile model above is comparable to that of the brownfield live production network, further demonstrating that even at the pre-deployment stage of the DCI network, reliable QoT estimation can be achieved through a set of design parameters combined with an accurate optical amplifier model. Such a model enables network operators to avoid unnecessary rework during the design, planning, and construction phases, thereby maximizing the utilization of network capacity and resources.

# 5

# Open-source dataset from live production network for QoT Estimation

## 5.1 Overview

In this Chapter, we introduce our open-source dataset from real optical transport networks, which is available on GitHub[138]. The comprehensive information required for QoT estimation is also provided on GitHub[138]. The network segments considered in the dataset consist of 6 OMSs and 25 OCHs as shown in **Figure 5.1**. All OCHs have data available for 162 h, with one data point per hour. Furthermore, we have also provided an additional 181 h of data for 6 OCHs after the change of link conditions, i.e., the gains of the EDFAs are the link, compared to the aforementioned time period. With this dataset, over 10,000 GSNRs can be estimated. We also provide real BER measurements and the corresponding BER-GSNR curves to serve as a benchmark for comparison with the estimated GSNRs. This section describes how our data can be utilized to establish the network topology and evaluate its performance.

## 5.2 Network Topology

This network is composed of 6 OMSs as shown in **Figure 5.1**(a). Each OMS is composed of a different number of three components: an optical line router (OLR) reconfigurable optical add-drop multiplexer (ROADM), an optical line amplifier (OLA), and an optical line equalizer (OLE). Each OLA and OLE is considered a single device, whose ID is labeled on top of the component. As OLRs are typically used in ROADMs, and a ROADM (one OLR device here) can have many OLRs (degrees), we use Dx to further identify an OLR, and the subscript x in Dx denotes the ID of the degree. For example, OMS 2 and OMS 3 share the same OLR device 10, and D1 is used in OMS 2 while D2 is used in OMS 3. The fiber type is G.652D, and its length is also shown in **Figure 5.1**. The lengths of OMSs range from 74 to 1249.7 km.

There are 25 OCHs in this network, and they are classified as four different OCH groups, which include OCHs from the same source to the same destination through the same route. The channel spacing is 100 GHz for all OCHs, and the modulation format is DP-QPSK with different baud rates. **Figure 5.1**(b) shows the path of each OCH group. For example, OCH Group 1 passes through OMS 1, OMS 2, OMS 3, and OMS 4 in sequence. Each OCH group contains a different number of OCHs, and their center frequencies are also denoted in **Figure 5.1**(b).

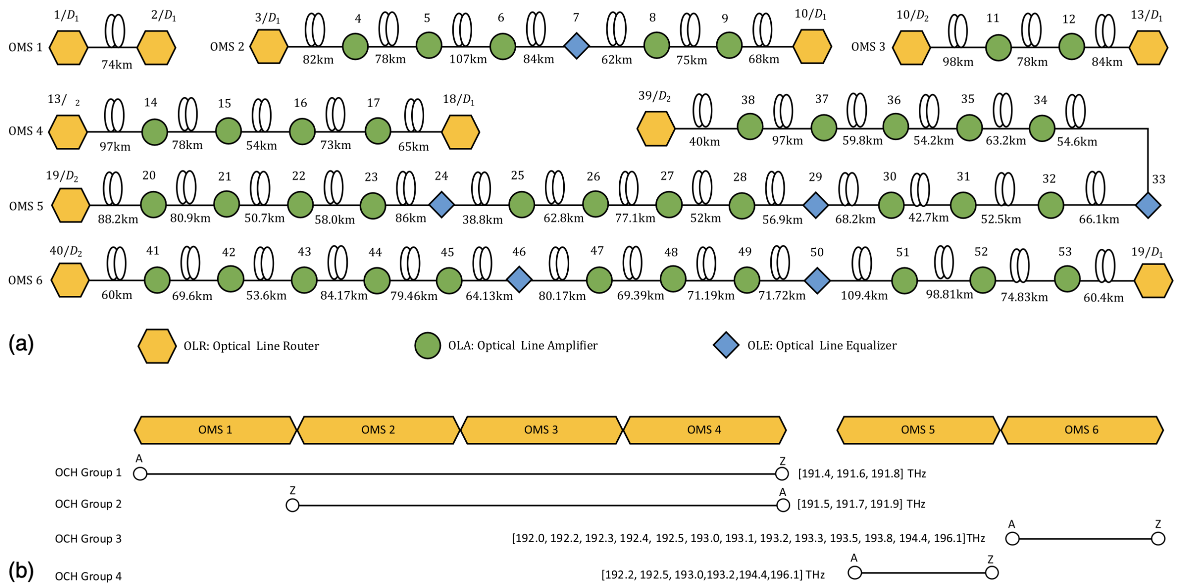



**Figure 5.1:** (a) Optical multiplex sections (OMSs) in the network. (b) Detailed information about optical channels (OCHs), including routing paths and center frequencies.

## 5.3 Performance Data

In addition to the network topology, it is necessary to know the key performance of network devices, including optical power values, the BER of the terminals, and specifications of different devices. Such information is introduced in this section by two parts: optical line systems and terminals.

- **Optical Line System**: Three types of devices are used in optical line systems: OLRs, OLAs, and OLEs. An optical line system may have all or some of the three line devices, depending on the system length and required functions. Two key performance data of the line devices, total optical power and channel power, are recorded in table $performance\_optical$ and $ocm$, which can be queried using the device ID and locator.

- **Optical Terminals**: In the table $performance\_elec$, performance of optical terminals (transponders) is saved. Each terminal is identified by the OCH group index, center frequency, and A or Z end by the column name as $ochgroup$, $center\_freuqency$, $side$. For each OCH, transponders at the two ends are denoted by A and Z. BER-OSNR curves of different kinds of transponders are measured under the B2B condition, and they are identified by the PN of the transponder. The baud rates and data rates of the transponders are provided in the table $ber - Gosnr$. In the table $performance\_elec$, statistical data of the pre-FEC BER is saved in the column $value$, including the max, min, and average value within each 15 min. It is worth noting that the time interval provided by us is also 1 h, and we only select the 15-min data within a 1-h period.

## 5.4 Summary

Our open-source data consists of 4 OCH groups and 25 OCHs. All OCH groups have data available for 162 h, with one data point per hour. Furthermore, we have also provided an additional 181 h of data for OCH Group 1, where the gains of the EDFAs were changed compared to the aforementioned time period. More than 10,000 GOSNRs can be calculated using our open-source data. Our open-source dataset addresses the scarcity of large, public, and well-structured datasets from real-world deployed networks, and provides a benchmark for comparing different QoT models in our research community.

# 6

# Conclusion and Prospect

## 6.1 Conclusion

The ever-growing demand to access cloud-based computational services such as machine learning (ML) applications running in data-center infrastructure, continues to drive the demands of transmission bandwidth and capacity on DCI and longhaul backbone networks even further. Hyperscalers need to reduce the cost of optical network design, deployment, and optimization at an accelerated pace to sustain this growth. This necessitates a transition towards fully autonomous optical networks that are predicated on the ability to automate the entire lightpath provisioning and operation life-cycle.

The modeling of physical layer impairment (PLI), or quality of transmission (QoT), is the critical enabler that provides the necessary physical layer awareness for a wide range of automated functions, from intelligent routing and dynamic provisioning to proactive maintenance and efficient capacity maximization. However, the trend of network disaggregation impose additional complexity in achieving accurate QoT model. One example is the complexity of system parameters characterizations such as signal power measurements of the amplifiers, OCMs and transceivers come from different vendors with varying levels of inaccuracies and even for the same model of the same vendor, there exist statistical fluctuations of their performance and parameter measurement accuracies. Large-scale dataset can be used for system calibrations and improving QoT estimation accuracy, but they are usually proprietary to optical hardware vendors and not network operators. In this connection, this thesis focus on addressing the input parameter uncertainty issues existing in commercial disaggregated optical networks for more accurate physical layer impairment (PLI) assessment.

This thesis first investigated the scenario of large-scale long-haul commercial optical network, and proposed effective input parameter refinement methods of signal power and gain profile for enhanced QoT estimation based on the real-time monitoring data in this scenario, which was combined with a data-driven optimization to minimize the location-specific and frequency-specific estimation errors. The proposed method can reduce the standard deviation of the QoT estimation error from 0.6337dB to 0.2377dB. The work demonstrated practical steps to improve QoT estimation accuracy in a long-haul network, reduce network margins and potentially increase network capacities without compromising reliability in realistic current and future mesh networks.

This thesis also investigated the scenario of large-scale metro DCI commercial optical networks, and proposed two applicable methods for improving the accuracy of QoT estimation for large-scale disaggregated DCI networks where transponder effects and ASE noise become dominant impairments rather than fiber nonlinearity in the long-haul. Using a real PM dataset from large-scale DCI networks, we trained ML models to learn the SNR incorporating various input-dependent transponder effects in 8 types of coherent transponders from 4 different vendors. The output of the trained ML model has been demonstrated to be able to estimate SNR for different types of TRX with good accuracy. Then, focusing specifically on the 400Gb/s coherent transponder that has been widely used in the DCI network understudy, we conducted extensive experiments on that type of transponder and obtained an input-dependent OSNR-SNR model by least-squares fitting the measurements. We also conducted extensive experiments to characterize the noise figures of 8 types of EDFAs from 2 different vendors, which is modeled as a function of gain, tilt, input power and frequency. Applying these two experimentally obtained models in the real PM dataset from the DCI networks, we showed that the RMSE of analytical model based SNR estimation from 0.662dB to 0.287dB which validated its effectiveness.

For metro DCI networks, we also addressed the gap between brownfield QoT estimation and greenfield QoT estimation. The absence of signal power monitoring presents a significant challenge for accurate greenfield QoT assessment. To overcome this, we propose a low-complexity polynomial fitting model to characterize the gain profiles of various EDFAs from two different vendors deployed in our DCI networks. The gain model is constructed using fully loaded measurement data within a typical parameter space, encompassing gain and tilt targets as well as input power levels ranging from minimum values to saturation points. Across nine types of EDFAs, the proposed model achieves a RMSE significantly below 0.083 dB, with a minimum RMSE reaching 0.038 dB. We further investigate the model's applicability to EDFA gain spectra subjected to WDM inputs with pre-emphasis. With straightforward refinements to the model's input parameters, an acceptable RMSE below 0.12 dB was achieved for two EDFAs operating in the C and L bands. The efficacy of this pre-emphasis-aware polynomial gain profile model is validated using real network data. This greenfield QoT model enables network operators to avoid costly rework during design, planning, and deployment phases, thereby maximizing network capacity utilization and resource efficiency.

This thesis also introduced an open-source dataset from real optical transport networks, which is available on GitHub[138]. The network segments considered in the dataset consist of 6 OMSs and 25 OCHs. All OCHs have data available for 162 h, with one data point per hour. Furthermore, we have also provided an additional 181 h of data for 6 OCHs after the change of link conditions, i.e., the gains of the EDFAs are the link, compared to the aforementioned time period. With this dataset, over 10,000 GSNRs can be estimated. Our open-source dataset addresses the scarcity of large, public, and well-structured datasets from real-world deployed networks, and provides a benchmark for comparing different QoT models in our research community.

In summary, this thesis investigated improved brownfield and greenfield QoT estimations, which provide accurate and adaptable physical layer models that enable accurate and fast assessment of signal impairments applicable in different network architectures and scenarios. For brownfield networks, our improved QoT estimation enhances the understanding and management of existing infrastructure, allowing for more informed upgrades and optimization. In greenfield scenarios, our novel EDFA gain profile modeling fills the critical gap caused by the lack of real-time signal power monitoring, enabling reliable QoT predictions from limited input data. Together, these advancements facilitate automated network design, planning, and operation by providing reliable models for QoT forecasting and resource allocation. Consequently, network operators can optimize capacity utilization, reduce unnecessary overprovisioning, and avoid costly manual interventions or rework. This leads to more efficient and automated network deployment and maintenance, lowering both capital and operational expenditures while supporting agile and scalable optical network evolution.

## 6.2 Prospects

The optimizations, algorithms, and dataset that have been introduced in this thesis have shown promising results in modeling the physical layer of real network scenarios, either in long-haul or in DCI networks. Nevertheless, there is always room for improvement. The prospects of this thesis are described below.

First, for the large-scale long-haul network, we have shown data-driven optimization of frequency-dependent bias can be an efficient method to minimize the estimation errors. However, more accurate characterization of physical parameters should also be performed by extensive experimental tests. For instance, the frequency-dependent NF profile of EDFAs of different types and vendors should be measured under different gain, tilt and input total power. In addition, OLEs are usually used in long-haul networks to flatten the power profile and counteract the SRS effects after a long-distance transmission. But it is composed of two EDFAs inside, which adds the difficulty of accurate parameter characterization and optimization. On the other hand, the proposed data-driven optimization of OMS-specific bias, though effective in reducing errors, was not related to a specified or deterministic explanation for this location-specific impairment. Some researchers have shown that "filtering penalty" can be the possible cause, which can be induced by a laser frequency shift at the Tx and bandwidth narrowing at the ADD direction of WSS[139]. Further experimental tests and validations should be first conducted to study the possible causes, and devise the corresponding optimization method for real network.

Second, for the metro DCI networks, we have shown that the ANN-based model can learn complicated patterns between the link OSNR and optical front-end penalty for multi-vendor mixed rate TRXs in this scenario. However, the generalization ability of the ANN model has not been extensively and comprehensively investigated. Approaches such as *drop-out*[140] and *deep ensembles*[141] can be applied to obtain a reliable model and prevent it from overfitting.

Third, for the amplifier gain profile modeling, we have proposed vendor-agnostic polynomial model and validated its effectiveness in multi-vendor multi-type EDFAs. However, the discrepancy between different EDFAs of the same type and same vendor have not been studied. The discrepancy may require new data-driven approaches to be modeled and then minimized, as such discrepancy may come from the internal material or manufacturing imperfections that are hard to physically modeled by the network operators.

After our open-source network dataset was publicly available in 2023[117,138], we found that researchers at Bell Labs in France have utilized it to develop their own input refinement techniques aimed at enhancing analytical model-based QoT estimation[142], and also to validate their power-aware coherent transceiver model[143]. It is gratifying to see our open-source dataset making a tangible impact within the research community. We hope this open-source dataset from real optical transport network will continue to serve as a benchmark for researchers focusing on physical layer modeling in optical networks.

# Bibliography


[1] B. Mukherjee, I. Tomkos, M. Tornatore, P. Winzer, and Y. Zhao, *Springer handbook of optical networks*. Springer Nature, 2020.

[2] E. Agrell, M. Karlsson, F. Poletti, S. Namiki, X. Chen, L. A. Rusch, B. Puttnam, P. Bayvel, L. Schmalen, Z. Tao, *et al.*, "Roadmap on optical communications," *Journal of Optics*, vol. 26, no. 9, p. 093001, 2024.

[3] "Optical Transport Network (OTN) Tutorial." https://www.itu.int/ITU-T/studygroups/com15/otn/OTNtutorial.pdf.

[4] A. D'Amico, B. Correia, E. London, E. Virgillito, G. Borraccini, A. Napoli, and V. Curri, "Scalable and disaggregated ggn approximation applied to a c+ l+ s optical network," *Journal of Lightwave Technology*, vol. 40, no. 11, pp. 3499–3511, 2022.

[5] Y. Liu, X. Liu, Y. Zhang, M. Cai, M. Fu, X. Zhong, L. Yi, W. Hu, and Q. Zhuge, "Building a digital twin of an EDFA for optical networks: a gray-box modeling approach," *Journal of Optical Communications and Networking*, vol. 15, no. 11, pp. 830–838, 2023.

[6] C. Cheng, "A global design of an erbium-doped fiber and an erbium-doped fiber amplifier," *Optics & Laser Technology*, vol. 36, no. 8, pp. 607–612, 2004.

[7] C. Giles and E. Desurvire, "Modeling erbium-doped fiber amplifiers," *Journal of Lightwave Technology*, vol. 9, no. 2, pp. 271–283, 1991.

[8] S. Kumar and M. J. Deen, *Fiber optic communications: fundamentals and applications*. John Wiley & Sons, 2014.

[9] N. B. Lê, "Advanced digital optical communications," 2015.

[10] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The gn-model of fiber non-linear propagation and its applications," *Journal of lightwave technology*, vol. 32, no. 4, pp. 694–721, 2013.

[11] T. Menzies, E. Kocagüneli, L. Minku, F. Peters, and B. Turhan, "Chapter 24 - using goals in model-based reasoning," in *Sharing Data and Models in Software Engineering* (T. Menzies, E. Kocagüneli, L. Minku, F. Peters, and B. Turhan, eds.), pp. 321–353, Boston: Morgan Kaufmann, 2015.

[12] 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," *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2616–2629, 2020.

[13] Y. Pointurier, "Design of low-margin optical networks," *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17, 2017.

[14] A. Mahajan, *Machine learning assisted QoT estimation for optical networks optimization*. PhD thesis, Universitat Politècnica de Catalunya, 2021.

[15] C. Xie, L. Wang, L. Dou, M. Xia, S. Chen, H. Zhang, Z. Sun, and J. Cheng, "Open and disaggregated optical transport networks for data center interconnects [Invited]," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C12–C22, 2020.

[16] H. Buglia, *Modelling and Optimisation of Ultra-wideband Optical Fibre Communication Systems*. PhD thesis, UCL (University College London), 2025.

[17] "Architecture of the optical transport network." https://handle.itu.int/11.1002/1000/15804.

[18] R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D84–D99, 2018.

[19] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*, pp. 195–211, Springer, 2000.

[20] A. Saleh, R. Jopson, J. Evankow, and J. Aspell, "Modeling of gain in erbium-doped fiber amplifiers," *IEEE Photonics Technology Letters*, vol. 2, no. 10, pp. 714–717, 1990.

[21] C. Giles and D. Di Gioanni, "Spectral dependence of gain and noise in erbium-doped fiber amplifiers," *IEEE Photonics technology letters*, vol. 2, no. 11, pp. 797–800, 1990.

[22] Y. Sun, J. Zyskind, and A. Srivastava, "Average inversion level, modeling, and physics of erbium-doped fiber amplifiers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 3, no. 4, pp. 991–1007, 1997.

[23] P. M. Becker, A. A. Olsson, and J. R. Simpson, *Erbium-doped fiber amplifiers: fundamentals and technology*. Elsevier, 1999.

[24] Y. Li, M. Zhang, and M. Tang, "A Data-effective Black-box EDFA Gain Model with Singular Value Decomposition," in *2022 Asia Communications and Photonics Conference (ACP)*, pp. 985–987, 2022.

[25] S. Bottacchi, *Noise and signal interference in optical fiber transmission systems: an optimum design approach*. John Wiley & Sons, 2008.

[26] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, no. 7, pp. 1074–1080, 2002.

[27] S. Cruzes, "Overview of quality of transmission estimation in optical networks," *Authorea Preprints*, 2024.

[28] D. Semrau, R. I. Killey, and P. Bayvel, "The gaussian noise model in the presence of inter-channel stimulated raman scattering," *Journal of Lightwave Technology*, vol. 36, no. 14, pp. 3046–3055, 2018.

[29] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *Journal of Lightwave technology*, vol. 30, no. 10, pp. 1524–1539, 2012.

[30] A. Bononi, P. Serena, N. Rossi, E. Grellier, and F. Vacondio, "Modeling nonlinearity in coherent transmissions with dominant intrachannel-four-wave-mixing," *Optics Express*, vol. 20, no. 7, pp. 7777–7791, 2012.

[31] R. Ayassi, *Artificial intelligence techniques to improve quality of transmission estimation in optical networks.* PhD thesis, Institut Polytechnique de Paris, 2023.

[32] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Express*, vol. 22, pp. 16335–16362, Jun 2014.

[33] D. Semrau, R. I. Killey, and P. Bayvel, "A Closed-Form Approximation of the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering," *Journal of Lightwave Technology*, vol. 37, no. 9, pp. 1924–1936, 2019.

[34] D. Semrau, E. Sillekens, R. I. Killey, and P. Bayvel, "A modulation format correction formula for the gaussian noise model in the presence of inter-channel stimulated raman scattering," *Journal of Lightwave Technology*, vol. 37, no. 19, pp. 5122–5131, 2019.

[35] Q. Fan, J. Lu, G. Zhou, D. Zeng, C. Guo, L. Lu, J. Li, C. Xie, C. Lu, F. N. Khan, and A. P. T. Lau, "Experimental comparisons between machine learning and analytical models for qot estimations in wdm systems," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2020.

[36] J. Lu, G. Zhou, Q. Fan, D. Zeng, C. Guo, L. Lu, J. Li, C. Xie, C. Lu, F. N. Khan, and A. P. T. Lau, "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems [Invited]," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B35–B44, 2021.

[37] T. Mano, A. D'Amico, E. Virgillito, G. Borraccini, Y.-K. Huang, K. Anazawa, H. Nishizawa, T. Wang, K. Asahi, and V. Curri, "Modeling transceiver BER-OSNR characteristic for QOT estimation in short-reach systems," in *2023 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, IEEE, 2023.

[38] T. Mano, Y.-K. Huang, G. Borraccini, E. Ip, A. D'Amico, Z. Wang, H. Nishizawa, G. Zussman, T. Chen, T. Wang, K. Asahi, D. Kilper, V. Curri, and K. Takasugi, "Modeling the Input Power Dependency of Transceiver BER-OSNR for QoT Estimation," in *Optical Fiber Communication Conference (OFC) 2024*, p. M1H.4, Optica Publishing Group, 2024.

[39] X. Ye, *Applications of Artificial Intelligence to Control and Analyze the Performance of Fiber-Optic Transmission Systems*. PhD thesis, Institut Polytechnique de Paris, 2023.

[40] 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)*, pp. 1–3, 2022.

[41] N. Morette, H. Hafermann, Y. Frignac, and Y. Pointurier, "Machine learning enhancement of a digital twin for wavelength division multiplexing network performance prediction leveraging quality of transmission parameter refinement," *Journal of Optical Communications and Networking*, vol. 15, no. 6, pp. 333–343, 2023.

[42] M. Björklund, "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)." RFC 6020, Oct. 2010.

[43] R. Enns, M. Björklund, A. Bierman, and J. Schönwälder, "Network Configuration Protocol (NETCONF)." RFC 6241, June 2011.

[44] O. Karandin, A. Ferrari, F. Musumeci, Y. Pointurier, and M. Tornatore, "Low-margin optical-network design with multiple physical-layer parameter uncertainties," in *European Conference and Exhibition on Optical Communication*, pp. Mo3B–2, Optica Publishing Group, 2022.

[45] Y. He, Z. Zhai, L. Wang, Y. Yan, L. Dou, C. Xie, C. Lu, and A. P. T. Lau, "Improved QoT Estimations through Refined Signal Power Measurements in a Disaggregated and Partially-loaded Live Production Network," in *Optical Fiber Communication Conference (OFC) 2023*, p. Tu2F.5, Optica Publishing Group, 2023.

[46] Y. He, Z. Zhai, L. Dou, L. Wang, Y. Yan, C. Xie, C. Lu, and A. P. T. Lau, "Improved QoT estimations through refined signal power measurements and data-driven parameter optimizations in a disaggregated and partially loaded live production network," *Journal of Optical Communications and Networking*, vol. 15, no. 9, pp. 638–648, 2023.

[47] J.-C. Antona, A. Carbó-Meseguer, and V. Letellier, "Evolution of high capacity submarine open cables," in *2019 Asia Communications and Photonics Conference (ACP)*, pp. 1–3, 2019.

[48] 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)*, pp. 1–3, 2021.

[49] G. Borraccini, V. Gatto, A. D'Amico, S. Straullu, F. Aquilino, A. Nespola, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Gain Profile Characterization and Modeling for Dual-Stage EDFA Abstraction and Control," *IEEE Photonics Technology Letters*, vol. 36, no. 2, pp. 107–110, 2024.

[50] Z. Wang, D. C. Kilper, and T. Chen, "Open EDFA gain spectrum dataset and its applications in data-driven EDFA gain modeling," *Journal of Optical Communications and Networking*, vol. 15, no. 9, pp. 588–599, 2023.

[51] A. Purkayastha, J. T. De Araujo, A. C. Meseguer, and J.-C. Antona, "EDFA Noise Figure Analysis in Non-Ideal Operating Conditions," in *2023 Asia Communications and Photonics Conference/2023 International Photonics and Optoelectronics Meetings (ACP/POEM)*, pp. 1–5, 2023.

[52] G. Borraccini, A. D'Amico, Y.-K. Huang, V. Curri, and T. Wang, "Characterization and Modeling of the Noise Figure Ripple in a Dual-Stage EDFA," in *2024 IEEE Photonics Conference (IPC)*, pp. 1–2, 2024.

[53] A. Castronovo, *EDFA noise figure and WSS DGD modeling.* PhD thesis, Politecnico di Torino, 2024.

[54] E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," *Journal of Optical Communications and Networking*, vol. 10, no. 2, pp. A298–A306, 2018.

[55] A. Ferrari, K. Balasubramanian, M. Filer, Y. Yin, E. L. Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "Assessment on the in-field lightpath QoT computation including connector loss uncertainties," *J. Opt. Commun. Netw.*, vol. 13, pp. A156–A164, Feb 2021.

[56] V. Curri, "GNPy model of the physical layer for open and disaggregated optical networking [Invited]," *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C92–C104, 2022.

[57] 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)*, pp. 1–3, 2021.

[58] J. Zhou, J. Lu, and C. Yu, "Improving the accuracy of QoT estimation with insertion loss distribution evaluation for C+L band transmission systems," *Journal of Optical Communications and Networking*, vol. 16, no. 1, pp. 12–20, 2024.

[59] J. Pesic, M. Lonardi, N. Rossi, T. Zami, E. Seve, and Y. Pointurier, "How Uncertainty on the Fiber Span Lengths Influences QoT Estimation using Machine Learning in WDM Networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2020.

[60] R. T. Jones, K. R. H. Bottrill, N. Taengnoi, P. Petropoulos, and M. P. Yankov, "Spectral power profile optimization of a field-deployed wavelength-division multiplexing network enabled by remote EDFA modeling," *Journal of Optical Communications and Networking*, vol. 15, no. 8, pp. C192–C202, 2023.

[61] J. Muller, S. K. Patri, T. Fehenberger, H. Griesser, J.-P. Elbers, and C. Mas-Machuca, "QoT estimation using EGN-assisted machine learning for multi-period network planning," *Journal of Optical Communications and Networking*, vol. 14, no. 12, pp. 1010–1019, 2022.

[62] N. A. Kaliteevskiy, V. Ivanov, P. Sterlingov, J. Downie, S. Makovejs, and J. Hurley, "Transponder implementation penalty-accounted gaussian-noise-based performance modeling of fiber-optic transmission systems," *Journal of Lightwave Technology*, vol. 38, no. 8, pp. 2253–2261, 2020.

[63] A. W. MacKay and D. W. Boertjes, "Field Learnings of Deploying Model Assisted Network Feedback Systems," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 01–03, 2022.

[64] C. Natalino, B. Shariati, P. Safari, J. K. Fischer, and P. Monti, "Analysis and mitigation of unwanted biases in ml-based qot classification tasks," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2024.

[65] H. Fawaz, F. Arpanaei, D. Andreoletti, I. Sbeity, J. A. Hernández, D. Larrabeiti, and O. Ayoub, "Reducing complexity and enhancing predictive power of ml-based lightpath qot estimation via shap-assisted feature selection," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–6, IEEE, 2024.

[66] S. Kozdrowski, P. Paziewski, P. Cichosz, and S. Sujecki, "A comprehensive study of machine learning application to transmission quality assessment in optical networks," *Applied Sciences*, vol. 13, no. 8, 2023.

[67] D. Azzimonti, C. Rottondi, and M. Tornatore, "Reducing probes for quality of transmission estimation in optical networks with active learning," *Journal of Optical Communications and Networking*, vol. 12, no. 1, pp. A38–A48, 2019.

[68] Z. Li, Z. Gu, J. Zhang, Y. Zhou, and Y. Ji, "Predictive uncertainty aware active learning for regression-based qot estimation in optical networks," in *2021 Asia Communications and Photonics Conference (ACP)*, pp. 1–3, 2021.

[69] O. Ayoub, S. Troia, D. Andreoletti, A. Bianco, M. Tornatore, S. Giordano, and C. Rottondi, "Towards explainable artificial intelligence in optical networks: the use case of lightpath qot estimation," *Journal of Optical Communications and Networking*, vol. 15, no. 1, pp. A26–A38, 2022.

[70] Z. Gao, S. Yan, J. Zhang, M. Mascarenhas, R. Nejabati, Y. Ji, and D. Simeonidou, "Ann-based multi-channel qot-prediction over a 563.4-km field-trial testbed," *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2646–2655, 2020.

[71] J. Müller, T. Fehenberger, S. K. Patri, K. Kaeval, H. Griesser, M. Tikas, and J.-P. Elbers, "Estimating Quality of Transmission in a Live Production Network using Machine Learning," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2021.

[72] F. Usmani, I. Khan, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Convolutional neural network for quality of transmission prediction of unestablished lightpaths," *Microwave and Optical Technology Letters*, vol. 63, no. 10, pp. 2461–2469, 2021.

[73] C.-Y. Liu, X. Chen, R. Proietti, and S. B. Yoo, "Performance studies of evolutionary transfer learning for end-to-end qot estimation in multi-domain optical networks," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B1–B11, 2021.

[74] Q. Wang, Z. Cai, A. P. T. Lau, Y. Li, and F. Nadeem Khan, "Invariant convolutional neural network for robust and generalizable qot estimation in fiber-optic networks," *Journal of Optical Communications and Networking*, vol. 15, no. 7, pp. 431–441, 2023.

[75] H. Gao, X. Chen, C. Lu, and Z. Li, "On the generalization of cognitive optical networking applications using composable machine learning," *Journal of Optical Communications and Networking*, vol. 16, no. 6, pp. 631–643, 2024.

[76] D. Chadha, *Optical WDM Networks: From Static to Elastic Networks.* John Wiley & Sons, 2019.

[77] C. Manso, R. Vilalta, R. Muñoz, N. Yoshikane, R. Casellas, R. Martínez, C. Wang, F. Balasis, T. Tsuritani, and I. Morita, "Scalability analysis of machine learning qot estimators for a cloud-native sdn controller¡? tex\break?¿ on a wdm over sdm network," *Journal of Optical Communications and Networking*, vol. 14, no. 4, pp. 257–266, 2022.

[78] A. A. Díaz-Montiel, J. Yu, W. Mo, Y. Li, D. C. Kilper, and M. Ruffini, "Performance analysis of qot estimator in sdn-controlled roadm networks," in *2018 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 142–147, 2018.

[79] N. Sambo, K. Christodoulopoulos, N. Argyris, P. Giardina, C. Delezoide, A. Sgambelluri, A. Kretsis, G. Kanakis, F. Fresi, G. Bernini, *et al.*, "Experimental demonstration of fully disaggregated white box including different types of transponders and monitors, controlled by netconf and yang," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, IEEE, 2018.

[80] Q. Wang, Z. Cai, and F. N. Khan, "Lifelong QoT prediction: an adaptation to real-world optical networks," *Journal of Optical Communications and Networking*, vol. 16, no. 11, pp. 1159–1169, 2024.

[81] T. Jimenez, J. C. Aguado, I. de Miguel, R. J. Durán, M. Angelou, N. Merayo, P. Fernandez, R. M. Lorenzo, I. Tomkos, and E. J. Abril, "A cognitive quality of transmission estimator for core optical networks," *Journal of Lightwave Technology*, vol. 31, no. 6, pp. 942–951, 2013.

[82] T. Panayiotou, G. Savva, I. Tomkos, and G. Ellinas, "Decentralizing machine-learning-based qot estimation for sliceable optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 7, pp. 146–162, 2020.

[83] A. D'Amico, S. Straullu, G. Borraccini, E. London, S. Bottacchi, S. Piciaccia, A. Tanzi, A. Nespola, G. Galimberti, S. Swail, and V. Curri, "Enhancing Lightpath QoT Computation With Machine Learning in Partially Disaggregated Optical Networks," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 564–574, 2021.

[84] H. Chouman, P. Djukic, C. Tremblay, and C. Desrosiers, "Forecasting Lightpath QoT with Deep Neural Networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2021.

[85] R. Ayassi, A. Triki, N. Crespi, R. Minerva, and M. Laye, "Survey on the Use of Machine Learning for Quality of Transmission Estimation in Optical Transport Networks," *Journal of Lightwave Technology*, vol. 40, no. 17, pp. 5803–5815, 2022.

[86] E. Le Rouzic, O. Renais, J. Meuric, T. Marcot, C. Betoule, G. Thouenon, A. Triki, M. Laye, N. Pelloquin, Y. Lagadec, E. Delfour, M. Ermel, J. Dost, and S. Turk, "Operator view on optical transport network automation in a multi-vendor context [Invited]," *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C11–C22, 2022.

[87] C. Delezoide, K. Christodoulopoulos, A. Kretsis, N. Argyris, G. Kanakis, A. Sgambelluri, N. Sambo, P. Giardina, G. Bernini, D. Roccato, A. Percelsi, R. Morro, H. Avramopoulos, E. Varvarigos, P. Castoldi, P. Layec, and S. Bigo, "Marginless Operation of Optical Networks," *Journal of Lightwave Technology*, vol. 37, no. 7, pp. 1698–1705, 2019.

[88] M. Bouda, G. Krishna, J. Krystofik, S. Oda, and P. Palacharla, "Towards vendor-agnostic real-time optical network design with extended Kalman state estimation and recurrent neural network machine learning [Invited]," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B21–B34, 2021.

[89] M. Filer, M. Cantono, A. Ferrari, G. Grammel, G. Galimberti, and V. Curri, "Multi-Vendor Experimental Validation of an Open Source QoT Estimator for Optical Networks," *Journal of Lightwave Technology*, vol. 36, no. 15, pp. 3073–3082, 2018.

[90] A. D'Amico, E. London, B. Le Guyader, F. Frank, E. Le Rouzic, E. Pincemin, N. Brochier, and V. Curri, "Experimental validation of GNPy in a multi-vendor flex-grid flex-rate WDM optical transport scenario," *Journal of Optical Communications and Networking*, vol. 14, no. 3, pp. 79–88, 2022.

[91] https://github.com/Telecominfraproject/oopt-gnpy/releases/tag/v2.7.

[92] J. Pesic, "Missing pieces currently preventing effective application of machine learning to qot estimation in the field," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2021.

[93] Y. Pointurier, "Design of low-margin optical networks," in *2016 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2016.

[94] P. Ramantanis, C. Delezoide, P. Layec, and S. Bigo, "Revisiting the calculation of performance margins in monitoring-enabled optical networks," *Journal of Optical Communications and Networking*, vol. 11, no. 10, pp. C67–C75, 2019.

[95] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrat, G. Grammel, G. Galimberti, and V. Curri, "GNPy: an open source application for physical layer aware open optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C31–C40, 2020.

[96] K. Kaeval, T. Fehenberger, J. Zou, S. L. Jansen, K. Grobe, H. Griesser, J.-P. Elbers, M. Tikas, and G. Jervan, "QoT assessment of the optical spectrum as a service in disaggregated network scenarios," *Journal of Optical Communications and Networking*, vol. 13, no. 10, pp. E1–E12, 2021.

[97] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018. Available at http://github.com/google/jax, version 0.3.13.

[98] "Terms and definitions for optical transport networks (OTN)." https://handle.itu.int/11.1002/1000/7337.

[99] I. Roberts, J. M. Kahn, J. Harley, and D. W. Boertjes, "Channel Power Optimization of WDM Systems Following Gaussian Noise Nonlinearity Model in Presence of Stimulated Raman Scattering," *Journal of Lightwave Technology*, vol. 35, no. 23, pp. 5237–5249, 2017.

[100] N. Newbury, "Pump-wavelength dependence of Raman gain in single-mode optical fibers," *Journal of Lightwave Technology*, vol. 21, no. 12, pp. 3364–3373, 2003.

[101] J. Cheng, L. Dou, C. Xie, *et al.*, "Configuration of Automatic Protection Switches in Data Center Interconnect Systems," in *2020 Opto-Electronics and Communications Conference (OECC)*, pp. 1–3, IEEE, 2020.

[102] I. Khan, M. Bilal, M. Umar Masood, A. D'Amico, and V. Curri, "Lightpath QoT computation in optical networks assisted by transfer learning," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B72–B82, 2021.

[103] J. Yu, Y.-K. Huang, S. Zhang, E. Ip, D. C. Kilper, T. J. Xia, and G. A. Wellbrock, "Neural-Network-Based G-OSNR Estimation of Probabilistic-Shaped 144QAM Channels in DWDM Metro Network Field Trial," in *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, pp. 1–3, IEEE, 2019.

[104] R. Morro, F. Lucrezia, P. Gomes, R. Casellas, A. Giorgetti, L. Velasco, E. Riccardi, A. C. Piat, A. Percelsi, J. Pedro, *et al.*, "Automated end to end carrier ethernet provisioning over a disaggregated WDM metro network with a hierarchical SDN control and monitoring platform," in *2018 European Conference on Optical Communication (ECOC)*, pp. 1–3, IEEE, 2018.

[105] M. Newland, R. Schmogrow, M. Cantono, V. Vusirikala, and T. Hofmeister, “Open optical communication systems at a hyperscale operator,” *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C50–C57, 2020.

[106] L. Galdino, D. Semrau, M. Ionescu, A. Edwards, W. Pelouch, S. Desbruslais, J. James, E. Sillekens, D. Lavery, S. Barnes, *et al.*, “Study on the impact of nonlinearity and noise on the performance of high-capacity broadband hybrid Raman-EDFA amplified system,” *Journal of Lightwave Technology*, vol. 37, no. 21, pp. 5507–5515, 2019.

[107] Y. He, K. Chandramouli, Z. Zhai, S. Chen, L. Dou, C. Xie, C. Lu, and A. P. Tao Lau, “QoT Estimation for Large-scale Mixed-rate Disaggregated Metro DCI Networks by Artificial Neural Networks,” in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2024.

[108] Y. He, Z. Zhai, S. Chen, H. Zhang, C. Xu, L. Dou, Y. Su, C. Lu, and A. P. T. Lau, “Transceiver Penalty and Amplifier Noise Figure Characterization for Accurate QoT Estimation in Hyperscale Disaggregated DCI Networks,” in *Optical Fiber Communication Conference (OFC) 2025*, p. M1J.6, Optica Publishing Group, 2025.

[109] Z. Zhai, L. Dou, S. Chen, H. Zhang, Y. He, A. P. T. Lau, and Y. Su, “Improving QoT Estimation Accuracy in Production Networks: A Data-Driven Approach to Address OLS Imperfections,” in *Optical Fiber Communication Conference*, pp. Tu3I–4, Optica Publishing Group, 2025.

[110] G. Borraccini, A. D’Amico, Y.-K. Huang, V. Curri, and T. Wang, “Characterization and Modeling of the Noise Figure Ripple in a Dual-Stage EDFA,” in *2024 IEEE Photonics Conference (IPC)*, pp. 1–2, 2024.

[111] G. Borraccini, A. D’Amico, S. Straullu, A. Nespola, S. Piciaccia, A. Tanzi, G. Galimberti, S. Bottacchi, S. Swail, and V. Curri, “Cognitive and autonomous QoT-driven optical line controller,” *Journal of Optical Communications and Networking*, vol. 13, no. 10, pp. E23–E31, 2021.

[112] G. Borraccini, S. Straullu, A. D’Amico, F. Aquilino, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, “Local and global optimization methods for optical line control based on quality of transmission,” *Journal of Optical Communications and Networking*, vol. 16, no. 5, pp. B60–B70, 2024.

[113] J. K. Perin, J. M. Kahn, J. D. Downie, J. Hurley, and K. Bennett, “Importance of Amplifier Physics in Maximizing the Capacity of Submarine Links,” *Journal of Lightwave Technology*, vol. 37, no. 9, pp. 2076–2085, 2019.

[114] R. D. Muro, “The $Er^{3+}$-fiber gain coefficient derived from a dynamic gain tilt technique,” *Journal of lightwave technology*, vol. 18, no. 3, p. 343, 2000.

[115] M. Newland, R. Schmogrow, M. Cantono, V. Vusirikala, and T. Hofmeister, “Open optical communication systems at a hyperscale operator [Invited],” *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C50–C57, 2020.

[116] E. Le Rouzic, O. Renais, J. Meuric, T. Marcot, C. Betoule, G. Thouenon, A. Triki, M. Laye, N. Pelloquin, Y. Lagadec, E. Delfour, M. Ermel, J. Dost, and S. Turk, “Operator view on optical transport network automation in a multi-vendor context [Invited],” *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C11–C22, 2022.

[117] Z. Zhai, L. Dou, Y. He, A. P. T. Lau, and C. Xie, “Open-source data for QoT estimation in optical networks from Alibaba,” *Journal of Optical Communications and Networking*, vol. 16, no. 1, pp. 1–3, 2024.

[118] J. Yu, S. Zhu, C. L. Gutterman, G. Zussman, and D. C. Kilper, “Machine-learning-based EDFA gain estimation [Invited],” *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B83–B91, 2021.

[119] 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)*, pp. 1–3, 2018.

[120] 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)*, pp. 1–4, 2020.

[121] Y. You, Z. Jiang, and C. Janz, “Machine Learning-Based EDFA Gain Model,” in *2018 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2018.

[122] A. D’Amico, *Analytical models and artificial intelligence for open and partially disaggregated optical networks*. PhD thesis, Politecnico di Torino, 2023.

[123] R. Di Muro, N. Jolley, S. Wilson, and J. Mun, “Experimental correlation of the extended-band EDFA with an optimized numerical model using a dynamic gain tilt technique,” *IEEE Photonics Technology Letters*, vol. 12, no. 9, pp. 1159–1161, 2000.

[124] F. C. N. O. Lima, A. A. B. Silva, L. V. W. J. Silva, T. C. C. Silva, E. A. Barboza, C. J. A. Bastos-Filho, L. D. Coelho, and J. F. Martins-Filho, “A Comparison of DGT and Power Mask for EDFA Modeling in GNPy,” in *2023 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, pp. 79–81, 2023.

[125] Y. Lan, L. Wang, and Z. Jiang, “EDFA gain profile prediction under off-target channel add/drop,” in *2025 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2025.

[126] C. De Boor, “A practical guide to splines,” *Springer-Verlag google schola*, vol. 2, pp. 4135–4195, 1978.

[127] V. M. C. Mathias, A. Carbo Meseguer, J.-C. Antona, A. Bononi, S. Grubb, O. Courtois, and V. Letellier, “Extension of the Measurement-Based Gain Model for non-Flat WDM Inputs and various pump currents,” in *2021 European Conference on Optical Communication (ECOC)*, pp. 1–4, 2021.

[128] K. Ishii, J. Kurumida, and S. Namiki, "Experimental Investigation of Gain Offset Behavior of Feedforward-Controlled WDM AGC EDFA Under Various Dynamic Wavelength Allocations," *IEEE Photonics Journal*, vol. 8, no. 1, pp. 1–13, 2016.

[129] C. Harvey, M. S. Faruk, and S. J. Savory, "Data-Driven Erbium-Doped Fiber Amplifier Gain Modeling Using Gaussian Process Regression," *IEEE Photonics Technology Letters*, vol. 36, no. 18, pp. 1097–1100, 2024.

[130] J. F. E. and, "On the Runge Example," *The American Mathematical Monthly*, vol. 94, no. 4, pp. 329–341, 1987.

[131] Z. Wang, Y.-K. Huang, S. Han, D. Kilper, and T. Chen, "Multi-span optical power spectrum prediction using cascaded learning with one-shot end-to-end measurement," *Journal of Optical Communications and Networking*, vol. 17, no. 1, pp. A23–A33, 2025.

[132] Y. Song, M. Zhang, Y. Zhang, Y. Shi, S. Shen, X. Tang, S. Huang, and D. Wang, "Lifecycle Management of Optical Networks With Dynamic-Updating Digital Twin: A Hybrid Data-Driven and Physics-Informed Approach," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 5, pp. 1767–1779, 2025.

[133] C. Sun, X. Yang, G. Charlet, P. A. Stavrou, and Y. Pointurier, "Digital twin-enabled multi-step strategies for autonomous power equalization in optical networks," *Journal of Optical Communications and Networking*, vol. 17, no. 7, pp. C41–C50, 2025.

[134] A. Srivastava, J. Zyskind, J. Sulhoff, J. Evankow, and M. Mills, "Room temperature spectral hole-burning in erbium-doped fiber amplifiers," in *Optical Fiber Communications, OFC.*, pp. 33–34, 1996.

[135] J. Sulhoff, A. Srivastava, C. Wolf, Y. Sun, and J. Zyskind, "Spectral-hole burning in erbium-doped silica and fluoride fibers," *IEEE Photonics Technology Letters*, vol. 9, no. 12, pp. 1578–1579, 1997.

[136] F. Flood, "Gain saturation behavior in L-band EDFAs," *IEEE Photonics Technology Letters*, vol. 12, no. 9, pp. 1156–1158, 2000.

[137] F. Flood, "Impact of pump and signal wavelength on inhomogeneous characteristics of L-band EDFAs," in *Optical Fiber Communication Conference. Technical Digest Postconference Edition. Trends in Optics and Photonics Vol.37 (IEEE Cat. No. 00CH37079)*, vol. 2, pp. 117–119 vol.2, 2000.

[138] "Alibaba-cloud transport system dataset," 2023. Available at https://github.com/alibaba/alibaba-cloud-transport-system.

[139] R. Ayassi, X. Yang, C. Sun, Y. Pointurier, and G. Charlet, "Filtering impairments-aware digital twin for snr prediction over network life-cycle," in *2025 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2025.

[140] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[141] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” *Advances in neural information processing systems*, vol. 30, 2017.

[142] A. Purkayastha, C. Delezoide, M. Lourdiane, C. Ware, and P. Layec, “Accuracy enhancement of an optical network digital twin based on open-source field data,” in *ECOC 2024; 50th European Conference on Optical Communication*, pp. 1033–1036, 2024.

[143] A. Purkayastha, C. Delezoide, V. Bajaj, F. Boitier, M. Lourdiane, C. Ware, and P. Layec, “Power-aware digital twin of coherent optical receiver,” in *2025 30th OptoElectronics and Communications Conference (OECC) and 2025 International Conference on Photonics in Switching and Computing (PSC)*, pp. 1–4, 2025.