Summary

This doctoral dissertation presents a comprehensive investigation into modeling techniques for two key components of optical communication systems, Indium Phosphide (InP) Mach–Zehnder Modulators (MZMs) and Erbium-Doped Fiber Amplifiers (EDFAs), within the emerging context of open, disaggregated, and software-defined optical networks. The overarching goal is to contribute to the development of accurate, scalable, and interpretable models capable of supporting real-time network emulation and optimization through the integration into digital twin (DT) platforms. The research is motivated by the increasing complexity and flexibility of optical networks, particularly as the industry moves toward open architectures and multi-vendor interoperability. In such scenarios, the traditional reliance on closed-form, vendor-specific models becomes inadequate. To address this, the thesis proposes a hybrid modeling paradigm that blends analytical physical models with machine learning (ML) approaches, offering a balance between physical interpretability and data-driven adaptability. This dual perspective enables not only accurate component-level modeling but also paves the way for effective integration into software-defined infrastructures where real-time performance and generalization capabilities are crucial.

The first part of the work focuses on the modeling and experimental validation of wideband InP dual-polarization IQ Mach–Zehnder Modulators, which are central devices in modern high-speed optical transceivers. Starting from a solid analytical foundation derived from the author's previous work, the model incorporates a realistic description of electro-optic transfer functions of the component. A full time-domain simulation framework is developed to replicate the modulator behavior under complex modulation formats and symbol rates. Experimental results on a commercial device, obtained through back-to-back and single-span transmission measurements, validate the model's predictions in terms of signal integrity metrics such as EVM and BER. The study further demonstrates how the integration of semiconductor optical amplifiers (SOAs) at the modulator terminals significantly enhances transmission performance in Data Center Interconnect (DCI) applications, eliminating the need for external booster EDFAs and enabling up to 1.2 Tbit/s throughput over metro-scale distances. Beyond physical modeling, the work explores advanced bias point control strategies for nested modulator architectures,

which are traditionally susceptible to drift and misalignment. A digital model of the modulator is constructed using deep learning techniques trained on experimental power measurements. This model enables automatic optimization of the modulator's operating point via a structure-agnostic algorithm based on bisection search and machine learning predictions. The approach is validated across multiple wavelengths and polarizations, achieving closed-loop control and significantly improving signal quality without requiring manual calibration.

The second part of the thesis turns attention to the modeling of EDFAs in real-world optical line systems, where signal degradation is dominated by ASE noise and amplifier response varies significantly under different spectral loads. In contrast to conventional white-box analytical models, which often lack flexibility in partially loaded systems, a novel grey-box modeling framework is introduced. This method enhances physical models with deep neural networks to independently correct signal and noise predictions based on telemetry and measured spectra. The grey-box model proves capable of accurate OSNR estimation even under sparse spectral loading conditions, maintaining mean absolute errors below 0.2 dB across hundreds of test cases. To reduce the dependency on large datasets, which is a key limitation of many ML approaches, the thesis incorporates Few-Shot Learning (FSL) techniques into the grey-box framework. After training a single reference EDFA model with 500 spectral configurations, the method successfully transfers knowledge to other amplifiers using only 4–8 new measurements per device. This drastically reduces measurement effort and proves the feasibility of scalable, data-efficient model deployment in partially disaggregated networks.

In parallel, a lightweight polynomial-based model is developed for noise figure estimation. This model uses polynomial coefficients to approximate the EDFA noise behavior as a function of gain, tilt, input power, and channel frequency. Despite its simplicity, it achieves prediction accuracy comparable to state-of-the-art ML models while requiring orders of magnitude fewer training samples and negligible computational overhead. The analysis further demonstrates that incorporating measurement data from multiple devices significantly improves generalization, while also revealing trends in prediction error linked to edge-channel effects and amplifier gain regimes. on constrained network controllers or embedded systems. The model has also been successfully integrated into a broader machine-learning-based framework for EDFA output spectrum prediction, where it replaces traditional heuristic approaches for noise figure estimation. This integration confirmed the model's practicality, improving prediction accuracy under dynamic operating conditions and demonstrating its potential for scalable deployment within full digital twin architectures.

Collectively, the research contributions in this dissertation establish a robust foundation for digital twin–enabled modeling of photonic components in open optical networks. By unifying analytical models with data-driven refinements and demonstrating their applicability across both experimental and simulated domains,

the work provides essential tools for the realization of intelligent, self-optimizing optical infrastructures. The methodologies developed herein are directly applicable to modern disaggregated environments and offer a pathway toward full life-cycle automation, including provisioning, monitoring, and dynamic optimization of optical transport networks.