POLITECNICO DI TORINO

Master's Degree in Computer and Communication Engineering

Master's Degree Thesis

Performance Analysis Of SWDM Systems For Datacenter Interconnects

Supervisors

Prof. Andrea CARENA

Dr. Ann Margareth ROSA BRUSIN

Candidate

Ali MAKKI
S306604

2023/2024

Summary

This thesis focuses on the statistical analysis and prediction of fiber performance in data center environments, specifically for short links. The study analyzes the relationships between various bandwidth metrics, such as the -3dB, -5dB, -10dB bandwidths, and the equivalent bandwidth, in predicting the maximum achievable fiber length (Lmax) and Bit Error Rate (BER). The dataset used for the analysis comprised 8 lasers, 4 wavelengths, and 3766 OM4 fibers, with measurements taken across multiple short distances commonly found in data centers.

Initially, a detailed correlation analysis was conducted to explore the relationships between these bandwidth metrics, Lmax, and BER. The equivalent bandwidth emerged as the most strongly correlated parameter with Lmax, particularly at shorter distances, such as 30m and 50m, where correlations reached as high as 96%. Additionally, the analysis was extended across a frequency range of 50 GHz to 300 GHz to better understand how frequency impacts this correlation. Despite these efforts, the correlation alone was insufficient to provide highly accurate predictions of Lmax.

To address this challenge, a Machine Learning (ML) model was developed to predict Lmax using the four key bandwidth parameters. The model successfully captured the complex relationships that could not be fully explained through statistical correlation alone. With an accuracy of 99.98%, the ML model significantly outperformed traditional predictive methods based solely on correlation analysis.

This work highlights the efficiency and practicality of using machine learning techniques in data center environments, where accurate and rapid predictions are essential for network performance evaluation and planning. The model developed in this thesis offers a highly effective alternative to time-consuming data extraction processes, enabling near-instant predictions of Lmax with high accuracy, thereby optimizing operational efficiency in data centers.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisors, Professor Andrea Carena and Professor Rosa Brusin Ann Margareth, for their expert guidance and unwavering support throughout the course of this research. Their insights and expertise were invaluable in shaping the direction and execution of my thesis on "Data Center Short Links."

I am also immensely thankful to the faculty and staff at the Politecnico di Torino. Their continuous encouragement and the intellectually stimulating environment they provided were fundamental to my academic growth.

Completing this thesis has been a challenging yet rewarding experience, and it would not have been possible without the support I received from my mentors and the broader university community. I am grateful to have had the opportunity to learn and conduct my research at such a prestigious institution.

Grazie

Table of Contents

List of Tables

List of Figures

Acronyms

ADC	Analog-to-Digital Converter
BER	Bit Error Rate
BERT	Bit Error Rate Test
BS	Simulation Bandwidth
BtB	Back-to-Back
DAC	Digital-to-Analog Converter
dB	Decibel
DFE	Decision Feedback Equalizer
DWDM	Dense Wavelength Division Multiplexing
FEC	Forward Error Correction
FFE	Feed Forward Equalizer
Gbps	Gigabits per Second
IoT	Internet of Things
IP	Internet Protocol
L_{max}	Maximum Achievable Link Length
LMS	Least-Mean-Square
MMF	Multimode Fiber
MLSE	Maximum-Likelihood Sequence Estimation
OMA	Optical Modulation Amplitude
OM4	Optical Multimode 4

PIN	Positive-Intrinsic-Negative
RIN	Relative Intensity Noise
RX	Receiver
SDN	Software-Defined Networking
SMF	Single-Mode Fiber
SNR	Signal-to-Noise Ratio
SWDM	Short Wavelength Division Multiplexing
TDE	Time Domain Equalization
TF	Transfer Function
TIR	Target Impulse Response
TIA	Transimpedance Amplifier
TX	Transmitter
VCSEL	Vertical-Cavity Surface-Emitting Laser
VOA	Variable Optical Attenuator
WDM	Wavelength Division Multiplexing

Chapter 1

Introduction

1.1 Introduction

Today's optical networks form the critical infrastructure for internet communications. The advent of fiber optics has revolutionized these networks by facilitating more efficient data transmission at higher rates and over greater distances, significantly outperforming traditional copper wire cables in terms of capacity.

As we move forward, the volume of internet traffic is expected to grow exponentially, driven by developments in the Internet of Things (IoT), cloud computing, 5G technology, and an increasing number of devices connecting to the network. Cisco's annual internet report projects a rise in networked devices from 18.4 billion in 2018 to 29.3 billion by 2023. This surge in network traffic is compelling network carriers, researchers, and those involved in network infrastructure to continuously seek new technological solutions to meet the rising demands for bandwidth.

The nature of internet services is both diverse and dynamic, necessitating a variety of bit rates and relying on multiple transmission protocols. This variability adds complexity to network management, pushing the industry to adapt and innovate continually to ensure efficient and robust internet connectivity.

Optical fibers are thin strands of glass or plastic that are capable of transmitting light signals over long distances with remarkably low loss. They are a cornerstone of modern telecommunications, enabling high-speed data transmission that forms the backbone of global communications networks. Unlike traditional metallic-based communication lines, optical fibers are immune to electromagnetic interference, which allows them to transmit data at the speed of light with minimal signal degradation. This characteristic makes them essential for internet backbones, cable television, and telephone systems, transforming how information is disseminated and how global economies operate .

Chapter 2

Literature Review

The rapid increase in data traffic, fueled by cloud computing, video streaming, and other bandwidth-heavy applications, has driven the need for advancements in data center architectures to handle higher data rates and improve efficiency. Central to this evolution are Vertical-Cavity Surface-Emitting Lasers (VCSELs) used in conjunction with multimode fiber (MMF), especially with the introduction of Short Wavelength Division Multiplexing (SWDM) technologies that enable transmission rates of 100 Gbps per wavelength. This literature review delves into the theoretical foundations, technological progress, and practical applications of these systems, highlighting the crucial role of statistical analysis in optimizing their performance.

2.1 Theoretical Foundations of MMF and VCSELs

The development of multimode fiber (MMF) is rooted in the pioneering work of Gloge and Marcatalli, who introduced the multimode theory of graded-core fibers in the 1970s. Their research identified the challenges associated with modal dispersion, where different light modes within the fiber travel at varying speeds, causing pulse broadening and limiting the fiber's bandwidth-distance capabilities [1]. Understanding these concepts is essential for grasping the constraints and optimization strategies necessary for high-speed data transmission in modern MMF systems.

Agrawal's research on fiber-optic communication systems is a cornerstone in understanding the intricate interplay of dispersion, attenuation, and nonlinearity in optical fibers [2]. His work offers a framework for addressing the challenges posed by increasing data rates, particularly in multimode environments where modal dispersion plays a significant role.

Yariv's book "Optical Electronics in Modern Communication" provides detailed insights into the functioning of VCSELs, focusing on the interaction between electronic drive signals and optical outputs. This knowledge is critical for designing VCSELs that can function effectively within MMF systems, especially at higher data rates [3].

The research by Kogelnik and Li on laser beams and resonators has laid down essential principles that continue to guide the design of laser sources, including VCSELs, for optical communication. Their work forms the basis for understanding how light behaves within optical resonators, which is crucial for developing high-performance VCSELs used in data centers [4].

Snyder and Love's "Optical Waveguide Theory" builds on these ideas, offering a rigorous mathematical approach to understanding light propagation in optical fibers. Their work is particularly significant for engineers designing multimode fiber systems, providing the theoretical foundation for optimizing fiber waveguides to reduce dispersion and enhance bandwidth [5].

2.2 Dispersion and Modal Challenges in MMF Systems

Effectively managing dispersion in MMF systems is critical to achieving reliable high-speed data transmission. Chromatic and modal dispersion are the primary factors that compromise signal integrity, especially at data rates nearing 100 Gbps per wavelength.

Gholami, Molin, and Sillard introduced an innovative method to counteract chromatic dispersion by leveraging modal dispersion in MMF-VCSEL systems. By carefully managing modal dispersion, they showed that it is possible to mitigate the effects of chromatic dispersion, thereby improving link performance [6]. This approach was further refined in physical models for 10 GbE optical communication systems, which provided more precise predictions of system behavior under various conditions [7].

Pimpinella and colleagues advanced the field by developing dispersion-compensated multimode fiber (DC-MMF). Their work focused on minimizing both modal and chromatic dispersion, thereby extending the reach and bandwidth potential of MMF systems. DC-MMF is particularly valuable in data centers where high-speed, long-distance links are crucial for efficient communication networks [5].

Yabre's comprehensive theory of dispersion in graded-index optical fibers offered deeper insights into the complex interactions between different forms of dispersion in multimode environments. His work remains a critical reference for those developing and optimizing MMF systems for high-speed data transmission [8].

Schicketanz contributed to the understanding of dispersion effects by developing methods to reduce bandwidth ambiguities in graded-index fibers. His approach involved using a weighted Gaussian lowpass filter on the fiber's transfer function to reduce the impact of dispersion on signal quality [9].

Ogawa's analysis of mode partition noise in laser transmission systems highlighted another significant challenge in multimode systems. Mode partition noise, which results from fluctuations in the power distribution among different modes, can lead to signal degradation, particularly in systems where modal stability is not guaranteed [10].

Agrawal et al. also examined the impact of dispersion penalties in lightwave systems using multimode semiconductor lasers, stressing the importance of managing these penalties to maintain signal integrity over long distances and at high data rates [11].

2.3 Practical Implementations and Industry Standards

The implementation of high-speed MMF links in data centers has been shaped by a series of industry standards that ensure system interoperability and performance.

The IEEE 802.3ba standard, established in 2010, set the specifications for 40 GbE and 100 GbE over MMF. This standard provides the essential framework for achieving reliable high-speed data transmission in data centers, including guidelines for physical media, signal quality, and testing [12].

The ANSI-INCITS 479 standard, introduced in 2011, further detailed the physical interface requirements for MMF systems, particularly those used in Fibre Channel networks. This standard addresses critical factors such as modal bandwidth and signal attenuation, which are vital for ensuring reliable data transmission at speeds up to 100 Gbps [13].

The TIA-492AAAD specification for laser-optimized OM4 MMF, released in 2009, was designed to enhance the performance of MMF links operating at 850 nm. This specification ensures that MMF systems can support high data rates by optimizing the modal bandwidth and minimizing dispersion [14].

The worst-case link model for optical physical media dependent (PMD) specifications, introduced in the IEEE 802.3z standard, has been crucial in ensuring the reliability of MMF links. This model takes into account variations in fiber and transmitter characteristics, providing a robust framework for designing and testing MMF systems under challenging conditions [15].

Dolfi's proposal to modify the ISI penalty in the current Gigabit Ethernet spreadsheet represents a critical advancement. As data rates increase, the need for accurate modeling of inter-symbol interference becomes more pressing, and Dolfi's modifications aim to enhance the precision of link performance predictions [16].

2.4 Advancements in VCSEL Technology and SWDM

VCSEL technology has continued to advance, enabling higher data rates and more efficient operation in MMF systems. Tatum et al. explored the role of VCSEL-based interconnects in both current and future data centers, emphasizing the benefits of VCSELs in terms of cost, power efficiency, and scalability. Their research underscores the importance of VCSEL technology in addressing the growing demand for high-speed data transmission [16].

The 100G-SWDM4 MSA technical specifications provide comprehensive guidelines for implementing SWDM in MMF systems, including the use of VCSELs operating at multiple wavelengths. These specifications are crucial for ensuring that SWDM systems can achieve the high data rates and reliability required in data center environments [17].

Experimental studies have confirmed the effectiveness of SWDM in extending the reach and capacity of MMF links. Parsons et al. demonstrated successful 100 Gbps SWDM transmission over 250 meters of OM5 and OM4+ multimode fibers, showcasing the potential of SWDM to meet the demands of modern data centers [18]. Similarly, Lyubomirsky et al. reported successful 100 Gbps SWDM4 transmission over 300 meters of wideband MMF, further validating the feasibility of SWDM in high-performance data center applications [19].

Sun et al. investigated SWDM PAM4 transmission over next-generation wideband multimode optical fibers, providing insights into the capabilities of PAM4 modulation in SWDM systems. Their research demonstrated that PAM4 could effectively double the data rate without the need for additional fiber, making it a promising approach for future data center deployments [20].

2.5 Techniques for Bandwidth Enhancement and Modal Control

Researchers have explored various techniques to further improve the performance of MMF links by controlling modal distribution and increasing bandwidth. The offset launch technique, examined by Raddatz, White, and Cunningham, is one such method. By strategically offsetting the launch position of light into the MMF, this technique selectively excites certain modes, thereby reducing modal dispersion and increasing the effective bandwidth of the link [21].

Pampaloni and Enderlein's primer on Gaussian, Hermite-Gaussian, and Laguerre-Gaussian beams provides valuable insights into how these beam profiles can be applied in optical communication systems to minimize losses and optimize data transmission efficiency. These beam profiles are particularly important for designing

VCSELs that efficiently couple into MMF, ensuring high performance in data center networks [22].

Dong et al. developed an innovative equalizer for 112 Gbps CAP-based data transmission over 150 meters of MMF links. Their research demonstrated the effectiveness of advanced equalization techniques in mitigating the effects of dispersion and inter-symbol interference, thereby enhancing the reliability and performance of high-speed MMF systems [23].

Lavrencik et al. investigated 4 $\lambda\times$ 100 Gbps VCSEL PAM-4 transmission over 105 meters of wideband multimode fiber. Their study showed that PAM-4 modulation, combined with VCSEL technology, could achieve high data rates over relatively short distances, making it a viable solution for data centers [24].

Karinou et al. explored 112 Gb/s PAM-4 optical signal transmission over 100 meters of OM4 multimode fiber. Their research highlighted the potential of PAM-4 modulation to significantly increase data rates without requiring additional fiber infrastructure, making it an attractive option for high-capacity data center interconnects [25].

2.6 Challenges and Future Directions

Despite substantial progress, challenges remain in ensuring reliable 100 Gbps per wavelength transmission over MMF. Issues such as dispersion penalties, mode partition noise, and inter-symbol interference are significant obstacles that must be addressed to ensure the long-term viability of MMF systems in high-speed applications.

Ogawa's analysis of mode partition noise in laser transmission systems highlighted the challenges of maintaining modal stability in multimode environments, which can lead to signal degradation [10]. Managing these noise factors is essential for ensuring consistent performance in data center networks.

As data centers continue to evolve, the importance of MMF and VCSEL technologies in enabling high-speed, cost-effective communication will remain central to meeting the growing demands of the digital age. Future research will likely focus on refining dispersion management techniques, developing more advanced VCSEL designs, and exploring new modulation formats to push the boundaries of what is possible with MMF systems.

Chapter 3

Multimode Fiber in Data Center Applications

Multimode fibers (MMFs) play a pivotal role in the infrastructure of modern data centers by facilitating the rapid transmission of massive amounts of data. They are commonly used to interconnect data storage systems and servers within the same facility. The key advantage of MMFs in these applications is their ability to transmit high data rates over short distances, typically up to 500 meters without significant loss of signal quality. This capability makes them ideal for the densely packed environment of data centers where servers and storage systems need to communicate with minimal latency and high bandwidth.

3.1 Advantages over Single-Mode Fibers (SMFs)

While single-mode fibers (SMFs) provide higher bandwidth and longer reach capabilities, MMFs are often preferred in data center scenarios for several reasons:

- Cost-Effectiveness: MMFs typically cost less than SMFs in terms of both the fiber itself and the associated optics. MMFs use less expensive light sources like LEDs or vertical-cavity surface-emitting lasers (VCSELs), which are more cost-effective compared to the laser sources needed for SMFs.

- Ease of Installation: The larger core size of MMFs is less susceptible to installation damage and allows for more relaxed alignment tolerances during connectorization, which simplifies the installation and maintenance processes.

- Bandwidth Sufficiency: For the typical distances in data centers, MMFs provide sufficient bandwidth for the applications being run. The data rate requirements of most current data center applications are well within the delivery capability of modern MMFs, especially with advancements in fiber design that reduce modal dispersion.

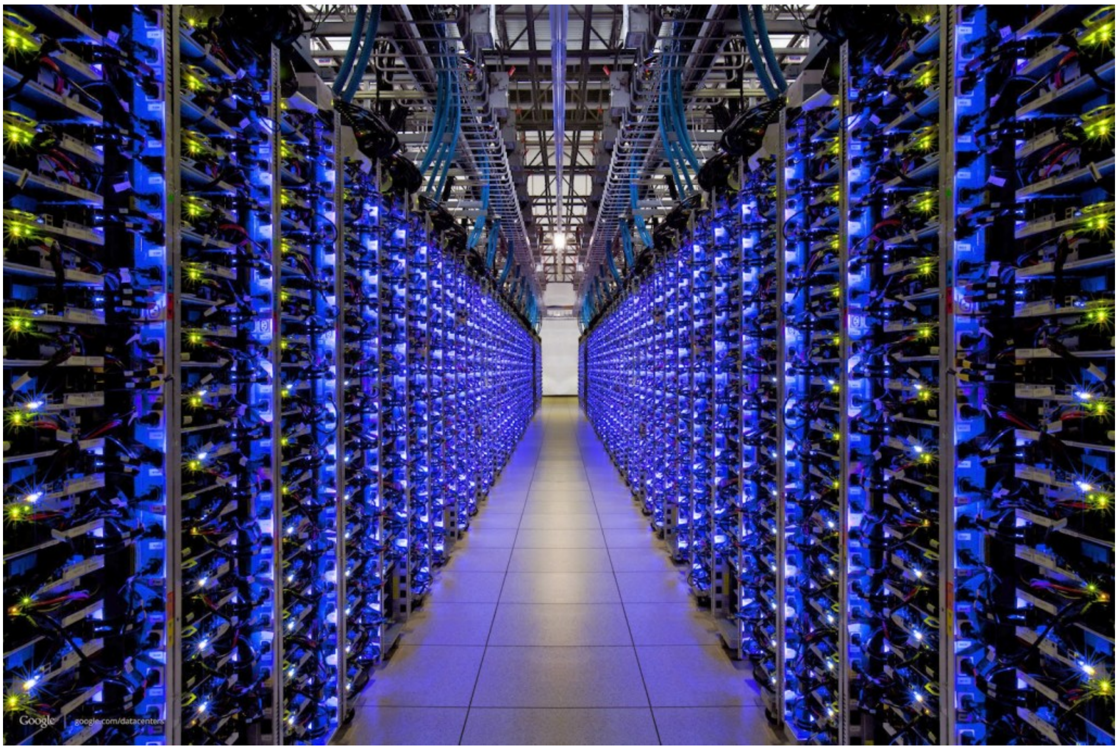


Figure 3.1: Data center

3.2 Future Trends in Data Center Networking

The landscape of data center networking is continuously evolving with the introduction of new technologies. Some of the significant future trends that may influence the use of MMF include:

- Software-Defined Networking (SDN): This technology allows for the dynamic management of network resources through software controls, facilitating more efficient data flow management. MMFs, with their high bandwidth capabilities, are well-positioned to benefit from SDN implementations by supporting flexible and scalable network configurations.

- Integration with Advanced Modulation Techniques: Technologies like advanced modulation formats and multiplexing techniques could extend the utility of MMFs even further by increasing the amount of data transmitted without altering the physical infrastructure.

- Photonic Integration: The development of photonic integrated circuits could revolutionize MMF applications by integrating multiple optical functions onto a single chip, thereby enhancing performance and reducing costs.

These advancements are poised to not only maintain but potentially expand the role of multimode fibers in data centers. As data center architectures evolve to incorporate cloud services, big data analytics, and the Internet of Things (IoT), the reliance on MMFs could further solidify, provided they continue to adapt to the increasing performance demands.

3.3 Purpose Of The Thesis

This thesis delves into the Statistical Analysis of 100 Gbps per Wavelength SWDM VCSEL-MMF Data Center Links, utilizing an extensive dataset generated through simulations on a comprehensive collection of OM4 fibers. The primary aim of this detailed study is to investigate and elucidate the complex relationships between the optical transfer function, the Bit Error Rate (BER), and the maximum achievable link length (L_{max}).

By harnessing advanced simulation tools, a large volume of data has been gathered, providing a solid foundation for our analytical endeavors. The focus is centered on unraveling the correlations between the transfer function's characteristics and key performance indicators such as BER and L_{max}. This involves a rigorous statistical analysis where various methodologies are applied to thoroughly examine the data. The exploration of these correlations is intended to identify pivotal patterns and trends that can significantly inform and enhance the design and optimization of optical data center links.

Once a robust correlation is established through this analysis, the findings will be utilized to develop a predictive simulator. This simulator will be designed to forecast BER and L_{max} based solely on the characteristics of the transfer function. The development of such a tool represents a pivotal advancement in our ability to predict and enhance the performance of SWDM VCSEL-MMF links in real-world applications. This proactive approach to simulation and prediction underscores the practical implications of our research, aiming to significantly boost data transmission efficiency and reliability in future generations of data center networks. Through this integrative study, we seek to contribute valuable insights and practical tools that pave the way for more efficient, reliable, and optimized data center connectivity solutions.

3.4 Types of Optical Fibers

Optical fibers can be broadly categorized into two main types: single-mode fibers (SMF) and multimode fibers (MMF).

- **Single-Mode Fibers (SMF)**: SMFs have a small core size (approximately 8 to 10 microns in diameter) and transmit infrared laser light (wavelengths of

1310 or 1550 nanometers). The smaller core size minimizes the distance light rays travel as they propagate, reducing signal attenuation and allowing the fibers to transmit information over longer distances. This makes SMFs ideal for long-distance telecommunications such as transcontinental or undersea cabling systems.

- **Multimode Fibers (MMF)**: MMFs have a larger core size (about 50 to 62.5 microns in diameter) which allows light rays to travel through the core via multiple paths, or modes. This ability makes them suitable for shorter distances as the light tends to scatter and disperse over longer stretches, leading to modal dispersion, a form of signal degradation. MMFs are typically used in applications such as within buildings or on campus environments where high bandwidth and short distances are the norms.

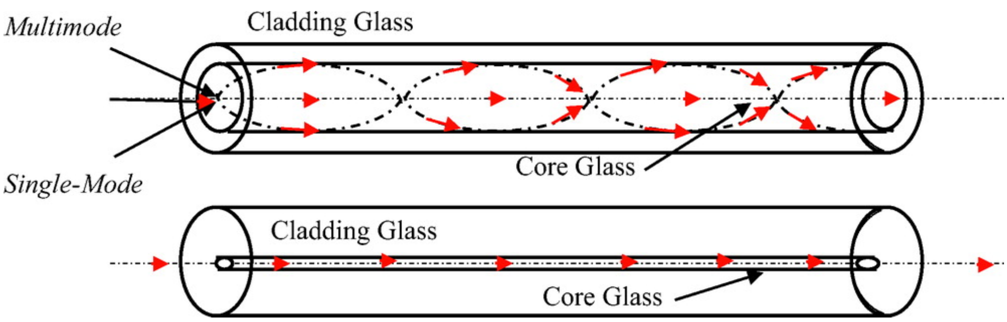


Figure 3.2: SMF vs MMF

3.5 Fiber Communication Systems

- Basic Principles of Fiber Optics

  Fiber optics utilize the physics of light propagation within thin, flexible fibers typically made of glass or plastic—to transmit data over long distances with minimal loss. The fundamental mechanism enabling this is total internal reflection. This occurs when light waves strike the boundary of a denser medium at an angle exceeding the critical angle, preventing the light from escaping and ensuring it follows the fiber's path. The phenomenon hinges on Snell's law, which relates the indices of refraction of the two media and the angles of incidence and refraction

3.6 Components of Fiber Communication Systems

A typical fiber optic communication system consists of several key components:

- Transmitters: These convert electrical signals into light signals. They usually incorporate lasers or LED sources that emit light when electrically stimulated.
- Optical fibers: The medium through which the light signals travel. These fibers are designed with a core and cladding layer; the core carries the light, and the cladding layer traps the light within the core using total internal reflection.
- Receivers: These convert light signals back into electrical signals. Photodiodes are commonly used for this purpose, as they generate an electrical signal in response to incoming light.
- Repeater: Placed at intervals along the fiber to boost the signal strength, repeaters are essential for maintaining signal integrity over long distances.

3.7 Advancements in Fiber Optic Technology

Recent technological advancements have significantly enhanced the capabilities of fiber optic systems. Dense Wavelength Division Multiplexing (DWDM) is a notable development, allowing multiple light wavelengths to be multiplexed onto a single fiber, thereby increasing the data transmission capacity manifold. Additionally, photonic switching and optical amplifiers have improved signal routing and boosting without requiring conversion back to electrical signals, facilitating more efficient and faster data transmission across networks.

By understanding and harnessing these advanced techniques, fiber optic technology continues to evolve, pushing the boundaries of data transmission capabilities and playing a pivotal role in the infrastructure of modern telecommunications.

3.8 Focus on Multimode Fibers

Construction and Properties of MMF Multimode fibers (MMFs) are characterized by their ability to propagate multiple light modes or paths concurrently through their core. This is facilitated by a core size significantly larger than that of single-mode fibers, typically ranging from 50 to 62.5 micrometers in diameter. The larger core size allows light rays to follow multiple paths through the fiber, which can lead to modal dispersion—a limitation in high-bandwidth or long-distance applications.

The structure of an MMF typically includes three key components:

- Core: The central part of the fiber where the light is transmitted. The larger core size supports multiple light modes.
- Cladding: A layer of material surrounding the core that has a lower refractive index, which confines the light within the core through total internal reflection.
- Buffer Coating: An outer protective layer that shields the fiber from physical damage and moisture, enhancing the durability and longevity of the fiber cable.

  The interaction between the core and cladding determines much of the fiber's optical characteristics, such as its numerical aperture, which influences the light-gathering ability and the angle at which light can be injected into the fiber.

3.9 Applications in Data Centers

MMFs are extensively used in data centers due to their efficiency in handling high-speed, short-range communications. This environment typically requires the transmission of a large volume of data across relatively short distances, where MMFs excel due to their higher capacity and lower cost compared to single-mode fibers. Their ability to transmit multiple light modes simultaneously makes them suitable for applications that require high data throughput over short distances, such as within a data center.

Challenges with MMF :

Despite their advantages in certain applications, MMFs face several challenges:

- Modal Dispersion: This occurs because different light modes travel different paths through the fiber, each at a slightly different speed. As a result, light pulses spread out in time, which can blur the signal over long distances or high data rates. This limits the effective bandwidth and the distance over which data can be transmitted without significant signal degradation.
- Bandwidth Limitations: The inherent properties of MMFs restrict their bandwidth over longer distances, compared to single-mode fibers. This is a critical limitation in expanding the role of MMFs in future high-speed network infrastructures.

Modal dispersion, in particular, has prompted the development of graded-index MMF, where the refractive index of the core decreases gradually from the center towards the cladding. This design helps to minimize modal dispersion by allowing light rays in different paths to travel at similar speeds, effectively increasing the bandwidth and enhancing the performance of MMF in data transmission.

These challenges necessitate ongoing advancements and innovations in fiber optic technology to enhance the capabilities of multimode fibers in a variety of communication environments.

3.10 Multimode Fiber in Data Center Applications

Role in Data Transmission

Multimode fibers (MMFs) play a pivotal role in the infrastructure of modern data centers by facilitating the rapid transmission of massive amounts of data. They are commonly used to interconnect data storage systems and servers within the same facility. The key advantage of MMFs in these applications is their ability to transmit high data rates over short distances, typically up to 500 meters without significant loss of signal quality. This capability makes them ideal for the densely packed environment of data centers where servers and storage systems need to communicate with minimal latency and high bandwidth.

3.11 Future Trends in Data Center Networking

The landscape of data center networking is continuously evolving with the introduction of new technologies. Some of the significant future trends that may influence the use of MMF include:

- Software-Defined Networking (SDN): This technology allows for the dynamic management of network resources through software controls, facilitating more efficient data flow management. MMFs, with their high bandwidth capabilities, are well-positioned to benefit from SDN implementations by supporting flexible and scalable network configurations.

- Integration with Advanced Modulation Techniques: Technologies like advanced modulation formats and multiplexing techniques could extend the utility of MMFs even further by increasing the amount of data transmitted without altering the physical infrastructure.

- Photonic Integration: The development of photonic integrated circuits could revolutionize MMF applications by integrating multiple optical functions onto a single chip, thereby enhancing performance and reducing costs.

These advancements are poised to not only maintain but potentially expand the role of multimode fibers in data centers. As data center architectures evolve to incorporate cloud services, big data analytics, and the Internet of Things (IoT), the reliance on MMFs could further solidify, provided they continue to adapt to the increasing performance demands.

Chapter 4

Research Objective and Methodology

The primary goal of this thesis is to investigate the relationship between the optical transfer functions of multimode fibers (MMFs) and key performance metrics such as Bit Error Rate (BER) and maximum achievable transmission length under various conditions. This study aims to provide insights into how alterations in the physical and optical properties of MMFs influence their efficiency and reliability in data transmission, particularly within high-demand environments like data centers.

This research utilizes a comprehensive dataset compiled from multiple sources, including experimental data collected from field tests and simulations. The statistics of the MMF is due to the different propagation speed of the modes induced by mode dispersion, which depends on the non-perfect fabrication of the MMFs. The selection of these particular data points is driven by their relevance to current industry standards and their ability to provide a broad understanding of MMF behavior across typical and extreme use cases. This diverse dataset ensures robustness in the findings, accommodating the variability encountered in practical applications.

- Source Variety: Data is sourced from both academic research facilities and telecommunications industry partners to cover a broad spectrum of MMF configurations and usage scenarios.

- Criteria for Inclusion: Data points were selected based on their relevance to the most commonly implemented MMF standards in contemporary data centers, as well as their potential to reveal insights about less understood phenomena in fiber optics, such as modal noise and differential mode delay.

4.1 Analytical Methods

The methodology section outlines the various techniques and tools employed to analyze the collected data:

- Statistical Analysis: Utilizes descriptive statistics to summarize data characteristics and inferential statistics to determine the relationships between MMF properties and performance metrics. Techniques such as regression analysis and correlation matrices are applied to understand and quantify these relationships.
- Machine Learning Models: Advanced predictive models, including neural networks , are developed to forecast MMF performance based on measured data. This approach allows for the extrapolation of the fibers' behavior under new conditions not explicitly represented in the dataset.
- Software and Tools: The analysis is supported by software such as MATLAB and Python, with libraries including SciPy for statistical tests, TensorFlow for machine learning, and custom scripts to handle data preprocessing and visualization.
- Validation Techniques: Cross-validation methods are employed to ensure the generalizability of the machine learning models. This includes dividing the dataset into training and testing sets to evaluate the model's predictive accuracy on unseen data.

4.2 SYSTEM SIMULATION

4.2.1 Simulation setup

This section outlines the system-level numerical simulator used to derive the core results of this thesis..[26] The simulation setup includes five main components: the digital transmitter (TX), VCSEL, multimode fiber (MMF), optical receiver (PIN+TIA), and digital receiver (RX), as shown in Figure **??**. DAC and ADC converters interface between the electro-optical channel and the digital TX/RX. A variable optical attenuator (VOA) adjusts the received optical power (ROP), which regulates the receiver's optical modulation amplitude (OMARX).

The digital TX generates a gray-coded PAM-4 signal, which is shaped with a Gaussian filter having a -3 dB bandwidth equal to 0.75 times the baud rate. FEC is applied, with two options: KP4-FEC (6.25% overhead, BER = 2×10^{-4}) and E-FEC (10.35% overhead, BER = 4×10^{-3}), resulting in net bit rates of 100 Gbps.

The VCSEL is modeled as a linear device with bandwidth limitations and relative intensity noise (RIN). The PAM-4 signal is converted into instantaneous laser power,

$P_{TX}(t)$, which is transmitted either in a back-to-back (BtB) configuration or via MMF, modeled by a transfer function $H_{MMF}(f)$. Chromatic dispersion values follow ITU-T G.651 standards.

At the receiver, shot and thermal noise are added to the optical power $P_{RX}(t)$. The signal is then filtered, downsampled to 2 samples per symbol (SpS), and passed through one of three equalization schemes:Feed Forward Equalizer (FFE), Decision Feedback Equalizer (DFE), or Maximum-Likelihood Sequence Estimation (MLSE). After equalization, the PAM-4 signal is decoded, and bit error rates (BER) are evaluated.[26]

Table 4.1: Simulation Parameters Used for Statistical Analysis of SWDM Systems

Parameter	**Value**	**Units**
ER	3.0	dB
RIN coefficient, Γ_{RIN}	-141.0	dBc/Hz
Received OMA, OMA_{RX}	-3.5	dBm
Received Optical Power, *ROP*	-1.75	dBm
IRND	20.0	pA/$\sqrt{\text{Hz}}$
R	0.5	A/W

Chapter 5

Preliminary Data Analysis

In the initial phase of handling a large dataset, one fundamental approach is to thoroughly explore the data by creating numerous visualizations. This process aids in understanding the underlying patterns and anomalies present in the data, facilitating more informed analysis later on.

For this study, several key datasets have been procured to support the investigation of 100 Gbps per Wavelength SWDM VCSEL-MMF Data Center Links. These datasets include:

- **MMF transferFunctions OM4 400m.mat**: This file contains the optical transfer functions (TFs) and the mode group delays (MGD) for each multimode fiber (MMF), specifically for fibers measuring 400 meters. Transfer functions are a key feature as they depict how the fiber's characteristics influence the signal's integrity over distance.

- **DataBER 106.25G OM4.mat**: This dataset comprises the Bit Error Rate (BER) observed at various lengths of fiber, assuming a transmission rate of 106.25 Gbps per wavelength. To accommodate different scenarios at the receiver (RX) side, three types of equalizers are considered: Feed Forward Equalizer (FFE), Decision Feedback Equalizer (DFE), and Maximum Likelihood Sequence Estimation (MLSE) equalizer. Each of these equalizers has a distinct impact on the performance of the fiber link, modifying how errors in the data are processed and corrected.

- **Data Lmax 106.25G OM4.mat**: This dataset provides the maximum achievable link length (Lmax) for the desired BER of 2×10^{-4} across the four SWDM channels and the three equalizers mentioned. Lmax is a crucial metric for network designers as it defines the maximum distance over which data can be reliably transmitted without exceeding the predefined error threshold.

5.1 Datasets Evaluation

First there is datasets common between all the three files which are :

- **Lambdas**: The vector of channel wavelengths [1× 4] which are [850,880,910,940] nanometers.
- **Lasers**: The vector of laser numbers [1 × 8] which are just 8 different lasers labeled from 1 to 8 : [1,2,3,4,5,6,7,8] .

Adding to this , each dataset contains more fields that will be presented below in details :

5.1.1 MMF_transferFunctions_OM4_400m.mat

This dataset contains the following fields:

- **freq**: Vector of frequencies over which the transfer functions are evaluated [301 × 1] which are from 0 to 300 Ghz.
- **Hf_MMF_OM4**: Matrix of transfer functions with size [Nlambda× Nlaser × Nfreq× NOM4],[4 × 8 × 301 × 3766] where NOM4 is the number of OM4 fibers in the dataset.
- **EMB_OM4**: Vector of effective modal bandwidths [1xNOM4].
- **taug_OM4**: Matrix of modal group delays with size [NlambdaxNOM4xNMG],[4 × 3766 × 19] where NMG is the number of mode groups.

5.1.2 DataBER_106.25G_OM4.mat

This dataset contains the following fields:

- **Lengths**: Vector of fiber distances [1× 11] which are {0, 30, 50, 70, 100, 150, 200, 250, 300, 350, 400} meters.
- **BitRate**: Gross bit rate equal to 106.25 Gbit/s .
- **EQtype**: Vector of cells containing the names of the considered equalizers [1xNEQs] wich is [1 × 3] .
- **BER_OM4**: Matrix of BER with size [NlambdaxNlaserxNEQsxNOM4xNdist] , [4 × 8 × 3 × 3766 × 11].

5.1.3 DataLmax_106.25G_OM4.mat

This dataset contains the following fields:

- **Lengths**: Vector of fiber distances [1 × 11] same presented in the file above .
- **BitRate**: Gross bit rate equal to 106.25 Gbit/s ..
- **EQtype**: Vector of cells containing the names of the considered equalizers [1 × 3].
- **BERtarget**: Target bit error rate equal to 2×10^{-4}.
- **Lmax_OM4**: Matrix of maximum reach at the BER target with size [NlambdaxNlaserxNEQsxNOM4] , [4 × 8 × 3 × 3766].

5.2 Step 1 : Computation of the Transfer Functions at Different Distances

In the initial phase of this study, the focus was on analyzing the transfer function, which is a four-dimensional dataset encapsulating wavelength, laser, frequency, and fiber characteristics. Specifically, the dataset encompasses transfer functions across four wavelengths: 850 nm, 880 nm, 910 nm, and 940 nm, spanning a frequency range from 0 to 300 GHz, and covering 3766 OM4 optical fibers.

A key observation from the analysis is that the fundamental shape of the transfer function, characterized by its lobes, remains consistent regardless of changes in fiber length. Instead of altering in shape, the transfer function exhibits changes in width—it narrows or widens corresponding to increases or decreases in fiber distance, respectively. This behavior underscores the scaling properties of the transfer function relative to fiber length.

To quantify this scaling effect, a scaling factor K is used, which is calculated based on the ratio of a reference distance (L_{ref}) to any desired distance (L). The formula for this scaling factor is given by:

$$K = \frac{L_{\text{ref}}}{L}$$

Utilizing this scaling factor, one can compute the transfer functions for different distances by simply adjusting the frequency axis of the transfer function measured at the reference distance of 400 meters. This process involves stretching the frequency axis by the factor K, allowing for a straightforward adaptation of the transfer function to various link lengths. This method provides a practical and efficient way to predict the optical characteristics of a VCSEL+MMF link over varying distances without the need for recalculating the entire transfer function from scratch.

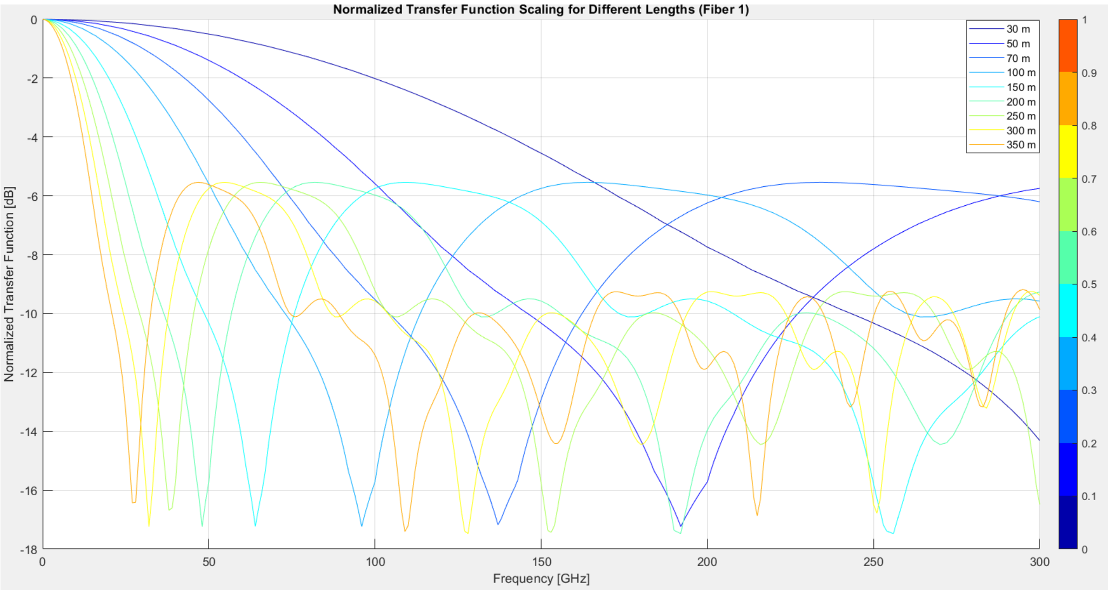


Figure 5.1: Transfer function versus frequency at different length for λ=850 nm and laser 1

Figure 5.1 shows how the transfer function varies as a function of frequency (0 to 300 GHz) for fiber lengths ranging from 30 meters to 350 meters. Each line represents the transfer function for a different fiber length, with the color corresponding to the length as indicated by the color bar. The plot reveals that, while the general shape of the transfer function remains consistent across different lengths, the function narrows for longer fibers and widens for shorter fibers, reflecting the scaling behavior of the transfer function with fiber length. This suggests that higher frequencies experience more attenuation as the fiber length increases, but the fundamental pattern of the function remains unchanged.

5.3 Step 2 : Correlation analysis

The ongoing data analysis aims to uncover correlations within the dataset, specifically focusing on relationships involving the transfer function. Initially, my interest centered on identifying any significant correlations between various parameters and the transfer function. To begin this exploration, I first computed the -3dB bandwidth for each fiber. This bandwidth is crucial as it represents the frequency at which the signal's power drops to 50% of its maximum value. I set specific indices for the wavelength and the laser, and then calculated the -3dB bandwidth across all 3766 fibers for each specified wavelength and laser combination, resulting in an array of 3766 -3dB bandwidth values, each corresponding to a single fiber.

To expand this analysis further, The B_{-3dB}, B_{-5dB}, B_{-10dB}, and equivalent bandwidth B_{eq} were calculated in all the cases , and each case in 1 specific wavelength and 1 specific laser .In this case there is 4 wavelengths [850,880,910,940] nanometers and 8 lasers labeled from 1 to 8 so 32 cases overall. This comprehensive approach was designed to explore which parameters might exhibit a notable correlation with the transfer function, providing deeper insights into the dynamics of the optical transmission.

Subsequently, the analysis was extended to encompass the maximum link length (Lmax) data. Lmax values are extracted for every fiber across all wavelengths and lasers in the dataset. This extraction resulted in a vast array of values, precisely $4 \times 8 \times 3766$ Lmax entries, reflecting the extensive range of conditions tested. In parallel, the bandwidths mentioned are computed for all fibers under every condition to ascertain which of these bandwidth measurements demonstrates significant correlations with both the transfer function and the Lmax values.

5.3.1 -3dB, -5dB, -10dB Bandwidth Calculation Algorithm

To calculate the bandwidths at -3dB, -5dB, and -10dB, the following algorithm is used:

1. Convert the transfer function (TF) to the decibel (dB) scale using:

$$TF_{\text{dB}} = 10 \times \log_{10}(|TF|)$$

2. Find the maximum value of TF_{dB}.

3. Identify the frequencies where the magnitude response is greater than or equal to $\max(TF_{\text{dB}}) - X$, where $X = 3$, 5, or 10 depending on the bandwidth type.

4. Compute the bandwidth as the difference between the highest and lowest frequencies within the range.

5.3.2 Equivalent Bandwidth Calculation Algorithm

To calculate the equivalent bandwidth, we follow these steps:

1. Compute the power by squaring the magnitude of the transfer function:

$$TF_{\text{squared}} = |TF|^2$$

2. Integrate over the frequency range to find the total power:

$$\text{total power} = \int_{\text{freq}} |TF|^2 \, d\text{freq}$$

3. Calculate the equivalent bandwidth as:

$$\text{equivalent bandwidth} = \frac{\text{total power}}{\max(|TF|^2)}/10^9$$

5.3.3 Correlation Calculation Method

To assess the relationship between the different bandwidth measurements and the bit error rate (BER) and the maximum achievable length for a fiber $L_{\max}$, we computed the Pearson correlation coefficients using MATLAB's `corr` function. The correlation matrix was calculated based on the following bandwidth metrics: B_{-3dB}, B_{-5dB}, B_{-10dB}, and the equivalent bandwidth B_{eq}, along with the selected BER values.

The procedure for calculating the correlation is as follows:

1. We first created a data matrix, denoted as `data_matrix`, where each column corresponds to one of the bandwidth measurements or the BER values:

$$\texttt{data\_matrix} = [B_{-3dB}, B_{-5dB}, B_{-10dB}, B_{eq}, BER].$$

2. The Pearson correlation coefficient, r, between two variables X and Y is calculated using the formula:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

   where $\bar{X}$ and $\bar{Y}$ are the means of the variables X and Y, respectively.

3. To ensure the correlation calculation is robust against missing data, we used the `'Rows', 'complete'` option, which excludes any rows containing `NaN` values from the calculation.

4. The result is a correlation matrix that quantifies the linear relationship between each pair of variables in the dataset. A value of $r = 1$ indicates a perfect positive linear correlation, $r = -1$ indicates a perfect negative linear correlation, and $r = 0$ indicates no linear relationship.

5. Finally, the correlation matrix is visualized using a heatmap to provide an intuitive understanding of the relationships between the different bandwidth features and the BER values.

The heatmap, displayed with labels for each variable, allows for quick visual inspection of the strongest and weakest correlations in the data.

5.4 Study at length equal to 400 meters

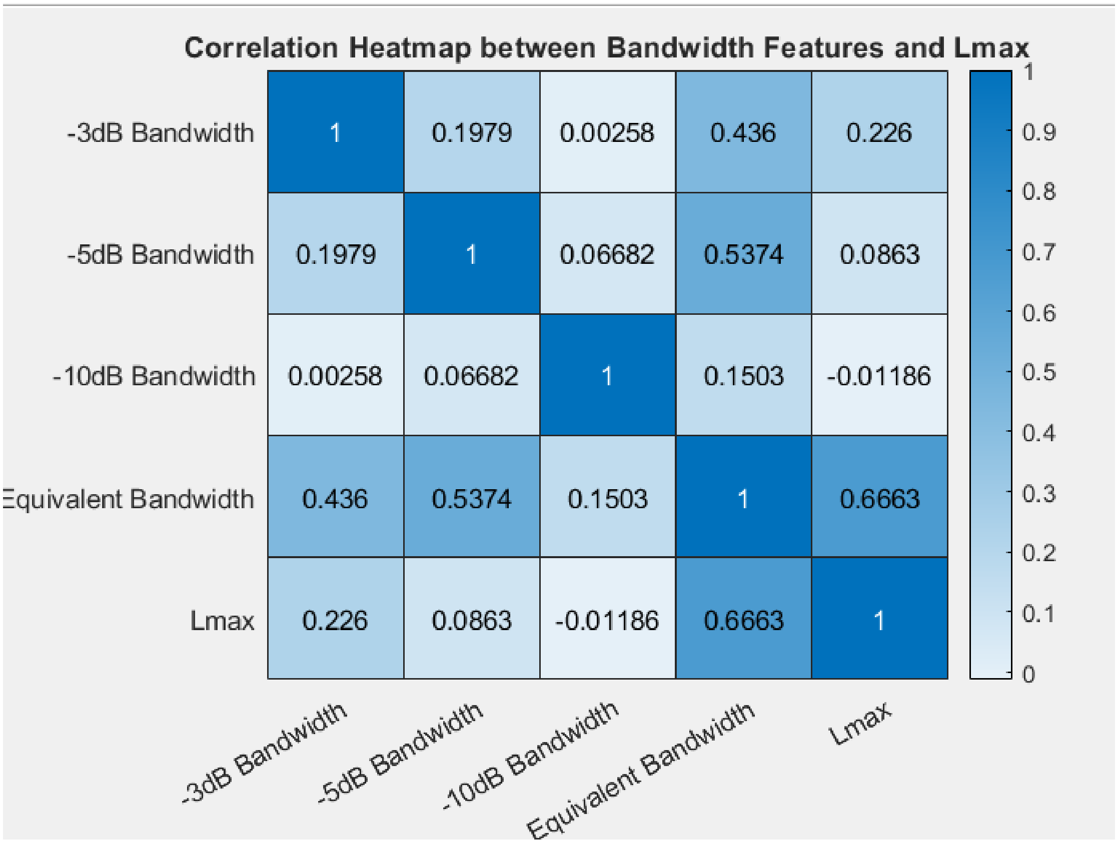


Figure 5.2: Correlation between -3dB , -5dB , -10 dB and Equivalent Bandwidth vs $L_{\max}$ for 3766 fibers at λ=850 nm and laser 1

Analysis :
Figure 5.2 the correlation between various bandwidth features and a parameter labeled $L_{\max}$ for 3766 fibers at a specific wavelength λ and laser . The features include the -3dB Bandwidth, -5dB Bandwidth, -10dB Bandwidth, and Equivalent Bandwidth. Here's a detailed analysis of what each part of the heatmap represents and the implications of these correlations:

Heatmap Interpretation

- **Diagonal (1.0)**: The diagonal values are all 1, which is expected as it represents the correlation of each bandwidth feature with itself.

  Since the study interest is just in the correlation of the different bandwdith with $L_{\max}$, these are the relevant correlation found :

- **Lmax and -3dB Bandwidth**: Correlation is 0.226, indicating a weak to moderate positive relationship.
- **Lmax and -5dB Bandwidth**: Correlation is very low (0.0863), suggesting minimal linear relationship.
- **Lmax and -10dB Bandwidth**: Correlation is slightly negative (-0.01186), indicating that there is essentially no relationship, and in some cases, it may even be slightly inverse.
- **Lmax and Equivalent Bandwidth**: Correlation is significant (0.6663), suggesting a strong positive relationship. This means that as the equivalent bandwidth increases, the Lmax, possibly representing maximum transmission length or loss, also tends to be higher.

Implications and Conclusions

Domainance of the Equivalent Bandwidth : The Equivalent Bandwidth appears to have a strong correlation with Lmax, indicating it might be a more comprehensive measure of bandwidth performance relative to Lmax than the specific -3dB, -5dB, or -10dB measures.

Taking into account the whole dataset

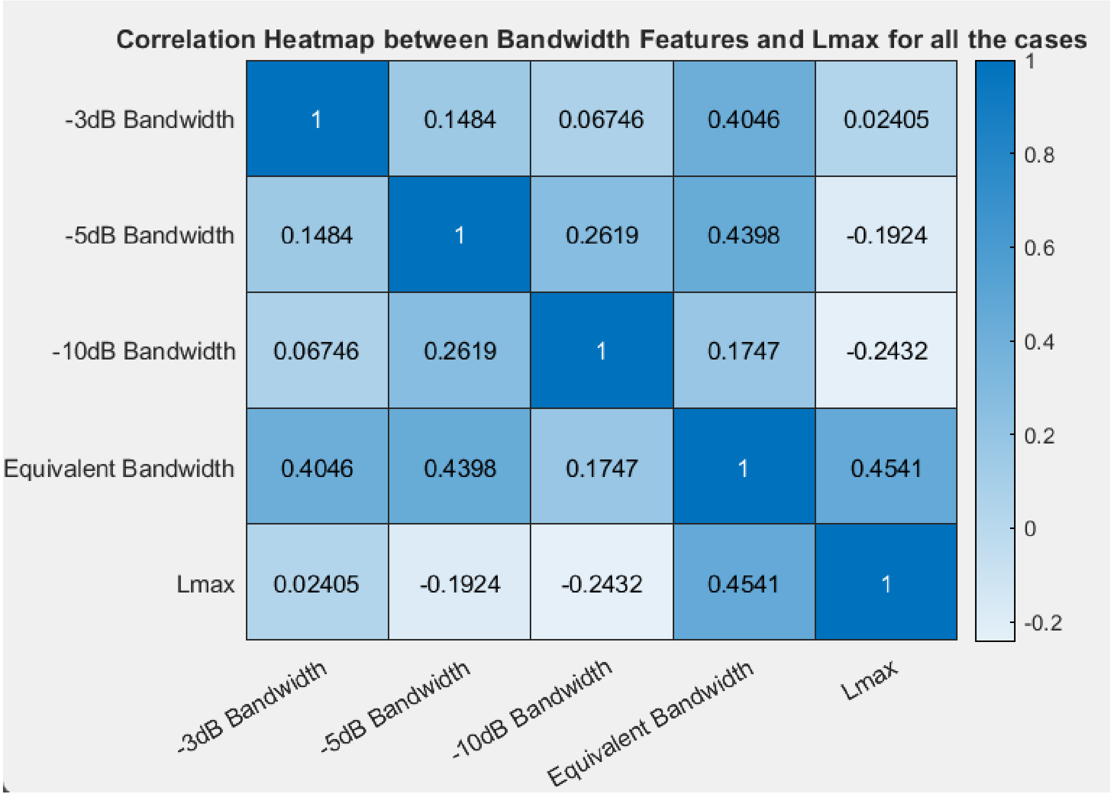


Figure 5.3: Correlation between -3dB , -5dB , -10 dB and Equivalent Bandwidth vs Lmax for the entire dataset of wavelengths and lasers

Analysis and Interpretation Figure 5.3 illustrates how bandwidth features correlate with Lmax (potentially maximum link length or loss characteristics) for the data encompassing all 120,125 cases which are for all possible combination of wavelength , laser and fiber so $4 \times 8 \times 3766$. Analysis :

- **Equivalent Bandwidth** shows the strongest positive correlation with $L_{\max}$, having a correlation value of 0.4541. This suggests a moderate positive linear relationship, meaning that as the equivalent bandwidth increases, $L_{\max}$ tends to increase as well. This indicates that the equivalent bandwidth is a reliable predictor of the maximum reach of the fiber.

- **-3dB Bandwidth** exhibits a very weak positive correlation with $L_{\max}$ at 0.02405. This indicates almost no relationship between the -3dB bandwidth and $L_{\max}$, suggesting that it is not a good predictor of the maximum reach.

- **-5dB Bandwidth** shows a weak negative correlation with $L_{\max}$ at -0.1924. This negative correlation suggests a slight inverse relationship, meaning that as the -5dB bandwidth increases, $L_{\max}$ tends to decrease slightly.
- **-10dB Bandwidth** has a negative correlation with $L_{\max}$ at -0.2432, indicating a moderate inverse relationship. As the -10dB bandwidth increases, $L_{\max}$ tends to decrease more noticeably than with the -5dB bandwidth. However, the overall correlation is weak, indicating that the -10dB bandwidth is not a strong predictor of maximum reach.

Summary: The equivalent bandwidth is the most significant predictor of $L_{\max}$, showing a moderate positive correlation. The -3dB, -5dB, and -10dB bandwidths show weak correlations, with the -3dB bandwidth having almost no influence on $L_{\max}$ and the -10dB bandwidth having the strongest negative relationship.

Proposed Study on BER and Bandwidth:

- Analyzing BER and Bandwidth: Investigating how BER correlates with bandwidth metrics (-3dB, -5dB, -10dB, and Equivalent Bandwidth) across different distances will highlight which bandwidth measure is most indicative of data integrity and transmission quality. This study could reveal important patterns useful for network management and optimization.
- Distance Impact: Since fiber optic properties such as dispersion and attenuation vary with distance, analyzing these correlations over varying distances could provide deeper insights into the optimal configurations of optical networks for different applications or conditions.

Conclusion:

The observed differences in correlation across dataset sizes underscore the complexity of optical network behaviors and the need for targeted studies under varied conditions. Moving forward with the BER analysis will potentially uncover more robust relationships that directly affect network performance and reliability.

Taking into account different approach

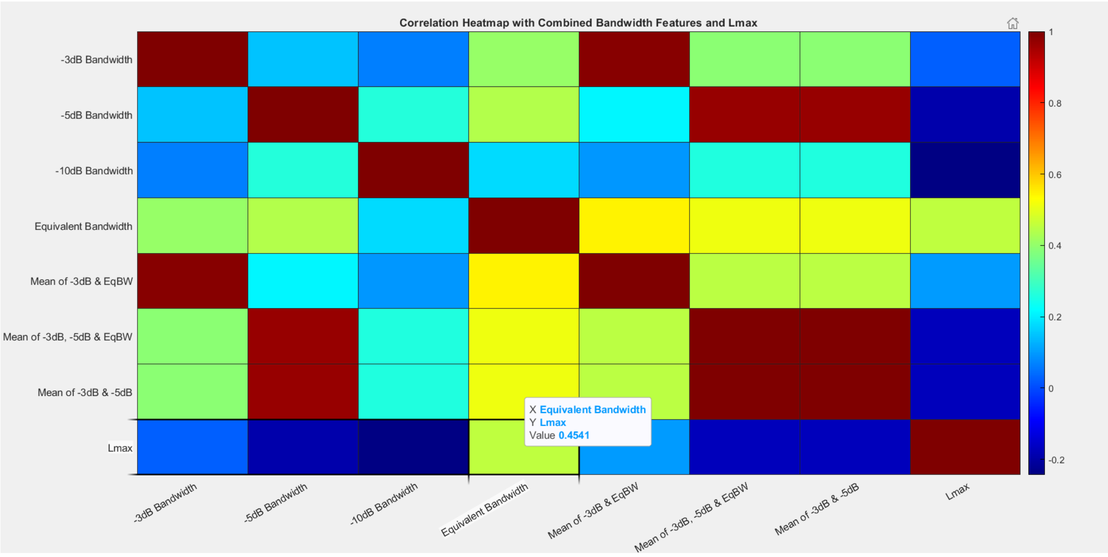


Figure 5.4: Correlation between -3dB , -5dB , -10 dB and Equivalent Bandwidth and different combination between them vs Lmax for 3766 fiber at λ= 850 nm and laser 1

Analysis and Interpretation

5.4 explores the relationship between various bandwidth measurements and their combined averages with $L_{\max}$. The goal was to identify stronger correlations by investigating both individual and combined metrics.

- **Individual Bandwidth Measurements:** -3dB, -5dB, -10dB Bandwidths, and Equivalent Bandwidth show moderate correlations with $L_{\max}$, with Equivalent Bandwidth having the highest at 0.4541.
- **Combined Metrics:** Averages of bandwidth metrics like -3dB and Equivalent Bandwidth, or -3dB, -5dB, and Equivalent Bandwidth, were explored. However, these combinations did not yield significantly higher correlations with $L_{\max}$ compared to individual metrics.

Correlation Observations

Moderate correlations were observed across the metrics, with none showing a strong predictive relationship with $L_{\max}$. The combined metrics did not improve upon the individual bandwidth measures in terms of correlation strength.

Implications and Future Directions

- The combined metrics did not enhance the correlation with $L_{\max}$, indicating that averaging bandwidth measurements may not provide additional predictive power.
- Future analysis should explore other optical characteristics or external factors (e.g., material properties, environmental conditions) that might better explain$L_{\max}$ variations.
- Machine learning models could be used to integrate multiple characteristics simultaneously for improved prediction of Lmax.

5.5 BER vs Bandwidth

Sine the correlation analysis between the B_{-3dB}, B_{-5dB}, B_{-10dB}, and equivalent bandwidth B_{eq} with Lmax was not sufficient to determine a possible correlation which is reliable . So , it was useful to also explore a new set of correlations, specifically examining how the Bit Error Rate (BER) correlates with various bandwidth metrics—namely the -3dB, -5dB, -10dB, and Equivalent Bandwidth—at different lengths. This shift is motivated by the need to understand more deeply how data integrity and transmission efficacy are influenced across varying distances, particularly in the context of fiber optics where signal quality can significantly degrade over length.

BER vs -3 -5 -10 dB and equivalent bandwidth for laser 1 , $\lambda = 850nm$ at 400 m

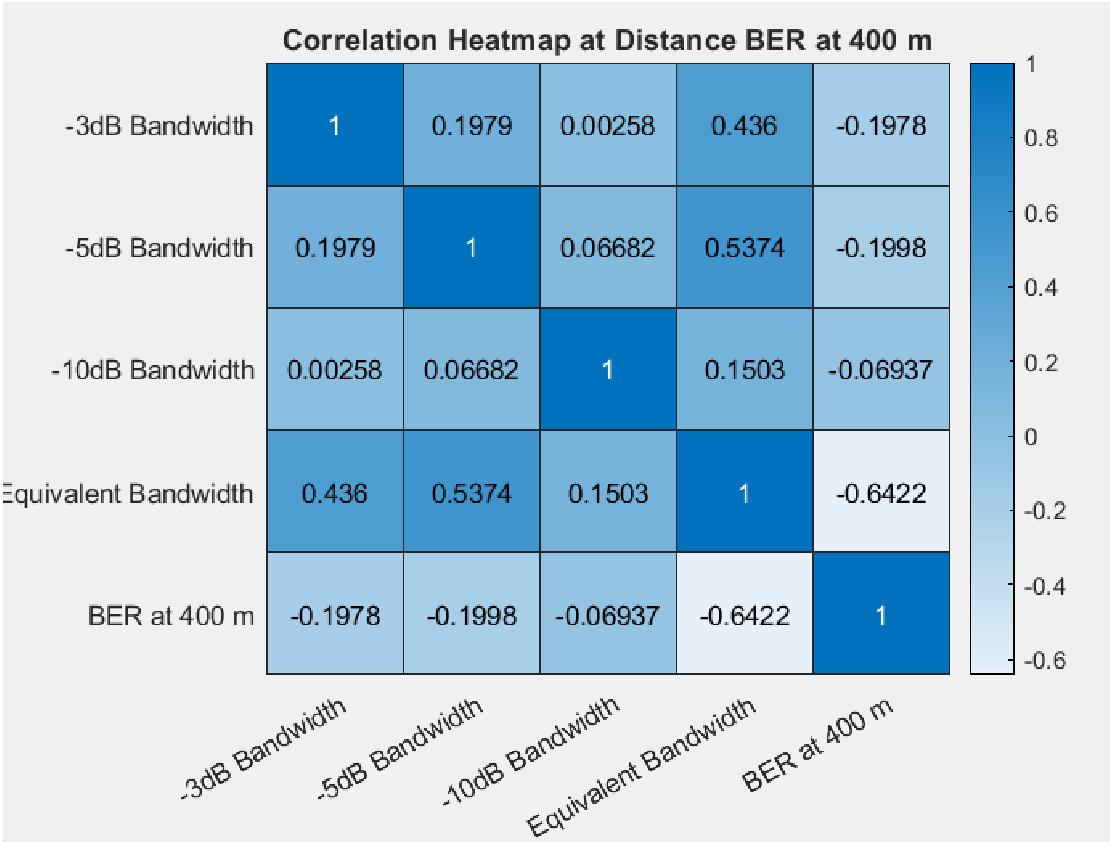


Figure 5.5: correlation at 400 m

Analysis and interpretation

The series of correlation heatmaps presented illustrate how the correlation between Bit Error Rate (BER) and various bandwidth metrics (-3dB, -5dB, -10dB, and Equivalent Bandwidth) varies over different distances from 30 meters to 400 meters. Here's a detailed analysis, focusing on how correlations change with distance and identifying any notable trends or shifts in the relationships:

Overview of Trends Across Distances:

- **Bandwidth Correlations:** The correlations between -3dB, -5dB, and -10dB bandwidths are consistently moderate to high across all distances, showing that these metrics are strongly related. The Equivalent Bandwidth maintains a steady moderate correlation with the other bandwidth metrics throughout the distance spectrum.

- **Bandwidth and BER Correlations:**
  - **30 to 100 meters:** Weak negative correlations between bandwidth metrics and BER are observed, with Equivalent Bandwidth showing the strongest correlation at -0.4415, indicating a slight improvement in BER with higher bandwidth.
  - **150 to 250 meters:** The negative correlation between Equivalent Bandwidth and BER becomes more pronounced, reaching -0.475 at 250 meters, suggesting that bandwidth has a stronger impact on BER over longer distances.
  - **300 to 400 meters:** The correlation stabilizes around -0.6422 at 400 meters, showing a more significant inverse relationship, implying that higher bandwidth increasingly helps reduce BER at these longer distances.

Key Observations:

- **Stronger Correlations Over Distance:** As distance increases, the negative correlation between bandwidth and BER becomes more pronounced, especially for Equivalent Bandwidth.

- **Consistency Among Bandwidth Metrics:** The stable correlations among -3dB, -5dB, and -10dB bandwidths suggest these are reliable transmission indicators, regardless of distance.

- **BER and Bandwidth Relationship:** Higher bandwidth has a stronger influence on improving BER at longer distances, where signal integrity is more challenging.

Conclusion:

The analysis confirms that bandwidth metrics, particularly Equivalent Bandwidth, have a stronger impact on reducing BER as transmission distance increases. This highlights the need to account for distance when optimizing optical communication systems. Graphs plotting the correlation between Equivalent Bandwidth and BER across all distances provide further clarity in understanding this relationship.

Then to make the interpretation easier , all the correlation values between the equivalent bandwidth and the BER at all the distances has been plotted in the graphs below :

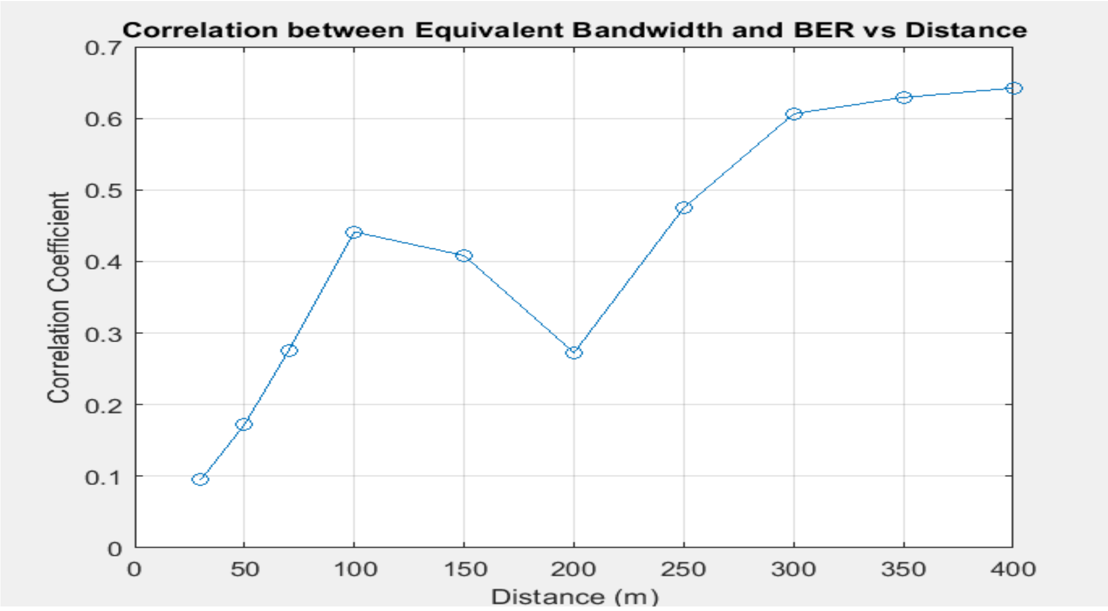


Figure 5.6: Absolute value of the correlation in 3766 cases for laser 1 and $\lambda = 850nm$

Analysis:

The plot shows the correlation between Equivalent Bandwidth and Bit Error Rate (BER) over distances from 30 to 400 meters, focusing only on positive correlations.

Analysis of the Positive Correlation Plot: - The correlation starts with a drop to 0.2 at 100 meters but rises steadily to 0.6 by 400 meters. - This trend suggests that as distance increases, the positive correlation between Equivalent Bandwidth and BER strengthens, indicating that higher bandwidth helps mitigate signal degradation and improves BER over longer distances.

Interpretation: - The plot shows that bandwidth plays an increasingly important role in reducing BER at longer distances. As distance grows, the correlation strengthens, reinforcing the significance of bandwidth in maintaining signal quality. - This analysis highlights the role of bandwidth in enhancing performance in long-distance fiber optic networks, suggesting that careful bandwidth management is critical in minimizing BER.

Next Phase:

Having focused on the positive correlation plot, the next phase will analyze a larger dataset of 120,125 cases with various wavelengths, lasers, and fibers. This expanded research will allow for a deeper understanding of the relationship between bandwidth and BER across different configurations, providing further insights for optimizing optical communication systems.

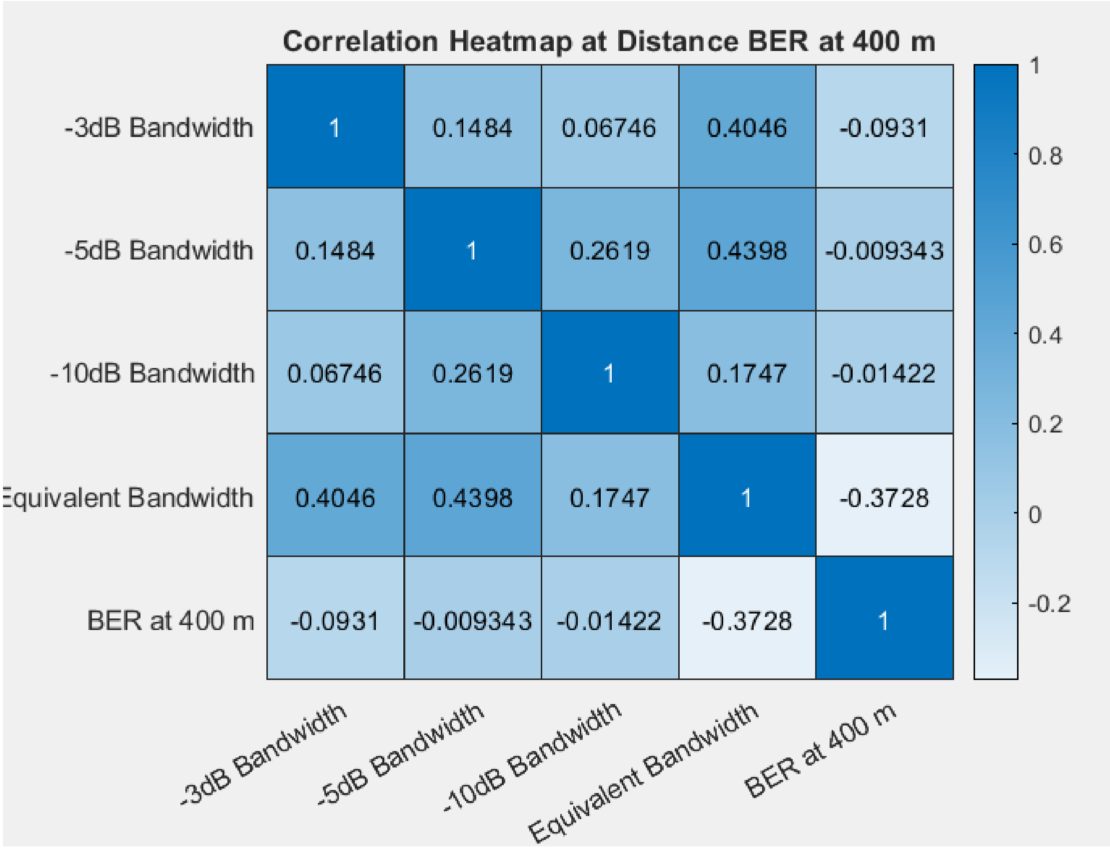


Figure 5.7: correlation at 400 m of 120512 case

Analysis and interpretation The series of correlation heatmaps provided explores the relationship between various bandwidth metrics (-3dB, -5dB, -10dB, and Equivalent Bandwidth) and Bit Error Rate (BER) at different distances (from 30 meters up to 400 meters). Here's an analysis of the correlation dynamics as distance increases, and how these might impact fiber optic communications:

General Trends Observed:

1. Stability in Bandwidth Metrics Correlation: - Across all distances, the internal correlations among the bandwidth metrics remain moderately positive. This consistency suggests that the fundamental properties of bandwidth as measured at different dB levels are relatively stable across various distances.

2. Correlation with BER: - At 30 meters: The correlations between bandwidth metrics and BER are all very weak, mostly hovering around zero. This indicates that at shorter distances, the influence of bandwidth on BER is minimal. - At 50 and 70 meters: As distance increases, there's a slight increase in the correlation values, but they remain weak. Notably, the Equivalent Bandwidth starts to show a slightly stronger negative correlation, suggesting that as distance increases, higher

bandwidth might begin to help reduce BER. - At 100 to 150 meters: The negative correlation between Equivalent Bandwidth and BER becomes more pronounced, reaching up to about -0.24. This trend suggests that over these intermediate distances, the effect of bandwidth on reducing BER becomes more apparent. - At 200 to 250 meters: This trend continues, with negative correlations becoming stronger. By 250 meters, the negative correlation with Equivalent Bandwidth shows notable improvement, indicating a significant relationship where higher bandwidth correlates with lower BER. - At 300 to 400 meters: The correlation between Equivalent Bandwidth and BER strengthens further, particularly at 400 meters where it reaches approximately -0.37. This shows that at longer distances, the role of bandwidth in influencing BER is crucial, and higher bandwidth can effectively reduce error rates.

Distance Impact on Correlation and Importance of Equivalent Bandwidth :

As the distance increases, the correlation between bandwidth metrics and BER strengthens, becoming more negative. This suggests that higher bandwidth offers better protection against BER at longer distances, likely mitigating the effects of increased signal degradation. Among the bandwidth metrics, Equivalent Bandwidth consistently shows the strongest negative correlation with BER, indicating its significance as a predictor of BER performance in fiber optic links. These findings imply that maintaining high Equivalent Bandwidth is crucial for minimizing BER in longer fiber optic links, enhancing reliability and efficiency in long-haul systems where signal integrity is critical.

Conclusion:

The correlation analysis across various distances provides valuable data on how bandwidth and BER are interrelated in optical fiber systems. As the distance increases, the role of bandwidth, especially the Equivalent Bandwidth, becomes increasingly vital in reducing BER. This relationship highlights the importance of optimizing bandwidth in the design and maintenance of fiber optic networks, particularly over longer distances where signal integrity challenges are greater.

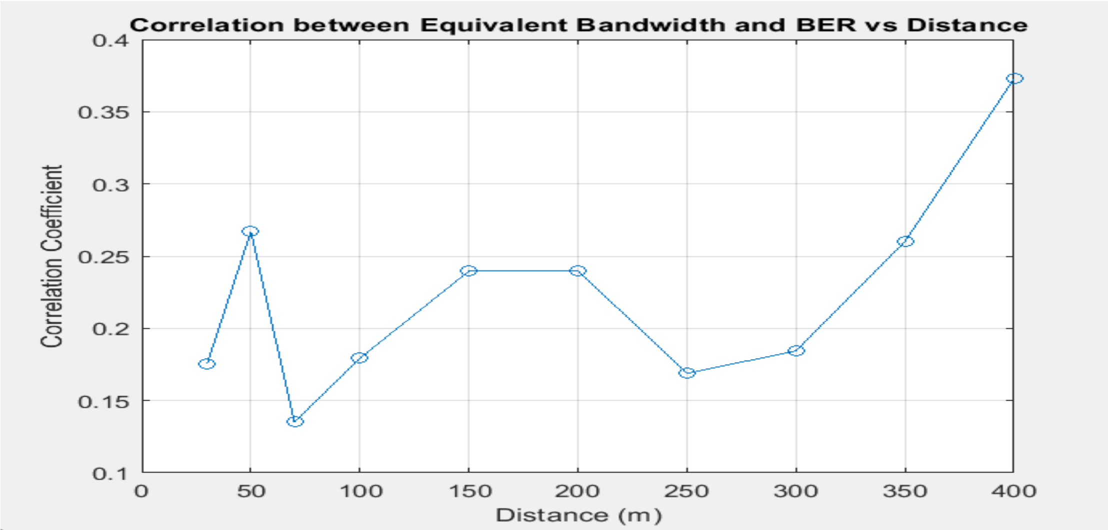


Figure 5.8: Positive correlation plot

5.6 Analysis

The plot shows the positive correlation between Equivalent Bandwidth and Bit Error Rate (BER) over distances ranging from 30 to 400 meters.

Key Observations:

- **Varying Peaks and Troughs:** The correlation fluctuates initially, peaking at 50 meters, dipping towards 100 meters, and stabilizing around 300 meters.
- **Strong Increase at Longer Distances:** Beyond 300 meters, the correlation strengthens significantly, reaching a peak at 400 meters. This indicates that higher bandwidth becomes more critical for maintaining BER at longer distances.

Implications:

- **Distance Matters:** As distance increases, the positive impact of Equivalent Bandwidth on BER becomes more evident, highlighting the importance of bandwidth management over long fiber optic links.
- **Optimization Potential:** The strong correlation at longer distances suggests opportunities for optimizing bandwidth to reduce BER and improve signal quality.

Conclusion: The plot demonstrates the importance of Equivalent Bandwidth in improving BER, especially at longer distances. Effective bandwidth management is essential for reducing BER in long-distance optical networks.

5.7 Comparison

This section compares the correlation between Equivalent Bandwidth and BER for two datasets: one limited (3766 fibers at fixed conditions) and one comprehensive (including all variations of lambda, laser, and fibers).

Limited Dataset:

- **Strong Negative Correlation:** The correlation starts strongly negative at -0.2 at 30 meters, quickly dropping to -0.6 at 50 meters.
- **Mid-Distance Recovery:** The correlation weakens around 100 to 150 meters before returning to a strong negative correlation, ending near -0.35 at 400 meters.

Comprehensive Dataset:

- **Less Extreme Initial Values:** The correlation starts less negative at -0.1 at 30 meters and decreases more gradually.
- **Stronger Decline at Longer Distances:** The correlation becomes more negative after 300 meters, reaching close to -0.7 at 400 meters.

Key Comparisons:

- **Fluctuations:** The limited dataset shows more variability, while the comprehensive dataset provides a smoother curve, reflecting a broader range of conditions.
- **Correlation and Distance:** In both datasets, the correlation becomes more negative with distance, but the comprehensive dataset shows a stronger correlation at longer distances, emphasizing the impact of distance on signal quality.

Conclusion: Both datasets confirm that distance negatively impacts BER and that bandwidth is crucial in mitigating this. The comprehensive dataset offers a broader and more reliable view of network behavior, making it useful for general network design and optimization. Future research should further explore specific fiber and wavelength configurations to improve network performance.

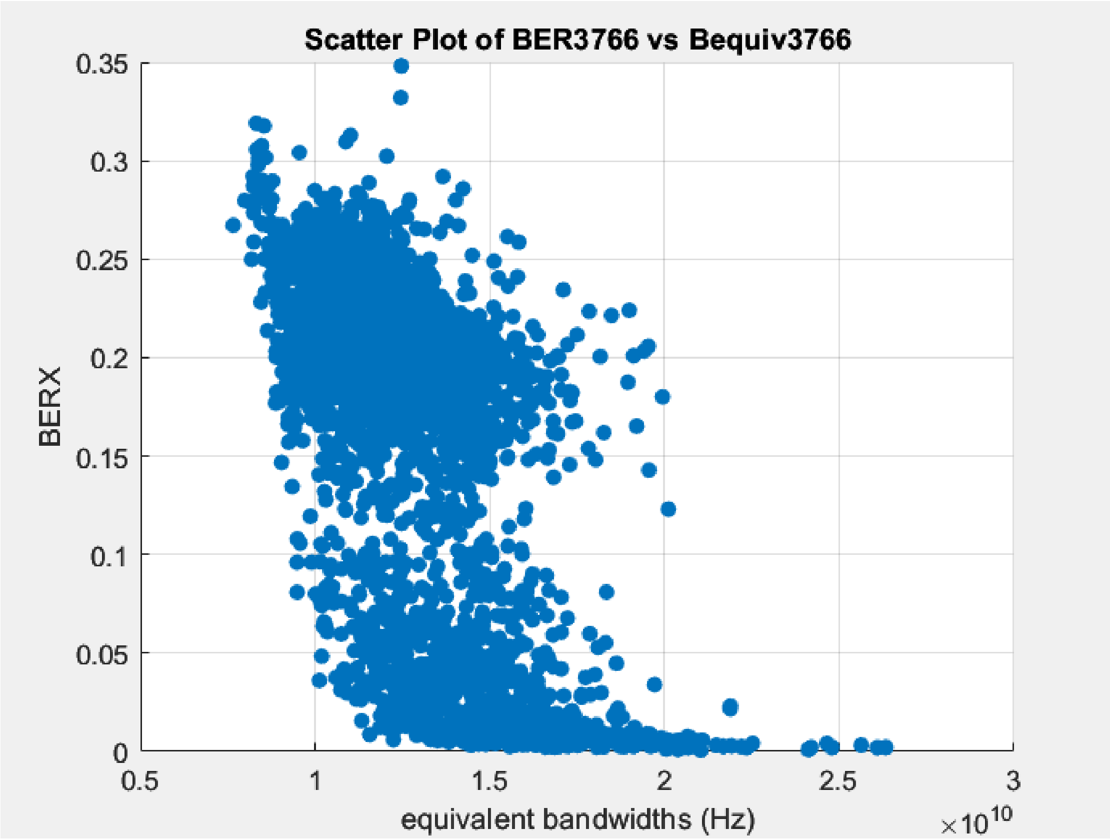


Figure 5.9: Scatter plot of Bequivalent3766 vs BER3766

This scatter plot shows the relationship between the equivalent bandwidth of 3766 fibers and the BER .

Analysis of Scatter Plot

- **Data Spread:** - The x-axis (equivalent bandwidth) ranges from 0.5 to 3×10^{10} Hz, showing a wide range of bandwidths across the 3766 fibers. - The y-axis (BER) ranges from 0 to 0.35, with most data points concentrated at lower BER values, indicating better performance.
- **Relationship between Variables:** - As equivalent bandwidth increases, BER remains low (below 0.15) until around 2×10^{10} Hz, after which a wider spread of BER values appears. - The increased variability in BER at higher bandwidths suggests that maintaining a low error rate becomes more difficult due to factors like noise or signal degradation.
- **Correlation Interpretation:** - The plot suggests a complex relationship between bandwidth and BER, without a simple linear trend. Other factors,

such as fiber quality or system configurations, likely influence BER beyond just bandwidth.

Conclusion: The scatter plot shows valuable insights into the relationship between bandwidth and BER, but the observed complexity indicates that a multi-variable approach is needed to fully optimize fiber optic network performance.

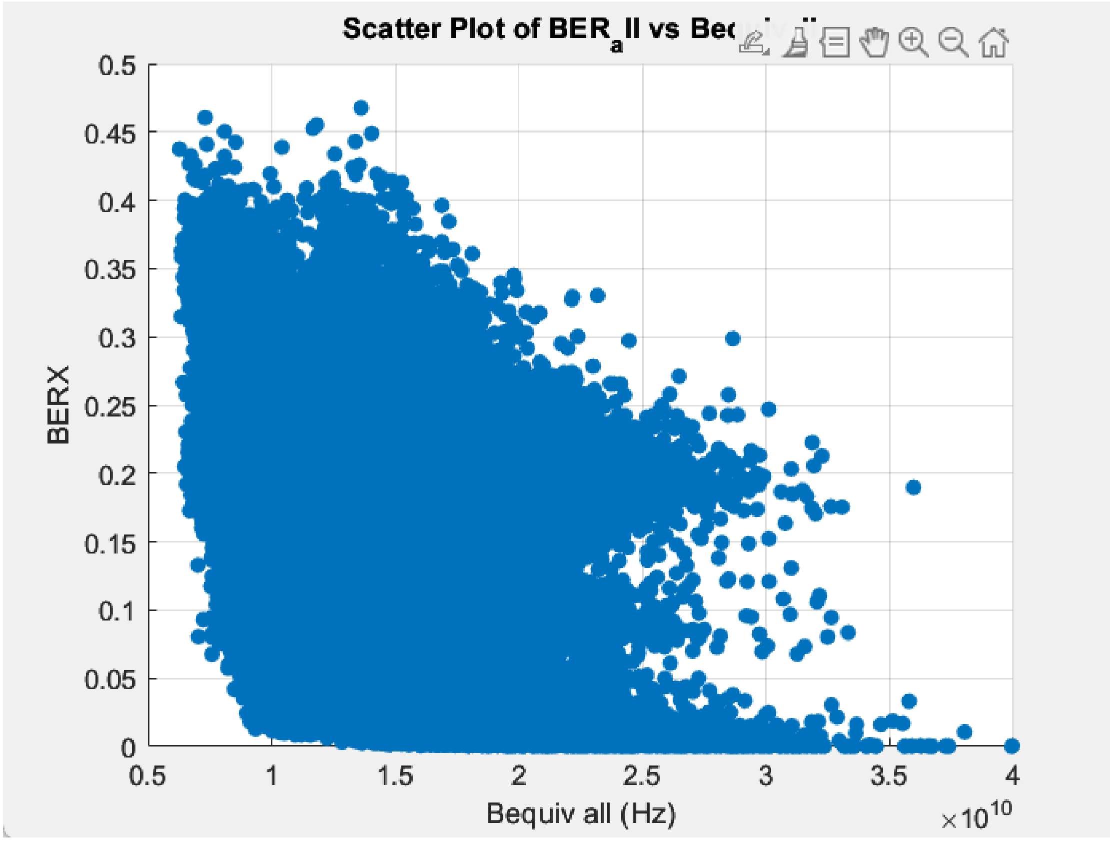


Figure 5.10: Scatter plot of B equivalent vs BER all

Analysis of Scatter Plot

This scatter plot illustrates the relationship between Bit Error Rate (BER) at 400 meters and equivalent bandwidth for all the cases in the data set which are $4(wavelength) \times 8(lasers) \times 3766(fiber)$ present in the dataset.

- **Horizontal Axis:** The x-axis represents equivalent bandwidth (in Hz), ranging from 0.5 to 4×10^{10} Hz, indicating diverse bandwidth conditions across fibers.
- **Vertical Axis:** The y-axis represents BER, ranging from 0 to 0.5, with increasing variability as bandwidth increases.

Data Distribution:

- **Lower Bandwidth:** At low bandwidth values (around 0.5×10^{10} Hz), BER is tightly clustered at lower values, suggesting better performance.
- **Middle Bandwidth:** At around 2.5×10^{10} Hz, BER begins to spread out, with most values below 0.3 but showing more variation.
- **Higher Bandwidth:** At higher bandwidths, BER spreads even further, with some values reaching 0.5, suggesting that very high bandwidths may lead to increased error rates.

Interpretation

- **Correlation Observation:** The plot suggests a non-linear relationship, where increasing bandwidth beyond a certain point does not improve BER and may even increase it. This indicates that factors like fiber quality or system noise play a role.
- **System Limitations:** The increased spread of BER at higher bandwidths may point to system limitations, such as component imperfections or difficulties maintaining signal integrity at high frequencies.
- **Optimal Bandwidth Range:** There may be an optimal bandwidth range where BER is minimized, offering a useful target for system design to balance speed and reliability.

Conclusion

While higher bandwidth typically increases data rates, this plot suggests a complex interaction with BER, especially at higher bandwidths. Future research should focus on identifying the causes of increased BER at high bandwidths and finding solutions, such as improved fiber types, signal processing techniques, or error correction, to enhance high-speed optical communication systems.

5.8 Further Analysis

After that we didnt catch any interesting correlation between the BER and the specidic bandwidths that we worked on , we moved the study to a new approach , which is to catch the relationship between Lmax , the maximum achievable reach of a fiber and the bandwidth.

Lmax vs Equivalent Bandwidth analysis

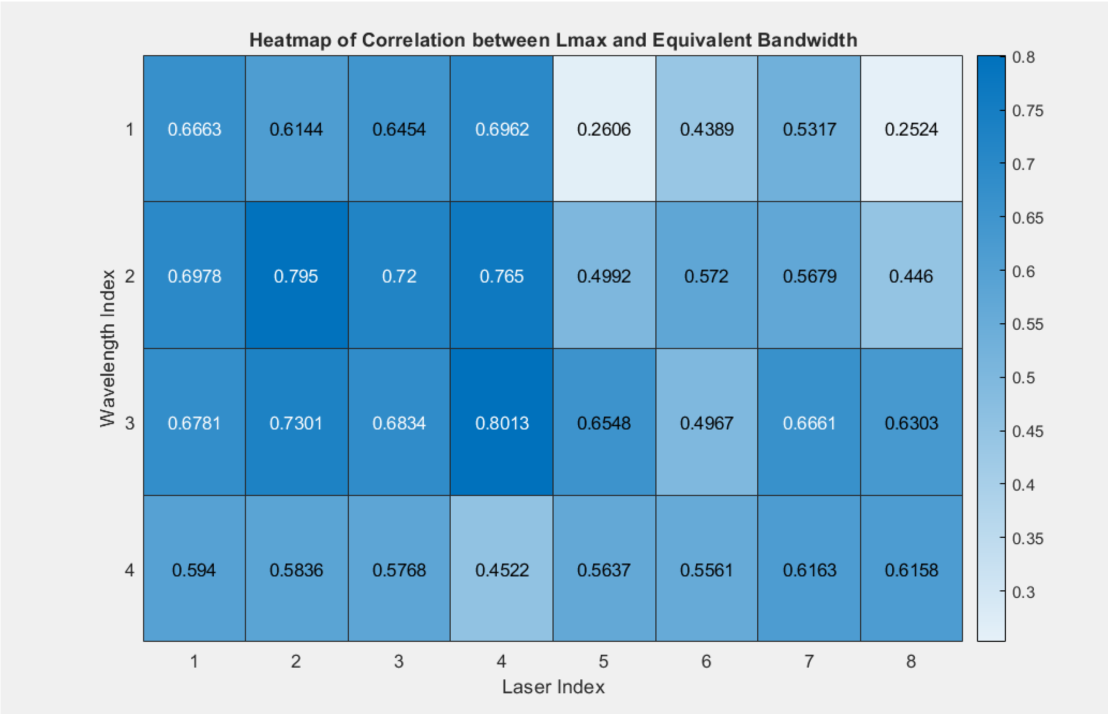


Figure 5.11: Lmax vs Beq correlation

This heatmap illustrates the correlation between maximum achievable length (Lmax) and equivalent bandwidth across different wavelength and laser combinations in a fiber optics setup. Each cell in the heatmap corresponds to a specific wavelength and laser combination, with wavelengths denoted by their indices along the vertical axis (1-4, corresponding to 850 nm, 880 nm, 910 nm, and 940 nm respectively) and laser settings denoted along the horizontal axis (1-8).

Analysis and Interpretation

General Trends: - The correlation values vary widely across different combinations of wavelength and laser, suggesting that the interaction between these parameters significantly affects the relationship between Lmax and equivalent bandwidth. - Higher correlation values indicate a stronger linear relationship between the maximum length and bandwidth, implying that for these settings, as bandwidth

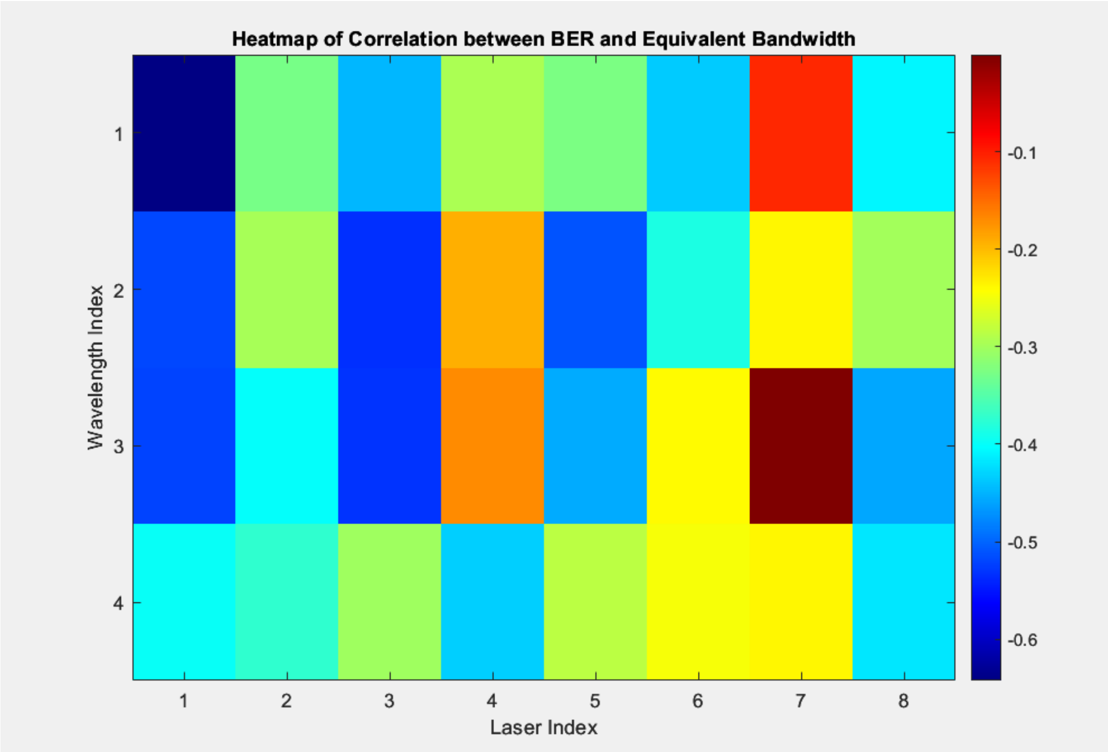


Figure 5.12: Lmax vs Beq correlation - heat map

increases, the achievable distance before reaching a BER threshold also tends to increase.

Specific Observations:

- **Wavelength 1 (850 nm):** Shows generally moderate to high correlations, with values like 0.6663 and 0.6962, suggesting a strong linear relationship in some laser settings. This might imply that at this wavelength, bandwidth is a good predictor of Lmax for certain lasers.
- **Wavelength 2 (880 nm):** Displays the highest correlation values across all wavelengths with a peak at laser 2 (0.795). This suggests that this particular wavelength and laser setting combination might be optimal for predicting Lmax based on bandwidth.
- **Wavelength 3 (910 nm):** Also shows relatively high correlations, especially with laser settings 4 and 7 (0.8013 and 0.6661, respectively), indicating robust predictive power of bandwidth over maximum length.
- **Wavelength 4 (940 nm):** The correlation values tend to be lower compared to other wavelengths, particularly with laser setting 4 (0.4522), indicating that

at this wavelength, bandwidth might not be as effective a predictor of Lmax.

Implications:

- **Optimization:** For systems where wavelength and laser settings can be chosen, configurations like Wavelength 2 with Laser 2 may offer the most reliable operation based on bandwidth capabilities.

- **Design Considerations:** Understanding these correlations can help in designing systems that maximize transmission distance while maintaining desired BER levels. Engineers can select wavelengths and laser combinations that correlate strongly to optimize system performance.

- **Predictive Modeling:** These correlations could be utilized to develop predictive models that estimate Lmax based on bandwidth measurements, potentially simplifying system design and testing.

Limitations:

- **Linearity Assumption:** The correlations presented assume a linear relationship; however, the actual relationship might be non-linear, especially at higher data rates or longer distances.

- **Variable Dependencies:** The correlations might also depend on other variables not considered here, such as fiber type, environmental conditions, and system noise, which could affect the practical applicability of these results.

This analysis underscores the importance of considering wavelength and laser settings in designing and optimizing fiber optical systems for data transmission over varying distances. Further investigation with more complex models might provide deeper insights into these relationships.

5.9 BER vs Equivalent bandwidth analysis

5.10 Further Analysis

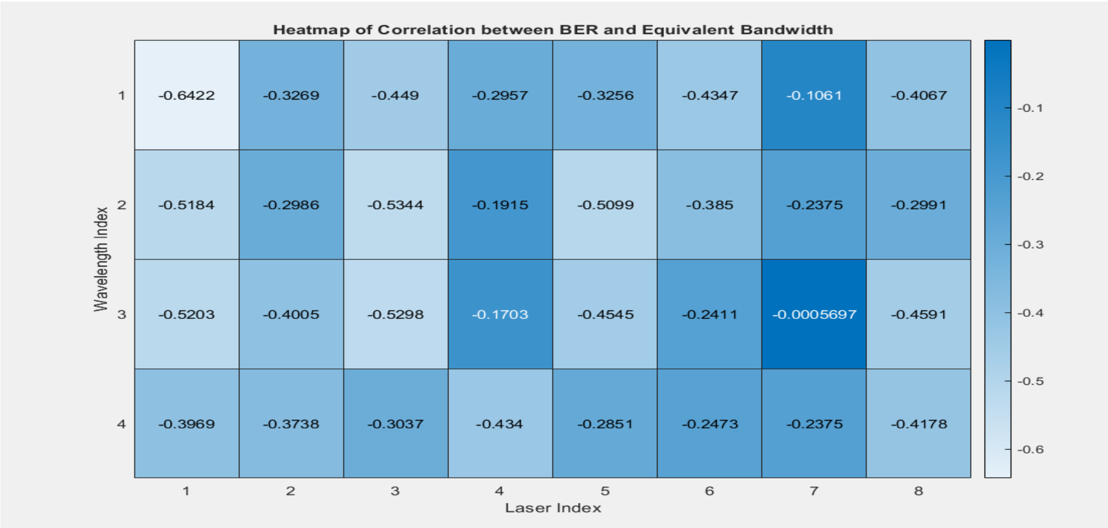


Figure 5.13: BER vs Beq correlation

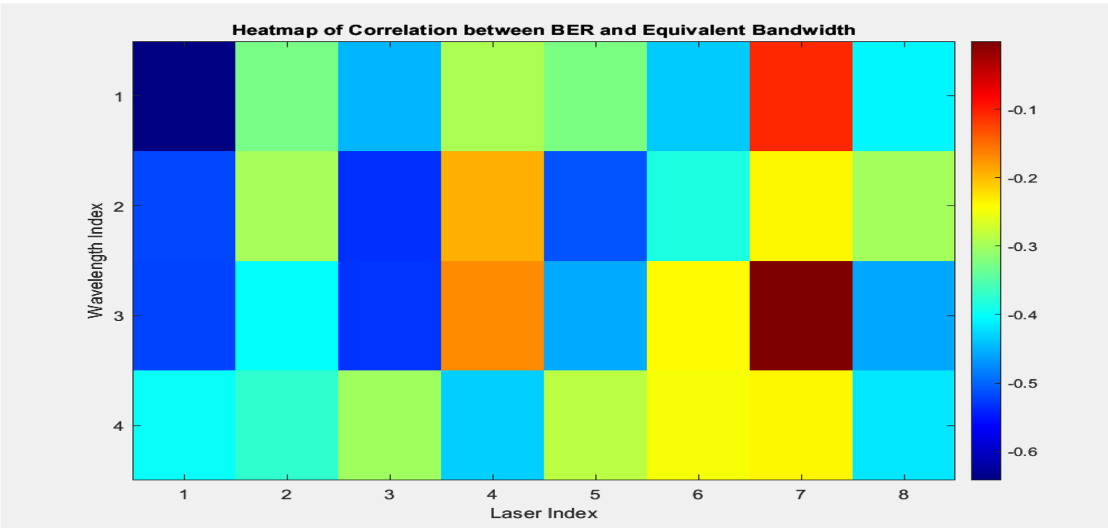


Figure 5.14: BER vs Beq correlation - heat map

The heatmap above represents the correlation coefficients between Bit Error Rate (BER) and Equivalent Bandwidth for different combinations of wavelengths and lasers in an optical fiber communication system. Each cell represents a specific pairing of wavelength (vertical axis: 1-4, corresponding to 850 nm, 880 nm, 910 nm, and 940 nm respectively) and laser (horizontal axis: 1-8). The color scale indicates the strength and direction of the correlation, with darker blue tones representing negative correlations, closer to -1.

Analysis and Interpretation:

General Observations: - All correlation values are negative, indicating an inverse relationship between BER and equivalent bandwidth across all tested wavelengths and laser configurations. This suggests that as the equivalent bandwidth increases, the BER tends to decrease, which is typical in communication systems where higher bandwidth allows for more efficient data transmission and error handling capabilities. - The magnitude of these negative correlations varies, suggesting that the strength of the inverse relationship between BER and bandwidth differs depending on the specific laser and wavelength configuration.

Specific Insights by Wavelength:

- **Wavelength 1 (850 nm):** Shows a range of correlations from moderate to strong negative values, suggesting that for this wavelength, increases in bandwidth are generally associated with significant improvements in BER.

- **Wavelength 2 (880 nm):** Also displays moderate to strong negative correlations. Particularly, some laser settings like laser 2 show stronger correlations (e.g., -0.5344), indicating that at this wavelength, bandwidth is a reliable indicator of BER performance.

- **Wavelength 3 (910 nm):** This wavelength exhibits similar patterns, with some settings (e.g., laser 3) demonstrating stronger negative correlations, suggesting effective bandwidth utilization in reducing BER.

- **Wavelength 4 (940 nm):** Generally, this wavelength shows weaker negative correlations compared to the others, indicating that while there is still an inverse relationship, the effect of bandwidth on BER is less pronounced at this wavelength.

Conclusion: The analysis highlights the crucial role of bandwidth in managing BER across various laser and wavelength settings in optical fiber systems. Systems engineers and network planners can use these insights to enhance the design and operation of fiber optic networks, ensuring more reliable and efficient data transmission.

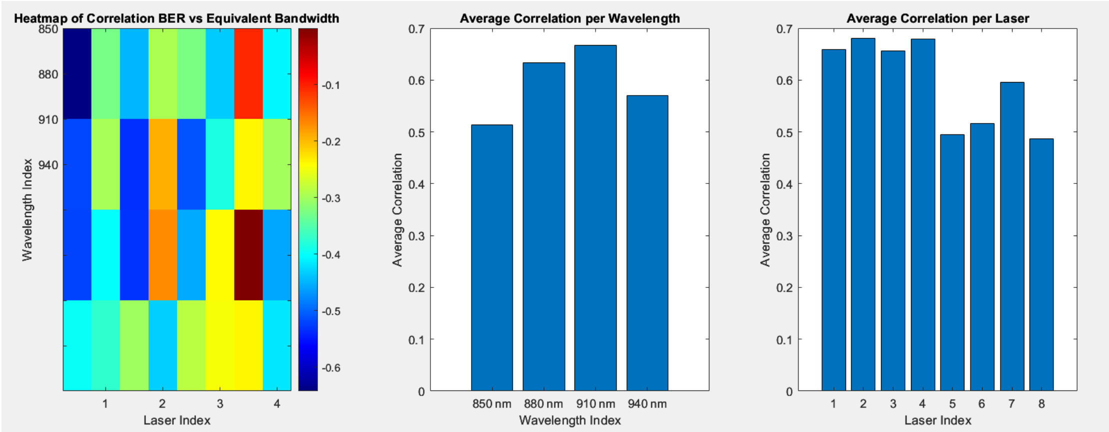


Figure 5.15: BER vs B EQ all.png

This visualization package includes a heatmap depicting the negative correlation between Bit Error Rate (BER) and Equivalent Bandwidth across different wavelengths and lasers, accompanied by bar charts showing the average negative correlations for each wavelength and laser.

Heatmap Analysis:

- **Overall Negative Correlation:** The heatmap prominently displays negative correlations across almost all wavelength and laser combinations. This suggests that as the Equivalent Bandwidth increases, the BER typically decreases, indicating better performance.
- **Variability Across Settings:** Some combinations, particularly at certain wavelengths like 910 nm with laser index 1, show deeper reds, indicating stronger negative correlations. Conversely, some areas such as the 940 nm wavelength with laser index 3 show milder negative correlations (lighter colors).

Average Correlation per Wavelength:

- **Trend Across Wavelengths:** The correlation appears to be generally stronger at lower wavelengths (850 nm and 880 nm) and decreases slightly as the wavelength increases. This trend suggests that the shorter wavelengths may be more susceptible to variations in bandwidth in influencing BER.
- **Possible Optical Characteristics:** The stronger negative correlation at shorter wavelengths might be due to the optical properties of the fibers and the interaction with these specific wavelengths, which could be more effectively managed through bandwidth adjustments.

Average Correlation per Laser: - **Variability Among Lasers:** The correlations are relatively high for the first four lasers, indicating a consistent inverse relationship between BER and bandwidth. However, there is a noticeable drop in the correlation strength for lasers 5 and 6, followed by a partial recovery in lasers 7 and 8. This pattern may indicate differences in how each laser's characteristics interact with the optical fiber properties, affecting how bandwidth influences BER.

Interpretations and Implications:

- **Impact of Bandwidth on BER:** The negative correlations underscore the importance of managing bandwidth to optimize BER, particularly in certain wavelength and laser setups. This could have practical implications in settings where precise control over bandwidth can lead to significant improvements in signal integrity and overall system performance.
- **Design and Optimization:** For optical communication system designers, understanding which laser and wavelength combinations are more sensitive to bandwidth changes can help in optimizing system configurations for improved BER. This insight is crucial for designing systems that need to operate over varying distances and conditions, as it helps in selecting the right components and settings for specific operational requirements.
- **Further Investigation Needed:** The variations in correlation strength across different setups suggest that further detailed studies could be beneficial. These studies could explore the underlying physical or technical reasons for these variations, potentially leading to more refined strategies for managing bandwidth and BER in fiber optic communications.

Overall, these insights could guide enhancements in fiber optic systems, emphasizing the critical role of managing bandwidth to minimize BER, especially in applications where data integrity is paramount.

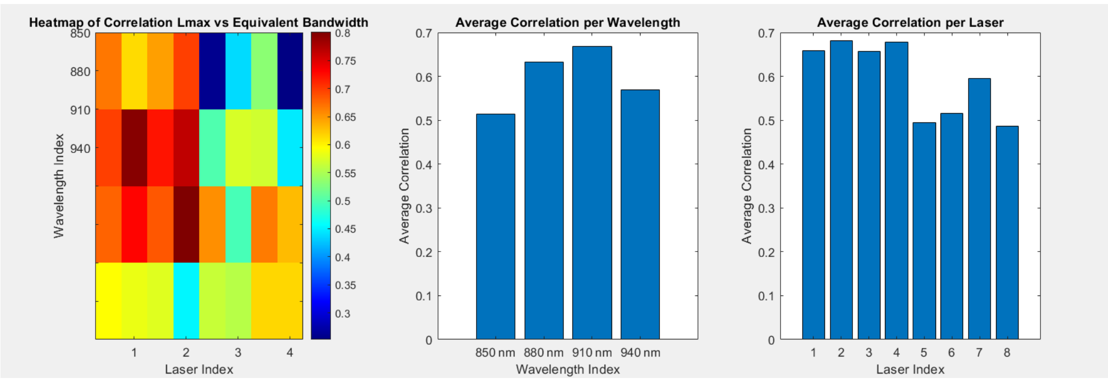


Figure 5.16: Lmax vs B EQ all

This visualization consists of three components: a heatmap showing the correlation between Lmax and Equivalent Bandwidth for different combinations of wavelength and laser indices, and two bar charts representing the average correlations across all lasers for each wavelength and across all wavelengths for each laser, respectively.

Heatmap Analysis: -Variability Across Combinations: The heatmap indicates varying degrees of correlation between Lmax and Equivalent Bandwidth across different laser and wavelength combinations. Some cells show high correlation values (close to 0.8), indicating a strong linear relationship between Lmax and bandwidth, while other cells show significantly lower correlations. **High and Low Correlation Regions:**

- **High Correlation:** The highest correlations are observed predominantly with certain laser settings across various wavelengths, suggesting that specific laser settings might be more consistent in their interaction with the optical fibers over these wavelength ranges.
- **Low Correlation:** Notably, some laser and wavelength combinations, especially those in laser index 5 and partially in 6, exhibit lower correlations. This might imply optical characteristics or interactions that are less predictable or are influenced by other factors not directly related to the bandwidth measurements.

Average Correlation per Wavelength: - Decreasing Trend: The bar chart indicates a generally high correlation across all wavelengths, with slight decreases as the wavelength increases from 850 nm to 940 nm. This trend might suggest that shorter wavelengths interact with the fiber characteristics in a way that more consistently correlates with Lmax compared to longer wavelengths.

Average Correlation per Laser: - Variability Among Lasers: The correlation values are relatively high for most lasers but show a dip around lasers 5 through 7. This suggests some lasers might be better at maintaining consistent relationships between bandwidth and maximum reach than others. It could be due to specific properties of the lasers or their settings.

5.11 Detailed Analysis

5.11.1 All data on 30 m :

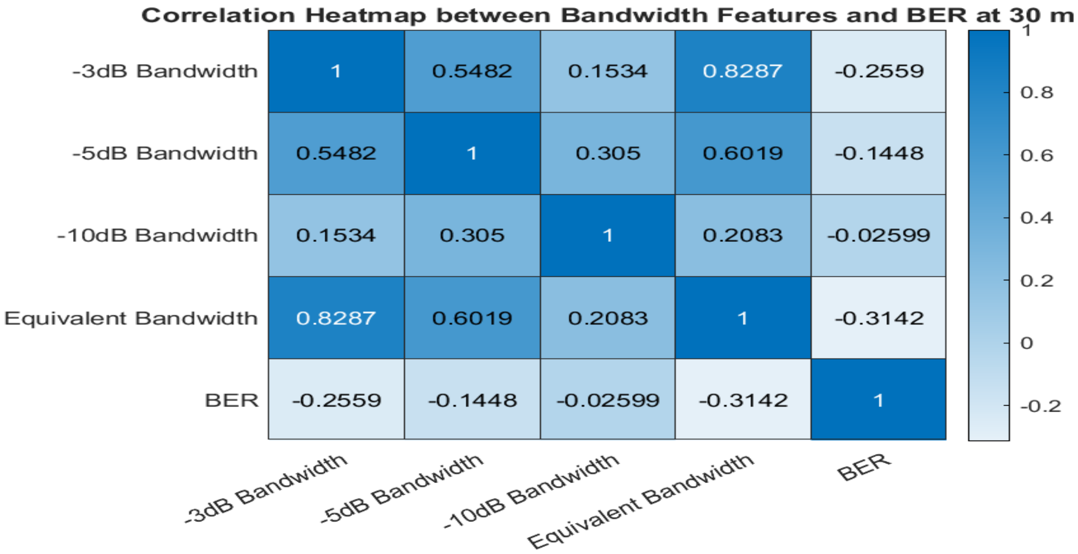


Figure 5.17: correlation heatmap between Bandwidth features and BER at 30 m

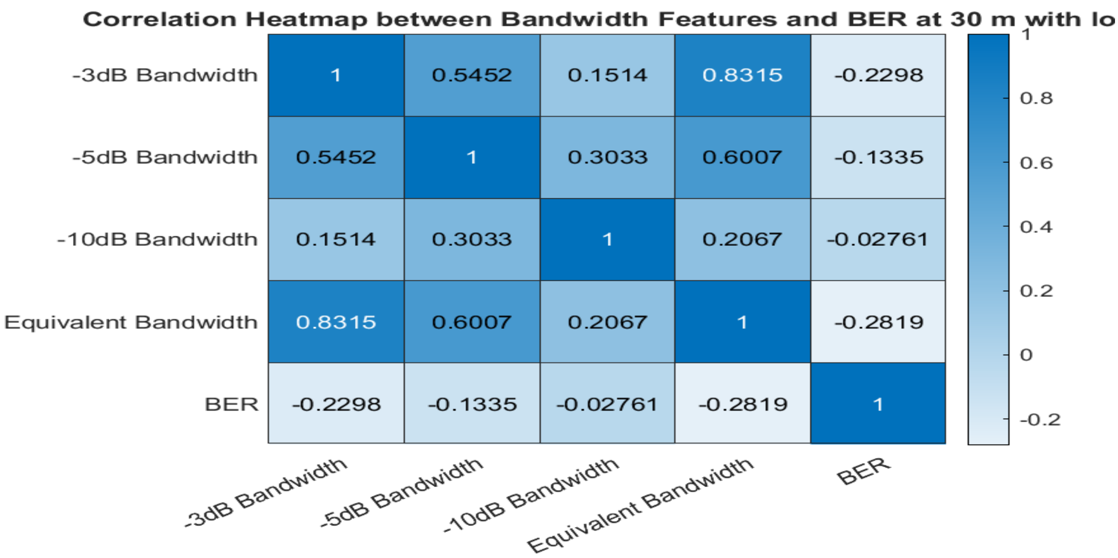


Figure 5.18: correlation heatmap between Bandwidth features and BER at 30 m with log scale

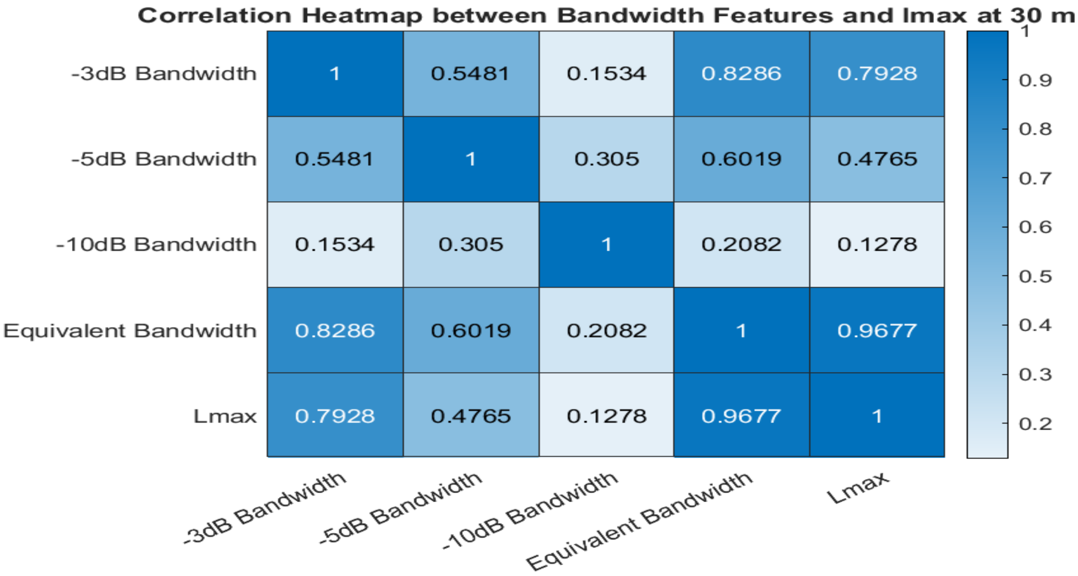


Figure 5.19: correlation heatmap between Bandwidth features and Lmax at 30 m

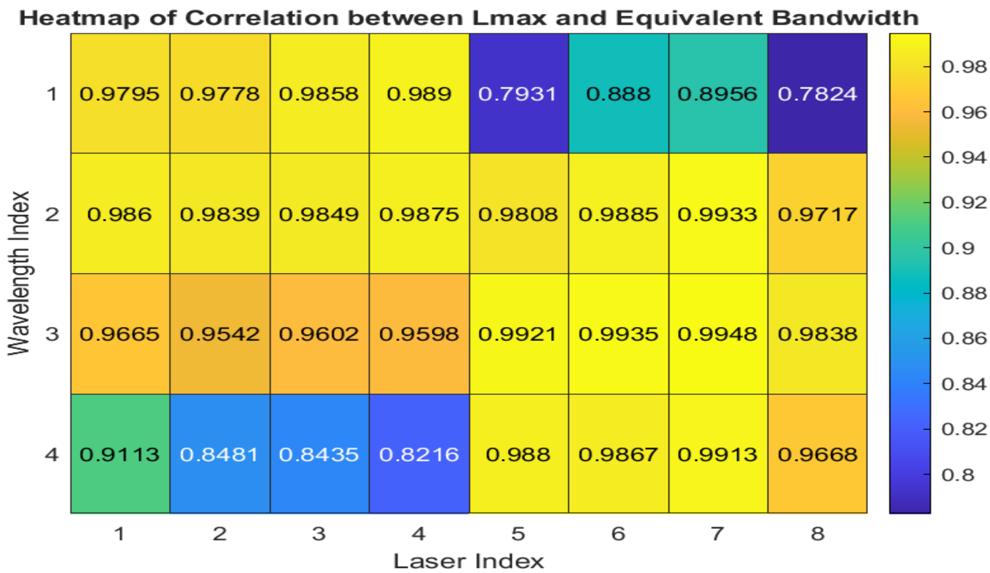


Figure 5.20: correlation heatmap between Lmax and the Equivalent Bnadwith for each laser indices 1 to 8 and wavelength λ = 850,880,910,940 nm showed as indice 1 to 4 at 30 meters

5.12 Summary and Analysis

Correlation Heatmap Between Bandwidth Features and BER at 30 meters:

- The equivalent bandwidth shows a strong positive correlation with -3dB Bandwidth (0.8287).
- BER negatively correlates with all bandwidth features, with the strongest negative correlation seen with equivalent bandwidth (-0.3142).

Correlation Heatmap Between Bandwidth Features and BER at 30 meters (Log Scale):

- The positive correlation between equivalent bandwidth and -3dB Bandwidth remains high (0.8315).
- BER's negative correlation with equivalent bandwidth weakens slightly (-0.2819).

Heatmap of Correlation Between BER and Equivalent Bandwidth (Across Wavelengths and Lasers):

- The heatmap shows a broader correlation across different wavelengths and lasers.
- Strong correlations are observed, especially at the first wavelength, where equivalent bandwidth and -3dB Bandwidth maintain high correlations.

Correlation Heatmap Between Bandwidth Features and Lmax at 30 meters:

- The equivalent bandwidth shows a very high correlation with Lmax (0.9677).
- As with BER, the equivalent bandwidth maintains strong correlations with other bandwidth features.

Heatmap of Correlation Between Lmax and Equivalent Bandwidth (Across Wavelengths and Lasers):

- Strong positive correlations are consistently observed between equivalent bandwidth and Lmax across different wavelengths and lasers.

5.13 Key Insights

Equivalent Bandwidth as a Key Metric:

- Equivalent bandwidth consistently shows strong correlations with both BER and Lmax, making it a crucial metric for analyzing system behavior.

Negative Correlation of BER with Bandwidth Features:

- BER negatively correlates with bandwidth features, suggesting that increasing bandwidth improves performance by reducing BER.

Consistency Across Wavelengths and Lasers:

- The correlation between equivalent bandwidth and other metrics (BER and Lmax) is consistent across wavelengths and laser indices, indicating reliable predictive behavior.

5.14 Recommendations

Focus on Equivalent Bandwidth:

- Given its strong correlation with key performance metrics, equivalent bandwidth should be prioritized in system optimization and analysis.

Model Development:

- Predictive models should be developed using equivalent bandwidth as a core feature to enhance system efficien

5.14.1 All data on 50 m :

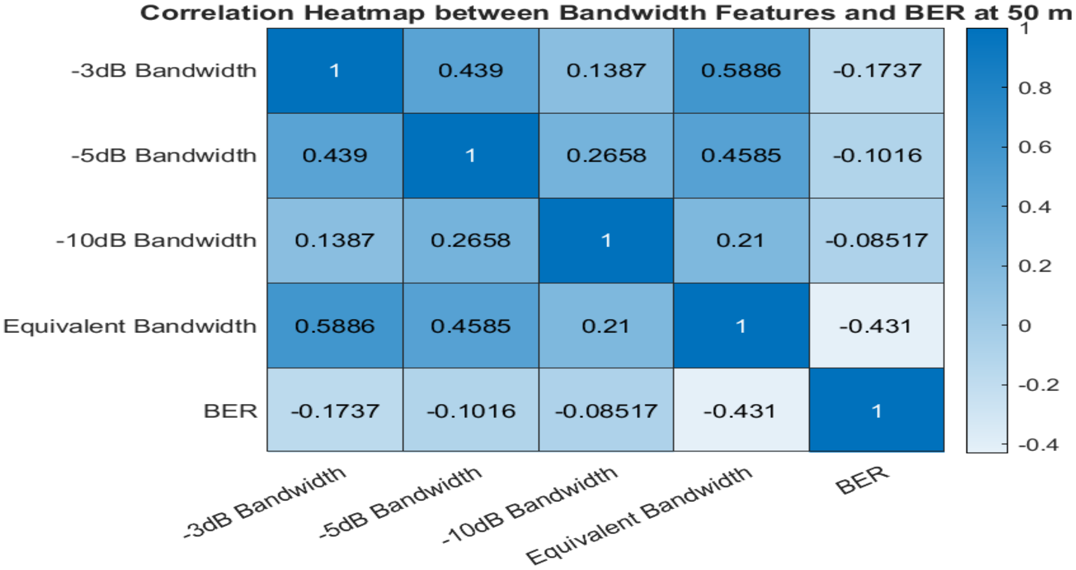


Figure 5.21: Correlation Heatmap Between Bandwidth Features and BER at 50 meters

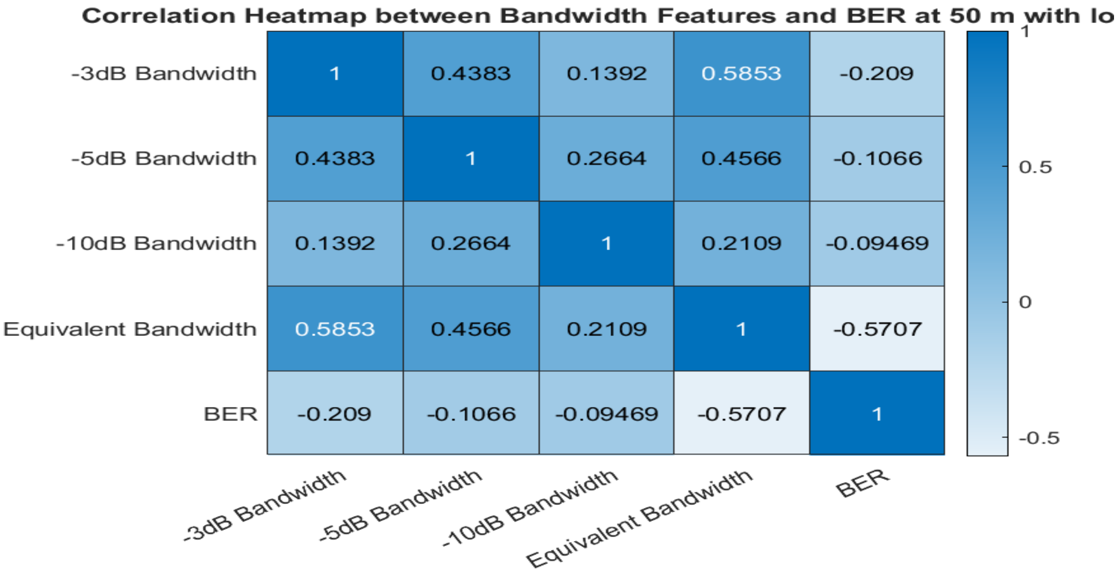


Figure 5.22: Correlation Heatmap Between Bandwidth Features and BER at 50 meters (Log Scale)

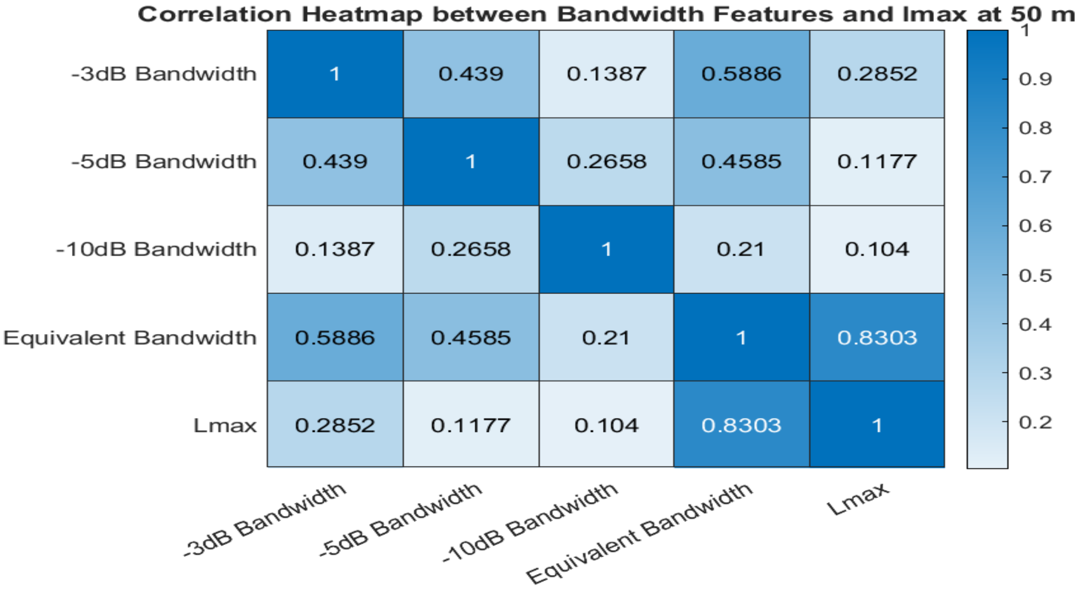


Figure 5.23: Correlation Heatmap Between Bandwidth Features and Lmax at 50 meters

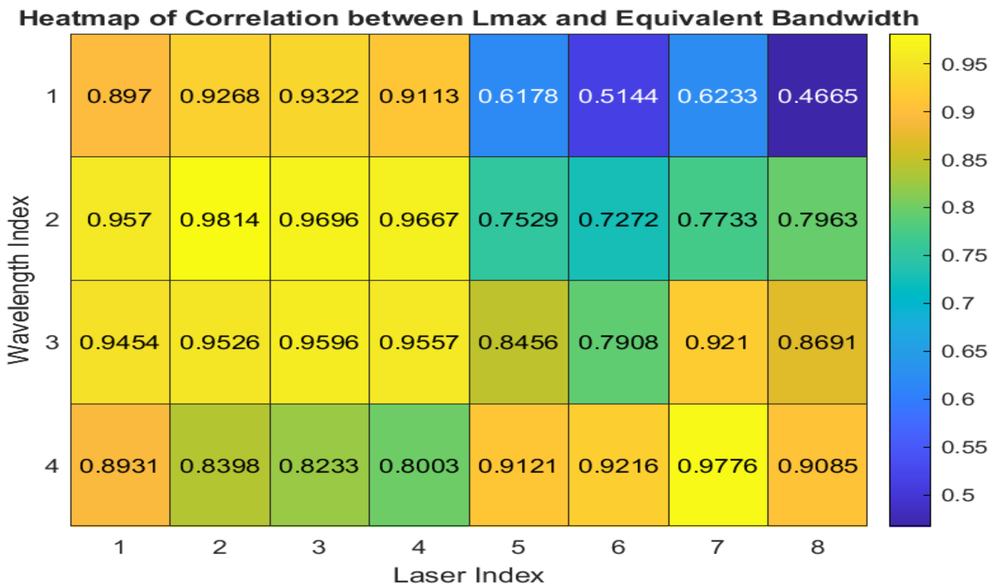


Figure 5.24: correlation heatmap between Lmax and the Equivalent Bnadwith for each laser indices 1 to 8 and wavelength λ = 850,880,910,940 nm showed as indice 1 to 4 at 50 meters

5.15 Analysis of the Correlation Matrices at 50m

1. Correlation Heatmap between Bandwidth Features and BER at 50m:

- **Equivalent Bandwidth vs BER:** Shows the strongest negative correlation (-0.431), indicating that as equivalent bandwidth increases, BER decreases, improving system performance.
- **-3dB Bandwidth vs BER:** Weaker negative correlation (-0.1737) compared to equivalent bandwidth.
- **Overall Correlations:** Equivalent bandwidth shows moderate to strong correlations with other bandwidth features, especially -3dB bandwidth (0.5886).

2. Correlation Heatmap between Bandwidth Features and log10(BER) at 50m:

- **Equivalent Bandwidth vs log10(BER):** Stronger correlation (-0.5707) than with BER, suggesting log-transformed BER offers clearer insights.
- **-3dB Bandwidth vs log10(BER):** Shows a moderate negative correlation (-0.209), slightly stronger than with BER.
- **Overall Correlations:** Equivalent bandwidth remains the feature with the strongest correlation to log10(BER).

3. Heatmap of Correlation between BER and Equivalent Bandwidth (Across Wavelengths and Lasers):

- **Wavelength and Laser Variation:** Strong correlations are observed across different wavelengths and lasers, with equivalent bandwidth consistently correlating with BER.
- **Key Observations:** The strongest correlations occur for certain wavelength-laser combinations, highlighting areas for bandwidth improvements.

4. Correlation Heatmap between Bandwidth Features and Lmax at 50m:

- **Equivalent Bandwidth vs Lmax:** Shows a very strong positive correlation (0.8303), suggesting higher equivalent bandwidth is linked to higher Lmax values.
- **-3dB Bandwidth vs Lmax:** Weaker positive correlation (0.2852).
- **Overall Correlations:** Equivalent bandwidth is a critical factor for maximizing Lmax.

5. Heatmap of Correlation between Lmax and Equivalent Bandwidth (Across Wavelengths and Lasers):

- **Wavelength and Laser Variation:** Strong correlations persist across different wavelengths and lasers, indicating the robustness of equivalent bandwidth as a predictive feature.

- **Key Observations:** The strongest correlations appear for specific wavelength-laser combinations, highlighting optimal conditions for maximizing Lmax.

5.16 Overall

- **Equivalent Bandwidth as a Critical Metric:** It consistently shows strong correlations with both BER and Lmax, making it a crucial parameter for performance optimization.

- **Improvement in BER and Lmax:** Increasing equivalent bandwidth could improve both BER and Lmax, enhancing system performance.

- **Consistency Across Conditions:** The strong correlations across different wavelengths and lasers suggest that equivalent bandwidth is a reliable predictor of system behavior.

- **Log Transformation of BER:** Log10(BER) offers a slightly clearer relationship with bandwidth features, suggesting its usefulness for further analysis.

- **Focus on Optimal Combinations:** Identifying the wavelength and laser combinations with the strongest correlations can guide targeted optimizations for performance improvements.

5.17 At 70 m

After noticing that the most important and significant corrrolation is between the Lmax and the equivalent bandwidth let's focus on it :

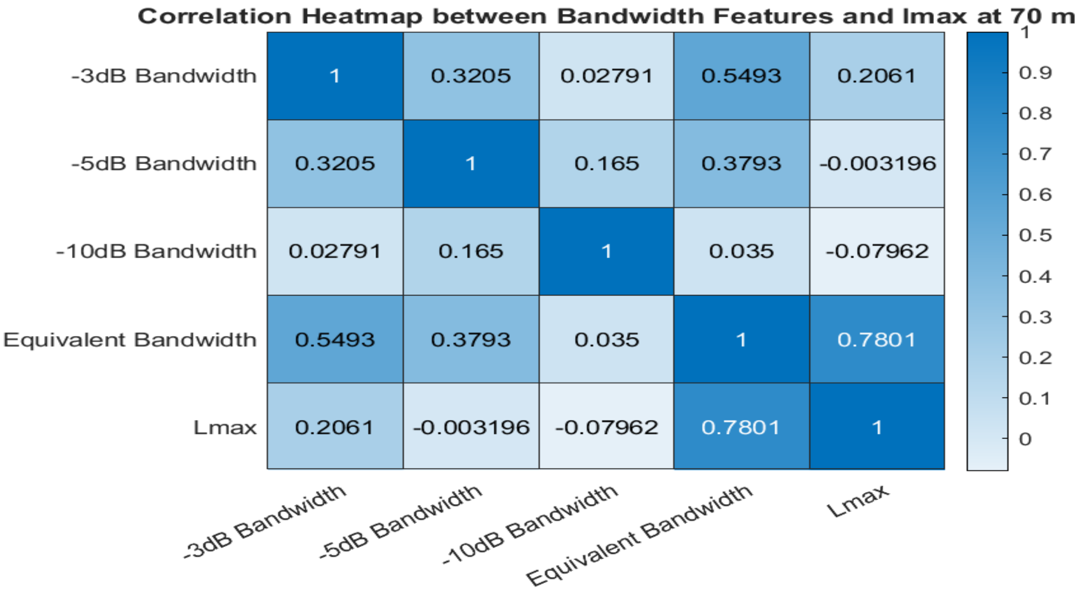


Figure 5.25: correlation heatmap between Bandwidth features and Lmax at 70 meters

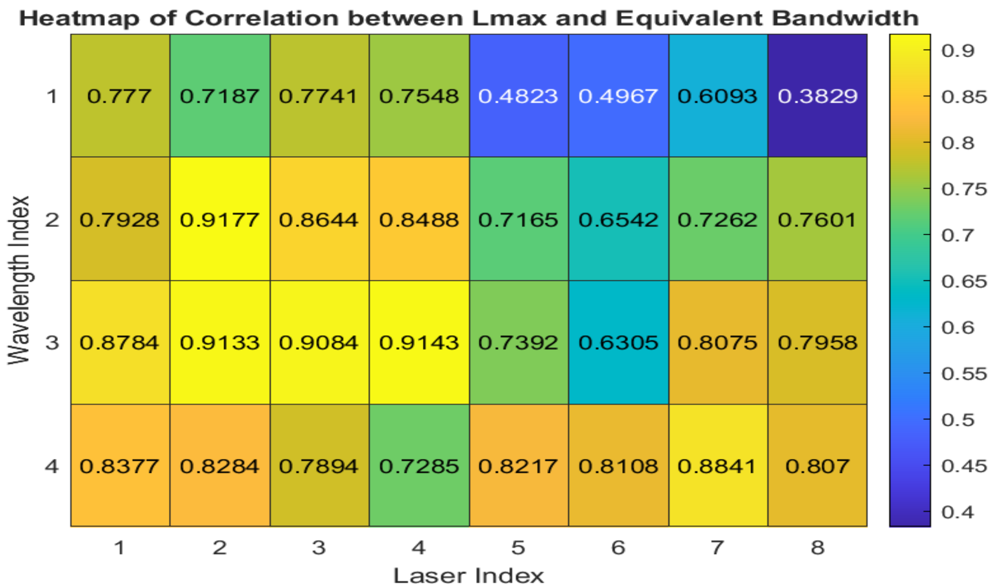


Figure 5.26: correlation heatmap between Lmax and the Equivalent Bnadwith for each laser indices 1 to 8 and wavelength λ = 850,880,910,940 nm showed as indice 1 to 4 at 70 meters

Analysis : Analysis :

1. Correlation Heatmap between Bandwidth Features and Lmax at 70m

Equivalent Bandwidth vs Lmax: The equivalent bandwidth shows a strong positive correlation with Lmax (0.7801), indicating that higher equivalent bandwidths are associated with higher Lmax values. However, as distance increases, the strength of this correlation tends to decrease, suggesting that the impact of equivalent bandwidth on Lmax becomes less significant at longer distances.

-3dB Bandwidth vs Lmax: The correlation between -3dB bandwidth and Lmax is positive but weaker (0.2061), reinforcing that equivalent bandwidth plays a more important role in determining Lmax.

Overall Correlations: The strong correlation between equivalent bandwidth and Lmax at shorter distances makes it a critical factor for optimizing Lmax. However, as the distance grows, the correlation weakens, indicating that other factors may start influencing Lmax more significantly.

Wavelength and Laser Variation: While the correlation between equivalent bandwidth and Lmax remains consistent across different wavelengths and lasers, this relationship becomes less robust as the distance increases, highlighting the diminishing influence of bandwidth on Lmax over longer distances.

Key Insights:

- **Equivalent Bandwidth as a Critical Metric:** At shorter distances, equivalent bandwidth shows a strong correlation with Lmax, making it a key metric for performance optimization. However, as the distance increases, its predictive power diminishes.

- **Diminishing Correlation with Distance:** As the distance increases, the correlation between equivalent bandwidth and Lmax weakens, suggesting the need to consider additional factors at longer distances to maximize Lmax.

- **Optimization Considerations:** While equivalent bandwidth remains important, optimizing the system for longer distances may require a more comprehensive approach, including other metrics that gain importance as distance increases.

5.18 All data on 150 m :

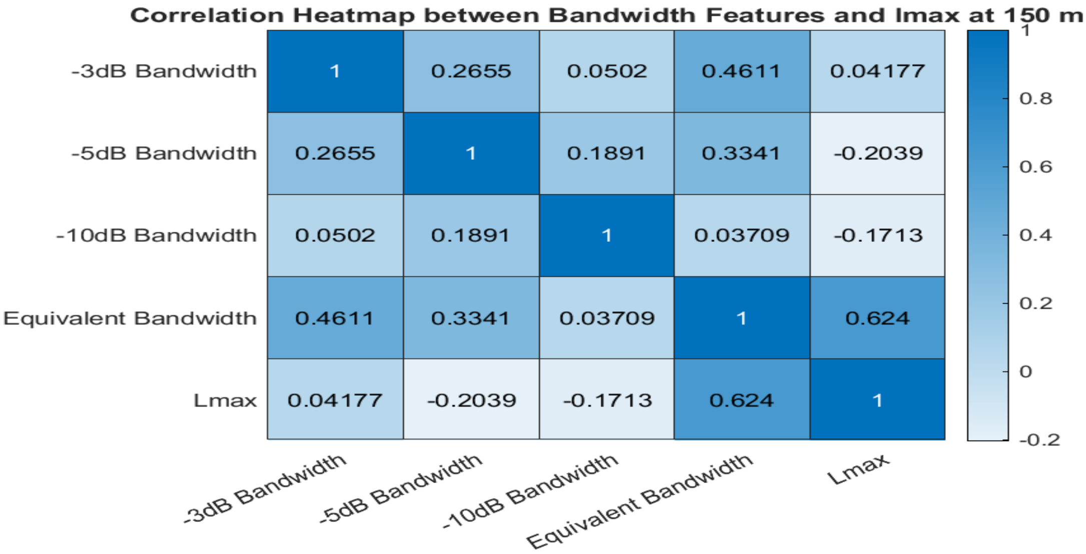


Figure 5.27: correlation heatmap Between Bandwidth features and Lmax at 150 meters

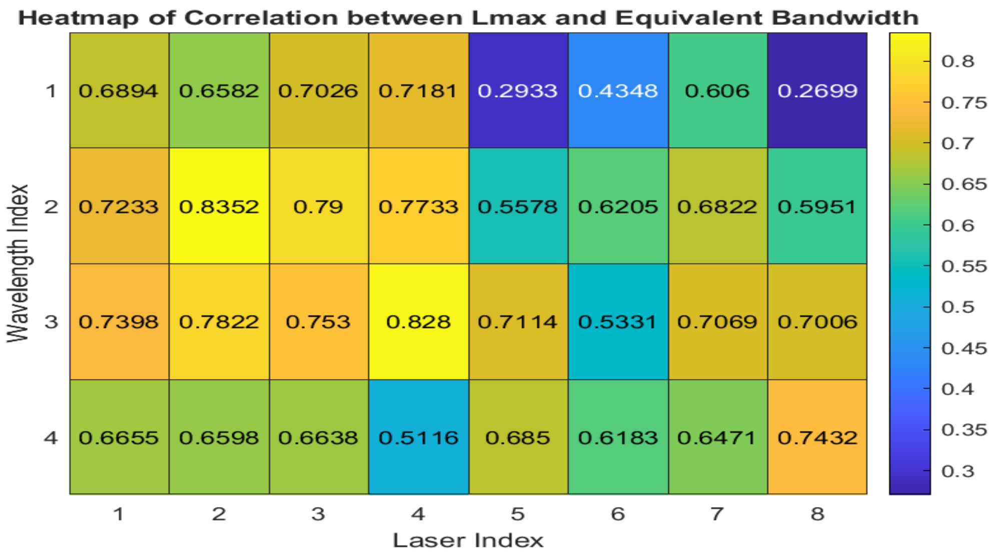


Figure 5.28: correlation heatmap between Lmax and the Equivalent Bnadwith for each laser indices 1 to 8 and wavelength λ = 850,880,910,940 nm showed as indice 1 to 4 at 150 meters

5.19 Analysis of the Correlation with Lmax at Increasing Distances

Equivalent Bandwidth vs Lmax: As we move from shorter to longer distances, the correlation between equivalent bandwidth and Lmax consistently decreases. At 100 meters, the equivalent bandwidth showed a strong positive correlation with Lmax, but as the distance increased to 150 meters, the correlation weakened to 0.624. This trend continues as distances are further increased, demonstrating a steady decline in the strength of the relationship between equivalent bandwidth and Lmax.

Progressive Distance Analysis: The analysis at 150 meters shows a significant drop in the correlation compared to 100 meters, and as we proceed to 200 meters, 300 meters, and eventually 400 meters, the correlation continues to diminish. While equivalent bandwidth remains positively correlated with Lmax, its predictive power weakens as the transmission distance grows, indicating that factors beyond bandwidth become more influential at greater distances.

-3dB Bandwidth vs Lmax: At 150 meters, the correlation between -3dB bandwidth and Lmax is minimal (0.04177), much weaker compared to equivalent bandwidth. This pattern of weak correlation holds true as distances increase, further reinforcing that -3dB bandwidth plays a less significant role in determining Lmax at longer distances.

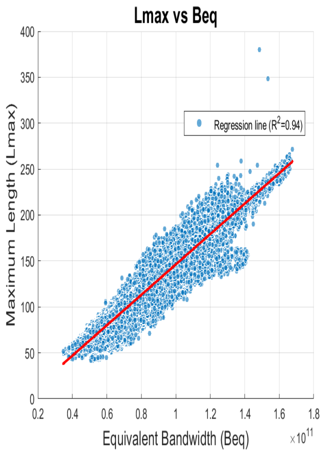
Overall Observations: As distance increases, the strong positive correlation between equivalent bandwidth and Lmax at shorter distances begins to decline. The equivalent bandwidth remains a critical factor for maximizing Lmax at shorter distances, but its influence diminishes beyond 150 meters. This suggests that while bandwidth is essential, other factors begin to dominate at greater distances, reducing the impact of equivalent bandwidth on Lmax performance.

Wavelength and Laser Variations: Even though the correlation between Lmax and equivalent bandwidth remains consistent across various wavelengths and lasers, the strength of these correlations weakens with increasing distance. This consistency at shorter distances suggests that equivalent bandwidth is a reliable predictor, but as distance grows, its effectiveness diminishes across different configurations.

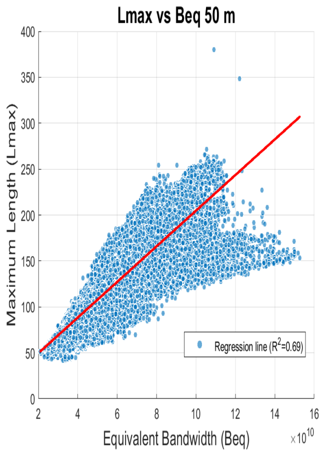
5.20 Key Insights and Recommendations

- **Decreasing Correlation with Distance:** As the distance increases from 100m to 400m, the correlation between equivalent bandwidth and Lmax progressively weakens. This indicates that while bandwidth is crucial at shorter distances, its impact reduces as distance increases.
- **Equivalent Bandwidth as a Critical Metric at Short Distances:** Equivalent bandwidth remains a key factor for optimizing Lmax at shorter distances (up to 150m). However, its influence on Lmax becomes less significant at longer distances, suggesting the need to explore additional factors.
- **Further Optimization at Longer Distances:** As distance increases, optimizing Lmax will require more than just bandwidth enhancements. A comprehensive approach, considering other variables, is necessary to maintain high Lmax values.
- **Wavelength and Laser Combinations:** At shorter distances, specific wavelength and laser combinations yield stronger correlations with Lmax. These combinations should be targeted for further optimization to enhance performance.

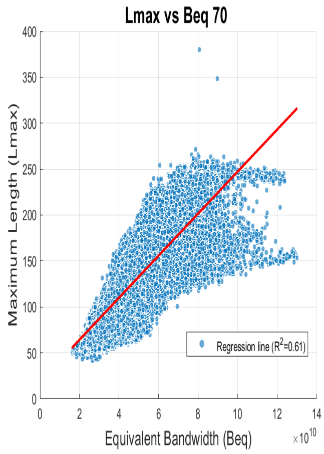
5.21 Lmax vs B equivalent



(a) Lmax vs Beq 30m

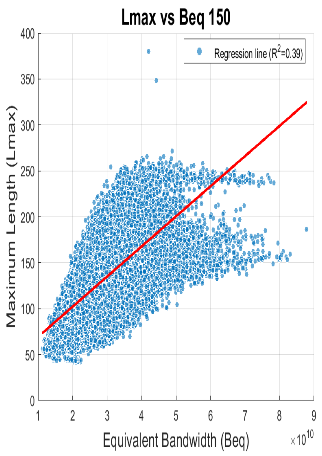


(b) Lmax vs Beq 50m

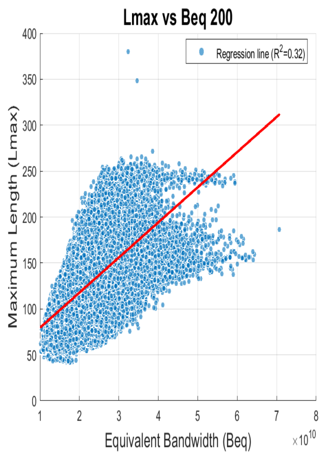


(c) Lmax vs Beq 70m

(d) Lmax vs Beq 100m



(e) Lmax vs Beq 150m



(f) Lmax vs Beq 200m

Figure 5.29: Lmax vs Beq for different distances

5.22 Analysis and Interpretation

The set of scatter plots provided illustrates the relationship between Maximum Achievable Link Length (Lmax) and Equivalent Bandwidth (Beq) for different distances of multimode fibers (MMF). The regression lines, along with their respective R^2 values, are included in each plot to show the strength and direction of the linear correlation between Lmax and Beq.

5.22.1 Key Observations

General Trend:

- Across all plots, there is a positive linear correlation between Equivalent Bandwidth (Beq) and Maximum Length (Lmax). This suggests that as Beq increases, Lmax generally increases as well.

- The strength of this relationship is indicated by the R^2 value in each plot.

Regression Line Fit (R^2 Values):

- At 30m: $R^2 = 0.94$, indicating a very strong linear relationship. This suggests that 94% of the variance in Lmax can be explained by the variance in Beq at this distance.

- At 50m: $R^2 = 0.69$, indicating a moderately strong linear relationship.

- At 70m: $R^2 = 0.61$, still a moderate linear relationship but weaker than at 50m.

- At 100m: $R^2 = 0.47$, indicating a moderate but noticeably weaker linear relationship.

- At 150m: $R^2 = 0.39$, showing a weak linear relationship.

- At 200m: $R^2 = 0.32$, indicating a weak linear relationship.

Effect of Distance on Correlation:

- As the distance increases from 30 meters to 200 meters, the R^2 values decrease. This indicates that the relationship between Beq and Lmax weakens as the distance increases. At shorter distances, Beq is a much stronger predictor of Lmax, while at longer distances, other factors may increasingly influence Lmax, diminishing the predictive power of Beq.

Scatter Density:

- The scatter points are more tightly clustered around the regression line at shorter distances (30m and 50m), which is consistent with higher R^2 values.

- As distance increases, the scatter of the points becomes more pronounced, indicating higher variability and a weaker relationship between Beq and Lmax.

5.22.2 Conclusion

Short Distances (30m to 50m): At shorter distances, Equivalent Bandwidth (Beq) is a highly reliable predictor of Maximum Achievable Link Length (Lmax). The high R^2 values suggest that improvements in Beq can significantly extend Lmax, making it a critical parameter in short-range MMF applications.

Longer Distances (100m to 200m): As the distance increases, the influence of Beq on Lmax diminishes. The lower R^2 values suggest that other factors, such as modal dispersion, attenuation, and possibly noise, start to play a more significant role in determining Lmax. This implies that for longer MMF links, relying solely on Beq to predict performance might be insufficient, and a more comprehensive analysis considering additional parameters is necessary.

5.23 Cosine Similarity Matrix

A **cosine matrix** refers to a matrix where the entries represent the *cosine similarities* between vectors. Each element of the matrix corresponds to the cosine of the angle between two vectors in a high-dimensional space. The cosine similarity is a measure of similarity between two non-zero vectors by calculating the cosine of the angle between them. Mathematically, for two vectors A and B, the cosine similarity is given by:

$$\text{cosine\_similarity}(A, B) = \frac{A \cdot B}{\|A\|\|B\|}$$

where $A \cdot B$ is the dot product of the vectors, and $\|A\|$ and $\|B\|$ are their magnitudes (norms). The result ranges from -1 to 1, where 1 means the vectors are identical, 0 means they are orthogonal (no similarity), and -1 means they are diametrically opposite.

In short, cosine matrices are valuable because they provide a scalable way to compare data points based on their orientation (rather than magnitude), making them useful in various applications requiring similarity measurement.

In the figures attached below (5.30, 5.31), the cosine matrices of all the bandwidths B_{-3dB}, B_{-5dB}, B_{-10dB}, and equivalent bandwidth B_{eq}, multiplied by the distance we are working on (ranging from 30 m to 400 m), are presented. Each cosine matrix contains 10 vectors, and we aim to verify the theory that suggests the product of the bandwidth vector and distance is a constant, denoted as K.

5.24 Logic used behind the cosine matrices

Algorithm 1 Checking if Bandwidth-to-Length Ratio (B × L) is Constant

```
Initialize bandwidth matrices for different distances (30m, 50m, ..., 400m).
Normalize the bandwidth matrices by converting to GHz.
for each bandwidth matrix (3dB, 5dB, 10dB, Equivalent) do
    Calculate the correlation between bandwidth and Lmax.
    Compute Mean Absolute Percentage Error (MAPE) between the bandwidth
    vectors.
    Visualize MAPE using a heatmap.
    Calculate and visualize cosine similarity between bandwidth vectors to mea-
    sure similarity.
end for
Plot the bandwidth vectors and Lmax to visually check their alignment. =0
```

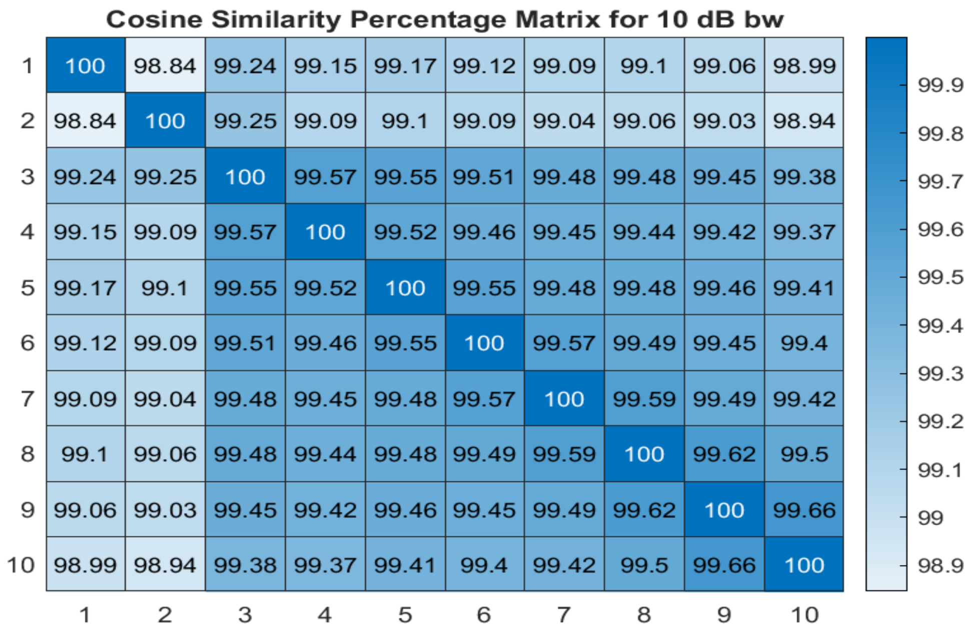


Figure 5.30: Cosine Similarity Matrix -10 dB

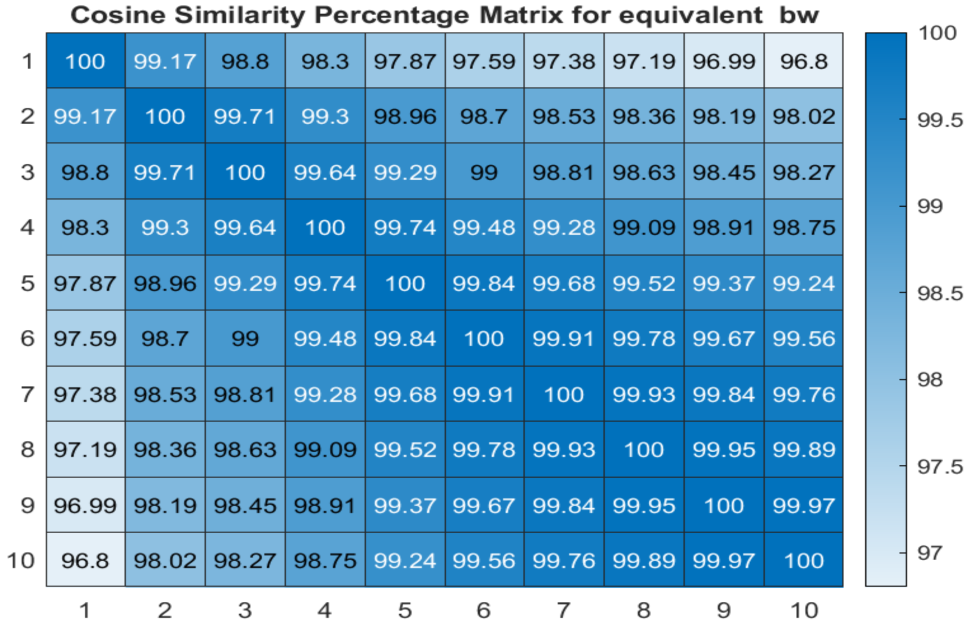


Figure 5.31: Cosine Similarity Matrix equivalent

Explanation of the Algorithm

The following algorithm outlines the procedure for verifying the consistency of the bandwidth-to-length ratio, $B \times L$, across varying distances. Specifically, this algorithm explores whether the product of bandwidth metrics (such as B_{-3dB}, B_{-5dB}, B_{-10dB}, and equivalent bandwidth B_{eq}) and fiber lengths ranging from 30m to 400m is constant. The method involves normalizing the bandwidth, computing correlations, and using the cosine similarity to evaluate the relationships between bandwidth vectors and L_{max}. Visual analysis is incorporated via heatmaps and cosine similarity plots to assess whether the bandwidth metrics and length exhibit consistent behavior. The Mean Absolute Percentage Error (MAPE) is used as a key indicator of error in bandwidth predictions.

- **Steps 1-2:** The bandwidth matrices for various distances are initialized and normalized by converting the bandwidth values to GHz, ensuring that they are comparable across different metrics.
- **Steps 3-8:** For each bandwidth metric (B_{-3dB}, B_{-5dB}, B_{-10dB}, and B_{eq}), the correlation with L_{max} is calculated. The MAPE is computed to assess prediction accuracy between bandwidth vectors, and the results are visualized using heatmaps. Additionally, cosine similarity is computed between vectors to measure the alignment between bandwidth and L_{max}.

- **Step 9:** Finally, the alignment of the bandwidth vectors and L_{max} is visually inspected to verify the consistency of the $B \times L$ ratio.

5.25 Analysis and Interpretation

5.25.1 Correlation with L_{max}

The correlation coefficients between the bandwidth vectors at different distances and the maximum achievable link length (L_{max}) were calculated. This analysis helps to understand how well each bandwidth measure at different distances predicts L_{max}).

- **3 dB Bandwidth:** The correlation coefficients between the 3 dB bandwidth vectors (A_30 to A_400) and L_{max} indicate the strength of the relationship between each bandwidth measure and L_{max}. Higher correlation coefficients suggest a strong predictive capability for L_{max}.

- **5 dB, 10 dB, and Equivalent Bandwidth:** Similar correlations were calculated for the 5 dB, 10 dB, and equivalent bandwidths (B, C, D vectors). The correlation matrices identify which bandwidth measure and distance best correlate with L_{max}, providing insights into optimal bandwidth measures for predicting fiber performance.

5.25.2 Cosine Similarity Analysis

Cosine similarity between the bandwidth vectors at different distances was calculated, ranging from -1 to 1, with 1 indicating maximum similarity. These values were then converted into percentage values and visualized using heatmaps.

- **Cosine Similarity for 3 dB Bandwidth:** The cosine similarity percentage matrix for 3 dB bandwidth shows how similar the bandwidth vectors are across different distances. High percentages (close to 100%) suggest that the 3 dB bandwidth is consistent across various lengths, indicating that the product $B \times L$ is constant.

- **Cosine Similarity for 5 dB, 10 dB, and Equivalent Bandwidth:** Similar cosine similarity analyses were performed for the 5 dB, 10 dB, and equivalent bandwidths. The heatmaps visually confirm how closely the bandwidth vectors at different lengths resemble each other.

5.25.3 Key Observations

- **Constant Product $B \times L$:** The cosine similarity matrices show high percentages, indicating that the product $B \times L$ remains relatively constant across different distances. This suggests an inverse relationship between bandwidth and fiber length in these multimode fibers.

5.25.4 Conclusion

The analysis confirms that the product of bandwidth and length remains constant across different fiber lengths for the tested multimode fibers. The cosine similarity, MAPE, and correlation analyses provide insights into how bandwidth measures at different distances relate to L_{max}. These findings are crucial for optimizing fiber optic systems, especially where consistent performance across varying distances is essential.

5.26 Analysis of MMF transfer function with respect to equivalent bandwidth

The algorithm adopted aims to find and plot the transfer functions for two specific values, namely the maximum and minimum points, from a set of bandwidth data but for same equivalent bandwidth. The aim of this study is see that at the same equivalent bandwidth we can have 2 different transfer functions. So the algorithm begins by defining the values to compare and initializes arrays to store the indices that correspond to these values in a 4D matrix called `Lmax_OM4`, which is reduced to a 3D matrix (`Lmax_first_equalizer`). For each value (maximum and minimum), it computes the difference between the value and elements in the matrix, searching for the closest match within a defined tolerance. If a match is found, the corresponding indices (wavelength, laser, and fiber) are saved. After identifying the indices, the algorithm extracts the corresponding transfer functions from another 4D matrix (`new_Transfer_fct_30`). These transfer functions are normalized, converted to decibels (dB), and then plotted on the same graph, showing how the maximum and minimum points behave across frequencies.

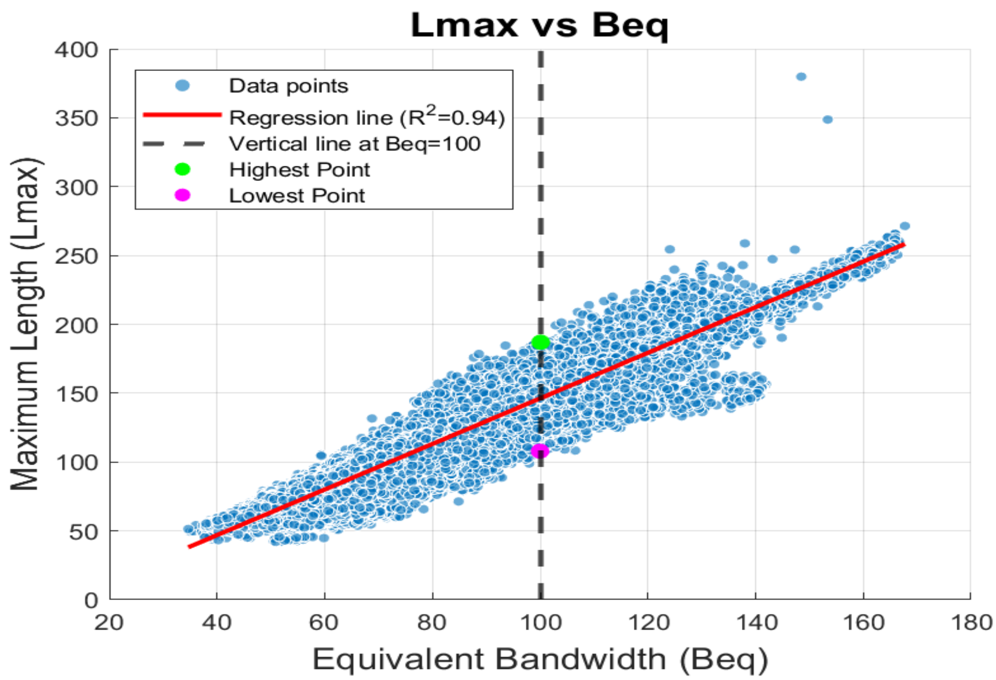


Figure 5.32: Beq vs Lmax at 30 m

Analysis and Interpretation

In Figure 5.32, the scatter plot illustrates the relationship between the Maximum Achievable Link Length (Lmax) and the Equivalent Bandwidth (Beq) at a distance of 30 meters. The plot includes a regression line with an R^2 value of 0.94, indicating a strong positive linear correlation between Lmax and Beq. This implies that as the Equivalent Bandwidth increases, the Maximum Length generally increases as well.

A vertical dashed line is drawn at Beq = 100 Gbps, marking the specific bandwidth of interest. The highest and lowest Lmax points within this bandwidth range are highlighted on the plot in green and pink, respectively. These points are critical as they are used to extract the corresponding transfer functions for further analysis shown in the figure 5.54

This analysis is done in order to see the transfer functions for the same equivalent bandwidth and to understand the behavior that we are seeing for the same equivalent bandwith wich is at 100 Ghz and to see if we can extract the the features for each point , the highest and the lowest intersection points between Equivalent bandwidth and the maximum achievable bandwidth not only on the 30 m but in all the cases from 30 m until 400m which are 30,50,70,100,150,200,250,300,350,400 meters.So now for the same equivalent bandwidth we have 2 different distance reach which are Highest and Lowest length reached by the fiber.

5.26.1 Intersection and Transfer Function Correlation

The intersection between the 100 Gbps equivalent bandwidth (marked in Figure 5.53) and the scatter plot is crucial for identifying the corresponding transfer functions. By selecting the highest and lowest Lmax points at Beq = 100 Gbps, we can examine the specific fiber characteristics that contribute to these extreme cases. The transfer functions plotted in Figure 5.33 highlight how the frequency response of the fiber differs between the cases of maximum and minimum link length, thereby providing insights into the factors that influence Lmax under specific bandwidth conditions.

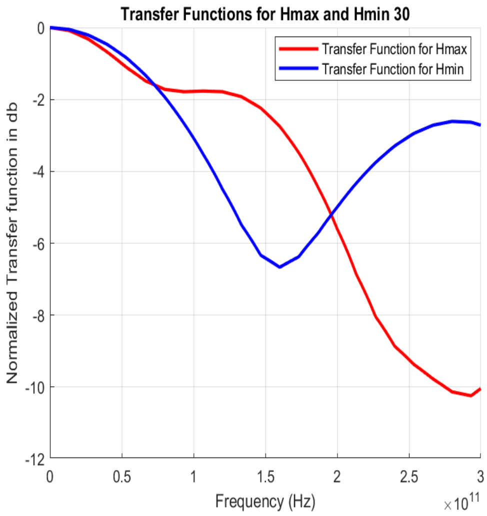
Algorithm followed :

The algorithm begins by defining the values that need to be compared, specifically the maximum and minimum points. It then initializes arrays to store the indices that correspond to these points in a 4D data matrix. The matrix is reduced to a 3D slice for easier comparison, and a tolerance level is defined to determine acceptable differences when comparing the values. For each value (maximum and minimum), the algorithm calculates the absolute difference between the elements in the matrix and the target value, finding the closest match within the defined tolerance. Once the closest match is found, the corresponding indices (wavelength, laser, and fiber) are stored. The algorithm then extracts the transfer functions corresponding to these indices from another 4D matrix. These transfer functions are normalized, converted to decibels (dB), and plotted on the same graph to visualize how the maximum and minimum points behave across different frequencies. Finally, the plot is saved as an image for further analysis. The algorithm is designed to analyze and compare transfer functions efficiently by identifying the relevant indices and visualizing the results.

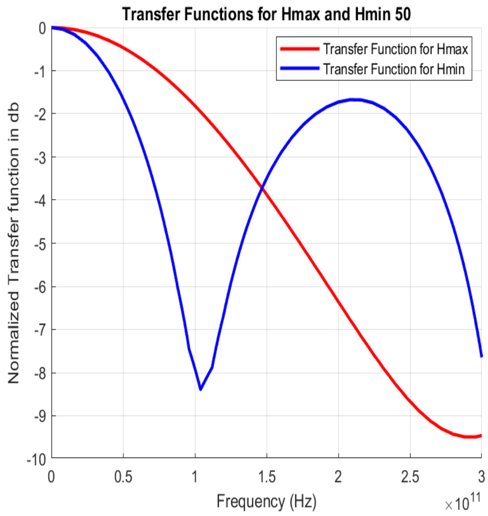
As it is clear , the results shows that for the same equivalent bandwidth we can see different transfer functions.

5.27 Final results

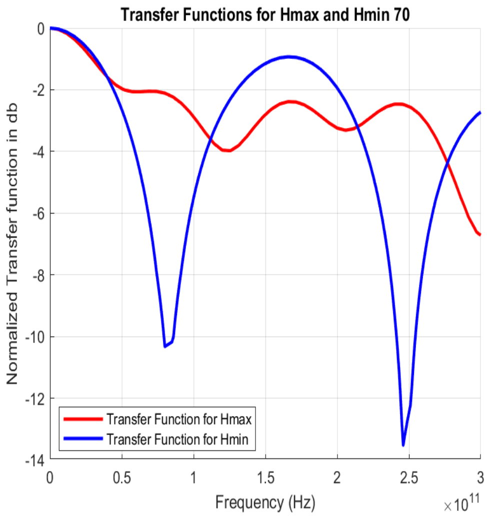
A new approach to catch more relevant correlation , since the equivalent bandwidth give the best correlation with $L_{\max}$, the new study is to calculate the correlation between the equivalent bandwidth at different frequencies , so we calculated the correlation between both quantities at 50 , 100 , 150 , 200 and 250 Ghz , and we ignored the frequency of 300 GHz since we did the study before and all this approach is done in all the distance that we are interested in from 30 to 300 meters . The correlation heatmap below shows the results in figures 5.34 5.35 5.36 5.37 5.38 5.39 5.40 5.41 .



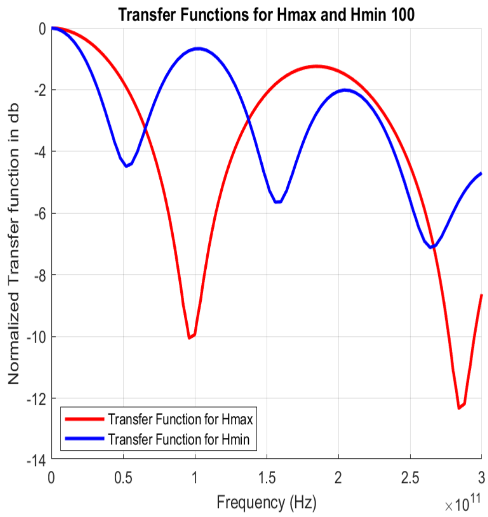
(a) Transfer Functions Hmax Hmin 30 m



(b) Transfer Functions Hmax Hmin 50 m



(c) Transfer Functions Hmax Hmin 70 m



(d) Transfer Functions Hmax Hmin 100 m

Figure 5.33: Transfer Functions of the to extracted points at the same equiavlent bandwidth at 30, 50, 70, 100 meters

Comprehensive Analysis and Interpretation from 30 Meters to 300 Meters

This comprehensive analysis focuses on the correlation between the maximum achievable reach (Lmax) of an optical fiber and the equivalent bandwidths at various frequencies (50 GHz, 100 GHz, 150 GHz, 200 GHz, and 250 GHz) over distances ranging from 30 meters to 300 meters. The goal is to understand how these relationships evolve with increasing distance and to identify which frequency bands are most predictive of Lmax.

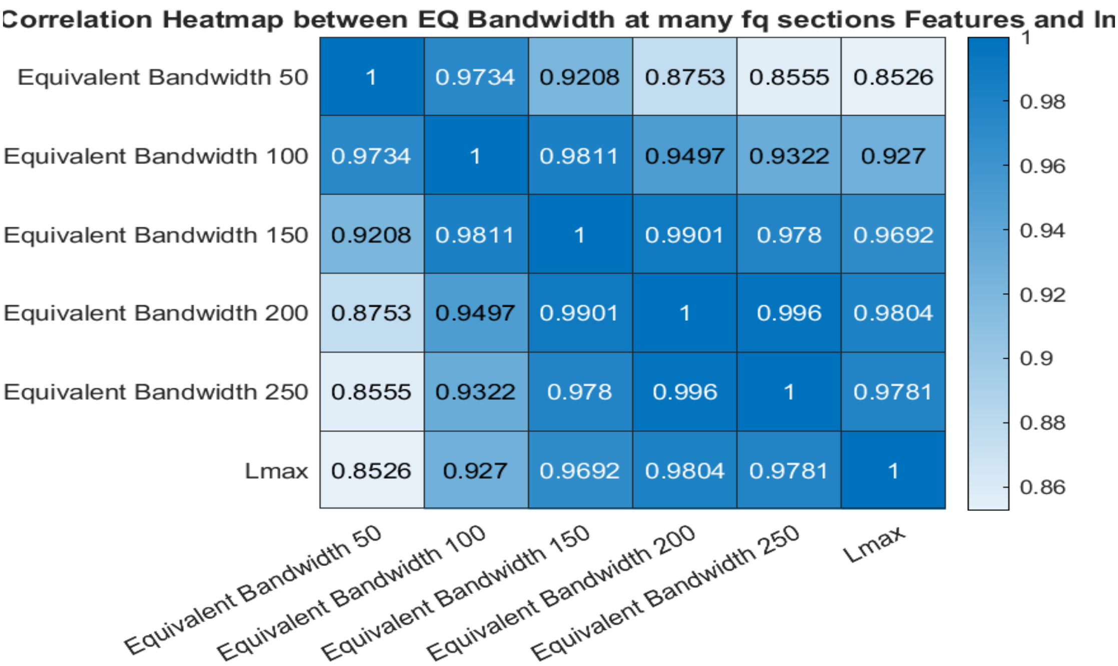
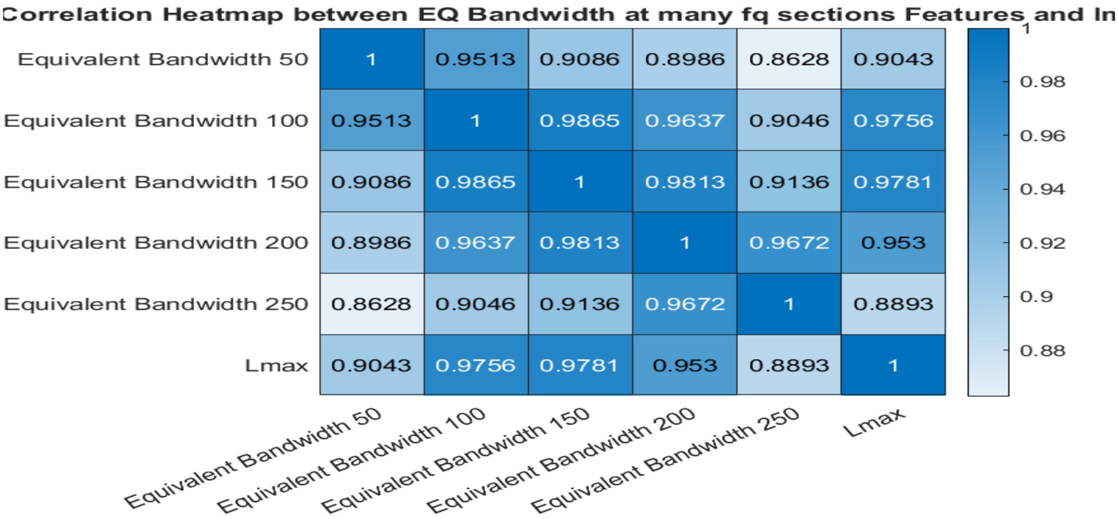


Figure 5.34: correlation heatmap Lmax BEQ 30

Correlation Analysis at 30 Meters

Figure 5.34 shows that at 30 meters, the equivalent bandwidths at different frequencies exhibit strong correlations with each other, with the highest correlations observed between 150 GHz and 200 GHz (0.9901).

Lmax shows a strong correlation with higher frequency bandwidths, particularly at 200 GHz (0.9804) and 250 GHz (0.9781). This indicates that at shorter distances, higher frequency bandwidths are critical for predicting Lmax.

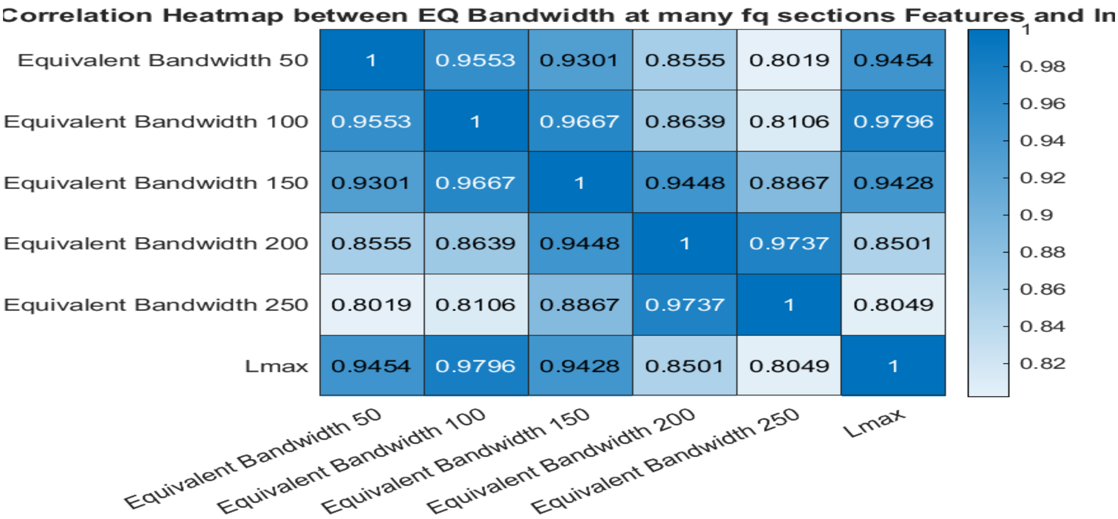


	Equivalent Bandwidth 50	Equivalent Bandwidth 100	Equivalent Bandwidth 150	Equivalent Bandwidth 200	Equivalent Bandwidth 250	Lmax
Equivalent Bandwidth 50	1	0.9513	0.9086	0.8986	0.8628	0.9043
Equivalent Bandwidth 100	0.9513	1	0.9865	0.9637	0.9046	0.9756
Equivalent Bandwidth 150	0.9086	0.9865	1	0.9813	0.9136	0.9781
Equivalent Bandwidth 200	0.8986	0.9637	0.9813	1	0.9672	0.953
Equivalent Bandwidth 250	0.8628	0.9046	0.9136	0.9672	1	0.8893
Lmax	0.9043	0.9756	0.9781	0.953	0.8893	1

Figure 5.35: correlation heatmap Lmax BEQ 50

Correlation Analysis at 50 Meters

Figure 5.35 shows that as the distance increases to 50 meters, the correlations among the equivalent bandwidths remain strong but show slight decreases compared to 30 meters. Lmax continues to correlate strongly with the equivalent bandwidths, particularly at 150 GHz (0.9781) and 100 GHz (0.9756). However, the correlation with higher frequency bandwidths, such as 250 GHz, decreases slightly, indicating a shift towards the importance of mid-range frequencies.



	Equivalent Bandwidth 50	Equivalent Bandwidth 100	Equivalent Bandwidth 150	Equivalent Bandwidth 200	Equivalent Bandwidth 250	Lmax
Equivalent Bandwidth 50	1	0.9553	0.9301	0.8555	0.8019	0.9454
Equivalent Bandwidth 100	0.9553	1	0.9667	0.8639	0.8106	0.9796
Equivalent Bandwidth 150	0.9301	0.9667	1	0.9448	0.8867	0.9428
Equivalent Bandwidth 200	0.8555	0.8639	0.9448	1	0.9737	0.8501
Equivalent Bandwidth 250	0.8019	0.8106	0.8867	0.9737	1	0.8049
Lmax	0.9454	0.9796	0.9428	0.8501	0.8049	1

Figure 5.36: correlation heatmap Lmax BEQ 70

Correlation Analysis at 70 Meters

Figure 5.36 shows that at 70 meters, correlations among equivalent bandwidths continue to weaken, especially between lower and higher frequency bandwidths. Lmax maintains strong correlations with lower frequencies, particularly 50 GHz (0.9454) and 100 GHz (0.9796), while the correlation with 250 GHz drops significantly to 0.8049. This suggests that lower and mid frequencies are becoming more dominant in predicting Lmax as distance increases.

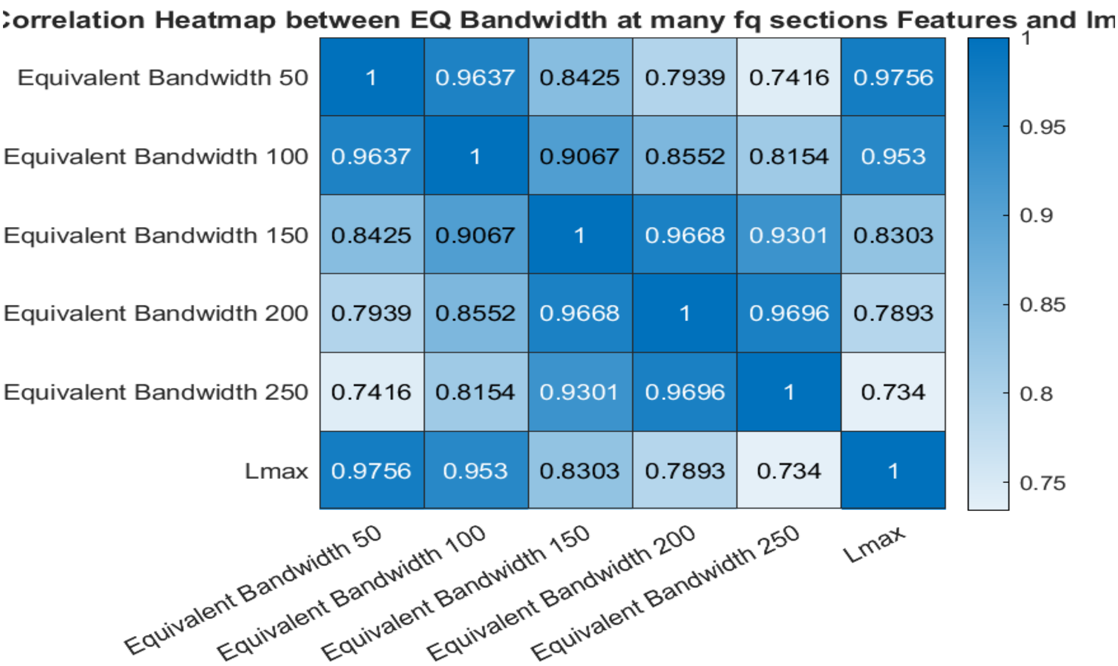


Figure 5.37: Transfer Functions Hmax Hmin 100

Correlation Analysis at 100 Meters

Figure 5.37 shows that at 100 meters, the correlations between equivalent bandwidths, especially between lower and higher frequencies, weaken considerably. The strongest correlation with Lmax is now at 50 GHz (0.9756), highlighting a shift towards lower frequencies as the most significant predictors. The correlation with 250 GHz drops further to 0.734, indicating a marked reduction in the predictive power of higher frequencies.

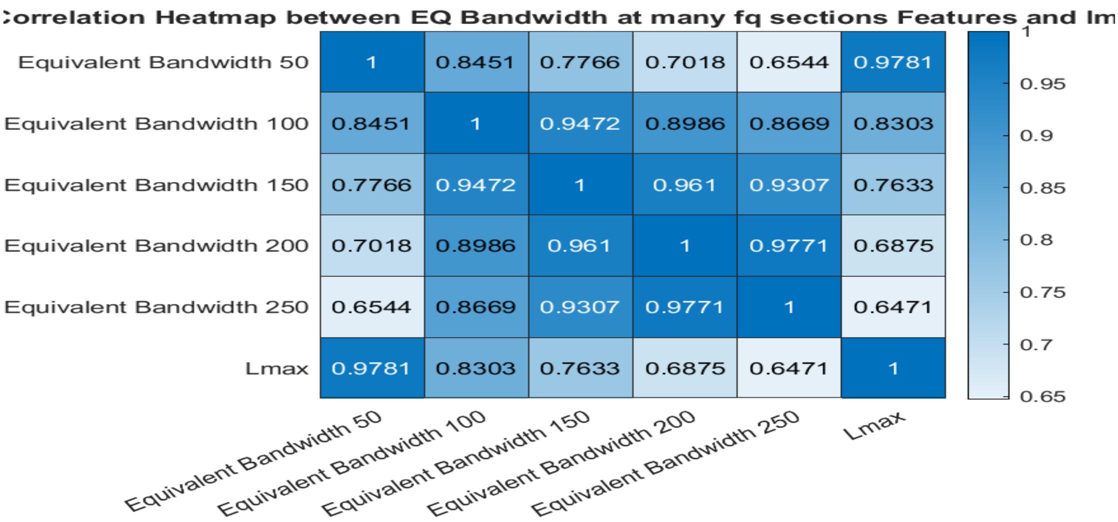


Figure 5.38: correlation heatmap Lmax BEQ 150

Correlation Analysis at 150 Meters

Figure 5.38 shows that as the distance extends to 150 meters, correlations across bandwidths weaken further, with the strongest correlations observed between 150 GHz and 200 GHz (0.961). Lmax shows the highest correlation with 50 GHz (0.9781), but correlations with all other frequencies continue to decline, particularly with 250 GHz (0.6471). This suggests that lower frequencies are increasingly important, while higher frequencies become less reliable for predicting Lmax.

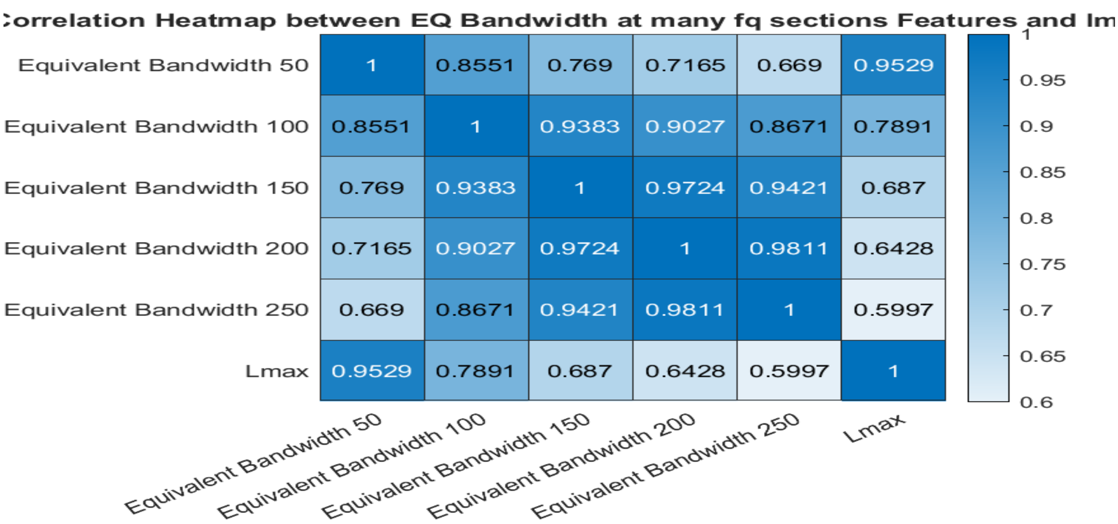


Figure 5.39: correlation heatmap Lmax BEQ 200

Correlation Analysis at 200 Meters

Figure 5.39 shows that at 200 meters, the correlation between equivalent bandwidths continues to diminish, especially between 50 GHz and higher frequencies. The strongest correlation with Lmax remains at 50 GHz (0.9529), while the correlation with higher frequencies, especially 250 GHz, drops significantly to 0.5997. This underscores the growing importance of lower frequencies as predictors of Lmax as distance increases.

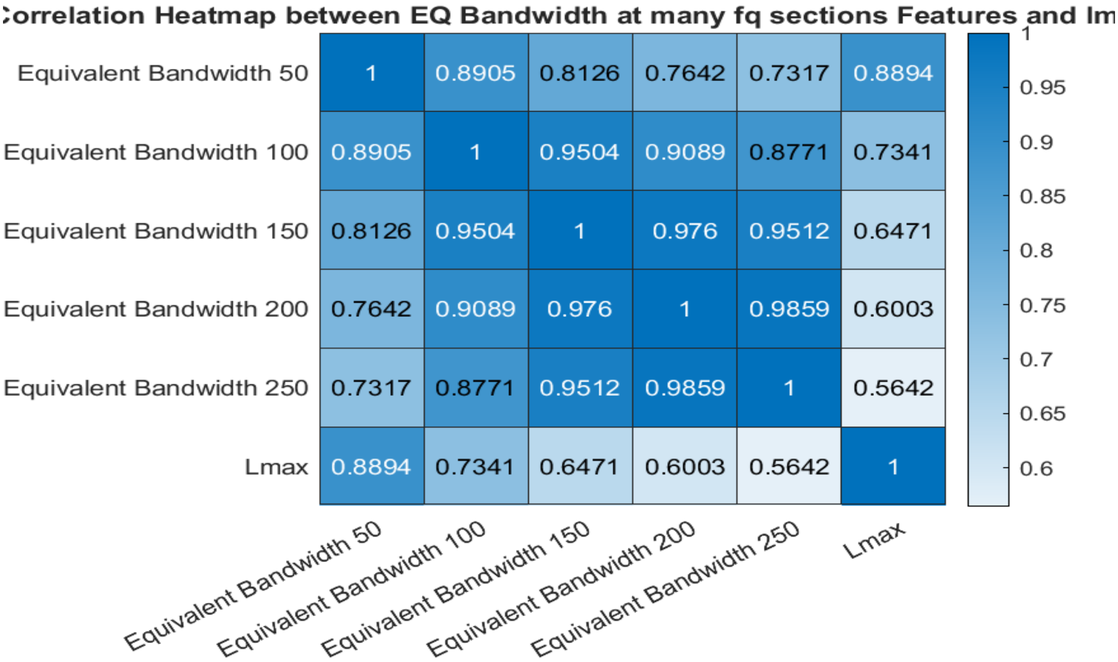


Figure 5.40: correlation heatmap Lmax BEQ 250

Correlation Analysis at 250 Meters

Figure 5.40 shows that at 250 meters, the correlations between equivalent bandwidths weaken even further, with the highest correlation still between 200 GHz and 250 GHz (0.9859). The strongest correlation with Lmax is at 50 GHz (0.8894), but this is weaker compared to shorter distances. The correlation with 250 GHz drops to 0.5642, indicating a significant reduction in the predictive power of higher frequencies.

Correlation Analysis at 300 Meters

Figure 5.41 shows that at 300 meters, correlations across all frequency bands weaken significantly. The correlation between Lmax and any equivalent bandwidth is now much lower. The highest correlation with Lmax is at 50 GHz (0.8303), but this represents a substantial decrease compared to previous distances. The correlation with 250 GHz drops to a very low 0.5385, indicating minimal predictive power at this distance.

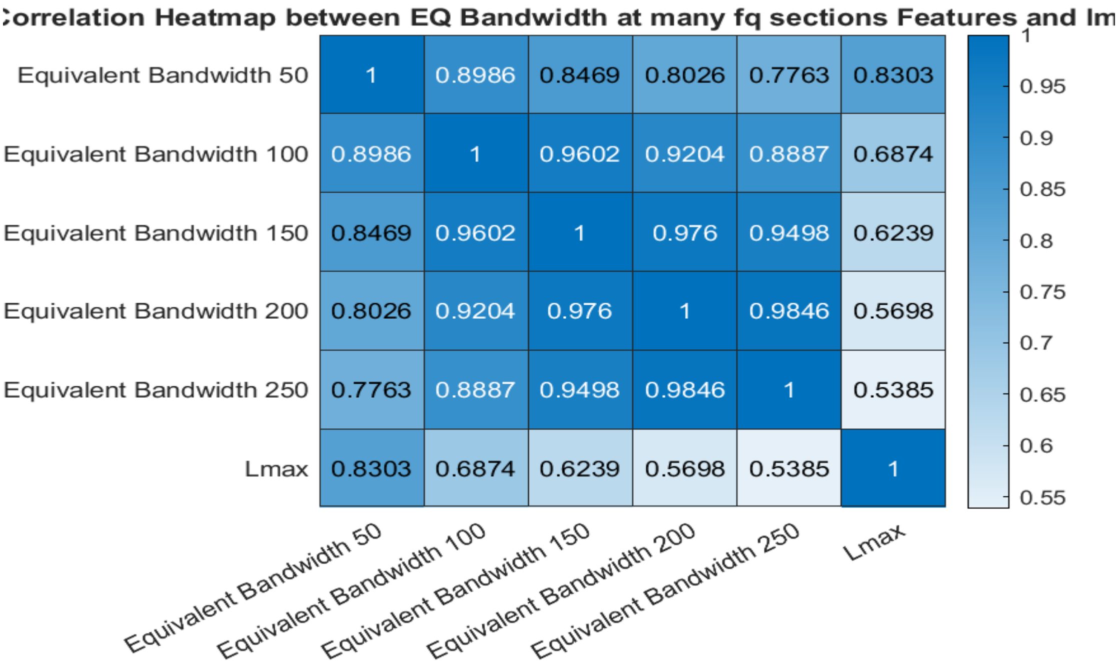


Figure 5.41: correlation heatmap Lmax BEQ 300

Summary of Key Findings

Increasing Importance of Lower Frequencies: As the distance increases from 30 meters to 300 meters, the importance of lower frequencies (particularly 50 GHz) in predicting Lmax becomes increasingly evident. The correlation between Lmax and higher frequencies diminishes significantly, indicating that lower frequencies are more reliable for predicting maximum reach at longer distances.

Decreasing Predictive Power of Higher Frequencies: Higher frequency equivalent bandwidths (200 GHz and 250 GHz) exhibit a steep decline in their correlation with Lmax as distance increases. By 300 meters, these frequencies have very low predictive power, making them unreliable for predicting maximum achievable reach.

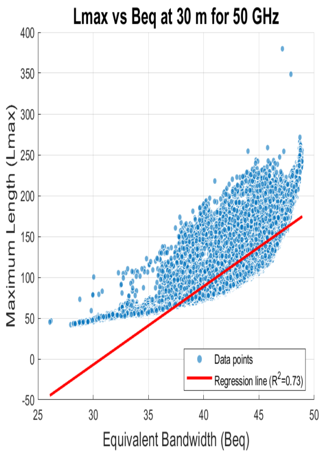
Overall Decline in Predictive Accuracy: The general trend across all distances shows that as the distance increases, the fiber's performance becomes more variable and less predictable. This is reflected in the declining correlations between Lmax and equivalent bandwidths across the frequency spectrum.

Conclusion

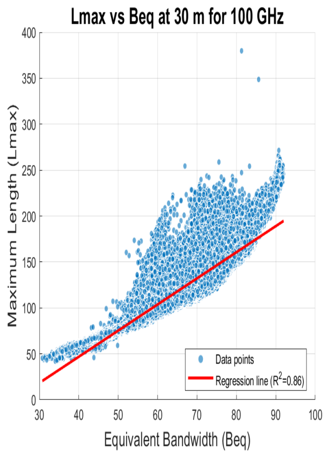
The analysis from 30 meters to 300 meters highlights the shifting importance of equivalent bandwidth frequencies in predicting the maximum achievable reach (Lmax) of an optical fiber. As distance increases, lower frequencies become more dominant in their predictive power, while higher frequencies lose their reliability. These findings are crucial for optimizing fiber performance, particularly in long-distance applications where maintaining accurate predictions of Lmax is essential.

5.28 Scatter Plots of Equivalent Bandwidth vs $L_{\mathbf{max}}$

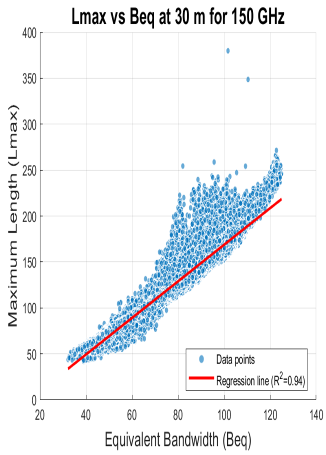
At distance equal to 30 meters :



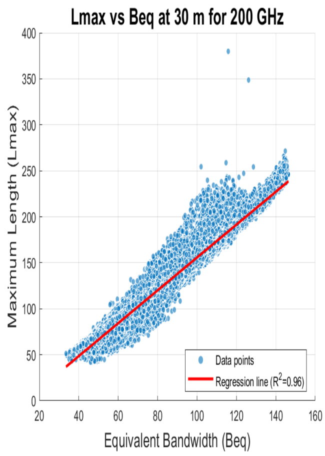
(a) Lmax vs Beq integrated up to 50 GHz at 30 m



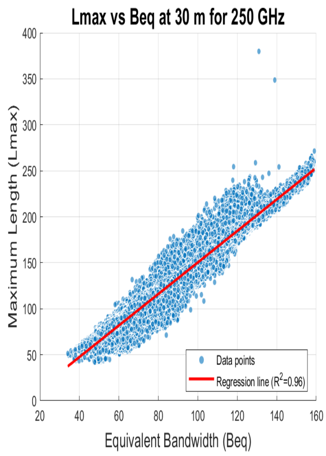
(b) Lmax vs Beq integrated up to 100 GHz at 30 m



(c) Lmax vs Beq integrated up to 150 GHz at 30 m



(d) Lmax vs Beq integrated up to 200 GHz at 30 m



(e) Lmax vs Beq integrated up to 250 GHz at 30 m

Figure 5.42: Lmax vs Beq for different GHz ranges

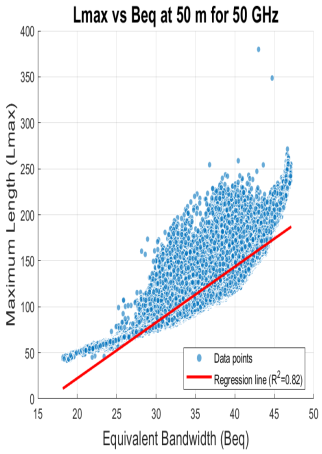
General Interpretation Across All Frequencies

Trend: As the frequency increases from 50 GHz to 250 GHz, the correlation between Lmax and Beq generally strengthens, as indicated by increasing R^2 values. This suggests that at higher frequencies, Beq becomes a more accurate predictor of the maximum achievable reach (Lmax).

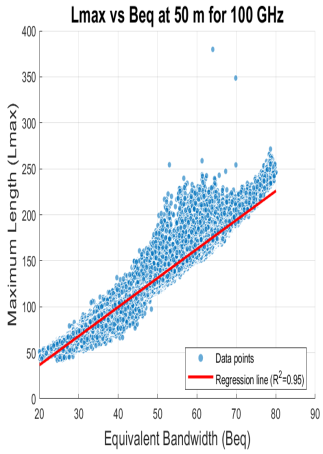
The highest predictive power is observed in the 200 GHz and 250 GHz ranges, with R^2 values of 0.96. These frequencies show the least scatter and the strongest linear relationships.

At lower frequencies (50 GHz and 100 GHz), while there is still a positive correlation between Lmax and Beq, the relationship is weaker, as evidenced by the lower R^2 values (0.73 at 50 GHz and 0.86 at 100 GHz). This indicates that other factors may have a more significant impact on Lmax at these frequencies.

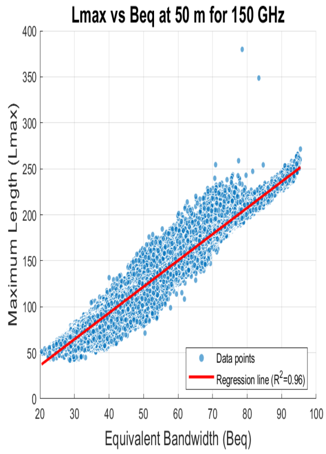
At distance equal to 50 meters :



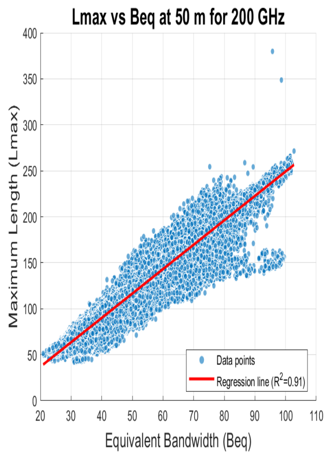
(a) Lmax vs Beq integrated up to 50 GHz at 50 m



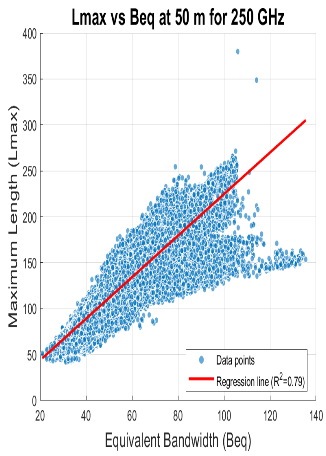
(b) Lmax vs Beq integrated up to 100 GHz at 50 m



(c) Lmax vs Beq integrated up to 150 GHz at 50 m



(d) Lmax vs Beq integrated up to 200 GHz at 50 m



(e) Lmax vs Beq integrated up to 250 GHz at 50 m

Figure 5.43: Lmax vs Beq 50 for different GHz ranges

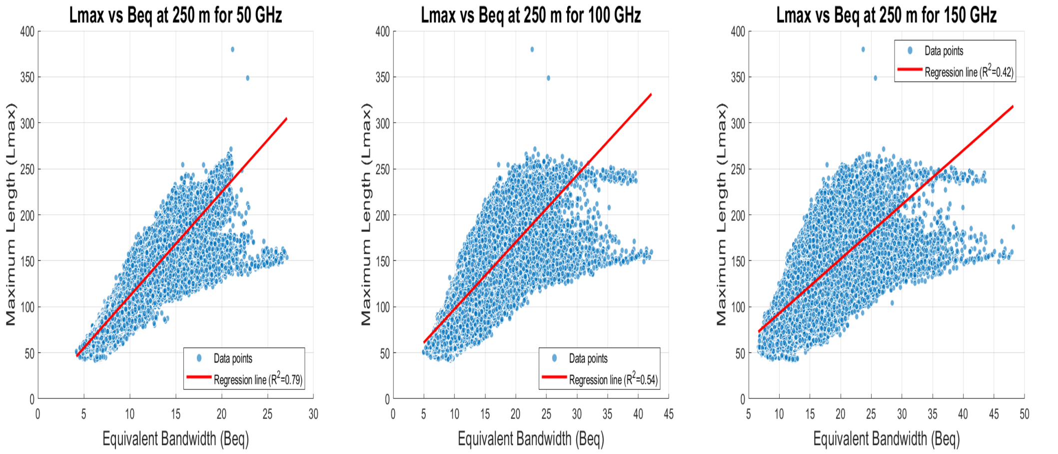
General Interpretation Across All Frequencies

Trend: As the frequency increases from 50 GHz to 150 GHz, the correlation between Lmax and Beq strengthens, reaching a peak R^2 value of 0.96 at 150 GHz. This suggests that Beq becomes a more accurate predictor of Lmax as the frequency increases up to this point.

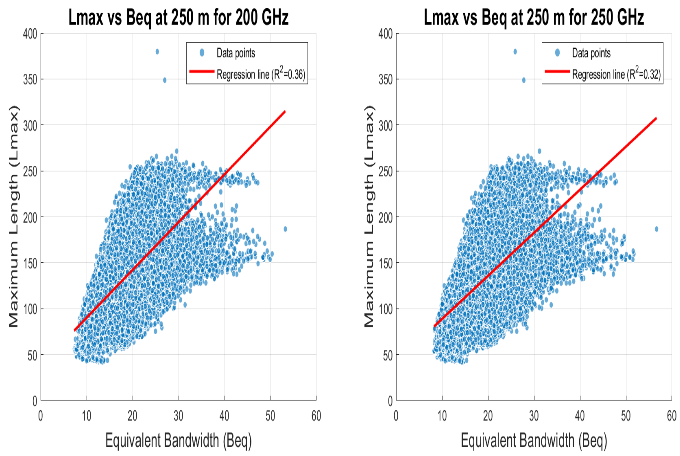
The highest predictive power is observed at 150 GHz with an R^2 value of 0.96, showing the least scatter and the strongest linear relationship. However, as the frequency increases further to 200 GHz and 250 GHz, the predictive power slightly decreases, indicated by lower R^2 values (0.91 and 0.79, respectively).

At higher frequencies (200 GHz and 250 GHz), while there is still a positive correlation between Lmax and Beq, the relationship weakens, as evidenced by the increased scatter and lower R^2 values. This suggests that other factors may have a more significant impact on Lmax at these frequencies, especially at 250 GHz, where the correlation drops to 0.79.

At distance equal to 250 meters :



(a) Lmax vs Beq 250 50 GHz **(b)** Lmax vs Beq 250 100 GHz **(c)** Lmax vs Beq 250 150 GHz



(d) Lmax vs Beq 250 200 GHz **(e)** Lmax vs Beq 250 250 GHz **(f)**

Figure 5.44: Lmax vs Beq 250 for different GHz ranges

5.28.1 Analysis and Interpretation at 250 m

50 GHz:

The strongest correlation is observed here with an R^2 value of 0.79. This indicates that at 250 meters, this frequency is the most reliable predictor of Lmax.

100 GHz:

The correlation decreases with an R^2 value of 0.54. While it still shows a positive relationship, it's significantly less reliable than 50 GHz.

150 GHz:

The correlation continues to weaken with an R^2 value of 0.42, indicating a further reduction in predictive power.

200 GHz and 250 GHz:

The R^2 values are 0.36 and 0.32, respectively, suggesting very weak correlations. These higher frequencies are not reliable for predicting Lmax at 250 meters.

Conclusion: At 250 meters, the lower frequency of 50 GHz remains the most effective for predicting the maximum achievable reach (Lmax) of the fiber. As the frequency increases, the reliability of predicting Lmax decreases significantly.

5.29 Analysis of Correlation Between Equivalent Bandwidth and $L_{\max}$

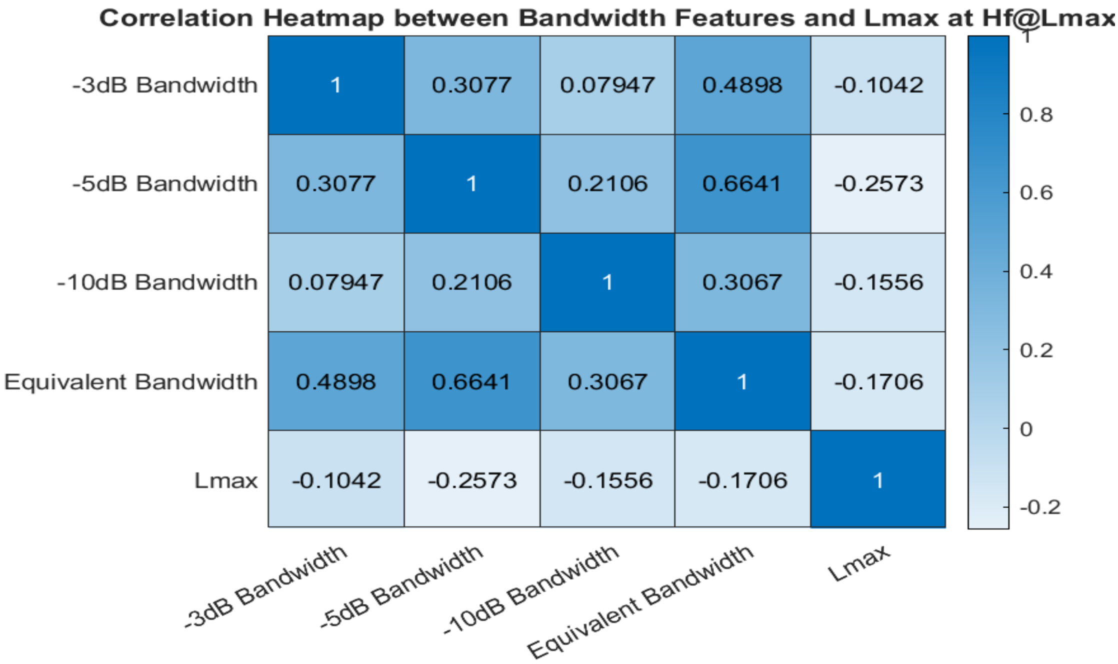


Figure 5.45: Correlation Heatmap between Bandwidth Features and Lmax at Hf at Lmax

Figure 5.45 shows the relationship between Equivalent Bandwidth and $L_{\max}$ (maximum achievable link length).

5.29.1 Key Observation

- The correlation coefficient between Equivalent Bandwidth and $L_{\max}$ is negative (-0.1706), indicating that as Equivalent Bandwidth increases, there is a slight decrease in $L_{\max}$.

5.29.2 Interpretation

The negative correlation suggests that higher Equivalent Bandwidth does not correspond to an increase in $L_{\max}$. In fact, increasing the Equivalent Bandwidth may result in a minor reduction in the maximum achievable link length. This finding indicates a limit to the effectiveness of Equivalent Bandwidth in extending fiber performance, highlighting the need to consider other factors when optimizing for $L_{\max}$.

5.29.3 Conclusion

Although Equivalent Bandwidth is an important metric for fiber performance, its negative correlation with $L_{\max}$ suggests that relying solely on increasing bandwidth may not significantly improve the maximum link length. Further analysis is needed to identify other influencing factors.

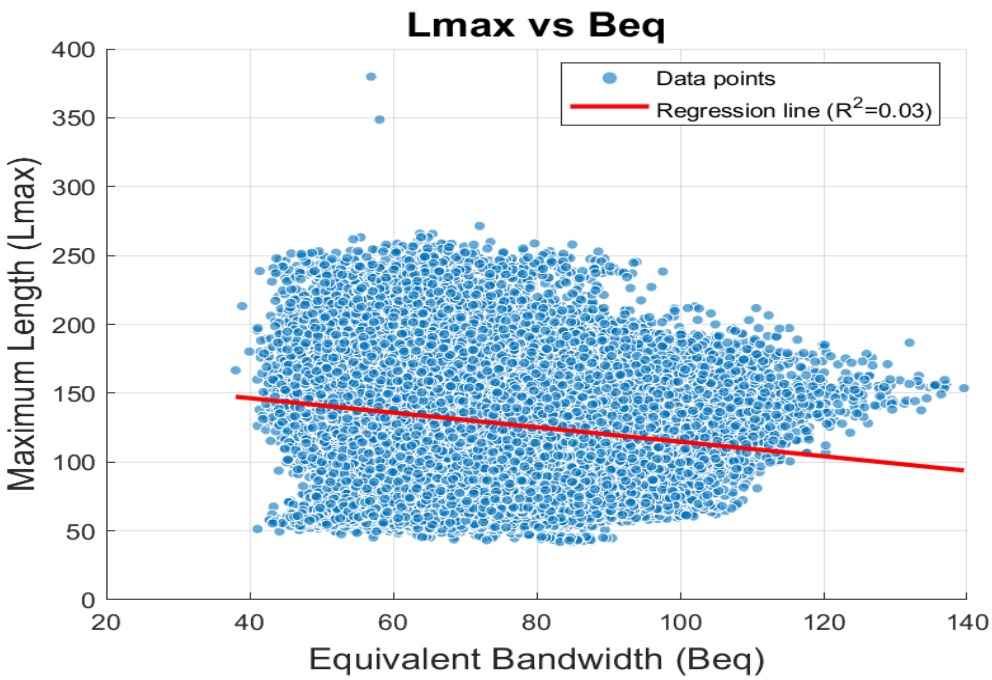


Figure 5.46: Lmax vs Beq Hf at Lmax

5.30 Interpretation of Lmax vs Beq

The scatter plot illustrates the relationship between the maximum achievable link length ($L_{\max}$) and the equivalent bandwidth (Beq). The blue data points represent the measured values, while the red regression line provides an indication of the overall trend.

5.30.1 Key Observations

- The regression line shows a slight negative slope, suggesting a weak inverse relationship between $L_{\max}$ and Beq. As Beq increases, there is a small tendency for $L_{\max}$ to decrease.
- The R^2 value of 0.03 indicates that equivalent bandwidth (Beq) accounts for only a small portion of the variation in $L_{\max}$, meaning that other factors likely play a more significant role in determining $L_{\max}$.

5.30.2 Interpretation

The weak correlation suggests that equivalent bandwidth (Beq) is not a strong predictor of $L_{\max}$. While higher bandwidth may slightly reduce $L_{\max}$, the relationship is weak, implying that other factors need to be considered for optimizing fiber performance and predicting $L_{\max}$ accurately.

Chapter 6

Machine Learning

6.1 Neural Network Algorithm Explanation

Machine Learning for Lmax Prediction

Previous studies failed to establish strong correlations between bandwidth measures (-3dB, -5dB, -10dB, and equivalent bandwidth) and Lmax using traditional methods. Due to the complex, nonlinear nature of these relationships, machine learning, specifically artificial neural networks (ANNs), was adopted as a more powerful solution.

How Neural Networks Work

Neural networks consist of layers of neurons that process inputs and make predictions. Key steps include:

- **Input Layer**: Takes the four bandwidth vectors as input.
- **Hidden Layers**: Neurons process the input using weighted sums and activation functions to model complex patterns.
- **Output Layer**: Outputs the predicted Lmax.
- **Training**: The model adjusts weights to minimize prediction error using training data.
- **Optimization and Evaluation**: Hyperparameters are tuned, and the model is tested on new data to ensure accuracy.

This approach allows better predictions of Lmax by capturing nonlinear relationships between the features and labels.

6.1.1 Step 1: Data Preparation

Data Preparation

The data preparation process involves loading, preprocessing, and splitting the dataset for training, testing, and validation. The process starts by selecting a `.mat` file through a file dialog, which contains the necessary data vectors: `vector_3dB_cleaned`, `vector_5dB_cleaned`, `vector_10dB_cleaned` , `vector_equiv_cleaned`, and `Lmax_all_vector_cleaned` , each vectors is of size 120507 elements wich are the number of all the cases that can be seen from 4 λ wavelength , 8 lasers and 3766 OM4 fibers after cleaning. These vectors represent the input features (bandwidths) and the target variable (Lmax). The input vectors are stacked horizontally to form a feature matrix, X, and the target variable, y, is set to the Lmax vector.

Next, the features in X are standardized using the `StandardScaler` to ensure consistent scaling across all input data, which is important for improving model performance. The target variable y remains in its original form without scaling.

After preprocessing, the data is split into two equal datasets: D_1 and D_2, each containing 50% of the data. The first dataset, D_1, is further split into training (80%) and testing (20%) subsets. These subsets are used for training the machine learning model and evaluating its performance, respectively. The second dataset, D_2, is reserved for validation, which will be used to assess the model's generalization on unseen data during hyperparameter tuning. Finally, a success message is displayed, confirming the successful loading and splitting of the data.

6.1.2 Step 2: Defining the Neural Network Structure

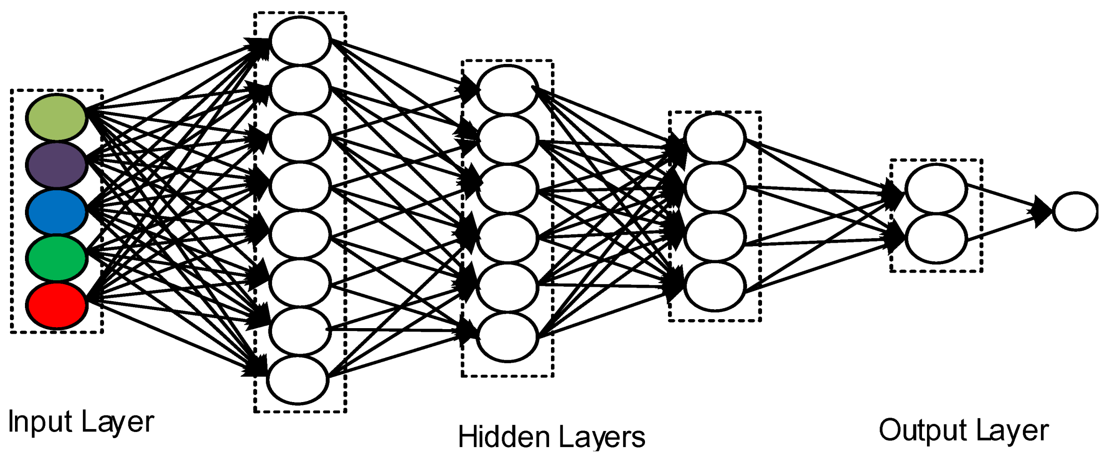


Figure 6.1: ANN Model

General Idea : Figure 6.1 shows the architecture of an artificial neural network (ANN) model. It consists of three main parts: the input layer, hidden layers, and output layer. The input layer comprises multiple nodes, each representing an input feature (in this case, four input vectors). The hidden layers are composed of several neurons that process the inputs through weighted connections, applying transformations to capture complex patterns in the data. The output layer contains a single neuron representing the final prediction (for example, in a regression problem, this would be a continuous value like Lmax). The model learns by adjusting the weights of connections between neurons during training to minimize the prediction error.

Architecture of the ANN Model:
The architecture of the artificial neural network (ANN) used consists of an input layer of 4 inputs: the $-3\,\text{dB}, -5\,\text{dB}, -10\,\text{dB}$ and equivalent bandwidth vectors, three hidden layers, and an output layer, which consists of the vector of $L_{\max}$ to be predicted. The first hidden layer contains between 128 and 512 neurons (with ReLU activation), the exact number determined by Optuna during the hyperparameter optimization process. The second hidden layer contains between 64 and 256 neurons, and the third hidden layer contains between 32 and 128 neurons. Each hidden layer includes L2 regularization to prevent overfitting and batch normalization to improve training stability. Dropout, with a rate between 0.1 and 0.3, is applied after each hidden layer for further regularization. The output layer contains a single neuron with a linear activation function, appropriate for regression tasks, as it predicts a continuous value. The model is compiled using the Adam optimizer, with the learning rate and L2 regularization factor also optimized by Optuna.

The ReLU (Rectified Linear Unit) activation function is used in the hidden layers due to its ability to avoid the vanishing gradient problem, which can occur with activation functions like sigmoid or tanh. ReLU allows for faster training by keeping the gradient consistent for positive values, while also being computationally efficient due to its simple thresholding mechanism. The dropout layers, with a rate between 0.1 and 0.3, are applied to reduce overfitting. By randomly deactivating a fraction of neurons during training, dropout forces the model to learn more robust features, ensuring that it doesn't rely too heavily on any specific neuron. The range of 0.1 to 0.3 is commonly used as it provides a balance between retaining enough information for learning while preventing overfitting.

6.1.3 Step 3: Model Compilation

Relevant Code:

```
# Compile the model with the Adam optimizer and tuned
  hyperparameters
```

```
        model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=learning_rate),
                      loss='mean_squared_error',
                      metrics=['mean_absolute_percentage_error'])
```

Explanation: The model is compiled with the Adam optimizer, which adapts the learning rate dynamically during training. The loss function used is `mean squared error (MSE)`, which is commonly used for regression tasks. The `mean absolute percentage error (MAPE)` is also tracked as an additional metric to evaluate prediction accuracy.

6.1.4 Step 4: Training the Neural Network

Relevant Code:

```
# Early stopping and learning rate scheduling to improve generalization
        early_stopping = EarlyStopping(monitor='val_loss', patience=15, restore_best_weights=True)
        lr_scheduler = ReduceLROnPlateau(monitor='val_loss', factor=0.1, patience=5, min_lr=1e-6, verbose=1)

        # Cross-validation setup (KFold)
        kf = KFold(n_splits=5, shuffle=True, random_state=42)
        losses = []

        for train_idx, val_idx in kf.split(X_train):
            X_train_fold, X_val_fold = X_train[train_idx], X_train[val_idx]
            y_train_fold, y_val_fold = y_train[train_idx], y_train[val_idx]

            model.fit(X_train_fold, y_train_fold, epochs=200, batch_size=32,
                      validation_data=(X_val_fold, y_val_fold),
                      callbacks=[early_stopping, lr_scheduler], verbose=0)

            loss, _ = model.evaluate(X_val_fold, y_val_fold, verbose=0)
            losses.append(loss)
```

Explanation: The model is trained using cross-validation, where the data is split into 5 folds. During each iteration, the model is trained on 4 folds and validated on the remaining fold. Early stopping is used to prevent overfitting by halting training when validation performance stops improving. The learning rate scheduler reduces the learning rate dynamically when the validation loss plateaus.

6.1.5 Step 5: Hyperparameter Optimization with Optuna

Relevant Code:

```
# Set up the Optuna study and optimize the objective function
        study = optuna.create_study(direction='minimize')
        study.optimize(objective, n_trials=1)  # Adjust the number of trials

        # Get the best trial
        best_trial = study.best_trial
        messagebox.showinfo("Optimization Completed", f'Best trial: {best_trial.params}')

        # Train the final model based on the best trial
        build_and_train_best_model(best_trial.params)
```

Explanation: Optuna is used to optimize the hyperparameters by running multiple trials. In each trial, the objective function defines the model's architecture with a different set of hyperparameters. The best-performing hyperparameters, which minimize the validation loss, are selected for final training.

6.1.6 Step 6: Training the Best Model

In this step, the best model is built and trained based on the optimal hyperparameters determined by Optuna. The architecture consists of an input layer, three hidden layers, and an output layer. Each hidden layer is configured with the optimal number of neurons, determined by the best hyperparameters (‘n_units_1‘, ‘n_units_2‘, and ‘n_units_3‘), and uses the ReLU activation function for non-linearity. L2 regularization is applied to prevent overfitting, and batch normalization is included to improve training stability. A dropout layer, with a dropout rate between 0.1 and 0.3, is added after each hidden layer to further reduce overfitting. The output layer consists of a single neuron with a linear activation function, suitable for regression tasks. The model is compiled using the Adam optimizer with a learning rate determined through Optuna, using ‘mean_squared_error‘ as the loss function and ‘mean_absolute_percentage_error‘ (MAPE) as the evaluation metric. Early stopping and learning rate scheduling are applied to ensure optimal training: early stopping halts the training process if the validation loss does not improve after 15 epochs, while the learning rate is reduced when validation loss plateaus for 5 epochs. The model is trained on 80% of the dataset, with 20% used for validation, over 200 epochs. After training, the model is evaluated on the test set (20% of the data), and its performance is measured using the test loss and MAPE. The model is further validated on the remaining 50% of the dataset, and both the test and validation results are displayed. Finally, the trained model is saved for future use, and the input data scaler is saved to ensure proper scaling for future predictions.

6.1.7 Step 7: Model Evaluation

The training and validation loss over epochs is plotted to visualize the model's learning progress. The test set predictions are compared with the actual values, and any outliers are identified to further assess the model's performance.

6.1.8 Step 8: Model Saving

Relevant Code:

```
model.save("my_trained_model.keras")
```

Explanation: After the model has been trained and evaluated, it is saved as a `.keras` file, allowing future reuse without the need for retraining.

6.1.9 Results

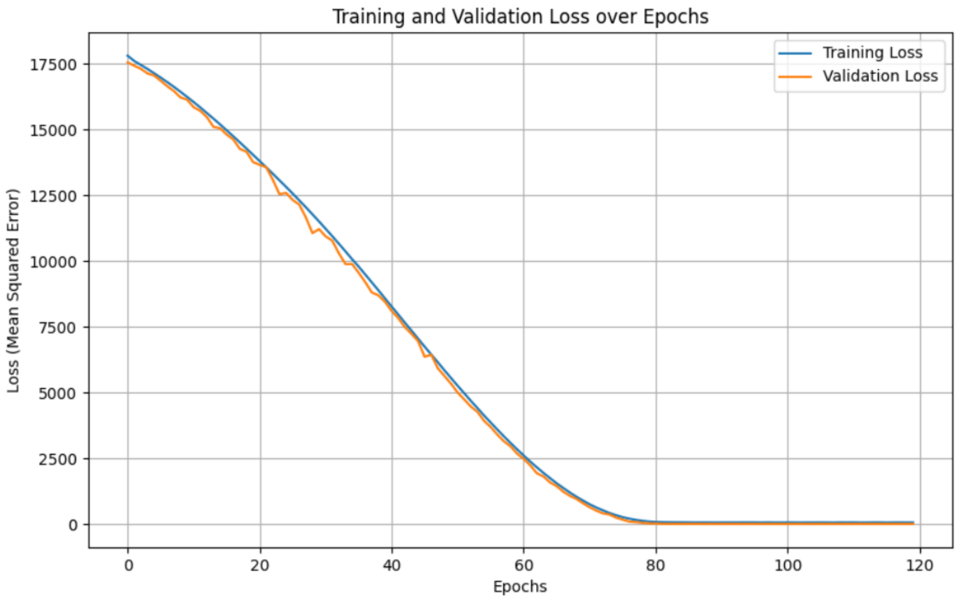


Figure 6.2: Training vs Validation Loss

6.1.10 Analysis of Training and Validation Loss over Epochs

The plot illustrates the training and validation loss (measured as Mean Squared Error) over the epochs during the training process of the neural network. The neural network was trained using the following hyperparameters: three hidden layers, each with a varying number of neurons — between 128 and 512 for the first hidden layer, between 64 and 256 for the second hidden layer, and between 32 and 128 for the third hidden layer, all activated using the ReLU function. A dropout rate between 0.1 and 0.3 was applied to prevent overfitting, and L2 regularization was used to further improve generalization. The learning rate was optimized using Optuna, and the Adam optimizer was employed for training.

Key observations from this plot are as follows:

- **Convergence of Loss:** Both the training loss and validation loss exhibit a sharp decrease during the early epochs, indicating that the model is learning and reducing error effectively in the initial training stages. After approximately 60 epochs, the loss stabilizes, indicating that the model has converged.

- **Consistent Performance:** The validation loss closely follows the training loss throughout the epochs, which suggests that the model generalizes well to unseen data. There is no significant divergence between the training and validation curves, implying that the model is not overfitting.

- **Low Final Loss:** Towards the end of training (after around 100 epochs), both the training and validation loss approach low values. This suggests that the model has learned to predict with high accuracy, minimizing error for both the training and validation datasets.

- **Model Stability:** The consistency between training and validation loss across all epochs demonstrates that the model maintains stability during training. This indicates that the learning rate and regularization techniques such as early stopping, dropout, and L2 regularization are well-tuned.

Conclusion: The plot indicates a well-performing model with no signs of overfitting or underfitting. The training and validation loss convergence to low values suggests that the neural network is capable of making accurate predictions on the given dataset.

6.1.11 Scatter Plot Analysis

Figures 6.3 and 6.4 show the relationship between the predicted maximum fiber reach (*Predicted Lmax*) and the actual measured values (*Actual Lmax*). These plots help us evaluate the model's performance.

- **Figure 6.3 (With Outliers)**:
  - The outliers are the predicted data points that deviate from the actual data points for a threshold of 20%.
  - The green points represent the predicted values, while the red line corresponds to perfect predictions where the predicted Lmax equals the actual Lmax.
  - Most of the points cluster near the red line (90 % of the predicted data are less than 5 meters away fro mthe actual data), indicating that the model is making good predictions for the majority of the data points.
  - However, some points deviate significantly from the red line, particularly in the higher Lmax range (above 300). These represent **outliers**, where the model's predictions deviate significantly from the actual values.
  - To quantify this error, we can analyze the percentage of cases within specific error margins. For example, we can measure how many cases have an absolute error of less than 1 meter, 5 meters, or more. Additionally, it is important to consider the relative error, as the impact of a 1-meter

error is different for a maximum reach (Lmax) of 10 meters compared to a Lmax of 100 meters. Evaluating both the absolute and relative errors provides a more comprehensive understanding of the model's performance, especially when analyzing outliers.

- **Figure 6.4 (Without Outliers)**:
  - This plot shows the same data after filtering out the outliers. The blue points closely follow the red line, indicating highly accurate predictions in the absence of outliers.
  - Most of the points in the range of 50 to 250 are tightly aligned with the perfect prediction line, confirming that the model performs well for more than 90% of the dataset.

General Interpretation:

- The model performs well overall, with accurate predictions for most data points, as indicated by the proximity of points to the red line in both plots.
- The presence of outliers in Figure 6.3 suggests that there are specific cases where the ANN model fails in providing good predictions

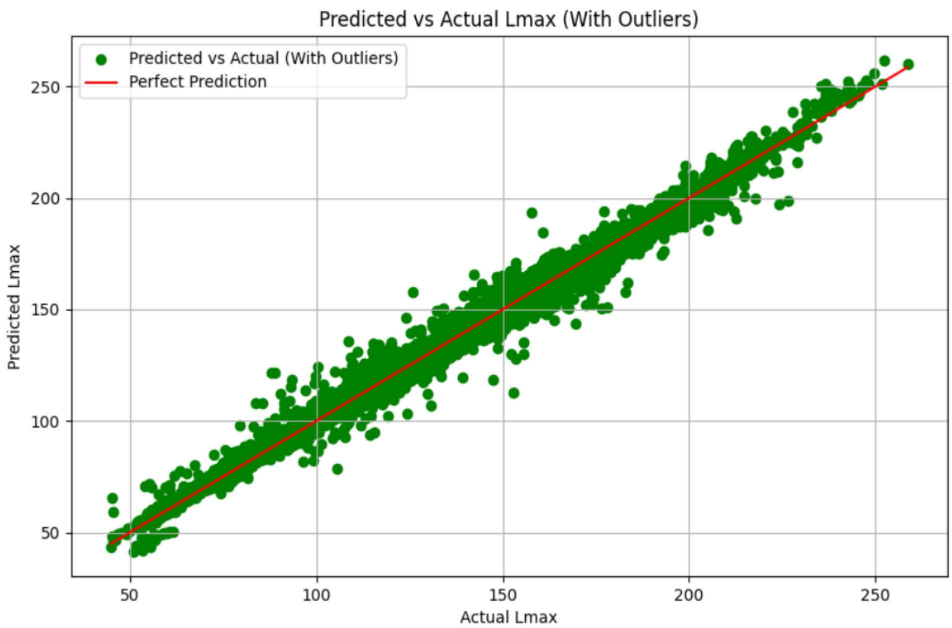


Figure 6.3: Scatter plots with outliers

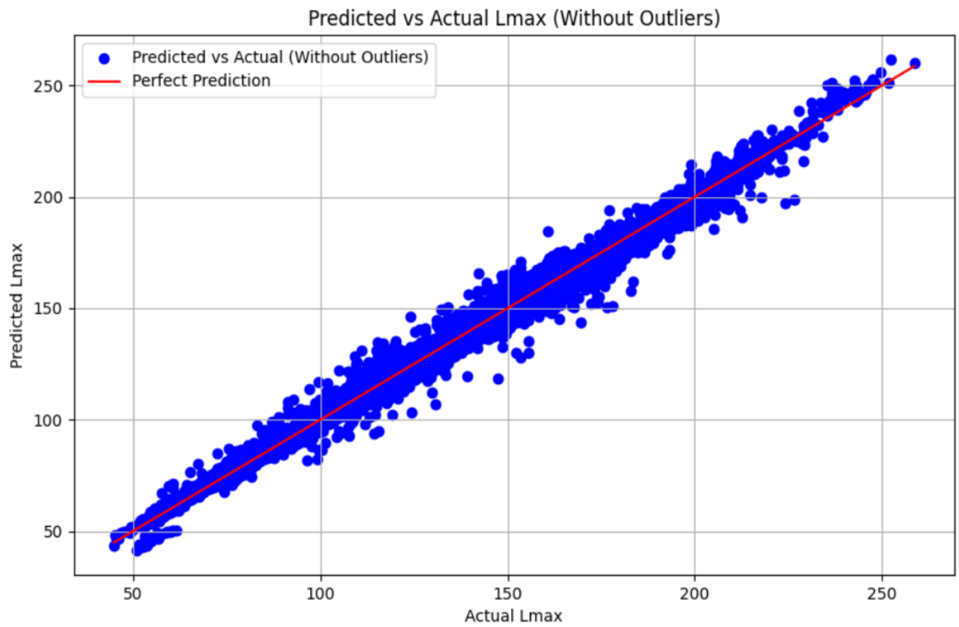


Figure 6.4: Scatter plots without outliers

6.1.12 Model Performance: Test and Validation Results

Test Results:

- **Test Loss:** The optimized test loss is 15.83, which indicates the degree of error between the actual and predicted $L_{\max}$. The lower the loss, the closer the predicted values are to the actual values.
- **Test MAPE:** The optimized test MAPE is 0.02%, suggesting a very low average percentage difference between the predicted and actual $L_{\max}$ values. This indicates highly accurate predictions during testing.
- **Test Accuracy:** The model achieved an accuracy of 99.98%, which signifies that the model's predictions are very close to the actual values in the test set, with minimal error.

Validation Results (D2):

- **Validation Loss:** The validation loss for the D_2 dataset is 16.87, which is slightly higher than the test loss, but still represents a small prediction error.
- **Validation MAPE:** The validation MAPE is 0.02%, consistent with the test MAPE, demonstrating that the model generalizes well and maintains high prediction accuracy on unseen data.

- **Validation Accuracy:** The validation accuracy is 99.98%, which matches the test accuracy, showing that the model performs consistently well across different subsets of data.

Interpretation: The results in table 6.1indicate excellent model performance, with both the test and validation accuracy nearing 100%. The low MAPE values further confirm the model's precision, as it produces minimal prediction errors for the maximum achievable link length (L_{max}). Additionally, the small difference between test and validation losses suggests that the model is not overfitting, and it generalizes well to new data.

Metric	**Value**
Best Trial:	
n_units_1	256
n_units_2	128
n_units_3	64
Dropout Rate	0.15
Learning Rate	$4.2161775820111514 \times 10^{-5}$
L2 Lambda	$1.9542051834034877 \times 10^{-6}$
Optimized Test Results:	
Test Loss	13.44979190826416
Test MAPE	0.02%
Test Accuracy	99.98%
Validation Results (D2):	
Validation Loss	14.506579399108887
Validation MAPE	0.02%
Validation Accuracy	99.98%

Table 6.1: Optimized Test and Validation Results

6.1.13 Error Margin

Analysis : Figure 6.5 illustrates the percentage of predictions within different absolute error margins: 1 meter, 5 meters, and 10 meters. The plot shows that approximately 25% of the predictions have an error of less than or equal to 1 meter, indicating that a quarter of the predictions are highly accurate. Furthermore, around 90% of the predictions fall within a 5-meter error margin, suggesting that the model is making reliable predictions for the majority of the data points. Lastly, nearly all predictions (close to 100%) have an error of less than or equal to 10 to 15 meters, highlighting that the model performs consistently well in terms of accuracy, even with a slightly larger error margin. This indicates strong generalization capabilities of the trained model across the dataset.

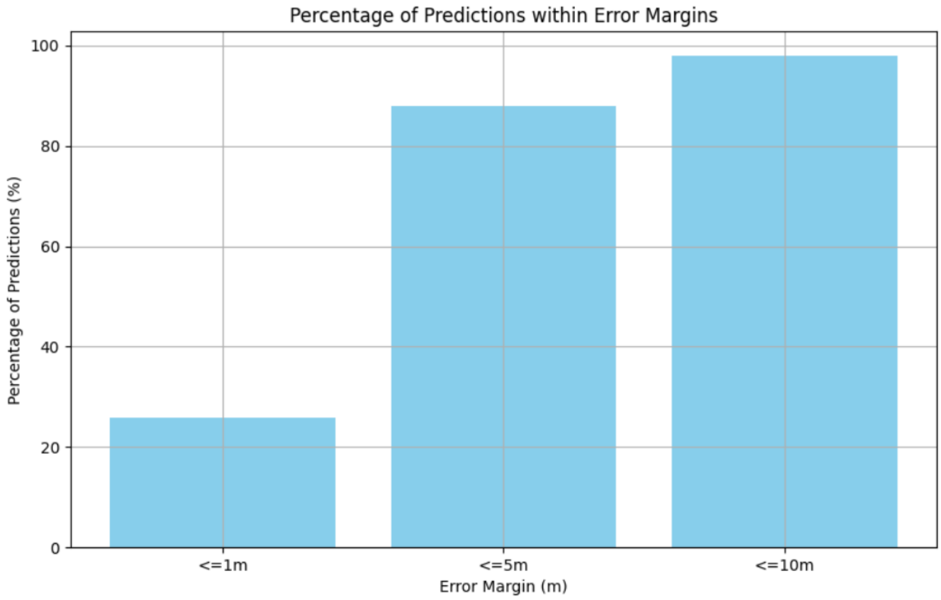


Figure 6.5: Error Margin (m)

6.1.14 Relative Error vs Actual Lmax

Analysis : Figure 6.6 depicts the relative error as a function of the actual Lmax values. The majority of the points are clustered below the 5% relative error threshold (represented by the green dashed line), indicating that the model performs with high accuracy for most data points. A smaller number of points exceed the 10% relative error threshold (represented by the red dashed line), particularly for lower Lmax values, suggesting that the model struggles more with smaller maximum reach distances. For larger Lmax values, the relative error tends to remain below 10%, showcasing the model's stability and consistency when predicting higher Lmax values. Overall, the model demonstrates strong generalization with a low relative error for the majority of predictions.

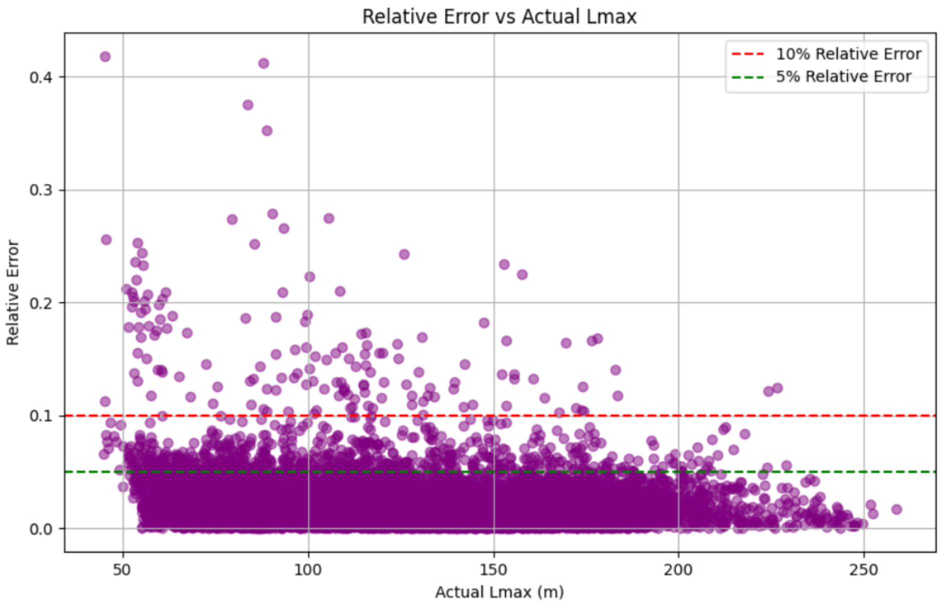


Figure 6.6: Relative Error vs Actual Lmax

6.1.15 Error ECDF

Analysis of ECDF of Errors between Actual and Predicted $L_{\mathbf{max}}$

Figure 6.7 presents the Empirical Cumulative Distribution Function (ECDF) of the absolute errors between the actual and predicted L_{max}. The x-axis represents the error magnitude, while the y-axis shows the cumulative probability.

- The curve indicates that the majority of errors are relatively small, as nearly 90% of the data points have an error of less than 10 meters.
- The ECDF quickly reaches a plateau, meaning most of the predicted L_{max} values are very close to the actual values, suggesting high accuracy in the model's predictions.
- Only a few data points exhibit larger errors (greater than 20 meters), implying that the model performs exceptionally well for most cases, with only approximetly 10 % outliers contributing to larger prediction discrepancies of 10 to 15 meters .

Overall, the ECDF highlights the model's strong predictive performance, with a high proportion of predictions having minimal error, and only a few cases with significant deviation.

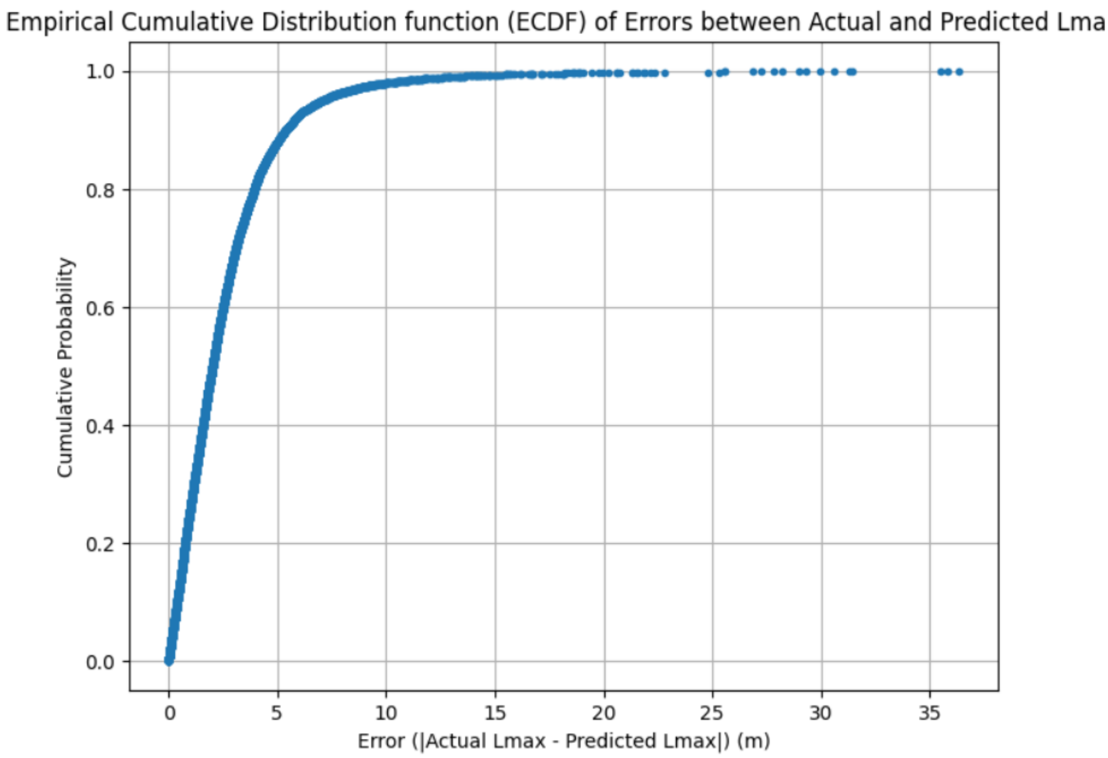


Figure 6.7: ECDF vs ΔL

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this work, a comprehensive correlation analysis was conducted between the -3dB, -5dB, -10dB bandwidths, the equivalent bandwidth and Lmax with BER. The results showed that the equivalent bandwidth had the strongest correlation with Lmax. Initially, the analysis focused on a single laser and fiber across 3766 fibers. This was later extended to the entire dataset, consisting of 8 lasers, 4 wavelengths, and 3766 OM4 fibers. The analysis across various distances revealed that shorter distances, such as 30m and 50m, exhibited a much stronger correlation, reaching up to 96%.

Further analysis was carried out to examine the correlation between the equivalent bandwidth and Lmax at different frequencies, ranging from 50 GHz to 300 GHz. Despite focusing on the equivalent bandwidth, the correlations were not sufficient to develop an accurate predictive model for Lmax based solely on bandwidth data. Consequently, machine learning techniques were employed to improve prediction accuracy.

An ML model was developed to predict the maximum achievable fiber length (Lmax) using four input parameters: the -3dB, -5dB, -10dB bandwidths, and the equivalent bandwidth. The equivalent bandwidth, being an integral measure of the transfer function over a defined range, effectively summarizes the fiber's frequency response. In addition to the equivalent bandwidth, key points from the transfer function were incorporated to provide a comprehensive view of the system's behavior. The ML model achieved an accuracy of 99.98% in predicting Lmax, significantly improving prediction reliability.

This approach has proven to be highly efficient, as it eliminates the need for time-consuming data extraction processes, which would otherwise take days. The ML model enables highly accurate predictions almost instantly, optimizing the time required for network performance evaluation and planning.

7.2 Future Work

While this study presents a promising approach for predicting Lmax using bandwidth parameters, several opportunities for future research remain. Firstly, exploring additional parameters beyond the current bandwidth metrics could potentially further enhance the model's accuracy and generalizability. Identifying other key fiber optic characteristics that may correlate with Lmax and BER would be beneficial, particularly in environments where varying fiber conditions or configurations exist.

Secondly, future work could focus on refining and optimizing the current ML model to improve both speed and accuracy. This might involve experimenting with more advanced machine learning algorithms, such as deep learning models, or employing techniques like hyperparameter tuning and cross-validation to achieve higher precision. Additionally, further efforts could be made to make the model more robust for real-world deployment, ensuring it performs well under diverse conditions encountered in data center environments.

Lastly, extending the research to different types of multimode fiber (MMF) or even single-mode fiber (SMF) could broaden the applicability of the model. Data centers are evolving rapidly, and new fiber technologies or architectures may emerge, each requiring specific adaptations to the ML model. By continuing to explore relevant parameters and configurations, we can enhance the model's predictive power and potentially uncover new insights into the behavior of fiber optics in high-speed data centers.

In conclusion, the model developed in this thesis not only saves time and resources but also opens the door for further optimization and expansion. By leveraging the predictive power of machine learning, we can continue to improve network design, performance analysis, and planning for data center environments, leading to more efficient and scalable systems in the future.

Bibliography

[1] D. Gloge and E. A. I. Marcatalli. «Multimode theory of graded-core fibers». In: *Bell Syst. Tech. J.* 52 (1973), pp. 1563–1578 (cit. on p. 2).

[2] G. P. Agrawal. *Fiber-Optics Communication Systems.* New York: Wiley, 2002 (cit. on p. 2).

[3] H. Li and K. Iga. *Vertical-Cavity Surface-Emitting Laser Devices.* Berlin, Germany: Springer-Verlag, 2003 (cit. on p. 3).

[4] D. M. Kuchta. «High capacity VCSEL-based links». In: *Proc. Opt. Fiber Commun. Conf.* Art. no. Tu3C.4. 2017 (cit. on p. 3).

[5] R. Pimpinella et al. «Dispersion compensated multimode fiber». In: *Proc. 60th Int. Wire and Cables Sym. and Conf.* Nov. 2011, 2011, pp. 1–10 (cit. on p. 3).

[6] J. M. Castro, R. Pimpinella, B. Kose, and B. Lane. «Investigation of the interaction of modal and chromatic dispersion in VCSEL–MMF channels». In: *J. Lightw. Technol.* 30.15 (2012), pp. 2532–2541 (cit. on p. 3).

[7] J. M. Castro, R. Pimpinella, B. Kose, P. Huang, A. Novick, and B. Lane. «Modal-Chromatic dispersion interaction effects for 850 nm VCSEL channels at 100 Gb/s per wavelength». In: *J. Lightw. Technol.* 39.7 (2021), pp. 2067–2076 (cit. on p. 3).

[8] G. Yabre. «Comprehensive theory of dispersion in graded-index optical fibers». In: *J. Lightw. Technol.* 18.2 (2000), pp. 166–177 (cit. on p. 3).

[9] D. Schicketanz. «Fitting of a weighted gaussian lowpass filter to the transfer function of graded-index fibres to reduce bandwidth ambiguities». In: *Electr. Lett.* 19.17 (1983), pp. 651–652 (cit. on p. 4).

[10] K. Ogawa. «Analysis of mode partition noise in laser transmission systems». In: *IEEE J. Quantum Electron.* 18.5 (1982), pp. 849–855 (cit. on pp. 4, 6).

[11] G. P. Agrawal et al. «Dispersion penalty for 1.3-µm lightwave systems with multimode semiconductor lasers». In: *J. Lightw. Technol.* 6.5 (1988), pp. 620–625 (cit. on p. 4).

[12] IEEE. *IEEE 40/100G Ethernet Standard*. IEEE 802.3.ba, Jun. 22, 2010. 2010 (cit. on p. 4).

[13] Fiber Channel Committee of INCITS. *ANSI-INCITS 479 Physical Interface -5 Standard*. Fiber Channel Committee of INCITS, ANSI, Jan. 2011. 2011 (cit. on p. 4).

[14] Telecommunication Industry Association. *TIA Specification for 850-nm Laser Optimized, OM4 MMF*. TIA-492AAAD, Sept. 2009. 2009 (cit. on p. 4).

[15] J. M. Castro et al. «Spectral dependence of multimode fiber modal bandwidth». In: *Proc. Eur. Conf. Opt. Commun.* 2018, pp. 1–3 (cit. on p. 4).

[16] J. A. Tatum et al. «VCSEL-Based interconnects for current and future data centers». In: *J. Lightw. Technol.* 33.4 (2015), pp. 727–732 (cit. on pp. 4, 5).

[17] *100G-SWDM4 MSA Technical Specifications, Optical Specifications*. SWDM MSA Technical Specifications Rev 1.0.1. [Online]. Available: `http://www.swdm.org/wp-content/uploads/2017/11/100G-SWDM4-MSA-Technical-Spec-1-0-1.pdf`. 2017 (cit. on p. 5).

[18] E. Parsons, M. Lanier, R. Patterson, and G. Irwin. «100 G SWDM transmission over 250 m OM5 and OM4+ multimode fibers». In: *Proc. Opt. Fiber Commun. Conf. Expo.* San Diego, CA, USA, 2018, pp. 1–3 (cit. on p. 5).

[19] I. Lyubomirsky et al. «100 G SWDM4 transmission over 300 m wideband mMF». In: *Proc. Eur. Conf. Opt. Commun.* Valencia, Spain, 2015, pp. 1–3 (cit. on p. 5).

[20] Y. Sun et al. «SWDM PAM4 transmission over next generation wide-band multimode optical fiber». In: *J. Lightw. Technol.* 35.4 (2017), pp. 690–697 (cit. on p. 5).

[21] L. Raddatz, I. H. White, and D. G. Cunningham. «An experimental and theoretical study of the offset launch technique for the enhancement of the bandwidth of multimode fiber links». In: *J. Lightw. Technol.* 16.3 (1998), pp. 324–331 (cit. on p. 5).

[22] F. Pampaloni and J. Enderlein. *Gaussian, Hermite-Gaussian, and Laguerre-Gaussian beams: A primer*. [Online]. Available: `http://arxiv.org/abs/physics/0410021v1`. 2004 (cit. on p. 6).

[23] X. Dong, N. Bamiedakis, D. George Cunningham, R. V. Penty, and I. H. White. «A novel equalizer for 112 Gb/s CAP-Based data transmission over 150 m MMF links». In: *J. Lightw. Technol.* 37.23 (2019), pp. 5937–5944 (cit. on p. 6).

[24] J. Lavrencik et al. «4 × 100 Gbps VCSEL PAM-4 transmission over 105 m of wide band multimode fiber». In: *Proc. Opt. Fiber Commun. Conf. Exhib.* Los Angeles, CA, USA, 2017, pp. 1–3 (cit. on p. 6).

[25] F. Karinou, N. Stojanovic, C. Prodaniuc, Z. Qiang, and T. Dippon. «112 Gb/s PAM-4 optical signal transmission over 100-m OM4 multimode fiber for high-capacity data-center interconnects». In: *Proc. 42nd Eur. Conf. Opt. Commun.* Dusseldorf, Germany, 2016, pp. 1–3 (cit. on p. 6).

[26] Pablo Torres-Ferrera, Giuseppe Rizzelli, Antonino Nespola, Jose M. Castro, Bulent Kose, Fabrizio Forghieri, Roberto Gaudino, and Andrea Carena. «Statistical Analysis of 100 Gbps per Wavelength SWDM VCSEL-MMF Data Center Links on a Large Set of OM3 and OM4 Fibers». In: *IEEE Journal on Lightwave Technology* (2024). Senior Member, IEEE (cit. on pp. 16, 17).