UNIVERSITY OF CALIFORNIA
SANTA CRUZ

LEARNING-BASED FRAMEWORKS FOR SPACE OPTICAL COMMUNICATIONS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Abdelrahman Elfikky

December 2024

The Dissertation of Abdelrahman Elfikky
is approved:

Professor Zouheir Rezki, Chair

Professor Hamid Sadjadpour

Professor Hao Ye

Peter Biehl
Vice Provost and Dean of Graduate Studies at UCSC.

Table of Contents

List of Figures

List of Tables

Abstract

Learning-Based Frameworks for Space Optical Communications

by

Abdelrahman Elfikky

This thesis addresses the significant challenges in optical wireless communica-tions in space, which are adversely affected by atmospheric turbulence, light attenuation, and detector noise, leading to degraded communication reliability. To mitigate these issues, this thesis develops a neural network-based channel estimator that is optimized across a wide range of signal-to-noise ratio levels. The proposed estimator achieves performance comparable to the minimum mean square error estimator while maintaining reduced computational complexity. Additionally, a novel autoencoder (AE) framework is introduced, incorporating advanced features such as layer normalization and multiple decoders. These enhancements improve receiver learning capabilities and bit error rate (BER) performance under both perfect and imperfect channel state information (CSI) conditions.

The AE framework presented in this thesis is designed to handle multiple code rates across diverse fading channels, making it a scalable and an adaptable solution for dynamic SOC environments. Furthermore, as the Poisson channel is the most accurate channel model for optical communication, this work addresses the non-differentiability of Poisson SOC channels by integrating the covariance matrix adaptation evolution strategy with AEs, achieving near-optimal BER performance without relying on Gaus-

sian approximations. We also propose enhanced AE designs for medium access control and transport layer settings, utilizing advanced techniques such as formulation layers to balance computational efficiency and performance.

The proposed solutions are evaluated using a system tool kit simulator for a downlink SOC channel connecting a geostationary satellite to a ground station. The results demonstrate that the NN-based channel estimator consistently outperforms state-of-the-art learning-based frameworks and achieves parity with minimum mean square error (MMSE) estimators. Similarly, the AE framework surpasses benchmark methods and popular convolutional coding techniques under both perfect and imperfect CSI conditions with various code rates. Together, the contributions of this thesis represent a significant advancement in the design of low-complexity, high-performance communication systems for space optical communications.

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to my advisor, Professor Zouheir Rezki, for his invaluable mentorship, encouragement, and unwavering support throughout my Ph.D. journey. His profound knowledge and thoughtful guidance have been instrumental in shaping both my research and personal growth, and I am forever grateful for his belief in me and his dedication to my success.

I am also deeply grateful to my committee members, Professor Hamid Sadjadpour and Professor Hao, whose expertise and insights have enriched my work immensely. Professor Sadjadpour's guidance and consistent support have been vital in navigating complex challenges, while Professor Hao's generous assistance and constructive feedback.

To my heroic mother, Mona EL Taraboulsi, and my remarkable father, Ahmed ElFikky—thank you for your boundless love, sacrifices, and for instilling in me the values of perseverance and hard work. To my beloved wife, Mai, and my wonderful children, Youssef and Mariam—your love, patience, and support have been my anchor through every step of this journey. I could not have achieved any of this without each of you by my side. I would like to express my heartfelt gratitude to my friend and undergraduate intern, Jorge Cortez, for his invaluable support in assisting me with both the technical report writings and the hardware project.

This thesis is a reflection of the unwavering support and belief from each of you, and I am honored to dedicate it to you. Thank you, from the bottom of my heart.

List of Publications

1. A. Elfikky, M. Soltani and Z. Rezki, "End-to-End Learning Framework for Space Optical Communications in Non-Differentiable Poisson Channel," in *IEEE Wireless Communications Letters*, vol. 13, no. 8, pp. 2090-2094, Aug. 2024.

2. A. Elfikky and Z. Rezki, "Symbol Detection and Channel Estimation for Space Optical Communications Using Neural Network and Autoencoder," in *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 110-128, 2024.

3. A. Elfikky, M. Soltani and Z. Rezki, "Learning-Based Autoencoder for Multiple Access and Interference Channels in Space Optical Communications," in *IEEE Communications Letters*, vol. 27, no. 10, pp. 2662-2666, Oct. 2023.

4. A. Elfikky and Z. Rezki, "On the Performance of Autoencoder-Based Space Optical Communications," *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, 2022, pp. 1466-1471.

5. A. Alatawi, H. R. Sadjadpour, Z. Rezki, M. Moltafet and A. Elfikky, "Adaptive Deep Neural Network for Non-Stationary Wireless Channels," *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Leeds, United Kingdom, 2024.

Chapter 1

Introduction and Background

1.1 Motivation

Wireless communication has turned out to be a necessity for our day-to-day activities. When transmitting data, most current communication strategies rely on radio frequency (RF) technologies. Bandwidth scarcity is a serious concern due to the restricted RF spectrum and the ever-increasing demand for wireless data. Accordingly, it is essential to also take into consideration higher frequency spectrums such as the optical spectrum for wireless communication. When compared to RF communications, optical wireless communications (OWC) and space optical communications (SOC) offer several benefits over their RF counterparts, including lower transmission power, license-free spectrum, higher throughput, and cost-effective installation [1].

Traditional RF communications have been developed mostly based on models stemmed from first principles relying on physics, wave propagation, communica-

tion, and information theories, etc.. Reliability of this model-based approach has been also strongly supported by sophisticated measurement campaigns throughout the last 60 years. While model-based schemes have demonstrated success in SOC, they have inherent issues as each system component requires individual optimization, resulting in increased complexity of optimizing the transmitter and receiver separately. Therefore, space agencies aim to complement model-based schemes with learning frameworks to address these issues. For instance, advanced learning frameworks have demonstrated potential in decreasing bit error rate (BER) for symbol detection in SOC, albeit additional optimization is necessary to achieve the performance level of capacity-achieving/approaching codes. Also, constructing channel estimators with low complexity and adequate mean square error (MSE) in SOC is a challenging task due to the scarcity of realistic data sets. Some research studies have achieved good MSE results in SOC using learning frameworks, but their designs often involve high-complexity schemes that somehow undermine their applicability.

SOC stands out due to its ability to provide large information bandwidth, low transmitted power requirements, enhanced directionality, and immunity to jamming—critical attributes for modern communication systems. These features have driven global interest, with space agencies worldwide integrating SOC into a variety of practical applications. For instance, SOC systems have been successfully employed for communication between geostationary Earth orbit (GEO) satellites and ground stations. Remarkable demonstrations, such as the Mars laser connectivity project, have achieved data transmission rates of 10 Mbps between Earth and Mars, showcasing the potential of SOC

for interplanetary communication [2]. Additionally, the first-ever two-way optical communication between high-altitude aircraft and a GEO satellite and NASA's Laser Communication Relay Demonstration highlight SOC's transformative impact on near-Earth and deep-space missions [3].

One of the fundamental aspects of SOC is its reliance on intensity modulation and direct detection (IM/DD), a modulation technique that simplifies implementation while eliminating the need for more complicated coherent detection [3]. Laser diodes modulate the data by controlling light intensity, ensuring a proportional relationship between the transmitted signal and light intensity. At the receiver, a photodiode detects the signal and converts it directly into a current. However, SOC systems face unique challenges, such as the need for high photon efficiency and peak power capability in long-distance laser transmissions to achieve acceptable BER. The design of lasers with narrow linewidths, high beam quality, and low modulation rates is critical to overcome these hurdles [3]. Moreover, while SOC shares some characteristics with OWC, such as using lasers as optical transmitters, it operates under distinct constraints. For downlink channels, geometric beam divergence leads to signal loss, but atmospheric effects are relatively minimal, given the predominantly non-atmospheric propagation path. In contrast, uplink channels face significant challenges due to atmospheric turbulence, spatial and temporal fluctuations, and pointing instability caused by refractive index variations. During satellite-to-ground downlink transmissions, beam divergence loss dominates, with only minor beam steering variations contributing to additional loss. Notably, atmospheric turbulence effects are generally small for downlink propagation,

as the beam traverses a non-atmospheric path until reaching approximately 30 km from the Earth's surface [3].

In conclusion, the advancements in SOC are not merely theoretical pursuits—they are key enablers of future communication networks that span beyond Earth, offering unprecedented opportunities for secure, reliable, and high-capacity communication in both near-Earth and deep-space environments.

Researchers are increasingly exploring the use of deep learning (DL) techniques in physical layer communication networks. This integration has led to significant progress in coding and decoding, modulation classification, channel estimation, and equalization [4–8]. AEs have demonstrated strong performance as an end-to-end (E2E) approach in OWC, particularly in point-to-point SOC. A study on AE-based OWC systems [6] showed that AEs can outperform hamming codes for BER, even in the presence of imperfect channel state information (CSI). In [7], AEs were shown to optimize both transmitter and receiver components, achieving satisfactory BER performance across interference channel (IC).

1.2 Related State of the Art

In this section, we explore the latest advancements in symbol detection for SOC systems. Symbol detection plays a crucial role in ensuring reliable and efficient data transmission in SOC, where unique challenges such as atmospheric turbulence, signal attenuation, and pointing errors often degrade signal quality. Traditionally, sym-

bol detection in OWC systems has been achieved through classical methods, including channel modeling, modulation and coding techniques, and channel estimation. However, with the advent of machine learning (ML) and DL technologies, AI-based methods such as AEs and deep neural networks (DNNs) are now being leveraged to enhance symbol detection accuracy and robustness. These approaches aim to address the complexities of SOC by adapting to dynamic communication environments, providing improved performance over conventional methods.

We will briefly explain how these recent AI-driven techniques have significantly advanced symbol detection capabilities in SOC systems. In particular, end-to-end learning frameworks, including AEs, have emerged as powerful tools for structuring NNs tailored to the specific challenges of SOC, such as noise and turbulence. In addition, attention-based models and convolutional neural networks (CNNs) are used to improve channel estimation and noise suppression. These DL techniques, along with novel hybrid model-based methods, demonstrate promising results in reducing BER and achieving more resilient communication links in SOC. This shift towards AI-centric solutions marks a substantial progression in SOC symbol detection, aligning with the growing demands for more reliable and adaptive communication networks across space and terrestrial applications.

1.2.1 Symbol Detection

There exists a wide body of work related to OWC in general and SOC in particular. This work can be categorized mainly in the following areas: channel modeling,

modulation, and coding, channel estimation, and learning-based design leveraging artificial intelligence (AI) methods such as autoencoder (AE) and/or DNN [7, 9]. Since the scope of the current work contributes to all these areas, we briefly overview the most notable related state-of-the-art next.

1.2.1.1 Space Optical Channel Model

SOC channels are subject to a unique set of impairments such as pointing errors, beam divergence, atmospheric turbulence, and scattering which are not present in terrestrial channels. Pointing errors arise due to misalignments between the transmitter and receiver, often caused by platform vibrations, mechanical jitter, or atmospheric turbulence-induced beam wander. This misalignment reduces the received signal power and increases the bit error rate. Effective acquisition, tracking, and pointing systems are critical to maintaining alignment, particularly for long-range systems such as satellite-to-ground links in SOC. The severity of pointing errors is statistically modeled using distributions like Gaussian or Rician to predict and mitigate their effects [10]. Beam divergence occurs because as the optical beam propagates through the atmosphere, its energy spreads, and the power density at the receiver becomes reduced. While a narrow beam can minimize divergence losses, it requires precise alignment to avoid link failure due to pointing errors. Conversely, a wide beam reduces alignment sensitivity but increases geometrical losses. Balancing beam divergence involves careful consideration of the system's link distance, required reliability, and transmitted power [10]. Atmospheric turbulence and scattering add further complexities to SOC systems. Tur-

bulence is caused by temperature and pressure variations along the propagation path which leads to scintillation, beam wander, and spreading, significantly degrading signal quality. Typically, the Gamma-Gamma distribution is used to model the strong atmospheric turbulence regime, and the Log-normal distribution is used to model the weak turbulence regime. Scattering, whether Rayleigh or Mie, depends on the size of atmospheric particles, such as rain, fog, or dust [11] that cause additional attenuation and signal distortion. These impairments necessitate advanced mitigation strategies such as adaptive optics, diversity techniques, and channel coding to ensure robust communication over turbulent atmospheric channels.

In [12], the authors integrated a hybrid RF/FSO lunar communications system that employed micro satellites in a Low Earth Orbit (LEO) constellation for communications with lunar missions. During this implementation, the channel modeling for the entire system is performed in the Analytical Graphics System Tool Kit (STK) simulator. Moreover, the STK program allows the accessibility to the propagation delay, transmission loss, and signal-to-noise ratio (SNR) measurements. Furthermore, the STK program is utilized to configure two ground stations and two satellites for point-to-point communications in order to create an SOC system [13]. The authors in [14] consider utilizing a Log-normal distribution for OWC to accurately represent the atmospheric modeling in weak turbulence regime. On the other hand, the Gamma-Gamma distribution is more suitable for strong turbulence regime considering weather attenuation, random fluctuations in satellite position/orientation and atmospheric turbulence [15]. The authors in [16] proved that double Generalized Gamma distribution

is an appropriate statistical model to represent the irradiance fluctuations in strong and weak turbulence regimes for OWC, as it generalizes many existing turbulence channel models and fits excellently to the published plane and spherical waves simulation data. On the contrary, laser beam pointing errors arise when the transmitter and receiver are in motion, an accurate acquisition, tracking, and pointing system (APT) is necessary for proper reception of the signal in inter-satellite communication [17]. In the down-link SOC channel, the pointing error can be easily mitigated due to the capability and stability of the ground station [3].

1.2.1.2 Modulation and channel coding

Coherent communication techniques involving modulation and detection of the amplitude and phase of the optical carrier can be used for SOC. However, incoherent modulation as IM/DD is preferred due to its simplicity, cost-effectiveness, and ease of implementation while having an input constraint of real and positive-valued modulating signal (electrical current) [18]. It has been shown that the modulation scheme generated from the AEs-based OWC in [19] and [6] has a similar output constellation as the IM/DD when trained on the AWGN channel with and without fading conditions. On the other hand, for increasing the number of accessible modes in limited optical communication systems, the authors in [20] propose fractional modulation of spatial modes of Bessel-Gaussian laser beams. To accomplish high-resolution identification of fractional modes, a convolutional NN decoder is specifically used. Narrowing down to channel coding schemes in SOC, the convolutional codes have been shown to outperform the Hamming

and Bose-Chaudhuri-Hocquenghem (BCH) linear block codes for various code rates while maintaining the same order of complexity [21]. Authors in [6, 19] applied the channel coding schemes via deep learning (DL) AEs and achieved similar performance to the Hamming codes in OWC. Instead of adding redundant bits as conventional coding schemes, researchers utilize the AEs by applying the compression at the encoder and expansion at the decoder.

1.2.1.3 Channel estimation

Attention-based models have emerged as a transformative paradigm in deep learning, making notable inroads into various domains. Particularly in the realm of channel estimation, attention mechanisms have shown the potential to address some challenges in communication systems [22–24]. Authors in [22] proposed the Channelformer, a neural framework tailored for enhanced orthogonal frequency-division multiplexing (OFDM) channel estimation in downlink scenarios. This model capitalizes on self-attention for input precoding and seamlessly integrates multi-head attention with residual convolution. Alongside this, they have incorporated a novel weight pruning technique, driving the architecture towards a leaner, high-performance, low-latency solution that reduces parameters by up to 70%. In addition, authors in [24] put forth a non-local attention methodology explicitly for OFDM channel estimation in a Multiple Input, Multiple Output (MIMO) system. This NN-centric approach harnesses frequency-specific attributes in a data-driven strategy, paving the way for optimized pilot design and more accurate channel estimation compared to linear minimum mean

square error (LMMSE) estimation. On the other hand, for slow-fading channels where channel characteristics largely remain steady over time, the utilization of attention mechanisms may not be needed given the low number of pilots needed and the absence of interpolation schemes. Communications systems that rely on least square (LS) channel estimators tend to perform poorly in the low SNR regime [25]. This poor performance is due to the fact that the LS estimation process does not suppress the effect of noise. Compared to LS, minimum mean square error (MMSE) channel estimator mitigates the noise effect and achieves the optimal performance, in terms of mean square error (MSE) [26]. However, MMSE channel estimation requires computing the cross covariance matrix between the received signal and the time-domain channel, thus inducing an increased complexity [26]. To undertake this issue, the authors in [27] proposed a DL-enabled image denoising network to acquire knowledge from a huge set of training data and to compute an estimate of the massive MIMO visible light communication (VLC) channel, where the channel matrix is identified as a two-dimensional natural image since the channel has sparsity. Furthermore, it was shown in [28] that a NN with one hidden layer and sigmoid activation functions can be trained to get an accurate CSI estimates in Log-normal fading conditions. However, the system therein is not practical as it needs a separate NN for every training SNR. In [25], the authors propose employing only one NN to rectify the LS estimation error. The results in [25] show that their NN design outperforms LS estimator but it is simpler in implementation compared with [28]. Despite their accurate CSI prediction results, the authors in [25] relied on an unrealistic assumption that all the input samples are already known in advance for the testing

phase. This assumption will lead to significant delay in the processing of the signal in the wireless communication system. The design of channel estimator NN should have adequate performance on every code word, to fulfill the real-time requirement of 5/6G.

1.2.1.4 End-to-end communication systems

AEs are considered as unsupervised DNNs where the input and predicted output aim to be identical. The input is transformed into a compressed code referred to as the latent space, using the end-to-end learning concept, which can then be used to reconstruct the input data [29]. In the landmark paper [7], the AE-based communications system was introduced and showed adequate performance compared to uncoded modulations employing maximum likelihood detector. Their approach considered single and multi-user communications over fading RF channel, and led to the large amount of work in end-to-end communications systems that followed. In [7], the authors have shown that it is feasible to create a point-to-point communications system in which NNs handle all of the physical layer computation over a practical channel. Training a system as an AE is a good approach for any stochastic channel model; nevertheless, substantial effort is needed before the system can be employed for transmission over the air [9]. On the other hand, Turbo AE is a fully end-to-end cooperatively trained neural encoder and decoder inspired by the turbo coding scheme, and its performance under canonical channels is close to that of the convolutional codes when using small block lengths [30]. The authors in [29], proposed Turbo AE with average power constraints instead of peak intensity constraints required by OWC in general and SOC in partic-

ular for THz communications. In OWC systems, performance of the AEs has shown comparable performance to Hamming codes in point-to-point communications [19]. It should be noted that the study in [19] only assumed the presence of an additive white Gaussian noise (AWGN) channel and did not investigate the performance of AE in fading channels. The authors in [6] expanded the work in [19] and incorporated the turbulence channels, resulting in a performance that is comparable to that of Hamming codes using the MMSE estimator for both perfect and imperfect CSI. The MMSE estimator can be used with AEs, although this strategy would increase system computation complexity [26]. On the other hand, the DL models created in [6, 19, 26] perform worse in terms of BER than convolutional codes.

1.2.2 Non-differentiable Channels

In communication systems, deep learning-based AEs are gaining attention for their potential to jointly optimize the transmitter and receiver components in an end-to-end learning framework. A key requirement for training these systems using gradient-based methods, such as backpropagation, is the differentiability of all system components, including the channel model. However, certain practical channels, such as those characterized by Poisson statistics or with quantization effects, are inherently non-differentiable. This non-differentiability poses a significant challenge, as it prevents the straightforward application of gradient-based optimization techniques essential for training.

The challenge of non-differentiable channels has prompted the development of

alternative approaches to enable effective training under these conditions. Some methods attempt to approximate the non-differentiable channel with a differentiable model, such as using a Gaussian distribution to approximate a Poisson channel. While effective in specific scenarios, these approximations can fail to generalize across varying channel conditions. Recent advances, including model-free and reinforcement learning techniques, have been proposed to bypass the need for differentiable channels, enabling end-to-end optimization by approximating gradients or by using non-traditional training algorithms. These advancements pave the way for training AEs in real-world communication environments where non-differentiability is a critical factor. When training AEs, both the channel model and all the layers of the AE must be differentiable. This poses a challenge, as some channel models are non-differentiable. In SOC, the received optical signal is often very weak, leading to the consideration of photon counting statistics. The Poisson distribution accurately represents the probability distribution of the number of photon detections in a given time period. One downside is that it cannot be implemented as an AE channel on its own due to its non-differentiability. Nevertheless, a significant portion of DL literature tends to avoid considering the Poisson channel [1, 6, 19, 31] due to its non-differentiable nature, making it impractical for calculating gradients during the backpropagation process.

In [32], the authors address the non-differentiability of the Poisson channel by approximating it with a Gaussian distribution, which enables the use of gradient-based methods for training. This approach leverages the Central Limit Theorem, whereby the sum of a large number of independent Poisson-distributed events can be approximated

by a Gaussian distribution, especially in high photon-counting scenarios. The Gaussian approximation provides a differentiable substitute, allowing AE-based communication systems to train on channels that would otherwise be inaccessible due to the non-differentiable nature of the Poisson process.

However, this Gaussian approximation has limitations. While it works well under high photon-count conditions, where the Gaussian model aligns closely with the Poisson distribution, it becomes less accurate in low photon-count scenarios, which are common in low-signal or long-distance communication environments. In these cases, the discrete, photon-based nature of the Poisson process dominates, and the continuous Gaussian model fails to capture the true channel characteristics. As a result, using a Gaussian approximation may lead to suboptimal training and degraded system performance in scenarios that deviate significantly from the assumptions required for the approximation, highlighting the need for training methods that do not rely on such approximations.

1.2.3 Multiple Access/Interference Channels

Multiple Access Channels (MAC) and Interference Channels (IC) are two key multi-user communication scenarios where multiple users transmit information over a shared medium, leading to interference and competing demands on the channel. In a MAC, several users transmit to a single receiver, necessitating strategies for managing interference and resource allocation to maintain reliable communication for all users. Interference channels, on the other hand, involve multiple transmitters and receivers,

where each receiver is interested in the message from its corresponding transmitter, but may also receive interfering signals from other transmitters. Addressing interference and optimizing transmission and reception in both MAC and IC scenarios is crucial for enhancing the overall network performance, especially in terms of BER and computational efficiency.

MAC: A common scenario in wireless communications is when multiple transmitters send information to a single receiver, such as with the uplink between mobile phones and a base station, known as the multiple access channel (MAC). At the receiver's end, the received signal is represented as:

$$y = h_1 x_1 + n_1 + h_2 x_2 + n_2 + \cdots + h_m x_m + n_m \tag{1.1}$$

In order to decode each user's individual data bits, access schemes must be employed. Traditionally, schemes such as frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA) were used, followed by orthogonal multiple access (OMA) schemes like orthogonal frequency division multiple access (OFDMA) and single-carrier FDMA (SC-FDMA). Such orthogonal designs such as OFDMA and SC-FDMA provide the advantage of having no mutual interference among the receiving user equipment (UE), meaning it is simple to decode their respective messages with low error. In recent years however, much research has been done on non-orthogonal multiple access schemes (NOMA) in order to share the same resources in time, frequency, and/or code in order to acheive better spectral effi-

ciency. Generally, NOMA schemes are one of two sub-schemes: power-domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA). In PD-NOMA, different users' signals are transmitted with at different power levels, allowing for an iterative decoding process at the receiver.

IC: When two transmitter-receiver pairs are communicating at the same frequencies, the respective messages can interfere with each other at the receivers. The received signals at the two receivers can be represented as:

$$\begin{aligned} y_1 &= h_{11}x_1 + h_{12}x_2 + n_1 \\ y_2 &= h_{22}x_2 + h_{21}x_1 + n_2 \end{aligned} \tag{1.2}$$

where y_1, y_2 are the received signals at Receiver 1 and 2, respectively, x_1, x_2 are the respective transmitted signals, h_{11}, h_{22} are the channel coefficients for the direct paths between each transmitter-receiver pair, h_{12}, h_{21} are the channel coefficients for the cross paths, and n_1, n_2 is the additive noise at the receivers. This issue of multi-user interference becomes prevalent when dealing with dense wireless networks that have multiple transmitter-receiver pairs in close proximity to one another and are using the same frequency band.

Recent work has demonstrated the potential of AEs to optimize both the transmitter and receiver components in these multi-user settings. In [7], AEs achieved satisfactory BER performance in an interference channel by jointly optimizing the components across the IC. Similarly, [19] explored the feasibility of AEs in OWC systems over a MAC channel, where AEs were compared with traditional time-sharing methods.

Additionally, [33] introduced a sparse AE, incorporating sparsity regularization to improve BER performance while reducing computational complexity in the MAC setting. However, while these AE-based approaches show promise, they each have certain limitations. For instance, [19] restricts the AE model to an AWGN channel without fading, limiting its generalizability to real-world IC scenarios. In [33], although computational efficiency is prioritized through sparsity, the BER performance does not yet surpass that of model-based schemes. These challenges highlight the need for further refinement in AE designs to enhance both BER performance and computational efficiency across varying multi-user channels.

1.3 Thesis outline

The outline of this thesis is shown in Fig. 1.1 and is organized as follows: Chapter 2 covers the fundamentals of DL and AEs. Chapter 3 introduces the proposed AE model for symbol detection and channel estimation in SOC, along with results for a multi-code rate AE model specific to SOC. Chapter 4 presents the proposed AE model tailored for non-differentiable channel models. Chapter 5 details the AE model designed for multiple access and interference channels in SOC. Lastly, Chapter 6 concludes the thesis with a summary of findings and suggestions for future research directions.

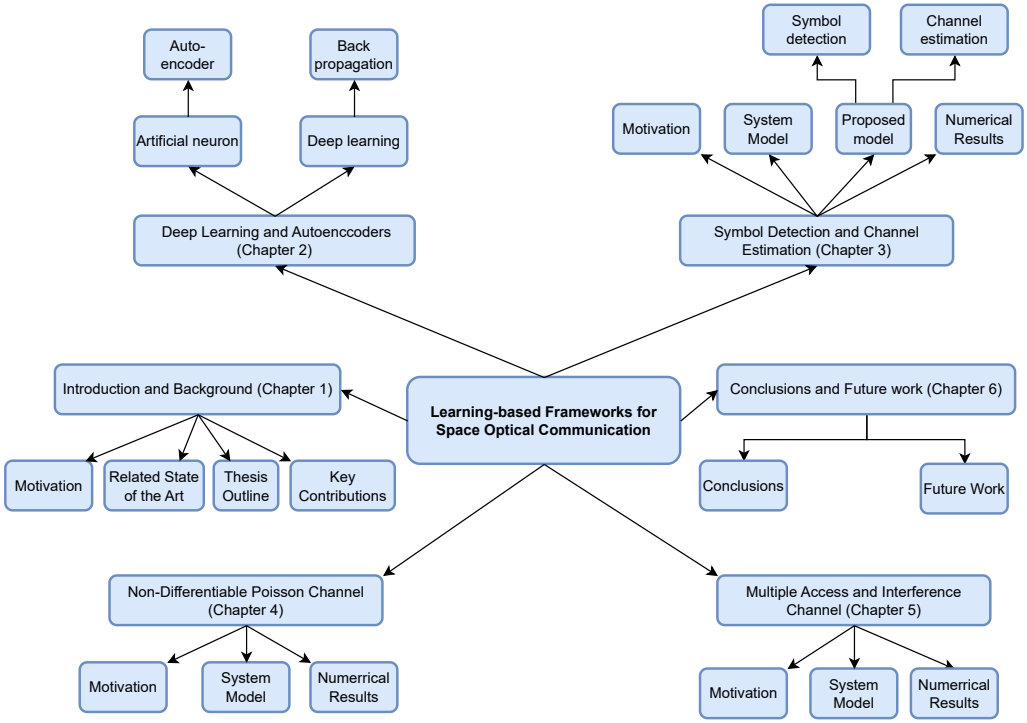


Figure 1.1: Orgnizational outline of the thesis.

1.4 Key contributions

The work in the thesis presents significant advancements in machine learning frameworks for SOC systems, focusing on low-complexity, high-performance solutions for symbol detection and channel estimation. By leveraging NNs and AE designs, this research addresses challenges in non-differentiable channels, scalability, and adaptability across various fading conditions. Key innovations include robust NN-based channel estimators, enhanced AE frameworks with normalization techniques, and novel approaches to optimize performance under imperfect CSI. These contributions collectively demonstrate improved BER performance, computational efficiency, and scalability in SOC

environments, pushing the boundaries of reliable data transmission. A summary of key contributions are listed below:

- Developed a single NN-based channel estimator robust across a range of SNRs, achieving performance comparable to the minimum mean square error (MMSE) estimator while simplifying complexity [34] (Chapter 3).
- Proposed a novel AE framework incorporating layer normalization for encoders and decoders, achieving superior BER performance under both perfect and imperfect CSI conditions [1, 34] (Chapter 3).
- Combined the proposed NN-based channel estimator and AE frameworks to deliver an integrated, low-complexity deep learning solution for symbol detection and channel estimation in SOC [34] (Chapter 3).
- Designed an AE framework capable of training and handling multiple code rates simultaneously across diverse fading channels, optimizing scalability and adaptability for SOC environments (Chapter 3).
- Addressed non-differentiability in Poisson SOC channels by integrating covariance matrix adaptation evolution strategy (CMA-ES) with AEs, achieving near-capacity BER performance without Gaussian approximations [35] (Chapter 4).
- Enhanced AE designs for MAC and IC settings, introducing normalization layers to significantly improve BER performance while maintaining computational efficiency [31] (Chapter 5).

Chapter 2

Deep Learning Autoencoders

2.1 Artificial neural networks

Neural networks (NNs) are the best known models in deep learning, however only NNs with a high enough number of 'hidden' layers can be regarded as 'deep' NNs (DNNs). DNNs allow for automatic feature extraction, and by adding more layers and more units, or neurons, within a layer, a deep NN can represent functions of increasing complexity [36, Ch. 6].

Feed-forward NNs (FNNs) form the basis of deep learning NNs. A FNN with L layers describes the mapping of an input vector $\mathbf{x}_0 \in \mathbb{R}^{N_0}$ to an output vector $\mathbf{x}_L \in \mathbb{R}^{N_L}$ given by $f(\mathbf{x}_0; \theta) : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$, where $\boldsymbol{\theta}$ is the set of all the parameters in the network $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_L\}$. The FNN computes the mapping for each layer ℓ iteratively:

$$\mathbf{x}_\ell = f_\ell(\mathbf{x}_{\ell-1}; \theta_\ell), \quad \ell = 1, \ldots, L \tag{2.1}$$

where (2.1) is the mapping carried out at the ℓth layer. The parameters for the ℓth layer is $\theta_\ell = \{\mathbf{W}_\ell, \mathbf{b}_\ell\}$, where $\mathbf{W}_\ell$ is the matrix representing the weights of the connections between neurons of layers $\ell - 1$ and ℓ, and $\mathbf{b}_\ell$ is a vector representing the bias of each neuron in the ℓth layer.

The goal of the FNN is to select a mapping function f that approximates an ideal mapping $f^\star$, learning the parameters θ that outputs a predicted value $\mathbf{x}_L$ that is as close to the ideal output $\mathbf{x}_L^\star$ as possible, known as minimizing the network's loss $L(\theta)$ which will be discussed below.

The FNNs hidden layers lie between the input layer ($\ell = 1$) and output layer ($\ell = L$), and they are what apply the transformations and feature extraction from the input vector $\mathbf{x}_0$. The number of hidden layers in the FNN is known as the 'depth' of the network, and generally speaking, having more hidden layers provides a greater ability for extracting higher-level information from the input data. However, having unnecessary hidden layers can lead to overfitting the model, meaning that it will learn the training data well, but will not be able to generalize to new, unseen data. Overfitting is observed as having a large difference between the training set error and the test set error [36, Ch. 5], and will be discussed later in this section.

2.2 Deep learning

2.2.1 Multilayer perceptron

Multilayer Perceptrons (MLPs) represent one of the foundational structures in deep learning, essential for nonlinear mappings between input and output data. Specifically, MLPs are employed in SOC for optimizing nonlinear functions associated with signal processing tasks such as channel estimation, symbol detection, and signal regeneration. This section delineates the mathematical structure and underlying principles of MLPs and their applicability to SOC systems.

An MLP is structured as a fully connected feedforward neural network composed of an input layer, multiple hidden layers, and an output layer. Given an input vector $\mathbf{x} \in \mathbb{R}^n$, the MLP model seeks to approximate the target function through a series of nonlinear transformations parameterized by weight matrices and bias vectors.

For a L-layer MLP with d_l neurons, the output of each layer l is computed as:

$$\mathbf{a}^{(l)} = \sigma\left(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}\right), \tag{2.2}$$

where $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$ and $\mathbf{b}^{(l)} \in \mathbb{R}^{d_l}$ are the weight matrix and bias vector of layer l, respectively, $\sigma(\cdot)$ denotes the non-linear activation function, and $\mathbf{a}^{(l-1)}$ represents the output (or "activation") of the previous layer. The output layer activation, $\mathbf{a}^{(L)}$, is the final output of the MLP.

The activation function $\sigma(\cdot)$ is critical for introducing non-linearity into the

model, allowing the MLP to capture complex functional mappings. Typical choices include the Rectified Linear Unit (ReLU):

$$\sigma(\mathbf{z}) = \max(0, \mathbf{z}), \tag{2.3}$$

or other functions such as the sigmoid or hyperbolic tangent (tanh), depending on the nature of the optimization problem.

2.2.2 Backpropagation and Gradient Descent Optimization

The MLP is trained by minimizing a loss function $\mathcal{L}$, which quantifies the discrepancy between the predicted output $\hat{\mathbf{y}} = \mathbf{a}^{(L)}$ and the true output $\mathbf{y}$ over N training samples. For regression-based tasks typical in physical layer optimization, MSE is a common choice:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 . \tag{2.4}$$

Gradient descent optimization is applied to adjust the weights and biases in each layer. The gradient of the loss with respect to the weights $\mathbf{W}^{(l)}$ and biases $\mathbf{b}^{(l)}$ in layer l is computed using the backpropagation algorithm, formulated as follows:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} = \frac{\partial \mathcal{L}}{\partial \mathbf{a}^{(L)}} \cdot \prod_{k=l+1}^{L} \sigma'(\mathbf{z}^{(k)}) \mathbf{W}^{(k)} \cdot \sigma'(\mathbf{z}^{(l)}) \mathbf{a}^{(l-1)T}, \tag{2.5}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} = \frac{\partial \mathcal{L}}{\partial \mathbf{a}^{(L)}} \cdot \prod_{k=l+1}^{L} \sigma'(\mathbf{z}^{(k)}) \mathbf{W}^{(k)} \cdot \sigma'(\mathbf{z}^{(l)}), \tag{2.6}$$

where $\sigma'(\cdot)$ denotes the derivative of the activation function. This chain of gradients

enables efficient weight updates for each layer during training, leading to convergence on the optimal weights for signal processing.

The backpropagation algorithm plays a crucial role in efficiently computing the gradients needed for training, making it one of the most important innovations in neural network optimization. The gradients calculated for each layer are influenced by the activation functions used in the network, with commonly used functions such as ReLU and sigmoid offering distinct trade-offs. ReLU, for instance, introduces sparsity in the network by zeroing out negative values, which can improve convergence speed but may suffer from the "dying neuron" problem if gradients become permanently zero. On the other hand, the sigmoid function provides smooth gradients but can lead to vanishing gradients in deep networks. Addressing these issues through careful selection or design of activation functions is crucial for ensuring effective signal processing in SOC tasks, where computational resources and performance constraints are stringent. The efficiency and robustness of the MLP training process are further enhanced by employing techniques like regularization and normalization. Regularization methods, such as L1/L2 penalties or dropout, help prevent overfitting by constraining the network's complexity or introducing stochasticity during training. Batch normalization, on the other hand, normalizes layer inputs, stabilizing the training process and enabling the use of higher learning rates. Together, these strategies ensure that the MLP generalizes well to unseen data, making it a reliable tool for SOC-based signal optimization tasks.

2.2.3 Autoencoder

While Multilayer Perceptrons (MLPs) are powerful for modeling complex non-linear mappings, they lack the inherent structure required to compress and reconstruct high-dimensional data efficiently. In the context of telecommunications, where the optimization of signal representation and noise robustness is paramount, an alternative architecture is necessary to derive low-dimensional representations that retain essential signal features. AEs provide such a framework, excelling in capturing latent structures within data, which is crucial for tasks in wireless communications.

As shown in Fig. 2.1, an AE is an unsupervised neural network model designed to learn a compressed representation of input data. It consists of two primary components: an *encoder* network that maps the input $\mathbf{x} \in \mathbb{R}^n$ to a lower-dimensional latent space $\mathbf{z} \in \mathbb{R}^k$ (with $k < n$), and a *decoder* network that reconstructs the original input from this latent space. Formally, the encoder and decoder functions are defined as:

$$\mathbf{z} = f_{\text{enc}}(\mathbf{x}) = \sigma\left(\mathbf{W}_{\text{enc}}\mathbf{x} + \mathbf{b}_{\text{enc}}\right), \tag{2.7}$$

$$\hat{\mathbf{x}} = f_{\text{dec}}(\mathbf{z}) = \sigma\left(\mathbf{W}_{\text{dec}}\mathbf{z} + \mathbf{b}_{\text{dec}}\right), \tag{2.8}$$

where $\mathbf{W}_{\text{enc}}$, $\mathbf{b}_{\text{enc}}$, $\mathbf{W}_{\text{dec}}$, and $\mathbf{b}_{\text{dec}}$ are the weights and biases of the encoder and decoder, respectively, and $\sigma(\cdot)$ is the activation function. The optimization objective for an AE is to minimize the reconstruction error, typically formulated as the MSE between the

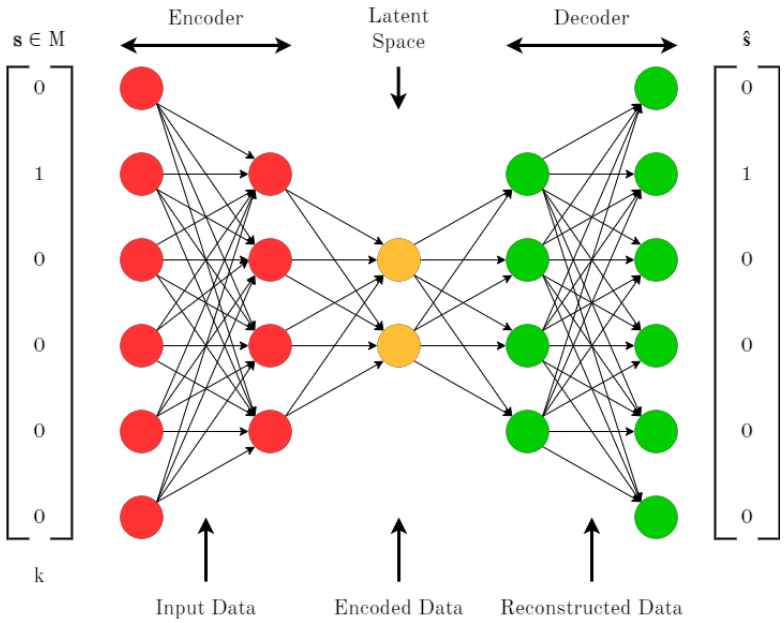


Figure 2.1: Structure of Standard AE in wireless Communications

input $\mathbf{x}$ and the reconstructed output $\hat{\mathbf{x}}$:

$$\mathcal{L}_{\mathrm{AE}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \,. \tag{2.9}$$

In SOC, AEs enable effective compression and feature extraction by identifying latent representations that are resilient to noise and channel distortions. This capability is particularly advantageous for channel equalization, where high-dimensional received signals may be projected into a low-dimensional space that preserves essential signal features while discarding noise. Additionally, the encoder-decoder structure allows the system to learn representations that are inherently robust, a critical factor in mitigating the effects of atmospheric turbulence, scintillation, and signal fading in optical communication channels.

2.2.4 End-to-End physical layer learning with Autoencoders

When integrated into an end-to-end framework, AEs can optimize both the transmitter and receiver parameters by structuring the latent space such that it minimizes channel-induced degradation. In a typical setup, the encoder network functions as a virtual transmitter, compressing the data prior to transmission, while the decoder serves as a virtual receiver, reconstructing the data post-channel. This approach not only improves data fidelity but also allows for adaptive modulation, where the latent space can dynamically adjust based on channel conditions.

By leveraging AEs in this capacity, the system can achieve a more effective end-to-end optimization of the physical layer, yielding superior performance in terms of signal fidelity, spectral efficiency, and resilience to channel impairments compared to standard MLP-based models.

Chapter 3

Symbol Detection and Channel Estimation for Space Optical Communications Using Neural Network and Autoencoder

3.1 Motivation

This chapter addresses three critical challenges in SOC systems, presenting novel contributions to advance their efficiency, adaptability, and scalability. One of the significant challenges in SOC systems is designing a NN-based channel estimator that performs as well as the MMSE estimator. Traditional approaches typically require creating separate NNs for each SNR level, which increases system complexity and computational overhead. This approach is impractical in dynamic communication scenarios

where channel conditions vary frequently. A scalable and efficient channel estimator is essential for maintaining reliable communication under such conditions. To address this, we propose a novel NN-based channel estimator trained for a single SNR. This solution eliminates the need for multiple NNs, reduces complexity, and ensures adaptability, achieving performance comparable to the MMSE estimator while being more efficient for real-world applications.

Another major challenge in traditional model-based communication schemes is the independent optimization of the transmitter and receiver. While effective in specific scenarios, this approach often results in higher system complexity and increased latency due to the iterative processes required to align the two components. This limitation becomes especially problematic in SOC systems operating under atmospheric turbulence, where rapid adjustments are necessary to maintain communication quality. In contrast, joint optimization of the transmitter and receiver offers a more streamlined approach. By treating the communication chain as a single, unified system, joint optimization reduces complexity, lowers processing delays, and improves adaptability to turbulence-induced impairments. In this chapter, we extend the principles discussed in Chapter 2 to develop an AE-based communication system. This end-to-end learning framework leverages deep learning techniques to enable joint optimization, providing robust performance under severe turbulence while significantly enhancing efficiency.

The third challenge lies in achieving scalability for multi-code rates. Traditional systems often require separate designs for each code rate, necessitating distinct AEs for different rates. This approach increases system complexity and limits scala-

bility, particularly in multi-user environments or scenarios requiring dynamic data rate adjustments. Scalability is critical for SOC systems to handle varying operational demands efficiently. To overcome this, we propose a single AE capable of operating across multiple code rates simultaneously. This unified design reduces complexity, enhances flexibility, and ensures efficient operation in dynamic communication environments, offering a practical solution for multi-user SOC systems.

By addressing these challenges, this chapter provides a comprehensive framework that advances the state of the art in SOC systems. Our contributions include a robust NN-based channel estimator for single SNR optimization, an AE-based joint optimization approach for the transmitter and receiver, and a scalable AE design for multi-code rates. Together, these innovations enable efficient, adaptable, and scalable solutions for the unique challenges of SOC systems operating under real-world conditions.

3.2 SOC channel model

We define the point-to-point downlink channel between GEO satellite and a ground station. Following this, we will describe a separate setup for a downlink channel between a LEO satellite and a ground station. The STK simulator facilitates precise channel modeling for the point-to-point SOC channels [13, 37, 38]. In the system, the ground station holds the receiver antenna gimbal and avalanche photo-detector. Additionally, the GEO satellite holds the laser transmitter and the gimbal for the transmitter

antenna. The gimbal system can be used to support and stabilize transmitters and receivers. The laser transmitter is modeled as a Gaussian beam model. The laser utilizes IM/DD, where the light intensity is modulated as an information-carrying signal, with data recovery accomplished by the detection of incoming light intensity. In addition, the generated modulating signal (current) is real and positive as a result of this procedure. This is a significant difference from RF coherent communications, where the modulated signal is complex-valued [18]. Furthermore, the modulated signal in IM/DD is peak-constrained for reasons of operation, safety, and illumination [18].

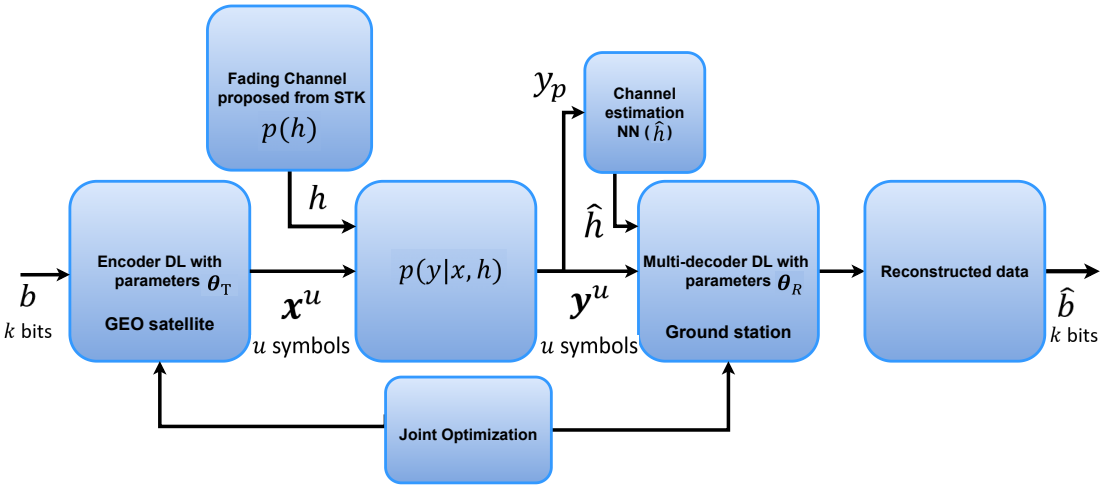


Figure 3.1: An overview of the system implementation for symbol detection and channel estimation for SOC channel. The transmitter at the GEO satellite employs an encoder based-GEO satellite to convert a stream of k bits $\boldsymbol{b}$ into a codeword $\boldsymbol{x}^u$ of u coded symbols. The encoded vector $\boldsymbol{x}^u$ satisfies the positivity and peak criterion conditions. The first symbol $\boldsymbol{x}^u(1) = x_p$ is assumed a pilot, which passes over a Log-normal fading channel verified from STK. At the receiver side (the ground station), the proposed channel estimator NN($\hat{h}$) utilizes the first element $\boldsymbol{y}^u(1) = y_p$ of the received sequence $\boldsymbol{y}^u \in \mathbb{R}^u$ in order to retrieve an estimated version of the channel gain $\hat{h}$. Afterwards, the muti-decoder AE makes use of $\hat{h}$ and the received sequence $\boldsymbol{y}^u$ to derive an estimate for the transmitted symbols $\hat{\boldsymbol{x}}^u$ and hence the recovered message $\hat{\boldsymbol{b}}$.

The Log-normal distribution is typically used to describe the weak atmospheric

turbulence regime and is the best distribution fitting that STK has recommended for the GEO to ground SOC channel. Changes in atmospheric temperature and pressure at various points along the signal's propagation are the cause of atmospheric turbulence [39]. The probability density function (PDF) for the Log-normal distribution of the channel gain is given by [14]

$$f_h(h) = \frac{1}{h\sigma_l\sqrt{2\pi}} \exp\left\{\frac{-(\ln h - \mu)^2}{2\sigma_l^2}\right\}, \tag{3.1}$$

where h represents the positive channel gain, μ represents the mean, and σ_l denotes the standard deviation. Next, we outline the downlink configuration from a LEO satellite to a ground station. Within this context, the presence of atmospheric turbulence leads to the scintillation effect, causing variations in the received signal power. Under conditions of strong turbulence, the Gamma-Gamma (GG) distribution emerges as a suitable model to represent the channel model in such scenarios [40, 41]. The GG model arises when we assume that the turbulence-induced log-intensity fluctuations can be described by the product of two statistically independent Gamma-distributed processes, typically associated with the strong turbulence effects.

The probability density function (pdf) of the Gamma-Gamma distribution is described as [40]

$$f_h(h) = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}} h^{\frac{\alpha+\beta}{2}-1}}{\gamma(\alpha)\gamma(\beta)} K_{\alpha-\beta}(2\sqrt{\alpha\beta h}), \tag{3.2}$$

where The parameters α and β represent the shape factors of the distribution, stemming from the individual shape parameters of the two Gamma distributions associated with

turbulence effects. The term $K_{\alpha-\beta}$ is identified as the modified Bessel function of the second kind with order $\alpha-\beta$, while $\gamma(\cdot)$ denotes the gamma function. Furthermore, the received sequence $\boldsymbol{y}^u$ is described as

$$\boldsymbol{y}^u = h\boldsymbol{x}^u + \boldsymbol{w}^u, \tag{3.3}$$

where $\boldsymbol{w}^u \sim \left(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I}_u\right)$ is the Gaussian noise and σ_w^2 is the noise variance. The vectors $\boldsymbol{y}^u$, $\boldsymbol{x}^u$, and $\boldsymbol{w}^u$ have dimensions of $\mathbb{R}^u$, where u represents the length of the sequence of symbols. In our model, we consider both perfect and imperfect CSI for the Log-normal fading channel.

The average amount of energy per bit to noise power spectral density ratio $\frac{E_b}{N_o}$ in on-off-keying (OOK) is given by [42]

$$\frac{E_b}{N_o} = \frac{A^2}{4\sigma_w^2}\frac{u}{k}, \tag{3.4}$$

where A is the peak intensity, k is the the message bits and u is the length of coded symbols.

3.3 Proposed End-To-End Learning-Based Design

As depicted in Fig. 3.1, we take into account an SOC system in which a transmitter located in the GEO satellite sends the message $b \in \mathcal{B}, \mathcal{B} = \{1, 2, \ldots, B\}$ to a certain receiver over a Log-normal fading channel. To model the channel, we use the

STK simulator, with the encoder on a GEO satellite and the receiver at a ground station. The message b is first fed into the DL encoder NN producing $\boldsymbol{x}^u$. The elements of $\boldsymbol{x}^u$ are represented as $x(i), 1 \leq i \leq u$, which meets both the peak and the non negativity constraints required by the optical channel's physical characteristics, i.e., $0 \leq x(i) \leq A$. The data rate is defined as $\frac{k}{u}$ bits/channel use, where $k = \log_2(B)$ bits are sent through u coded symbols. Additionally, the encoded vector $\boldsymbol{x}^u$ is transmitted through a SOC channel. The resulting sequence is denoted as $\boldsymbol{y}^u \in \mathbb{R}^u$. The received sequence which can be obtained in accordance with the probabilistic law given by

$$\mathcal{P}\left(\boldsymbol{y}^u \mid \boldsymbol{x}^u, h\right), \tag{3.5}$$

where $h \in \mathbb{R}_+$ denotes the optical fading channel produced by STK and it is considered to remain constant through the transmission of the sequence $\boldsymbol{x}^u$. The result of $P(\boldsymbol{y}^u \mid \boldsymbol{x}^u, h)$ is a conditional probability distribution that describes the likelihood of receiving a particular sequence $y^u \in \{y_1, y_2, y_3, \ldots, y_u\}$ given that the transmitted input sequence $x^u \in \{x_1, x_2, x_3, \ldots, x_u\}$ and the channel fading coefficient h. Understanding $P(\boldsymbol{y}^u \mid \boldsymbol{x}^u, h)$ allows you to build decoders that can perform as efficiently as possible given the characteristics of the communication channel, thereby enabling more reliable and robust communication systems [43]. In this chapter, we argue that the proposed channel estimator NN can be trained to acquire the knowledge of the transition probability law for an input-output model that could be governed by (2), or could also be more general as in (4) without an explicit law. The channel estimator NN is based on two inputs

with a single NN whose parameters are tuned across a wide range of training SNRs. Furthermore, we take into account a pilot-based channel estimation approach, wherein the pilot symbol x_p is used for channel estimation and is communicated as the first symbol $x(1)$ of the transmitted sequence, i.e., $x_p \triangleq x(1)$. For symbols' detection, we propose the AE structure and we consider 3 cases: AWGN (no fading), fading with perfect CSI at the receiver, and fading with imperfect CSI at the receiver. The proposed AE is developed with multiple decoders along with a layered structure of encoders and decoders that employs LN layers. Next, design details regarding the proposed NN-based estimation and the AE-based detection are discussed.

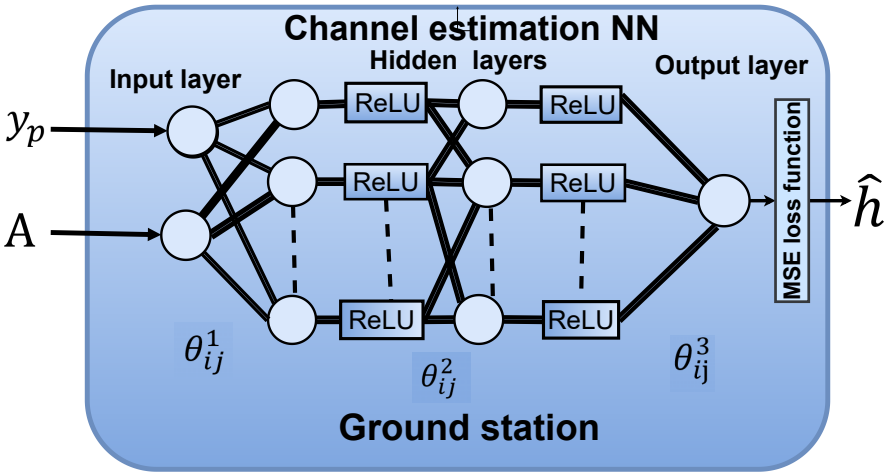


Figure 3.2: The implementation of the proposed NN used for channel estimation and located in the Geo satellite. The inputs are the received pilot y_p and the peak intensity A. The NN is composed of two FC hidden layers. Each neuron is followed by ReLU activation function for each layer. The output $\hat{h}$ is an estimated version of the channel gain

.

3.4 Proposed NN Design for Channel Estimation

In this section, we present the proposed channel estimator NN $\text{NN}(\hat{h})$. Additionally, we derive the mathematical expression for the MMSE estimator in Log-normal fading channel and apply it in both estimation and detection, as a benchmark. Although, the MMSE estimator provides the optimal performance in terms of MSE, this estimator has a considerable level of computational complexity and requires an explicit input-output model like the one in (2). On the contrary, the proposed channel estimator NN is capable of predicting the CSI and obtaining equal performance as the model-based MMSE estimator with far less complexity and without the need of an explicit input-output model. In addition, the proposed channel estimator NN relies on two inputs, and we train with a single NN whose parameters are adjusted across a wide range of training SNRs as opposed to generating a separate NN for each possible training SNR.

The proposed NN architecture: The proposed NN estimator is installed at the GEO satellite. It is composed of two fully connected (FC) hidden layers, a rectified linear activation unit (ReLU) activation function at each hidden layer, and a linear activation function at the output layer. As shown in Fig. 3.2, the NN has two inputs: the received signal $y_p \triangleq y(1)$ and the peak intensity A.

Training methodology: The following steps generate the training data used in channel estimation:

- We first generate the true channel coefficients based on Log-normal fading channel

from (1), h^n with $1 \leq n \leq N_s$, where N_s is the number of training samples.

- We distribute the peak intensities uniformly and randomly of the N_s samples. We then generate various peak intensity constraints $A \in [A_{min}, A_{max}]$ to cover a wide range of SNR values. In the training set, samples exhibiting high peak intensity values have a higher probability of occurring, while samples with low peak intensity are set to have a lower probability of occurring.

- The NN has two inputs: y_p and A. To generate the first element of the received pilot element $y_p^{(n)}$ for the n^{th} training sample, we substitute the corresponding peak intensity $A^{(n)}$ and the true channel coefficients $h^{(n)}$ in (2).

- The label of the training data tuple is based on two inputs as $\left\{(y_p^{(n)}, A^{(n)}), h^{(n)}\right\}$, where $(y_p^{(n)}, A^{(n)})$ is the input tuple to the NN and $h^{(n)}$ is the target value for the n^{th} training sample.

3.4.1 Learning Algorithm

The proposed channel estimator NN only makes use of two inputs, and we train with a single NN whose parameters are adapted across a wide range of training SNRs as opposed to creating a new NN for each possible training SNR. There are two phases to NN's learning process: training and testing. The network model must be trained in three steps before effective channel parameter estimation can be implemented. The first step is to select the data samples to utilize. Second, the gradient descent algorithm is used to calculate the partial derivative of the cost function by minimizing

the difference between the output value and the target value. Specifically, its value should be adjusted in the direction of the fastest descent of the error function, or the direction of the negative gradient. Third, when the training data for an epoch is finished, the validation data is used to determine the best model across all training iterations. In Fig. 3.2, θ_{ij}^{l} corresponds to the weight of the link between the j^{th} neuron in the $(l-1)^{th}$ layer and the i^{th} neuron in the l^{th} layer. The l^{th} layer pre-activation is represented by

$$z_i^{[l]} = \sum_j \theta_{ij}^{[l]} a_j^{[l-1]} + b_i^{[l]}, \tag{3.6}$$

where $b_i^{[l]}$ represents the bias of the i^{th} neuron in the l^{th} layer and $a_j^{[l-1]}$ is the activation of the j^{th} neuron in the $(l-1)^{th}$ layer. Employing the rectified linear unit (ReLU) activation function, the neuron output activation can be rewritten as

$$a_i^{[l]} = \mathrm{ReLU}(z_i^{[l]}) = \mathrm{ReLU}\left(\sum_j \theta_{ij}^{[l]} a_j^{[l-1]} + b_i^{[l]}\right), \tag{3.7}$$

At the start of the training, the initial point of the weights is selected as a random number drawn from a Gaussian distribution. Then, the state vector $\boldsymbol{z}^{[l]}$ can be obtained through each layer using the forward propagation formula as

$$\boldsymbol{z}^{[l]} = \boldsymbol{\Theta}^{[l]} \boldsymbol{a}^{[l-1]} + \boldsymbol{b}^{[l]}, \tag{3.8}$$

where $\boldsymbol{\Theta}^{[l]}$ is denoted as the weight matrix with i rows and j columns, $\boldsymbol{a}^{[l-1]}$ is the activation vector of dimension j in the $(l-1)^{th}$ layer and the bias vector of dimension

i in the l^{th} layer is denoted as $\boldsymbol{b}^{[l]}$.

Afterwards, $\boldsymbol{z}^{[l]}$ is fed into a ReLU activation function resulting the output vector $\boldsymbol{a}^{[l]}$ at layer l:

$$\boldsymbol{a}^{[l]} = \text{ReLU}(\boldsymbol{z}_i^{[l]}). \tag{3.9}$$

Each hidden layer applies a nonlinear ReLU function $f_a(x) = \max(0, x)$, after each neuron to enable the learning of complex, nonlinear relationships between the inputs and output. By employing network's hidden layers, inputs from the training data are extracted and then used to generate estimation results. The NN estimated channel gain at the final output layer L can be described as

$$\hat{h} = \left(\boldsymbol{\Theta}^{[L]}\boldsymbol{a}^{[L-1]} + \boldsymbol{b}^{[L]}\right), \tag{3.10}$$

where $\boldsymbol{\Theta}^{[L]}$ describes the connection weight matrix of the output layer, $\boldsymbol{b}^{[L]}$ represents the bias vector in the final output layer, and $\hat{h}$ denotes the estimated channel gain generated by the output of the entire NN. Then, the loss calculations follow the feed forward computations. The utlized loss function $\mathcal{L}(\hat{h}, h)$ is the normalized MSE which is the most suitable function in regression problems, defined as

$$\mathcal{L}(\hat{h}, h) = \frac{1}{N_s}\sum_{n=1}^{N_s}\left(\hat{h}^{(n)} - h^{(n)}\right)^2, \tag{3.11}$$

where $h^{(n)}$ is the true output of the n^{th} sample, $\hat{h}^{(n)}$ is the actual output provided by the NN of the n^{th} sample. Then, the objective of the proposed channel estimator NN

during the training stage is to minimize the training loss, which can be described as

$$\begin{aligned} \underset{\hat{h}}{\text{minimize}} \quad & \mathcal{L}(\hat{h}, h), \\ \text{subject to} \quad & 0 \leq \hat{h} < \infty. \end{aligned} \tag{3.12}$$

The detailed steps for the backpropagation process which minimizes the training loss are provided in Appendix B. The learning strategy of the proposed NN estimator is summarized in Algorithm 1.

Testing stage: The NN-based estimator utilizes the received signal $y_p \triangleq y(1)$ to obtain an estimate of the channel gain $\hat{h}$.

To demonstrate how the proposed channel estimator NN compares to the MMSE estimator, we derive the MMSE estimator in a log-normal fading channel. The MMSE objective function can be described as

$$\hat{h}_{\text{MMSE}} = \arg\min_{\hat{h}} \left\{ E\left[|\hat{h} - h|^2 \right] \right\}. \tag{3.13}$$

The MMSE algorithm is noise resistant and takes into account the influence of Gaussian noise on estimation performance, but it has a high computational complexity. The estimated channel gain for the MMSE estimator in the log-normal fading channel can be described as

$$\hat{h}_{\text{MMSE}} = \int_0^\infty \frac{e^{-\frac{(y-hA)^2}{2}} \exp\left\{\frac{-(\log h+\mu)^2}{0.18}\right\}}{\int_0^\infty e^{-\frac{(y-hA)^2}{2}} \frac{1}{h} \exp\left\{\frac{-(\log h+0.045)^2}{0.18}\right\} dh} dh. \tag{3.14}$$

The detailed steps for the derivation are provided in the Appendix C.

Algorithm 1 Proposed NN estimator

Require: N_s fading coefficients $(h_1, h_2, \ldots, h_{N_s})$, peak intensity constraints $A \in [A_{min}, A_{max}]$, batch size m, and learning rate η.
Ensure: $\hat{h}^{(i)} = h^{(i)}\ \forall i \in \{1, 2, \ldots, m\}$.

```
θ ← initialize neural network parameters.
repeat
    Draw m minibatch samples (h^(1), h^(2), ..., h^(m)) and m minibatch peak intensities (A^(1), A^(2), ..., A^(m)).
    for i ← 1 to m do
        y_p^(i) ← A^(i) h^(i) + w^(i)
        ĥ^(i) ← N_θ(y_p^(i), A^(i))
    end for
    Calculate minibatch loss: L ← (1/m) Σ_{i=1}^{m} (ĥ^(i) − h^(i))^2
    Calculate gradients: ∇_θ L ← ∂L/∂θ
    Update parameters: θ ← θ − η ∇_θ L
until convergence.
```

3.5 Symbol Detection

In this section, we present two subsections. The first subsection discusses the proposed AE for symbol detection with a single code rate, detailing its design and performance. The second subsection addresses the scalability of the AE, focusing on its ability to handle multiple code rates efficiently.

3.5.1 Proposed Autoencoder for Symbol Detection with a Single Code Rate

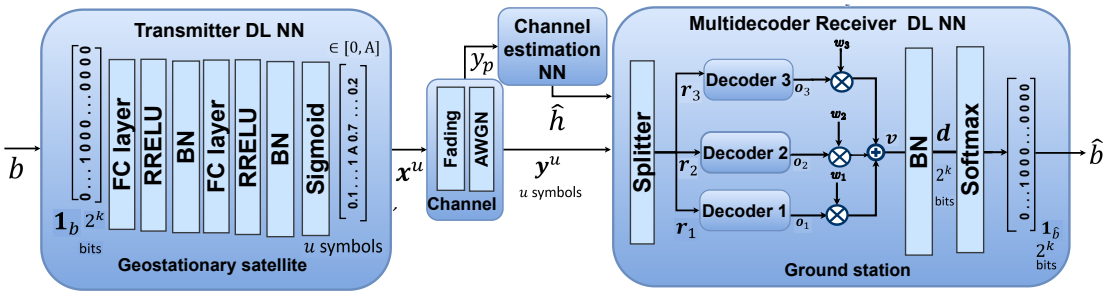


Figure 3.3: The proposed AE(k, u) architecture has a code rate of $R = k/u$, where $k = 7$ is the number of bits in the input message, and $u = 21$ is the length of the encoded message. The encoder is located on a GEO satellite, while the receiver is based at a ground station. The message b is represented by the one hot vector $\mathbf{1}_b$ of length $2^k = 128$. The input hot vector $\mathbf{1}_b$ is passed through a sequence of multiple dense layers in order to construct the encoded vector $\boldsymbol{x}^u$ of length $u = 21$. The normalization layer, the last layer of the transmitter, uses a weighted sigmoid $A \times$ sigmoid $(\cdot)$ to ensure that $\boldsymbol{x}^{\boldsymbol{u}}$ lies inside the interval $[0, A]$. The input to the receiver is the corrupted vector $\boldsymbol{y}^u$ that is produced when the encoded vector $\boldsymbol{x}$ is transmitted across the SOC channel. The receiver is composed of three decoders. The entire input hot vector with dimension 2^k is estimated independently by the three decoders. The first decoder's input vector $\boldsymbol{r}_1$ of length $u_1 = 7$ is fed into multiple dense layers and the output vector is denoted as $\boldsymbol{o}_1$. Additionally, the second and the third decoder map the vectors $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$, of length 7 each, into the output vectors $\boldsymbol{o_2}$ and $\boldsymbol{o_3}$, respectively. The length of $\boldsymbol{o}_{1,2,3}$ is equivalent to $M = 2^k = 128$. The vector $\boldsymbol{v}$ has the dimension (3×2^k) as a result of the concatenation of the output vectors $\boldsymbol{o}_1$, $\boldsymbol{o}_2$ and $\boldsymbol{o}_3$. Afterwards, vector $\boldsymbol{v}$ is fed into multiple layers. Finally, the estimated hot vector $\mathbf{1}_{\hat{b}}$ of dimension 2^k is then output from the softmax activation layer.

An AE can be described as an unsupervised NN that auto-learns how to compress the data efficiently via an encoding process. In addition to compressing data, the AE learns how to recreate the original data from the compressed form. Furthermore, the AE system can be expressed by the pair (k, u), where k and u are the number of message bits and the codeword length, respectively. The channel code rate is described

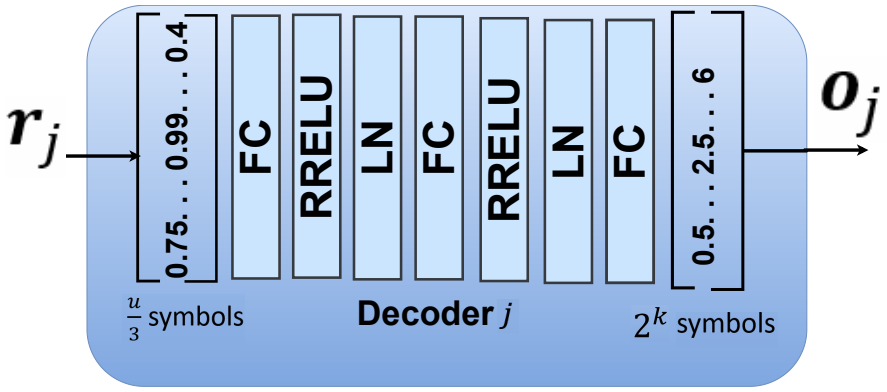


Figure 3.4: Proposed decoder architecture at the receiver in the ground station.

as $R = k/u$. The proposed AE(k, u) is illustrated in Fig. 3.3 for SOC system with code rate 1/3 without loss of generality. The receiver is based at a ground station, whereas the encoder is on a GEO satellite. The channel coding code rate is 1/3, where $k = 7$ and $u = 21$. The system is composed of three components: the transmitter, the SOC channel, and the receiver. First, the transmitter sends one out of M possible messages $b \in \mathcal{M}, \mathcal{M} = \{1, \ldots, M\}$ as one hot vector $\mathbf{1}_b$ of dimension 2^k bits. The transmitter then uses the mapping function $\boldsymbol{f} : \mathcal{M} \to \mathbb{R}^u$ to transform the input hot vector $\mathbf{1}_b$ into the encoded vector $\boldsymbol{x}^u$. The benefits of one-hot-encoding are that the output is binary rather than ordinal. The one-hot vector has all zero inputs, except one indexing a message $m \in \mathcal{M}$. The symbol vector $\boldsymbol{x}^u$ generated by the normalization stage of the transmitter satisfies the positivity and peak requirements for SOC. Then, it is transmitted through the SOC channel provided by STK. The SOC channel is constructed from both Log-normal fading and AWGN channel with zero mean and unit variance. Subsequently, the estimated hot vector $\mathbf{1}_{\hat{b}}$ is generated by the receiver, which uses a

multiple-decoder approach to recover the the message b from the corrupted vector $\boldsymbol{y}^u$.

Moreover, the transmitter model is based on FC layers, with LN layers occurring after each FC layer and a Randomized Leaky Rectified Linear Unit (RReLU) activation function in between. In order to generate more accurate models, AE can make use of RReLU activation, a non-saturated function that produces simultaneous activations associated with regression and classification [44]. The RReLU activation outperforms the Sigmoid and Tanh activations in terms of both training time and generalization capabilities [44].

In addition, for both the encoder and each decoder, we utilize LN on all of the hidden units in the same layer. LN is a technique to normalize the distributions of intermediate layers. It enables smoother gradients, faster training, and better generalization accuracy [30]. LN re-centers and re-scales the input vector $\boldsymbol{c}$ as

$$\boldsymbol{f} = \boldsymbol{g} \odot N(\boldsymbol{c}) + \boldsymbol{b}, \tag{3.15}$$

where the input vector to LN is denoted as $\boldsymbol{c} = (c_1, c_2, \ldots, c_D)$ of size D, $\odot$ is the dot product operand, and $\boldsymbol{f}$ is the output of a LN layer. Since it transforms the input distribution into a standard normal distribution (SND) with zero mean and a unit standard deviation, the standard score $N(\boldsymbol{c})$ is denoted as

$$N(\boldsymbol{c}) = \frac{\boldsymbol{c} - \mu}{\sigma}, \tag{3.16}$$

where, the mean and standard deviation of input $\boldsymbol{c}$ is μ and $\sigma = \sqrt{\frac{1}{D}\sum_{i=1}^{H}(c_i - \mu)^2}$, respectively.

While most existing learning-based frameworks only employ a single decoder at the receiver [19], [6], we employ a multi-decoder scheme. Using a set of several decoders and LN-based layered structure of both encoders and decoders, we found that the gradient descent can significantly improve the BER performance over the existing state-of-the-art models by minimizing the error loss function.

The BER improves when the error loss function decreases. When this occurs, our AE model provides predictions that are close to the actual data. Through the multi-decoder approach, more than one path can be employed to update encoder and decoder weights during training, resulting in a more robust model than that would be possible with a single decoder architecture. Furthermore, during the training phase, the encoder and decoder operate as a unified NN. This means that the backpropagation method can be employed simultaneously to compute error gradients for both components in every training iteration. This concurrent computation facilitates the combined training of the encoder and decoder. The feedback from backpropagation guides each layer on how to adjust its parameters to reduce the error in the cross-entropy loss function. Employing optimization strategies such as stochastic gradient descent, the parameters of both the encoder and decoder are refined. This iterative process continues until the error reaches the lowest possible value.

Figure 3.3 shows that the input to the first decoder is $\boldsymbol{r}_1$ of length $u_1 = 7$. Similarly, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$ correspond to the second and third decoder inputs. Every decoder

makes an independent prediction of the estimated input hot vector with dimension 2^k. Each decoder, as shown in Fig. 3.4, is built from a sequence of dense layers based on FC, RReLU, and LN layers, which is similar to the construction of the encoder. To estimate the input hot vector, each decoder maps the input vector $\boldsymbol{r_j}$ to the corresponding output vector $\boldsymbol{o_j}$ of length $M = 2^k$, where $j \in \{1, 2, 3\}$. The estimated vectors from each decoder are then combined into a vector $\boldsymbol{v}$ of dimension (3×2^k). Then, vector $\boldsymbol{v}$ is fed into FC, RReLU, and LN layers to produce a vector $\boldsymbol{d}$ of dimension 2^k. The softmax activation function is applied to the resultant vector $\boldsymbol{d}$ to get a probability vector over all possible messages $\boldsymbol{p}$ of length $M = 2^k$. The decoded message $\hat{b}$ is the index of the highest probability. A definition of the softmax function is:

$$\boldsymbol{p}(i) = \frac{e^{\boldsymbol{d}(i)}}{\sum_{t=1}^{M} e^{\boldsymbol{d}(t)}} \in [0, 1], \tag{3.17}$$

where $i \in \{1, 2, \ldots, M\}$. Cross-Entropy loss is a significant cost function for improving classification model precision. The cross entropy loss function can be described as

$$\mathcal{L} = -\sum_{i=1}^{M} \mathbf{1}_b(i) \log \boldsymbol{p}(i), \tag{3.18}$$

Our NN is trained at a fixed peak intensity A or a corresponding SNR according to (3). To determine which training peak intensity A value yields the lowest cross entropy loss, we investigate a wide variety of values throughout the AE training stage. The best value of training A for AWGN, perfect CSI, and imperfect CSI for a particular code rate will be demonstrated in the numerical results. In addition, training with

a peak intensity A higher than necessary is not promising because the network will only update its weights for the high SNR regimes, which might produce good results during training but poor results while testing. During the testing phase, we do not only assess our model's performance at the trained SNR but also across a broad range of SNRs. Tables 3.1 and 3.2 detail the relevant parameters for AE (7, 21) and AE (7, 14), respectively. In addition, we evaluate the AE (7,14) and AE (7,21) against convolutional codes and state-of-the-art learning-based frameworks in terms of BER performance. Algorithm 2 summarizes the learning strategy for the novel AE design.

Table 3.1: Encoder and decoder layers of AE (7,21) utilized in Fig. 3.3.

Modules	Number of layers	Input dimension of each layer	Output dimension of each layer
Encoder	2	$128, 100$	$100, 21$
First decoder	2	$7, 50$	$50, 128$
Second decoder	2	$7, 50$	$50, 128$
Third decoder	2	$7, 50$	$50, 128$

Table 3.2: Encoder and decoder layers of AE (7,14).

Modules	Number of layers	Input dimension of each layer	Output dimension of each layer
Encoder	2	$128, 100$	$100, 14$
First decoder	2	$9, 50$	$50, 128$
Second decoder	2	$5, 20$	$20, 20$

3.5.2 Scalable AE for multi-code rates

Scalability across multiple code rates is a critical attribute in modern communication systems, particularly in space optical communications, where adaptability to varying channel conditions and data requirements is essential. A single autoencoder ca-

Algorithm 2 Proposed AE training algorithm

Require: M messages $(1, 2, \ldots, M)$, transmitter peak intensity A, batch size m, learning rate η.

Ensure: $\text{argmax}(\boldsymbol{d}^{(i)}) = b^{(i)}\ \forall i \in \{1, 2, \ldots, m\}$

$E_\theta \leftarrow$ initialize encoder parameters.

$D_{\phi_1}, D_{\phi_2}, D_{\phi_3} \leftarrow$ initialize decoder units parameters.

$FC_\psi \leftarrow$ initialize FC layer parameters.

$LN_\upsilon \leftarrow$ initialize LN parameters.

repeat

  Draw m minibatch messages $(b^{(1)}, b^{(2)}, \ldots, b^{(m)})$.

  for $i \leftarrow 1$ to m **do**

    $\mathbf{1}_{b^{(i)}} \leftarrow \text{one\_hot\_vector}(b^{(i)})$ $\quad \{\mathbf{1}_{b^{(i)}} \in \{0,1\}^M\}$

    $\boldsymbol{x}^{(i)} \leftarrow E_\theta(\mathbf{1}_{b^{(i)}}, A)$ $\quad \{\boldsymbol{x}^{(i)} \in [0, A]^u\}$

    $\boldsymbol{y}^{(i)} \leftarrow \boldsymbol{x}^{(i)}\boldsymbol{h}^{(i)} + \boldsymbol{w}^{(i)}$ $\quad \{\boldsymbol{y}^{(i)} \in \mathcal{R}^u\}$

    $\boldsymbol{y}^{(i)} \leftarrow \boldsymbol{y}^{(i)}/\hat{\boldsymbol{h}}^{(i)}$

    $\boldsymbol{r}_1^{(i)}, \boldsymbol{r}_2^{(i)}, \boldsymbol{r}_3^{(i)} \leftarrow$ Split $\boldsymbol{y}^{(i)}$ into three segments of equal length.

    for $j \leftarrow 1$ to 3 **do**

      $\boldsymbol{o}_j^{(i)} \leftarrow D_{\phi_j}(\boldsymbol{r}_j^{(i)})$ $\quad \{\boldsymbol{o}_j^{(i)} \in \mathcal{R}^M\}$

    end for

    $\boldsymbol{v}^{(i)} \leftarrow \text{concatenate}(\boldsymbol{o}_1^{(i)}, \boldsymbol{o}_2^{(i)}, \boldsymbol{o}_3^{(i)})$ $\quad \{\boldsymbol{v}^{(i)} \in \mathcal{R}^{3M}\}$

    $\boldsymbol{d}^{(i)} \leftarrow LN_\upsilon(\text{RReLU}(FC_\psi(\boldsymbol{v}^{(i)})))$ $\quad \{\boldsymbol{d}^{(i)} \in \mathcal{R}^M\}$

    $\boldsymbol{p}^{(i)} \leftarrow \text{Softmax}(\boldsymbol{d}^{(i)})$ $\quad \{\boldsymbol{p}^{(i)} \in [0,1]^M\}$

  end for

  Calculate minibatch loss:

  $\mathcal{L} \leftarrow -\sum_{i=1}^{m}\sum_{t=1}^{M} \mathbf{1}_{b^{(i)}}(t)\log(\boldsymbol{p}^{(i)}(t))$

  Calculate gradients:

  $\nabla_\theta\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\theta}$, $\nabla_{\phi_1}\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\phi_1}$, $\nabla_{\phi_2}\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\phi_2}$,

  $\nabla_{\phi_3}\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\phi_3}$, $\nabla_\psi\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\psi}$, $\nabla_\upsilon\mathcal{L} \leftarrow \frac{\partial\mathcal{L}}{\partial\upsilon}$

  Update parameters:

  $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}$

  $\phi_1 \leftarrow \phi_1 - \eta\nabla_{\phi_1}\mathcal{L}$

  $\phi_2 \leftarrow \phi_2 - \eta\nabla_{\phi_2}\mathcal{L}$

  $\phi_3 \leftarrow \phi_3 - \eta\nabla_{\phi_3}\mathcal{L}$

  $\psi \leftarrow \psi - \eta\nabla_\psi\mathcal{L}$

  $\upsilon \leftarrow \upsilon - \eta\nabla_\upsilon\mathcal{L}$

until convergence

pable of handling multiple code rates offers significant advantages in terms of efficiency and practicality. By eliminating the need for separate models for each code rate, such a scalability greatly reduces system complexity, resource consumption, and hardware requirements. This consolidation not only minimizes computational overhead but also streamlines memory and processing needs, resulting in a more efficient and compact design.

The dynamic nature of communication channels in space optical systems demands robust adaptability. Atmospheric turbulence, scintillation, and fading introduce varying levels of noise and degradation, which require flexible solutions to maintain performance. A scalable autoencoder can seamlessly adapt to these conditions, ensuring reliable and robust communication without the need to switch between distinct models for different code rates. This adaptability enhances the system's ability to operate effectively across a wide range of scenarios, making it a crucial feature for real-world applications.

Integrating a single autoencoder for multiple code rates simplifies system architecture and integration, reducing the complexity of development and testing. With fewer distinct models to design and verify, the process becomes faster and more error-resistant. This uniformity benefits multi-user scenarios as well, where different users may require diverse data rates or levels of error protection. A scalable solution accommodates these varying needs efficiently, enhancing overall user experience and system flexibility.

From a cost perspective, scalability translates to significant savings in design,

implementation, and maintenance. By training and deploying a single model capable of handling multiple configurations, system developers can avoid the expense and labor of maintaining multiple models. This cost-effectiveness is particularly valuable in large-scale applications, such as satellite constellations or extensive ground station networks, where operational and maintenance efficiencies are paramount.

Scalability also ensures that communication systems remain future-proof. As standards evolve and new requirements emerge, a flexible autoencoder can adapt without necessitating an overhaul of the entire system. This adaptability ensures the longevity and relevance of the framework, providing a robust foundation for future developments. Additionally, training a single model for multiple code rates fosters shared learning, improving generalization, performance, and convergence during the training phase.

In the context of emerging technologies like intelligent reflecting surfaces and software-defined networks, scalable solutions align perfectly with the priorities of flexibility and adaptability. By addressing the critical need for scalability, systems equipped with a single multi-code rate autoencoder pave the way for more efficient, resilient, and forward-compatible communication frameworks. This capability is instrumental not only in enhancing current applications but also in preparing for the future of space optical communication technologies.

Accordingly, we trained our proposed scalable autoencoder using a parallel and sequential approach, shown in Fig. 3.5 and Fig. 3.6, respectively. In the parallel model, a shared layer module compresses the input bits from 128 to 100, then an encoder

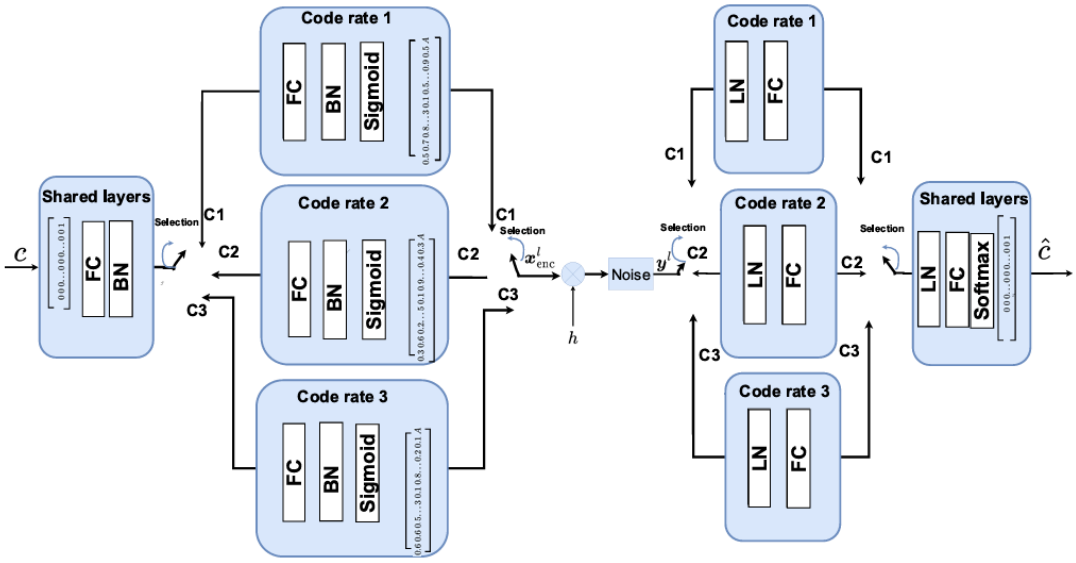


Figure 3.5: Parallel model structure for scalable AE for multi-code rate.

model is chosen that determines the code rate. In the sequential model, certain layers are skipped in the encoding/decoding corresponding to the code rate chosen.

Accordingly, we trained our proposed autoencoder using a sequential approach, achieving impressive results. The model was simultaneously trained for three code rates: 1/2, 1/4, and 1/3. Our primary objective was to develop a single autoencoder capable of effectively handling multiple code rates. In the numerical results section, we will present the performance of our model for a code rate of 1/2, along with comparisons to uncoded On-Off Keying and Convolutional code-based On-Off Keying.

3.6 Simulation Results

In this section, the proposed channel estimator NN is compared to the MMSE estimator and different state-of-the-art learning-based estimations. Then, in the pres-

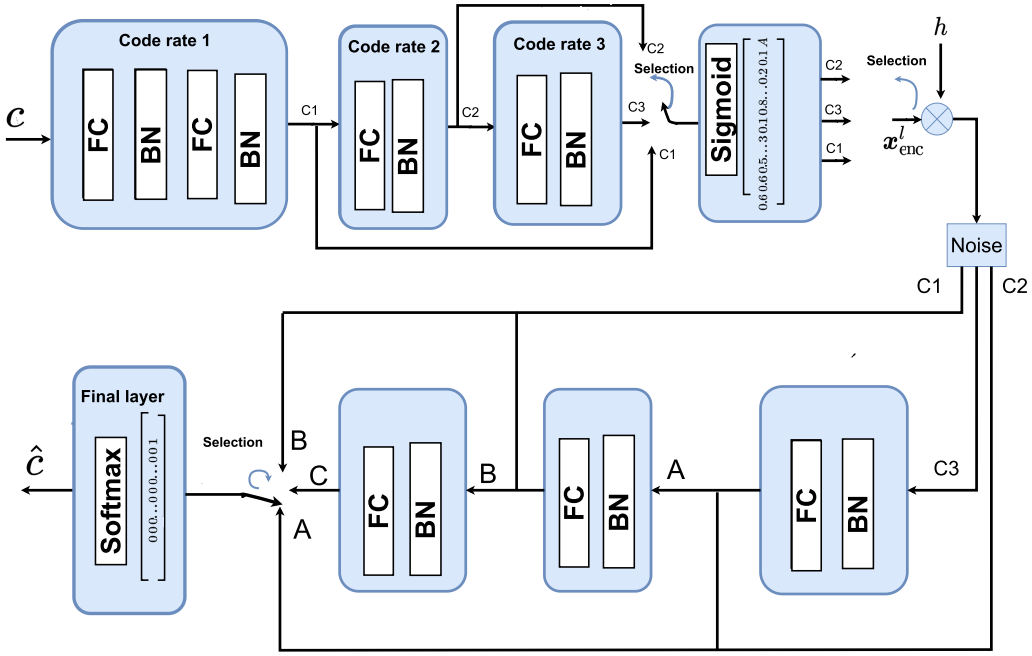


Figure 3.6: Sequential model structure for scalable AE for multi-code rate.

ence of AWGN and Log-normal fading channels, we compare the BER performance of AE-based SOC systems with the learning-based frameworks and convolutional codes at code rates 1/2 and 1/3 for perfect and imperfect CSI. Additionally, we train with a single NN whose parameters are adjusted across a wide range of training peak intensities. Following the procedures outlined in Section 3.3, the input tuple to the NN is based on two inputs $(y_p^{(n)}, A^{(n)}))$, where $y_p^{(n)}$ and $A^{(n)}$ are the pilot received sequence and corresponding peak intensity of the n^{th} sample, respectively. The distribution of the peak intensity A among the training samples is uniform except for the high peak

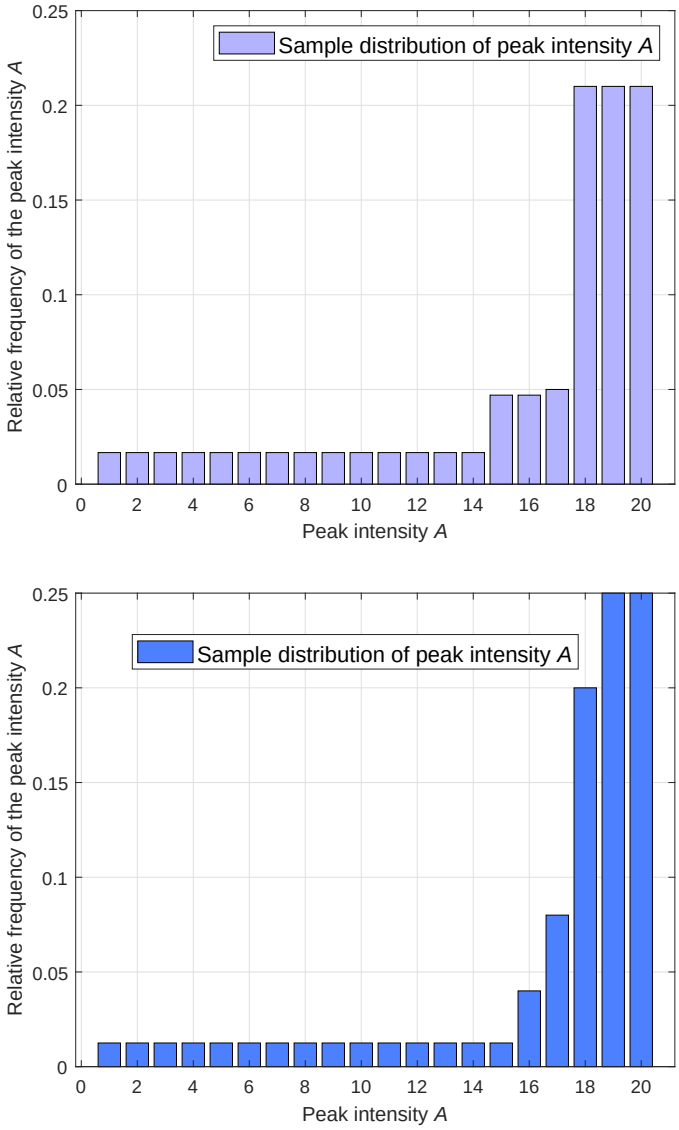
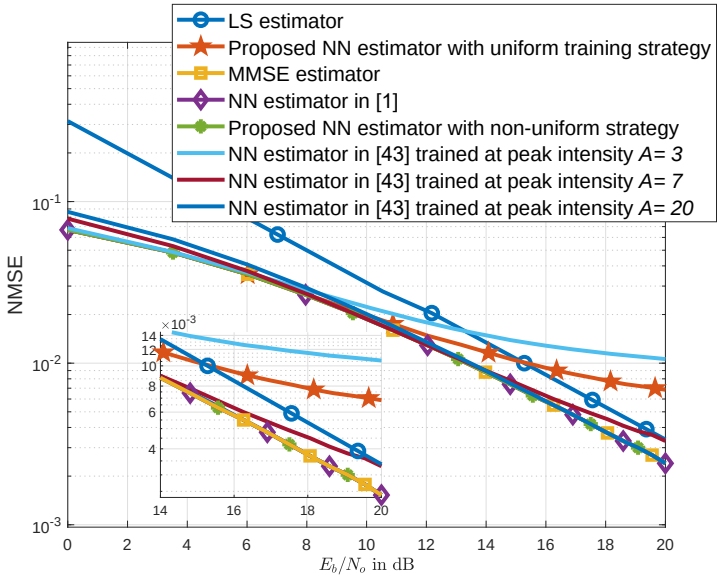
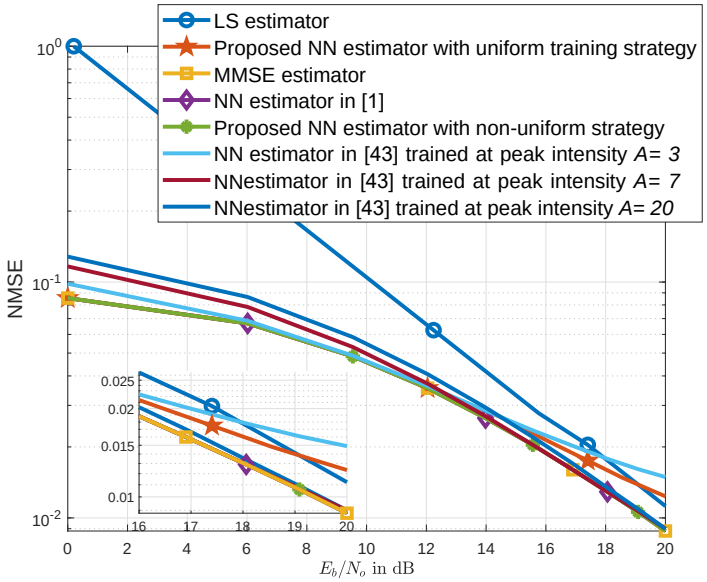


Figure 3.7: Constellation points of training peak intensity A versus probability of occurrence.: (a) Log-normal fading channel and (b) Gamma-Gamma fading channel.



(a)



(b)

Figure 3.8: The NMSE versus E_b/N_o of the proposed channel estimator NN compared with the MMSE estimator and learning based frameworks: (a) Log-normal fading channel and (b) Gamma-Gamma fading channel.

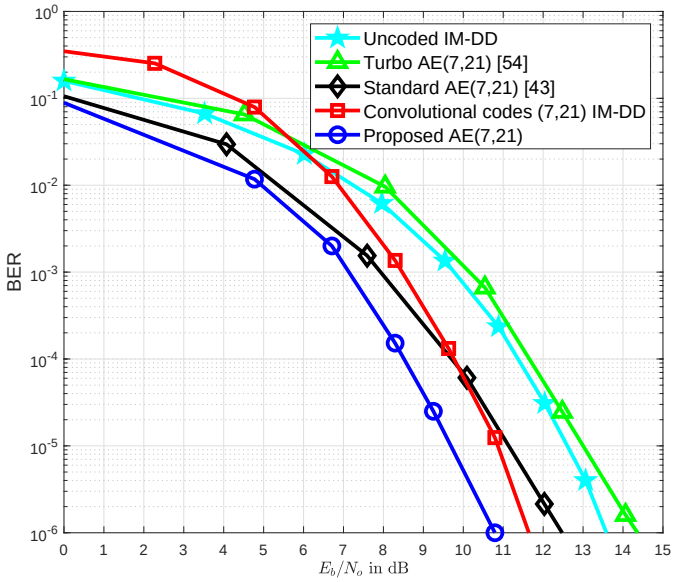


Figure 3.9: BER versus SNR for the proposed AE $(7, 21)$ compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate $1/3$ in a SOC channel with AWGN.

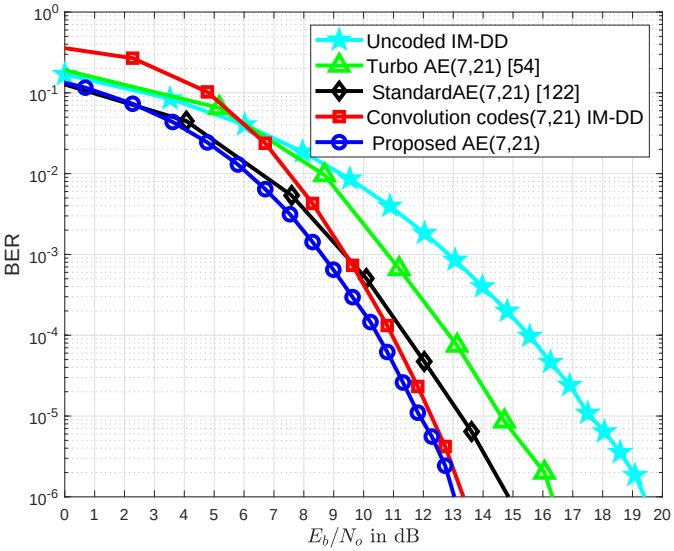


Figure 3.10: BER versus SNR for the proposed AE (7,21) compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate $1/3$ in a SOC channel with $\sigma = 0.3$ for perfect Log-normal channel.

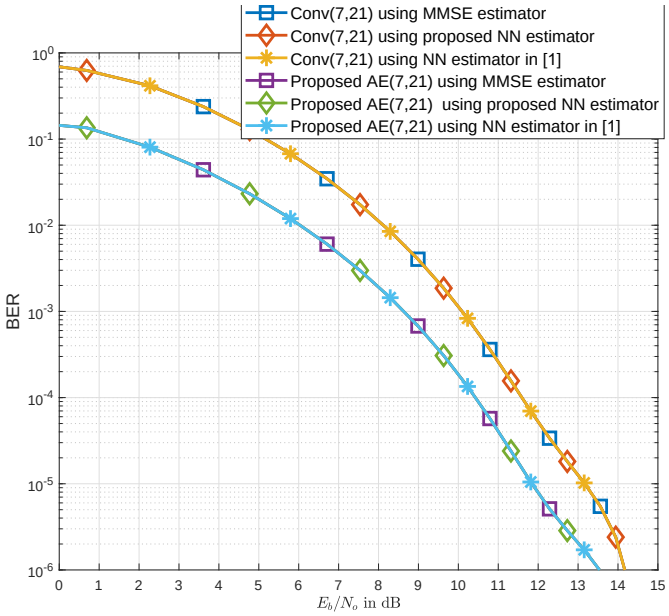
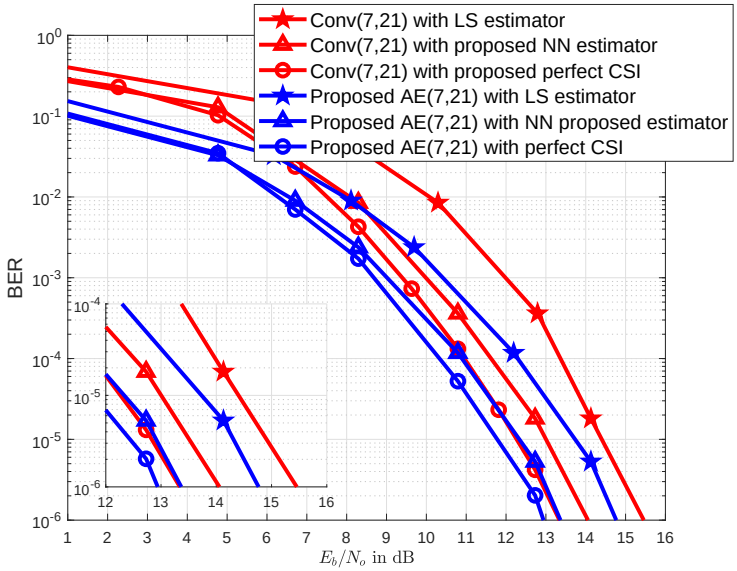


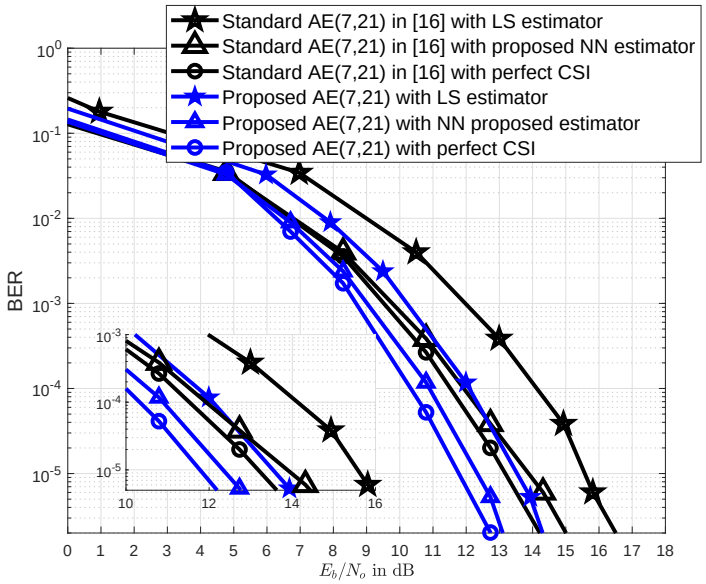
Figure 3.11: BER versus SNR for the proposed AE (7,21) compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate 1/3 in a SOC channel with $\sigma = 0.3$ for imperfect Log-normal channel.

intensities as depicted in Fig. 3.7. The batch size is 1000 and number of training, validation, testing samples are 40, 5, 10 million samples, respectively. The output of the NN is a single neuron representing the estimated channel gain $\hat{h}$.

As can be seen in Fig. 3.8, the proposed channel estimator NN exhibits the same normalized mean square error (NMSE) performance as the MMSE estimator and the channel estimator NN in [28] across all SNR ranges in the testing phase for a Log-normal fading channel. In contrast to [28], where a NN is designed for each SNR, we only need to develop a single NN that is optimized for all training SNR levels. Moreover, authors in [25] need to build three NN in order to achieve the MMSE estimator performance; one trained at $A = 3$ yields the best estimation from 0 to 7 dB, another trained



(a)



(b)

Figure 3.12: The BER versus SNR of the AE (7,21)-based detection in the existence of imperfect CSI against: (a) convolutional codes employing IM/DD and (b) benchmark learning frameworks for a SOC channel at a code rate of 1/3.

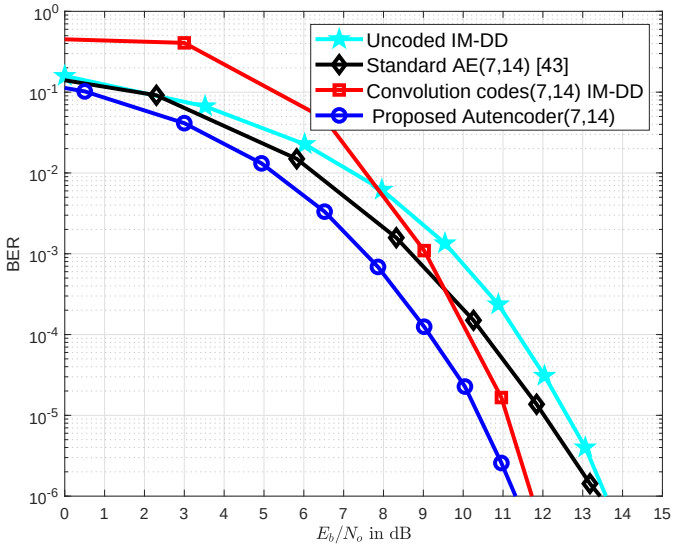


Figure 3.13: BER versus SNR for the proposed AE (7,14) compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate 1/2 in a SOC channel for AWGN channel.

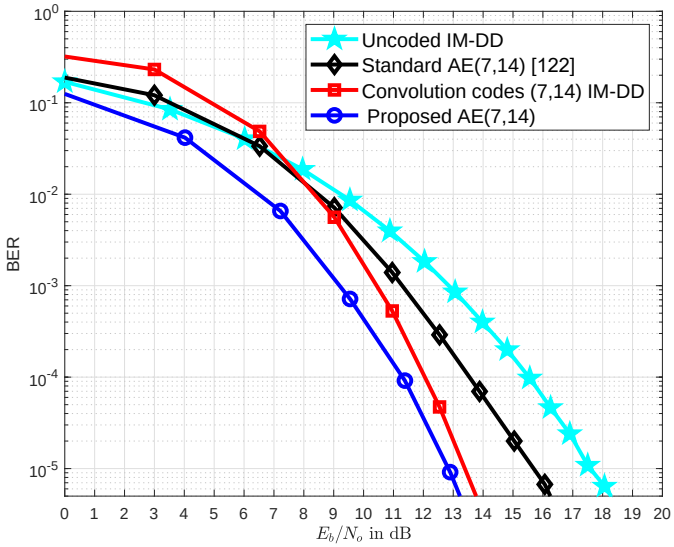


Figure 3.14: BER versus SNR for the proposed AE (7,14) compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate 1/2 in a SOC channel with $\sigma = 0.3$ for perfect CSI.

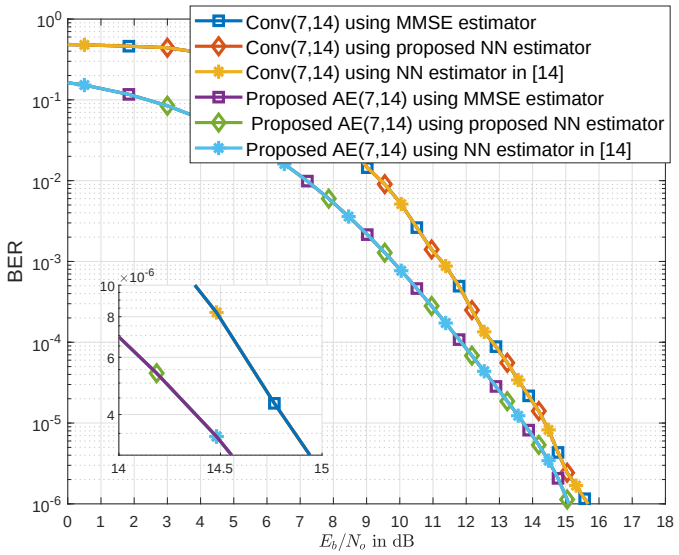
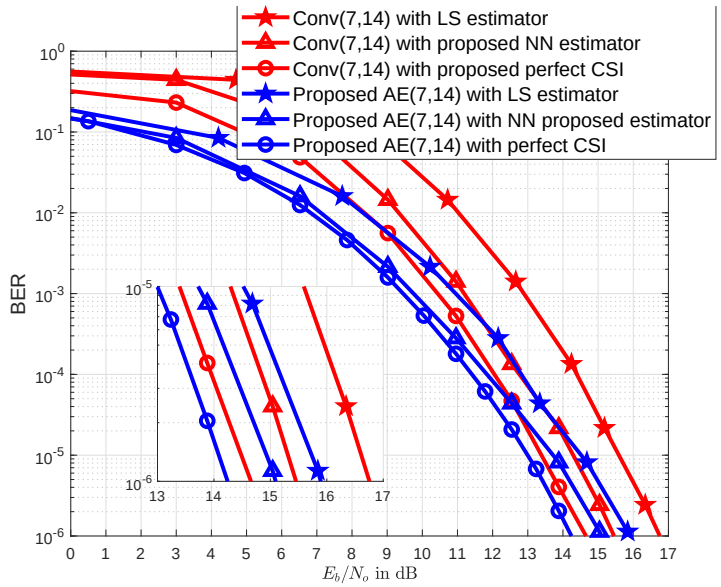


Figure 3.15: BER versus SNR in the existence of imperfect CSI for the proposed AE (7,14) compared with the convolutional codes employing IM/DD for code rate 1/2 in a SOC channel.

at $A = 7$ provides the estimation range from 7 to 14 dB, and another trained at $A = 20$ gives the estimation range from 14 to 20 dB. Their approach appears to yield good results with lower complexity compared to the MMSE estimator and channel estimator NN in [28]. However, it relies on the impractical assumption that the statistical data of testing samples are known in advance. The proposed single channel estimator NN outperforms the MSE performance of [25] and does not require any prior knowledge of the statistics of the testing samples nor does use multiple NNs. The proposed channel estimator NN achieves MSE improvement of 15% for SNR 6 dB over the estimation in [25] trained at peak intensity $A = 20$. When compared to the estimation in [25] trained at peak intensity 3, the proposed channel estimator NN yields MSE improvement of 37% in SNR 12 dB. We note that the MMSE estimator is designed using equation (13).



(a)

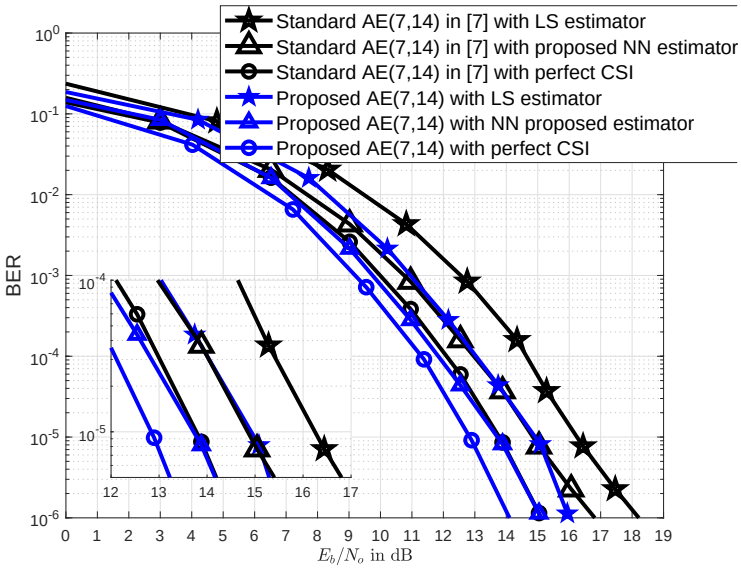


Figure 3.16: The BER versus SNR of the AE (7,14)-based detection in the existence of imperfect CSI against: (a) convolutional codes employing IM/DD and (b) benchmark learning frameworks for a SOC channel at a code rate of 1/2.

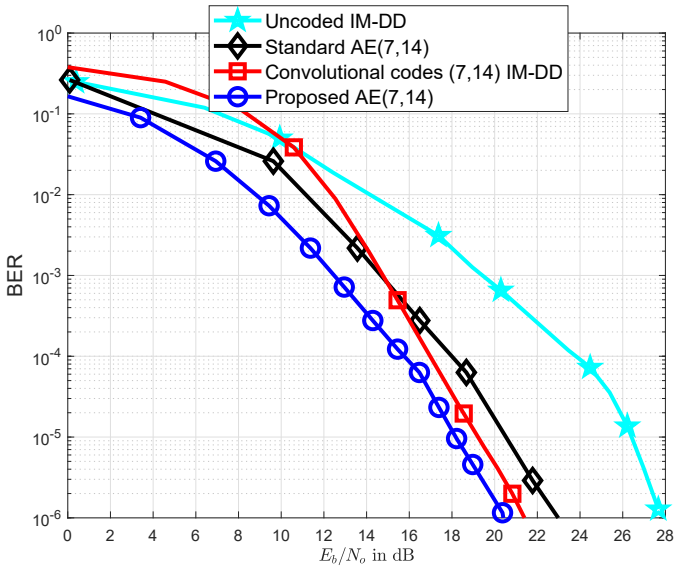
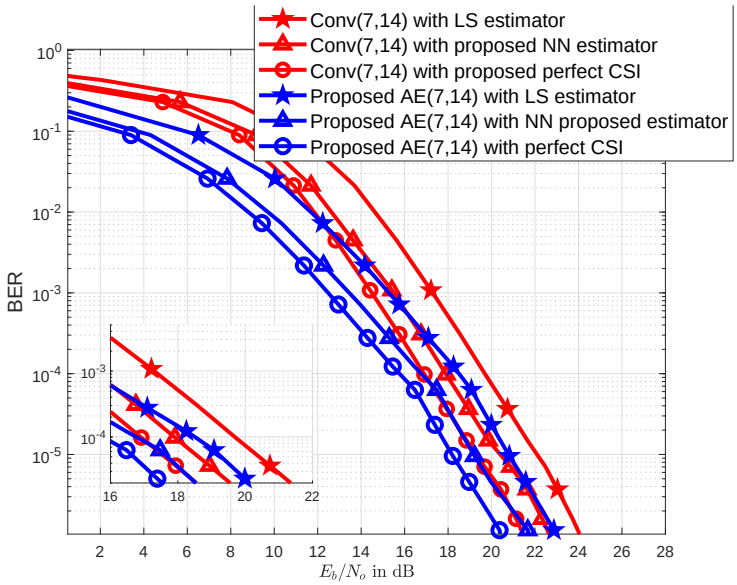


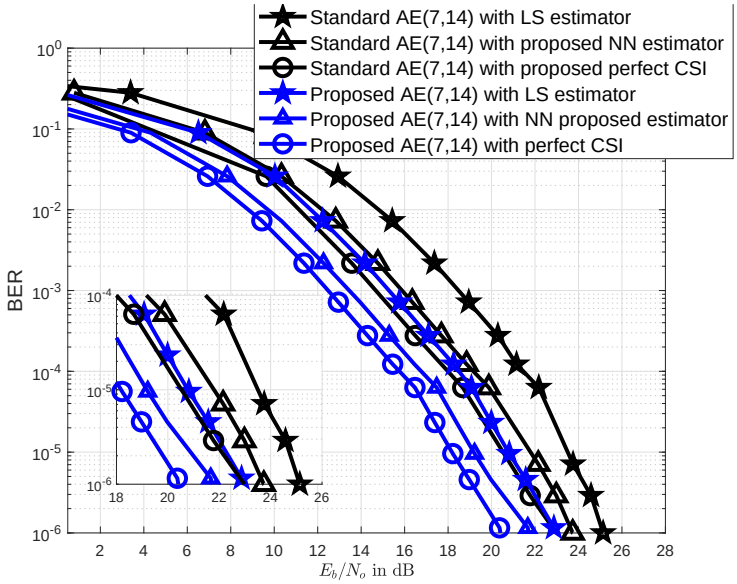
Figure 3.17: BER versus SNR for the proposed AE (7,14) compared to the convolutional codes using IM/DD and benchmark learning frameworks for code rate 1/2 in a SOC channel with Gamma-Gamma fading channel for perfect CSI.

Next, we demonstrate the BER performance of the proposed AE-based SOC at 1/2 and 1/3 coding rates. In addition, we compare the proposed AE model to both state-of-the-art learning-based approaches and model-based coding schemes. Figures 3.3 and 3.4 illustrate the simulation layout for the proposed AE. A total of 20,000,000 samples were used for training, and 10,000,000 used for testing. We accomplish both training stability and the effective learning weights by employing the Adam optimizer and a learning rate of 0.0001 throughout 100 training epochs. Convolutional codes using IM/DD at code rates of 1/2 and 1/3, as well as uncoded IM/DD, are implemented and compared in terms of BER with the proposed AE.

In addition, we evaluate our results against the benchmarking AE models as described in [6, 19, 30]. Although [19] demonstrates the viability of standard AE in



(a)



(b)

Figure 3.18: The BER versus SNR of the AE (7,14)-based detection in the existence of imperfect CSI against for Gamma-Gamma fading channel: (a) convolutional codes employing IM/DD and (b) benchmark learning frameworks for a SOC channel at a code rate of 1/2.

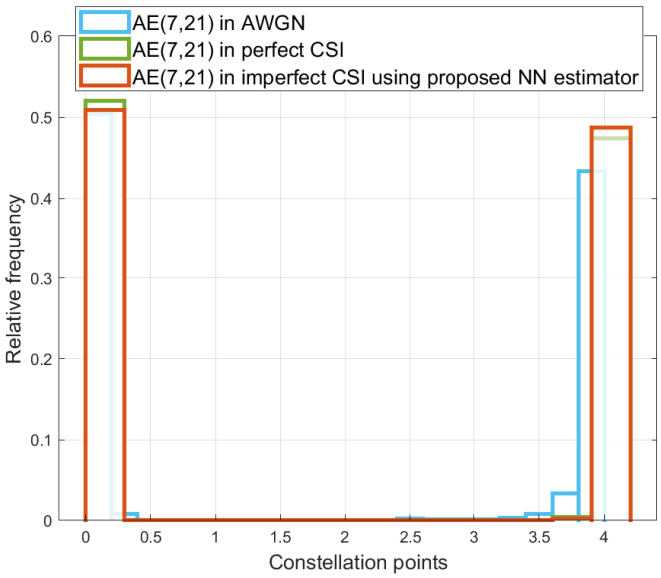


Figure 3.19: Constellation points against relative frequency generated by the proposed AE (7,21) with AWGN, perfect and imperfect CSI for peak intensity $A = 4$.

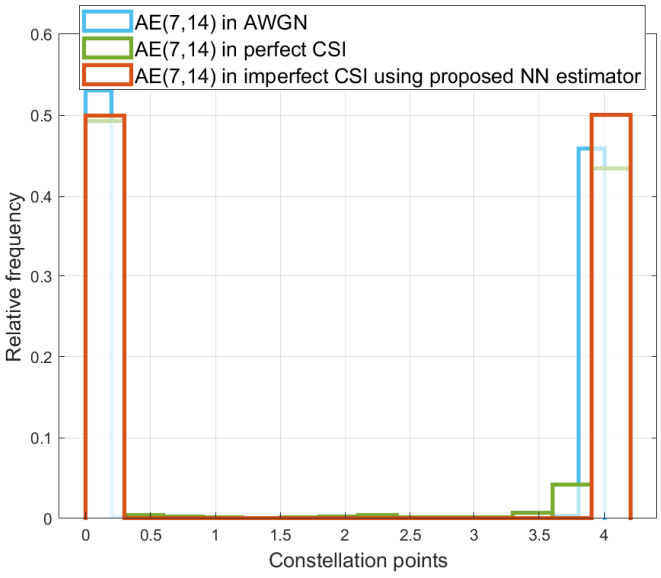


Figure 3.20: Constellation points against relative frequency developed by the proposed AE (7,14) with AWGN, perfect and imperfect CSI for peak intensity $A = 4$.

OWC channels under the assumption of an AWGN channel, they do not explore the performance of AE in fading channels. By extending the work of [19] to include turbulence channels, the authors of [6] were able to apply changes for standard AE to adapt with both perfect and imperfect CSI. In addition, the Turbo AE [30] performance in SOC was not satisfactory after optimizing the training SNR and switching to positive normalization, which is suitable for SOC. The proposed AE outperforms learning-based frameworks presented in [6, 19, 30] for code rates of 1/2 and 1/3, respectively. This improvement can be attributed to the utilization of a new layered structure that incorporates Log-normal (LN) encoders and decoders, along with a multi-decoder approach. The convolutional codes (7,21) in Fig. 3.9 are expressed by the polynomial $z^3 + z + 1, z^3 + z^2 + 1, z^3 + z^2 + 1$ [45].

As can be seen in Fig. 3.9, the AE (7,21)'s BER performance is 0.6 dB better than the convolutional codes at BER 10^{-6} for AWGN channel. For BER 10^{-4} the AE outperforms the Turbo AE and standard AE by 2.1 dB and 1.4 dB, respectively. Furthermore, at BER 10^{-4} for code rate 1/3, the proposed AE performance is superior than the uncoded SOC system employing uncoded IM/DD and a maximum likelihood decoder by 2.9 dB. The proposed AE (7,21) is developed in an AWGN channel with a training peak intensity $A = 3$. In Fig. 3.10, we observe that the proposed AE (7,21) achieves 0.3 dB better performance than the convolutional codes at a BER of 10^{-4} and 0.1 dB better performance at a BER of 10^{-6} when using a Log-normal fading channel with perfect CSI at the receiver. For BER 10^{-5}, it exceeds the performance of the learning-based framework of standard AE and Turbo AE by 1.1 dB and 2.1 dB,

respectively. The training peak intensity A employed with a Log-normal fading channel is set to 4.

In Fig. 3.11, the proposed AE (7,21)-based detection utilizing the MMSE estimator exhibits the same performance as the proposed channel estimator NN. Despite its superior estimation performance, the MMSE estimator involves high implementation complexity [26]. The same BER performance is also obtained when utilizing the channel estimator NN provided in [28] which uses a design for a NN for each training SNR. In contrast to [28], we only need to develop a single NN to achieve the same results. In both the low and high SNR regimes, convolutional codes (7,21) exhibit same BER performance when using the MMSE, [28], and the proposed channel NN estimator.

As can be seen in Fig. 3.12a, the proposed AE (7,21) outperforms the convolutional codes for a Log-normal fading channel with imperfect CSI at the receiver by 0.9 dB at a BER of 10^{-4} and by 0.6 dB at a BER of 10^{-6}, provided that both convolutional codes and AE are using the proposed channel estimator NN. Furthermore, the performance for BER 10^{-4} is 0.8 dB better than that of the learning-based frameworks of standard AE as depicted in Fig. 3.12b. The proposed AE (7,21) employing channel estimator NN is inferior by 0.5 dB compared with the perfect CSI case at BER 10^{-6}. Moreover, it has the same performance when utilizing [28] which use a training NN for each training SNR. The training peak intensity is set to $A = 4$ in the imperfect scenario.

Figure 3.13 demonstrates that the proposed AE yields a significant improvement of 1.6 dB over the standard AE at a BER of 10^{-4}. We also discover that for BER 10^{-6}, the AE's performance is 0.25 dB greater than that of the convolutional codes in

AWGN channel. Furthermore, the uncoded SOC system employing IM/DD is inferior by 2.3 dB at BER 10^{-6} compared to the proposed AE (7,14). For convolutional code (7,14) with 3 memory registers, we use the polynomials $z^3 + z, z^3 + z^2 + 1$. At BER 10^{-4}, the AE outperforms the convolution code (7,14) by 1 dB. As illustrated in Fig. 3.14, at a BER of 10^{-4}, the proposed AE (7,14) surpasses the standard AE by 1.6 dB with the presence of fading channels. Moreover, when compared to the convolutional code (7,14), the proposed AE (7,14) offers a 0.8 dB improvement at BER 10^{-4} and a 0.3 dB improvement at BER 10^{-6}.

In Fig. 3.15, the BER performance of convolutional code (7,14) using the MMSE estimator is identical to that of convolutional code (7,14) using the proposed channel estimator NN in [28]. Again, we achieve similar behavior as Fig. 3.11 when code rate 1/3 is used. While utilizing the estimator presented in [28], which employs a design for a NN for each training SNR, it achieves the same BER performance as convolutional code (7,14) utilizing the proposed channel estimator NN. The proposed AE (7,21) performs the same operations as convolutional codes, demonstrating that the BER is consistent across a wide range of SNR values, whether the proposed channel estimator NN or the MMSE estimator is used.

As illustrated in Fig. 3.16a, the proposed AE (7,14) with the proposed channel estimator NN only deviates from the perfect CSI case by 0.8 dB for a BER of 10^{-6}. Narrowing down to the imperfect CSI, the proposed AE outperforms the convolutional codes by 0.4 dB for BER $10^{-}6$. In Fig. 3.14b, we further investigate this behavior for standard AE and reveal that AE in [6] utilizing the proposed channel estimator NN

only differs from the perfect CSI case by 1.1 dB at a BER of 10^{-4}. For BER 10^{-6}, the proposed AE (7,14)-based detection achieves 1.6 dB better performance than the learning-based framework of standard AE. In contrast to the perfect CSI case, where the training peak intensity is $A = 5$, the training peak intensity is increased to $A = 6$ in the imperfect scenario. As illustrated in Fig. 3.19, the proposed AE (7,21) has roughly

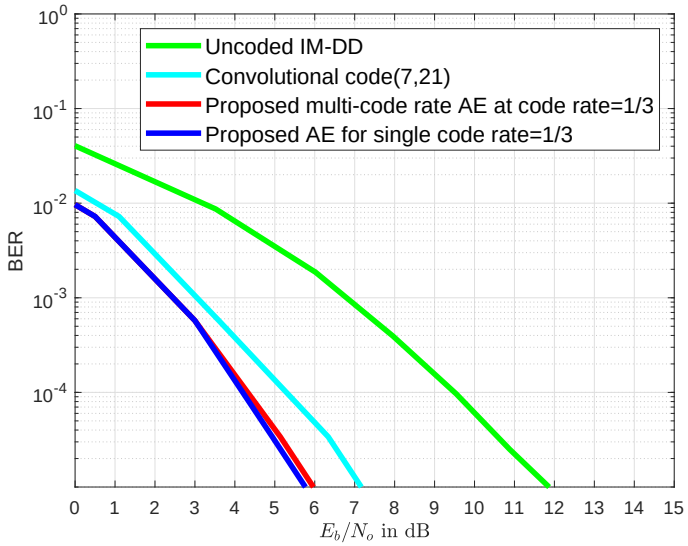


Figure 3.21: The BER versus SNR of the Proposed scalable AE -based detection in the existence of Proposed AE for single code rate=1/3, convolutional codes employing IM/DD and uncoded modulations.

learned an IM with constellation points located at 0 and $A = 4$ for both AWGN and perfect and imperfect CSI. Similarly, AE (7,14) has roughly learned IM with probability of occurrence mostly located at $A = 0$ and $A = 4$ as depicted in Fig. 3.20. Both Figs. 3.19 and 3.20 are trained and tested at $A = 4$. The results presented in this section demonstrate that the proposed channel estimator NN outperforms learning-based frameworks while performing as well as MMSE estimator in terms of MSE. The proposed AE for

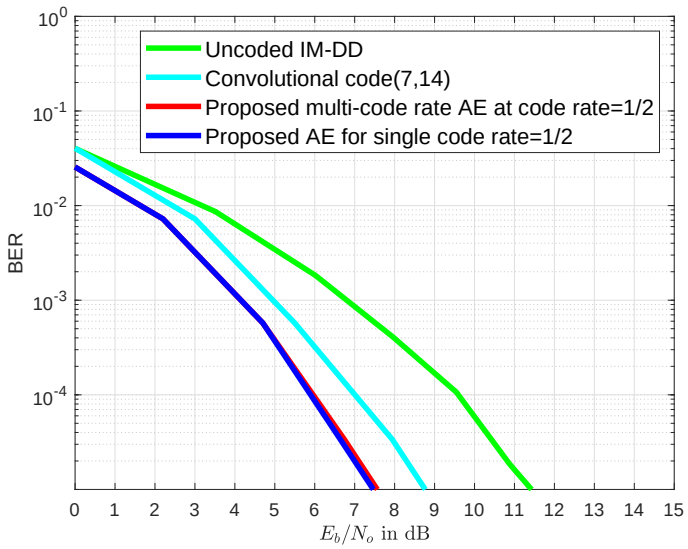


Figure 3.22: The BER versus SNR of the Proposed scalable AE -based detection in the existence of Proposed AE for single code rate =1/2, convolutional codes employing IM/DD and uncoded modulations.

both 1/2 and 1/3 code rates has learned encoding and decoding functions that outperform convolutional codes with IM/DD and learning-based frameworks in terms of BER for AWGN as well as perfect and imperfect CSI.

As illustrated in Fig. 3.21, at a code rate of 1/3, the proposed multi-code rate AE achieves performance nearly identical to that of the AE designed for a single code rate 1/3. Moreover, it consistently outperforms both convolutional code(7,21) and uncoded modulations across all SNR levels. At a BER of 10^{-4}, the proposed scalable AE at code rate 1/3 offers a 1.3 dB improvement over the convolutional code (7,21).

Similarly, as shown in Fig. 3.22, for a code rate of 1/2, the proposed AE surpasses the convolutional code (7,14) by 1 dB at a BER of 10^{-4}. Furthermore, the proposed scalable AE demonstrates a 2 dB performance advantage over uncoded mod-

ulations at a BER of 10^{-3}. This shows that the scalable AE achieves the same performance as the single code rate AE and has been validated across different code rates.

3.7 Concluding Remarks

This work presents a novel channel estimator NN that is optimized in a wide range of SNR levels in the training stage. The numerical results demonstrate that the proposed channel estimator NN outperforms learning-based frameworks and performs as the optimal MMSE estimator. Further, we propose an AE detection for creating an end-to-end communication system for SOC over AWGN and fading channels with perfect and imperfect CSI at the receiver. The proposed AE further employs multiple decoders and a stacked structure for building encoders and decoders that is based on LN. Compared to the state-of-the-art models, the innovative method can facilitate the training, which reduces the computation complexity. To the best of our knowledge, this is the first time that AE-based detection has been demonstrated to be superior than the state-of-the-art capacity-approaching convolutional codes in SOC. This study shows that the proposed AE holds considerable potential for use in future SOC systems that will benefit from more efficient coding, modulation, and decoding strategies.

Chapter 4

End-to-End Learning Framework for Space Optical Communications in Non-Differentiable Poisson Channel

4.1 Motivation

Training AEs requires both the channel model and AE layers to be differentiable, posing a challenge for SOC, where non-differentiable channel models like the Poisson channel often arise. In SOC, weak received optical signals necessitate the use of photon counting statistics, accurately modeled by the Poisson distribution, which represents the probability distribution of photon detections over time. Despite its accuracy, the Poisson channel cannot be directly implemented in an AE due to its non-differentiability, a limitation that many DL studies avoid by approximating it with differentiable models like Gaussian noise. However, such approximations often lack

generalizability across SOC scenarios.

The Poisson model excels in SOC by characterizing discrete photon arrival statistics, offering valuable insights into photon count probabilities. Conventional continuous models, by contrast, fail to capture this discrete nature. Nonetheless, its non-differentiability challenges the use of DL frameworks for gradient-based optimization. Prior work has addressed this limitation by approximating the Poisson channel with Gaussian distributions, but this approach compromises accuracy in various contexts.

In this work, we tackle the challenges of training AEs on Poisson channels without resorting to approximations. To handle the non-differentiability during back-propagation, we integrate CMA-ES with the proposed AE, enabling efficient gradient estimation for the Poisson channel. Furthermore, the AE employs BN layers in both encoders and decoders, ensuring regularization and consistent data distributions during training. Comparative evaluations reveal that our AE model achieves notable improvements in BER over non-gradient-based optimization models and standard AEs, performing on par with channel-capacity-approaching convolutional codes and uncoded modulation schemes. These advancements showcase the potential of addressing non-differentiability in SOC, providing robust and accurate solutions for photon-limited communication systems.

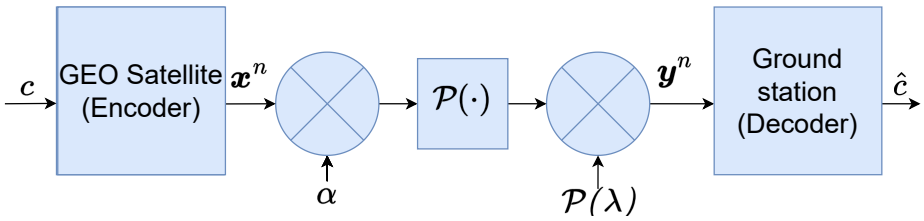


Figure 4.1: The system model for point-to-point SOC over the Poisson channel.

4.2 System Model

We model a geostationary satellite with a laser transmitter, and the detector is located at the ground station as shown in Fig. 4.1. The SOC channel is represented using the Poisson channel, known for its precision in capturing optical channel impairments. In low-received power conditions within the SOC, individual photons gain prominence. The Poisson model excels in such scenarios by effectively characterizing the statistical behavior of discrete photon arrival, offering insights into the probability of observing a specific photon count within a given timeframe. On the other hand, the conventional continuous models may not accurately reflect the discrete nature of photon interactions. The emitted light intensity transmitted from the laser terminal on the GEO satellite can be expressed as follows [46]:

$$x^{(i)} = \begin{cases} 0 \leq x^{(i)} \leq A \\ \mathbb{E}\left[x^{(i)}\right] \leq \mathcal{E} \end{cases}, \text{ for } i \in [1, n], \tag{4.1}$$

where n is the number of codeword symbols, $x^{(i)}$ follows the positivity, peak constraint of A, and average power constraint of $\mathcal{E}$. The optical wireless channel employing IM-DD is modeled by a Poisson distribution as follows [46]:

$$p_{Y|X}\left(y^{(i)} \mid x^{(i)}\right) = e^{-\left(\alpha x^{(i)}+\lambda\right)} \frac{\left(\alpha x^{(i)}+\lambda\right)^{y^{(i)}}}{y^{(i)}!}, \tag{4.2}$$

where $y^{(i)}$ is the channel output, $\alpha > 0$ represents the channel gain, and $\lambda \geq 0$ is the dark current rate of the photodetector. While the Poisson channel is an accurate model in SOC, it faces a challenge related to non-differentiability, which can complicate its implementation in DL modes. In the following section, we elaborate on how to address the non-differentiable issue in DL models.

4.3 Proposed AE model over Poisson channel

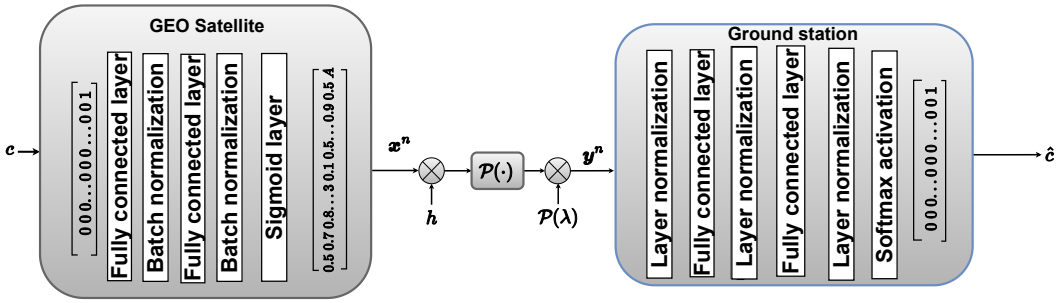


Figure 4.2: Diagram of proposed AE model with Poisson channel.

In this section, we describe the structure of the proposed AE, including the normalization layers. Following that, we discuss how we utilize a CMA-ES in conjunction with the proposed AE to effectively estimate the gradients of the Poisson channel.

First, the AE can be described as an unsupervised NN that auto-learns how to compress the data efficiently via an encoding process. In addition to compressing data, the AE learns how to recreate the original data from the compressed form. The AE

system can be expressed by the pair (k, n), where k and n are the number of message bits and the codeword length, respectively. The channel code rate is described as $R = k/n$. As shown in Fig. 4.2, the proposed AE consists of both a transmitter and a receiver neural network (NN), which are jointly optimized to streamline the learning process. A binary message c_i is taken from $\{1, ..., M\}$, where $M = 2^k$. The vector $\mathbf{1}_c$ used as input is the one-hot encoding of c. We perform one-hot encoding on the input message to ensure that the model is not biased toward any specific value. The input vector is then passed through the transmitter NN and encoded into the vector $\boldsymbol{x}^n$ of length n, which is used as input to the channel. In our study, we consider the Poisson distribution $\mathcal{P}(\lambda)$ to model a low-power SOC channel. After $\boldsymbol{x}^n$ passes through the channel, it is distorted into the noisy signal $\boldsymbol{y}^n$ and reaches the ground station. There, the receiver NN outputs its reconstruction of the original one-hot encoded vector, denoted as $\hat{\mathbf{1}}_c$. Unlike the standard AE model, we introduce normalization layers in between fully connected (FC) layers to reduce the effect of poor gradient exploding and increase the speed of convergence during training. The encoder utilizes batch normalization (BN), which is applied across each batch. The BN can be defined as

$$\mathrm{BN}_{\boldsymbol{\gamma},\boldsymbol{\beta}}\left(\boldsymbol{z}\right) = \boldsymbol{\gamma}\widehat{\boldsymbol{z}} + \boldsymbol{\beta}, \tag{4.3}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are considered as learnable parameters similar to the weights and the input vector to the BN layer $\boldsymbol{z} = \{z_1, z_2 \dots z_B\}$. The normalized output $\hat{\boldsymbol{z}} = \frac{\boldsymbol{z}-\mu_B}{\sqrt{\sigma_B^2+\epsilon}}$ where μ_B, σ_B are the mean and variance of the mini batch input data of size B, re-

spectively. Although BN has provided good results when employed in the encoder, it does not deliver satisfactory results in the decoder, mainly because of the changes in peak intensity during the testing stage. Resulting in a shift in the mean of the data distribution, it becomes necessary to adjust the normalization scheme in the decoder. Consequently, we employ layer normalization, which treats each sample independently. This normalization method ensures consistent performance in both training and testing phases for the decoder.

At the last layer on the decoder NN, we use a softmax activation function to determine the most likely value of the original message. The softmax activation is defined as

$$\boldsymbol{p}(i) = \frac{e^{\boldsymbol{d}(i)}}{\sum_{t=1}^{M} e^{\boldsymbol{d}(t)}} \in [0, 1]. \tag{4.4}$$

The goal of the AE during training is to construct an output that is identical to the input. This is achieved through minimizing the AE Cross-Entropy (CE) loss as follows [31]:

$$\mathcal{L} = -\sum_{i=1}^{M} \mathbf{1}_c(i) \log \boldsymbol{p}(i). \tag{4.5}$$

In addition, the Poisson channel considered in this work is not differentiable, so using the back-propagation algorithm along with gradient descent to tune the AE parameters becomes unfeasible. To handle the non-differentiability of the Poisson channel, we set the gradient of the channel to a constant, denoted as J, during back-propagation.

This constant J can be considered as a hyper-parameter and thus can be tuned using hyper-tuning optimization algorithms. In this study, we utilize the CMA-ES to optimize the hyper-parameters selection process. It was shown to be consistently outperforming other ES methods, especially on non-separable problems, functions that cannot be solved with a comparatively small number of function evaluations, or larger-dimensional search spaces [47]. Also, CMA-ES are stochastic, derivative-free methods utilized to solve non-linear or non-convex optimization problems. In CMA-ES, candidate solutions (often denoted as x) are produced in a stochastic manner in each iteration, based on the existing candidate solutions from the previous iteration. Subsequently, a selection process is conducted to determine which candidate solutions will serve as parents in the ensuing iteration, with the criteria usually being their value of the objective function, denoted as $f(x)$. As this process unfolds across successive iterations, candidate solutions with progressively improving 'f-values' are generated.

In CMA-ES, the normal probability distribution responsible for generating new candidate solutions, given the distribution parameters such as mean, variances, and covariances, represents the maximum entropy probability distribution over $\mathbb{R}^p$, where p denotes the number of parameters in the candidate solutions. In other words, this distribution reflects the sample distribution with the least amount of prior information embedded into it. During the k^{th} iteration, the process initiates by sampling $\beta > 1$ candidate solutions $\boldsymbol{x}_i \in \mathbb{R}^p$, where $i = [1, \ldots, \beta]$, from a multivariate normal distribution $\mathcal{N}(\boldsymbol{m}_k, \sigma_k^2 \boldsymbol{\Sigma}_k)$.

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{m}_k, \sigma_k^2 \boldsymbol{\Sigma}_k) \ , \tag{4.6}$$

where $\boldsymbol{m}_k \in \mathbb{R}^p$ is the distribution mean and current solution to the optimization problem, $\sigma_k > 0$ is the step-size, and $\boldsymbol{\Sigma}_k \in \mathbb{R}^{p \times p}$ is a symmetric and positive-definite covariance matrix. The candidate solutions $\boldsymbol{x}_i$ are evaluated using the objective function $f : \mathbb{R}^p \to \mathbb{R}$ and are subsequently sorted as follows [48]:

$$\{\boldsymbol{x}_{j:\beta} \mid j = 1 \dots \beta\} = \{\boldsymbol{x}_j \mid j = 1 \dots \beta\}, \tag{4.7}$$

where $f(\boldsymbol{x}_{1:\beta}) \leq \cdots \leq f(\boldsymbol{x}_{\mu:\beta}) \leq f(\boldsymbol{x}_{\mu+1:\beta}) \leq \cdots$, and $\mu \leq \beta/2$ is the number of best candidate solutions selected at each iteration. The mean is updated as follows:

$$\boldsymbol{m}_{k+1} = \sum_{r=1}^{\mu} w_r \boldsymbol{x}_{r:\beta}. \tag{4.8}$$

In this context, the positive recombination weights $w_1 \geq w_2 \geq \cdots \geq w_\mu > 0$ are selected such that their sum equals one. Conventionally, these weights are determined to satisfy the condition $1/\sum_{r=1}^{\mu} w_r^2 \approx \beta/4$.

The step-size σ_k is updated using cumulative step-size adaptation (CSA), sometimes also denoted as path length control [47].

$$\sigma_{k+1} = \sigma_k \times \exp(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\boldsymbol{p}_\sigma\|}{\mathrm{E}\|\mathcal{N}(0, I)\|} - 1 \right), \tag{4.9}$$

where $c_\sigma^{-1} \approx p/3$ is the backward time horizon for the evolution path $\boldsymbol{p}_\sigma$ and larger than

one, and d_σ is the damping parameter.

Finally, the covariance matrix is updated as follows [48]:

$$\boldsymbol{\Sigma}_{k+1} = \left(1 - p^2 - \frac{\mu_w}{p^2} + c_s\right)\boldsymbol{\Sigma}_k + c_1 \boldsymbol{p}_c \boldsymbol{p}_c^T + c_\mu \sum_{r=1}^{\mu} w_i \frac{\boldsymbol{x}_{r:\beta} - \boldsymbol{m}_k}{\sigma_k}\left(\frac{\boldsymbol{x}_{r:\beta} - \boldsymbol{m}_k}{\sigma_k}\right)^T \tag{4.10}$$

where $c_c^{-1} \approx p/4$ is the backward time horizon for the evolution path $\boldsymbol{p}_c$, which is greater than one, $\mu_w = 1/\left(\sum_{i=1}^{\mu} w_i^2\right)$ is the variance effective selection mass, and $c_s = \left(1 - \mathbf{1}_{[0,1.5\sqrt{p}]}\left(\|\boldsymbol{p}_\sigma\|\right)^2\right) c_1 c_c \left(2 - c_c\right)$. The indicator function $\mathbf{1}_{[0,1.5\sqrt{p}]}\left(\|\boldsymbol{p}_\sigma\|\right)$, yields a value of one if and only if $\|\boldsymbol{p}_\sigma\| \leq 1.5\sqrt{p}$.

4.4 Numerical Results

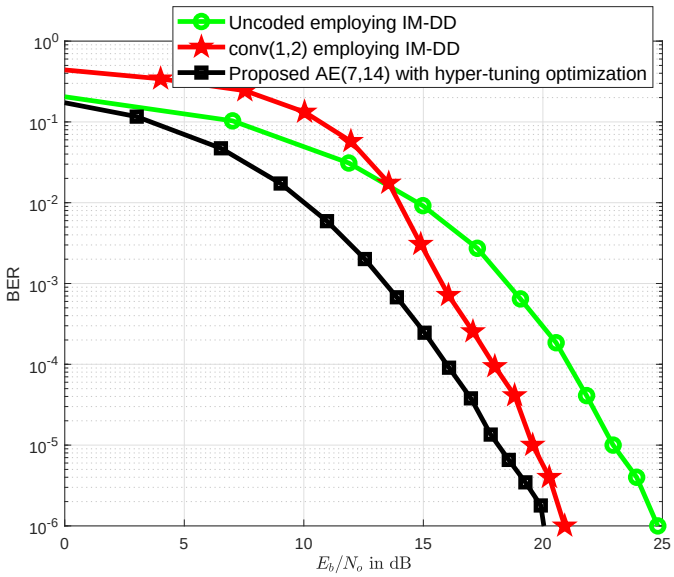


Figure 4.3: BER versus E_b for $\lambda = 1$ of the proposed AE with model-based frameworks.

Algorithm 3 CMA-ES

Require: number of parameters p, number of iterations v number of candidate solutions β, number of solutions selected at each iteration μ, recombination weights $w_1, w_2, \ldots, w_\mu$.

Ensure: minimize $f(\boldsymbol{x}_i)\ \forall i \in \{1, 2, \ldots, \mu\}$

$\boldsymbol{m}_0 \leftarrow$ initialize mean vector.
$\sigma_0 \leftarrow$ initialize step size.
$\boldsymbol{\Sigma}_0 \leftarrow$ initialize covariance matrix.
for $k \leftarrow 1$ to v **do**
  for $i \leftarrow 1$ to μ **do**
    $\boldsymbol{x}_i \leftarrow$ sample_multivariate_normal$(\boldsymbol{m}_{k-1}, \sigma_{k-1}^2 \boldsymbol{\Sigma}_{k-1})$
    $f_i = f(\boldsymbol{x}_i)$
  end for
  $\boldsymbol{x}_{1\ldots\mu} \leftarrow \boldsymbol{x}_{s(1)\ldots s(\mu)}$
  with $s(i) = \text{argsort}\,(f_{1\ldots\mu}, i)$ {sort solutions}
  $\boldsymbol{m}_k \leftarrow$ update_m$(x_1, \ldots, x_\mu)$ {move mean to better solutions}
  $\boldsymbol{p}_{\sigma k} \leftarrow$ update_ps$(\boldsymbol{p}_{\sigma k-1}, \sigma_{k-1}^{-1} {\boldsymbol{\Sigma}_{\boldsymbol{k-1}}}^{-1/2} (\boldsymbol{m}_k - \boldsymbol{m}_{k-1}))$ {update isotropic evolution path}
  $\boldsymbol{p}_c \leftarrow$ update_pc $\left(p_c, \sigma^{-1} (\boldsymbol{m}_k - \boldsymbol{m}_{k-1}), \left\|\boldsymbol{p}_{\sigma k-1}\right\|\right)$ {update anisotropic evolution path}
  $\boldsymbol{\Sigma}_k \leftarrow$
  update_cov$(\boldsymbol{\Sigma}_{k-1}, \boldsymbol{p}_{ck}, (\boldsymbol{x}_1 - \boldsymbol{m}_{k-1})/\sigma_{k-1}, \ldots, (\boldsymbol{x}_\mu - \boldsymbol{m}_{k-1})/\sigma_{k-1})$ {update covariance matrix}
  $\sigma_k \leftarrow$ update_step $(\sigma_{k-1}, \|\boldsymbol{p}_{\sigma k}\|)$ {update step-size}
end for

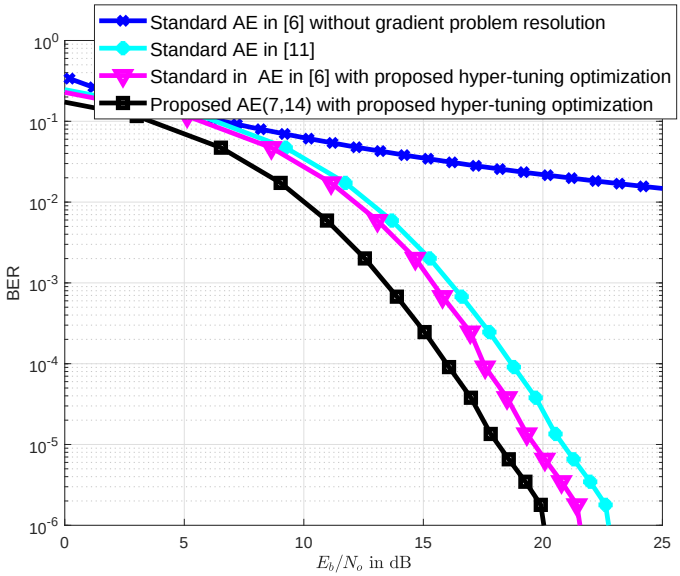


Figure 4.4: BER versus E_b for $\lambda = 1$ of the proposed AE with learning frameworks.

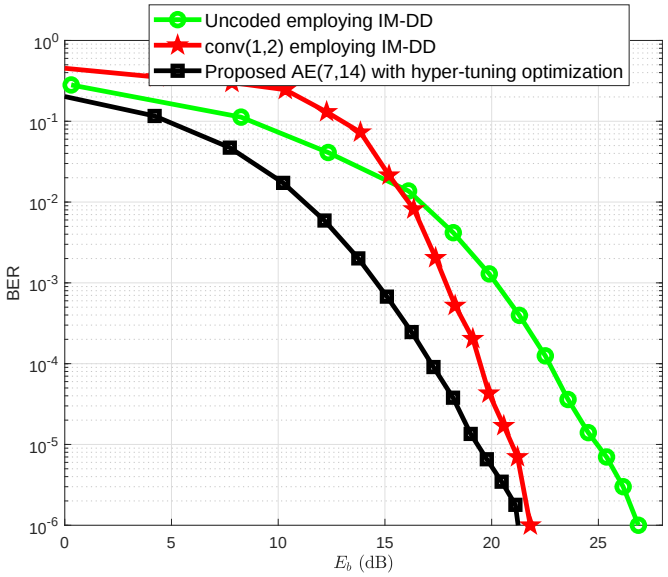


Figure 4.5: BER versus E_b for $\lambda = 2$ of the proposed AE with model-based schemes

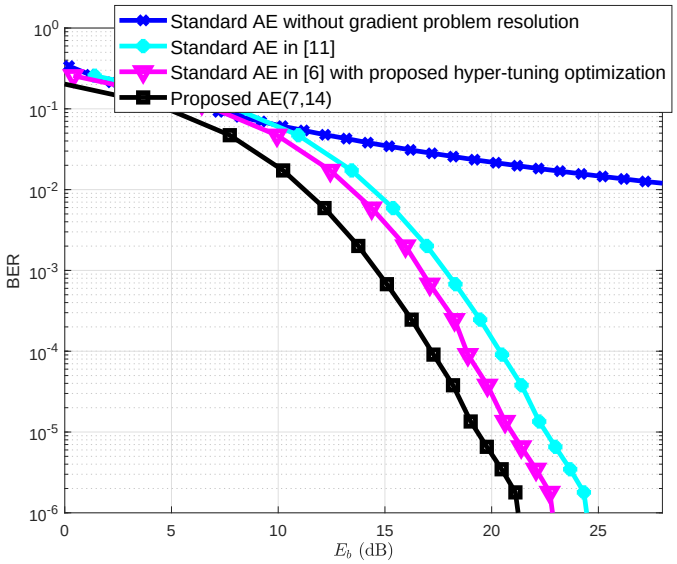


Figure 4.6: BER versus E_b for $\lambda = 2$ of the proposed AE with learning frameworks.

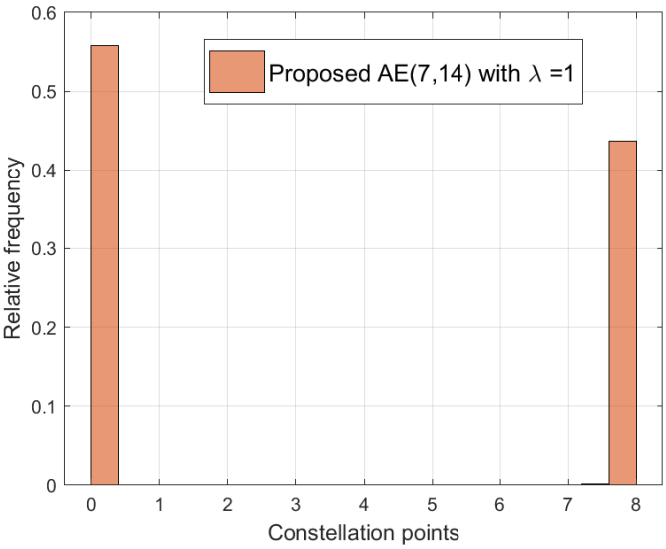


Figure 4.7: The constellation points versus the relative frequency: Proposed AE with CMA-ES algorithm,

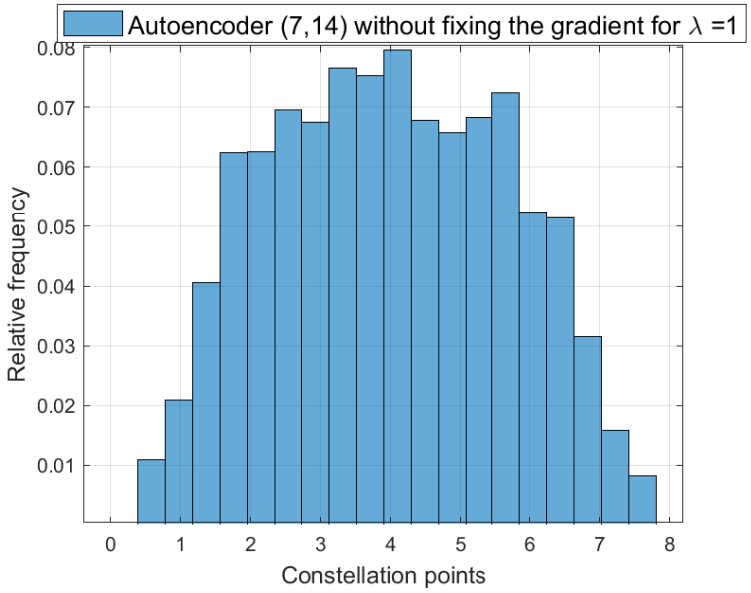


Figure 4.8: The constellation points versus the relative frequency: AE without fixing the channel gradient.

In this section, we evaluate the performance of our proposed AE with normalization layers and gradient fix against several channel coding schemes for code rate $R = \frac{1}{2}$ across the Poisson channel. The proposed AE presented in Fig. 4.2 is trained over 25 epochs with 8 million training samples. Not only utilize the CMA-ES algorithm for the channel gradient of the Poisson channel, but also for the learning rate, peak intensity, and batch size. The hyper-tuning parameters, optimized by the CMA-ES algorithm for $\lambda = 1$, are as follows: $A = 7.3$, learning rate $= 0.00045$, batch size $= 32$, and Poisson channel gradient $= 1.23$. For $\lambda = 2$, the optimized hyper-tuning parameters are $A = 7.8$, learning rate $= 0.00012$, batch size $= 32$, and Poisson channel gradient $= 1.8$. To evaluate training loss, we utilize the CE loss and for updating the weights we utilize Adam optimizer. Finally, in testing, we utilize 5 million samples and the

optimal learnable parameters proposed from the training stage. Figs. 4.3 and 4.4 show the result of testing models on the Poisson channel with $\lambda = 1$, while Figs. 4.5 and 4.6 refer to the results with $\lambda = 2$.

In Figs. 4.3 and 4.5, the proposed AE is compared to the standard AE(7, 14), uncoded IM/DD, proposed AE(7, 14) without gradient fix, and convolutional codes. These results show that the proposed AE achieves better performance than model-based and other DL approaches across the Poisson channel. In Fig. 4.5, at BER 10^{-5}, the proposed AE has an improvement of about 1.5 dB over convolutional codes. Additionally, Fig. 4.3 shows that the proposed AE is better by at least 1 dB over all other encoding schemes tested. Through these baselines, the proposed AE is shown to be more effective across the Poisson channel than both the model and learning-based frameworks.

In Figs. 4.4 and 4.6, the proposed AE is compared to the standard AE(7, 14) and uncoded IM/DD again, as well as the sparse AE(7, 14) and standard AE with gradient fix. In Fig. 4.4, at BER 10^{-5}, the proposed AE is better by about 1 dB over the standard AE with gradient fix. Similarly, Fig. 4.6 also shows an improvement of approximately 1 dB over the standard AE. The better performance of the proposed AE over the standard AE with gradient fix proves that introducing normalization layers into the AE model greatly improves the accuracy. Finally, since the AE is shown to have relatively low E_b for each given BER, it can be more efficiently implemented in hardware for SOC than other methods, as the proposed AE can get accurate results using less power.

As shown in Fig. 4.7, upon calculating the channel gradient for the Poisson channel, the histogram illustrates that the modulation utilized by the proposed AE resembles on-off keying modulation at its peak intensity $A = 8$. Conversely, when employing the AE without estimating the channel gradient as illustrated in Fig. 4.8, the histogram displays a randomly distributed pattern of the modulated signal from the AE. This randomness substantiates the notable degradation in BER performance of the AE when the gradient of the Poisson channel is not estimated.

4.5 Conclusion

This chapter explores the utilization of an AE model within a point-to-point SOC scenario, considering the impact of a practical channel model referred to as the Poisson channel. While the Poisson channel model effectively characterizes the SOC, its non-differentiable attributes pose challenges for DL models. A novel non-gradient-based optimization framework has been employed to estimate the channel gradient, effectively tackling the non-differentiable nature of Poisson channels. Moreover, our AE integrates normalization layers in both the encoding and decoding modules. The numerical results demonstrate that the proposed AE outperforms both state-of-the-art learning frameworks and model-based schemes in BER performance within the Poisson channel. Our model presents a promising solution that incorporates the Poisson channel into deep learning models without relying on approximations or transformations.

Chapter 5

Learning-Based Autoencoder for Multiple Access and Interference Channels in Space Optical Communications

5.1 Motivation

The study of MAC and IC channels is fundamental in SOC systems, particularly under the influence of Log-normal fading. These channels are essential for enabling efficient communication in multi-user environments where resources such as power and bandwidth are shared. SOC systems, which operate under unique constraints like Log-normal fading, weak turbulence, and high path loss, require a thorough understanding of MAC and IC models to address inter-user interference, optimize resource utilization,

and ensure reliable data transmission.

In MAC channels, multiple users transmit to a single receiver, such as a satellite receiving data from various ground stations. The efficient operation of MAC systems in SOC ensures fair resource allocation, effective scheduling, and minimized inter-user interference, which are vital for applications like satellite broadband and remote sensing. IC channels, on the other hand, represent a more complex scenario with multiple transmitters and receivers, leading to significant interference between links. This is especially challenging in SOC, where Log-normal fading amplifies signal variations, making robust interference management critical for maintaining performance.

The importance of studying MAC and IC in SOC extends to practical applications in satellite IoT, collaborative sensing, and inter-satellite communication. Advancements in these channels facilitate improved spectral efficiency, reduced latency, and enhanced resilience to Log-normal fading, enabling SOC systems to meet the growing demand for reliable and high-capacity multi-user communication. Furthermore, addressing the complexities of interference in Log-normal fading conditions supports the scalability of SOC systems, ensuring efficient operation across diverse applications.

This chapter presents an AE model for multi-user SOC systems operating under Log-normal fading in the weak turbulence regime. The proposed model incorporates batch normalization in encoders and layer normalization in decoders, enabling effective gradient-based optimization while mitigating the challenges posed by Log-normal fading and inter-user interference. Numerical results demonstrate that the AE achieves a 1 dB performance improvement over state-of-the-art frameworks with a 20% reduc-

tion in computational complexity. Moreover, the model's superior BER performance in both MAC and IC scenarios underscores its potential to enhance inter-user interference mitigation, optimize resource allocation, and advance the scalability of SOC systems.

5.2 SOC Channel Model

A MAC channel is established between K GEO satellites that transmit signals to a shared receiver located on the ground station. The use of system tool kit (STK) simulator enables accurate modeling of the SOC channels [38]. In the system, the ground station holds the receiver antenna gimbal and avalanche photo-detector. Unlike RF coherent communication, the modulated signal in intensity-modulation and direct-detection (IM/DD) is real and non-negative. Additionally, the signal is peak-constrained in SOC for operation and safety regulations [46]. The Log-normal distribution is commonly used to describe weak atmospheric turbulence, recommended by STK for the GEO to ground SOC channel [1]. In our model, we take into consideration both additive white Gaussian noise (AWGN) and slow fading channels in the multi-user SOC channel. In this particular configuration, we assumed the Gaussian channel; however, when operating at low power levels, a better channel model would be a Poisson channel [46].

5.3 Multi-User SOC Channel Based on Autoencoders

The notation (k, L) is used to denote the proposed AE model, where k represents the number of message bits and L represents the codeword length. The proposed AE(k, L) in the SOC system with rate $R = k/L$. The system consists of three distinct modules: the transmitter, receiver, and channel layers. There are K independent transmitters and one shared receiver in this MAC channel. Each transmitter is located in a GEO satellite, while the common receiver is the ground station. Over a SOC channel, transmitter $i \in \{1, 2, ..., K\}$ send the message b_i to the common receiver in the ground station, where $b_i \in \mathcal{M} = \{1, \ldots, M\}$, and $M = 2^k$.

Transmitter: The transmitter starts by selecting b_i, which is one of M possible messages and then converts it into a one-hot vector $\mathbf{1}_{b_i}$ of size M, which has a 1 in the message index position and 0's elsewhere. Utilizing a one-hot encoded vector ensures equal significance for all messages since they are all in binary rather than ordinal format. Then, using the mapping function $\boldsymbol{u} : \mathcal{M} \to \mathbb{R}^L$, each transmitter converts the input one-hot vector $\mathbf{1}_{b_i}$ into the encoded vector $\boldsymbol{x}_i^L$. Obviously, each encoder applies both modulation and channel coding simultaneously. The codebook consists of all possible codewords generated by the encoder of the AE, i.e., the set $\{\boldsymbol{x}_i^L\}$, with cardinality 2^k. When the transmitter normalization stage outputs the symbol vector $\boldsymbol{x}_i^L$, it meets the non-negativity and peak conditions for SOC. To meet the constraints $0 \leq \boldsymbol{x}_i^L(l) \leq A$ with $l = 1, \ldots, L$, each transmitter applies a normalization layer to the transmitted symbols using the weighted sigmoid function. We examine a two-user MAC channel

depicted in Fig. 5.1, without loss of generality. Fully connected (FC) layers provide the basis of the transmitter model, with a subsequent BN layer added directly after each FC layer. The BN normalizes the layer's inputs for each mini-batch, which both shortens the training time and keeps the learning process stable. The gradient explosion issues are also lessened by the BN, according to [49]. Until the weighted sigmoid normalization step in the encoder, it is clear that no input scaling takes place. Since the input scaling does not change between training and the testing, BN parameters learned during training will lead to the same superior performance throughout testing.

SOC Channel Model: The normalized vector $\boldsymbol{x}_i^L$ for both transmitters are fed into the SOC channel. The SOC channel is composed of two components: an AWGN channel, and a Log-normal fading model with a standard deviation of σ. The channel parameters are provided by the STK simulator discussed earlier. The input to the neural networks on the receiver's side is represented as $\boldsymbol{y}^L$.

$$\boldsymbol{y}^L = \sum_{i=1}^{K} h_i \boldsymbol{x}_i^L + \boldsymbol{w}^L, \tag{5.1}$$

where $\boldsymbol{w}^L \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and the Log-normal fading coefficient, represented by h_i for the transmitter i.

Common Receiver: Lastly, the receiver-based ground station processes the corrupted vector $\boldsymbol{y}^L$ and generates the estimated one-hot vector $\mathbf{1}_{\hat{b}_i}$. The common receiver includes K decoders, each is composed of multiple FC layers followed by Layer Normalization (LN). Since the received symbols $\boldsymbol{y}^L$ will be scaled with different A values

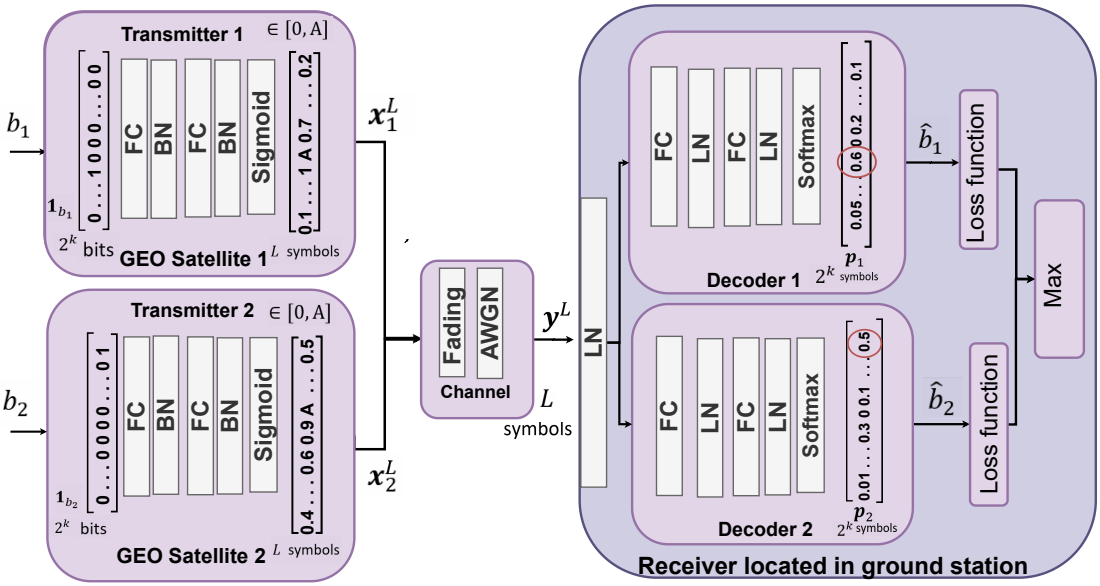


Figure 5.1: Proposed AE architecture in a two-user MAC channel.

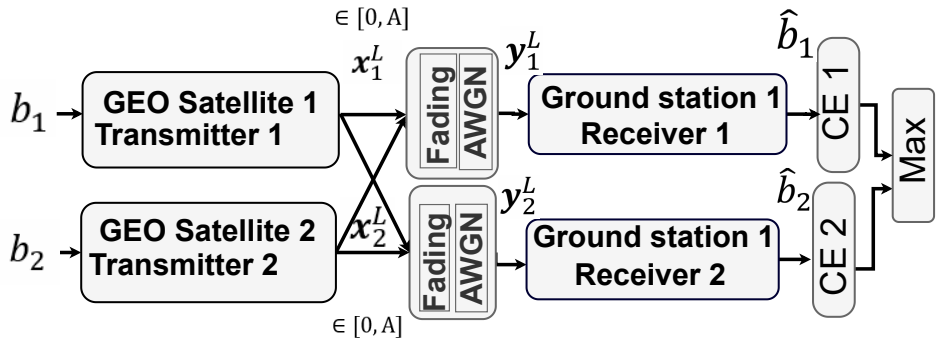


Figure 5.2: The proposed AE architecture in a two-user IC setting.

than the one used in training, using BN in the decoder can not aid to standardize the data, since the testing input scale differs from the training due to varying A values. BN's learned parameters can not work well in testing unless A matches the trained value.

To prevent this issue, the receiver could wait and collect a certain amount of samples in order to produce the desired results. This implies that the BN will continue to work in the same manner during both training and testing. However, this assumption is impractical in wireless communications. Accordingly, we need to design the system for deployment to deal with received codewords independently without the need for a significant number of samples before the processing begins. Therefore, LN is introduced in the decoders, which operates sample by sample, in order to maintain efficient standardization across the entire AE system. The process of LN involves re-centering and re-scaling its input. Since LN works in the same manner in both training and testing phases, the effect of scaling the input to the decoder is mitigated, and trained decoder weights are effective in testing. Additionally, the corrupted vector $\boldsymbol{y}^L$ undergoes a single LN unit before being inputted to the decoders. This guarantees consistent scaling and enhances system performance during testing across various SNR values. The cross-entropy (CE) loss function for each AE is defined as

$$C_j = -\sum_{r=1}^{M} \mathbf{1}_{b_i}(r) \log \boldsymbol{p}_j(r), \tag{5.2}$$

where $\mathbf{1}_{b_i}(r) \in \{0, 1\}, i = j$ is the r^{th}value in the input one-hot vector $\mathbf{1}_{b_i}$. The softmax

is applied at the last layer of the j^{th} decoder producing the probability value $\boldsymbol{p}_j(v)$ and it can be expressed as follows

$$\boldsymbol{p}_j(v) = \frac{e^{\boldsymbol{q}_j(v)}}{\sum_{t=1}^{M} e^{\boldsymbol{q}_j(t)}}, \tag{5.3}$$

where $\boldsymbol{q}_j$ is the j^{th} decoder final layer output. By employing the min-max algorithm, it is possible to decrease the error probability as follows

$$\max\left(\Pr\left\{b_1 \neq \hat{b}_1\right\}, \Pr\left\{b_2 \neq \hat{b}_2\right\}, ..., \Pr\left\{b_K \neq \hat{b}_K\right\}\right). \tag{5.4}$$

The use of the min-max algorithm can lead to a decrease in error probability by selecting a decision rule that aims to minimize the maximum possible loss [19]. In light of this, in a given step, only the weights corresponding to the maximum loss are updated, while all other weights are left intact. Overall, the system's loss function is given by

$$C = \max\left(C_1, C_2, ..., C_K\right). \tag{5.5}$$

Algorithm 1 summarizes the learning strategy for the novel AE design in the MAC channel.

Interference channel: The idea of MAC AE can be easily expanded to incorporate multiple transmitters and receivers that utilize a shared channel. Both transmitter-receiver pairs are implemented as NNs and the difference with respect to the MAC AE is that the encoded vectors $\boldsymbol{x}_i^L, \forall i \in \{1, 2, ..., K\}$ now interfere at the

receivers, resulting in the noisy observations

$$\boldsymbol{y}_j^L = \sum_{i=1}^{K} h_{ij}\boldsymbol{x}_i^L + \boldsymbol{w}_j^L, \forall j \in \{1, 2, ..., K\} \tag{5.6}$$

The Log-normal fading coefficient, represented by h_{ij}, refers to the coefficient that connects the i^{th} transmitter with the j^{th} receiver, where i and $j \in \{1, 2, ..., K\}$. The target for receiver j now is to estimate the message b_i transmitted by the i^{th} transmitter, while ignoring any other messages that may interfere with it. Furthermore, since the decoders are located at different ground stations, each decoder will have a separate LN unit at its input to negate the effect of scaling $\boldsymbol{y}_i^L$ with different SNR values during testing. Without loss of generality, we consider here the two-user IC as shown in Fig. 5.2. Transmitter 1 wants to communicate message $b_1 \in \mathcal{M}$ to receiver 1 while transmitter 2 wants to communicate message $b_2 \in \mathcal{M}$ to receiver 2.

5.4 Simulation Results

In this section, we present the symbol detection performance evaluated for the proposed AE for code rate $R = 1/3$ in two different scenarios: MAC and IC with two users. Also, the proposed AE architecture follows the layout given in Figs. 5.1 and 5.2. The learnable parameters are tuned using the stochastic gradient descent (SGD) algorithm with the Adam optimizer and a learning rate of 0.0001. The model is trained using $5,000,000$ randomly generated samples for 20 epochs with the objective of minimizing the training loss. Finally, the system BER performance is evaluated

Algorithm 4 AE training algorithm in a K-user MAC channel

Require: M messages $(1, 2, \ldots, M) \in \mathcal{M}$, transmitter peak intensity A, batch size m, learning rate η.

Ensure: $\text{argmax}(\boldsymbol{q}_j^{(t)}) = b_i^{(t)}\ \forall t \in \{1, 2, \ldots, m\}, i, j \in \{1, 2, \ldots, K\}$

$\theta_1, \ldots, \theta_K \leftarrow$ initialize encoder units parameters.
$\phi_1, \ldots, \phi_K \leftarrow$ initialize decoder units parameters.
$\upsilon \leftarrow$ initialize LN layer parameters.
repeat
  Draw m minibatch samples for each user $((b_1^{(1)}, \ldots, b_K^{(1)}), \ldots, (b_1^{(m)}, \ldots, b_K^{(m)}))$.
  for $t \leftarrow 1$ to m **do**
    for $z \leftarrow 1$ to K **do**
      $\mathbf{1}_{b_z^{(t)}} \leftarrow \text{one\_hot\_vector}(b_z^{(t)})$ $\quad \{\mathbf{1}_{b_z^{(t)}} \in \{0,1\}^M\}$
      $\boldsymbol{x}_z^{(t)} \leftarrow E_{\theta_z}(\mathbf{1}_{b_z^{(t)}}, A)$ $\quad \{\boldsymbol{x}_z^{(t)} \in [0, A]^L\}$
    end for
    $\boldsymbol{y}^{(t)} \leftarrow \sum_{z=1}^{K} \boldsymbol{x}_z^{(t)} \boldsymbol{h}_z^{(t)} + \boldsymbol{w}^{(t)}$ $\quad \{\boldsymbol{y}^{(t)} \in \mathcal{R}^L\}$
    $\boldsymbol{y}^{(t)} \leftarrow \boldsymbol{y}^{(t)} / \hat{\boldsymbol{h}}^{(t)}$
    for $z \leftarrow 1$ to K **do**
      $\boldsymbol{q}_z^{(t)} \leftarrow D_{\phi_z}(LN_\upsilon(\boldsymbol{y}^{(t)}))$ $\quad \{\boldsymbol{q}_z^{(t)} \in \mathcal{R}^M\}$
      $\boldsymbol{p}_z^{(t)} \leftarrow \text{Softmax}(\boldsymbol{q}_z^{(t)})$ $\quad \{\boldsymbol{p}_z^{(t)} \in [0,1]^M\}$
    end for
  end for
  for $z \leftarrow 1$ to K **do**
    $C_z \leftarrow -\sum_{t=1}^{m} \sum_{r=1}^{M} \mathbf{1}_{b_z^{(t)}}(r) \log(\boldsymbol{p}_z^{(t)}(r))$
  end for
  $C = \max(C_1, C_2, \ldots, C_K)$
  for $\tau \leftarrow \theta_1, \ldots, \theta_K, \phi_1, \ldots, \phi_K, \upsilon$ **do**
    $\nabla_\tau C \leftarrow \frac{\partial C}{\partial \tau}$
    $\tau \leftarrow \tau - \eta \nabla_\tau C$
  end for
until convergence

Table 5.1: Layers design of the AE(k, L).

Module	Layer	Input shape	Output shape	Number of parameters
Transmitter	Fully connected	M	100	$100(M+1)$
	Batch normalization	100	100	400
	Fully connected	100	L	$101L$
	Batch normalization	L	L	$4L$
	Weighted sigmoid	L	L	0
Normalizer	Layer normalization	L	L	$2L$
Decoder	Fully connected	L	100	$100(L+1)$
	Layer normalization	100	100	200
	Fully connected	100	M	$101M$
	Layer normalization	M	M	$2M$

over $1,000,000$ testing samples. The computational complexity of the system can be optimized by assigning the appropriate number of learnable parameters for each layer. We make our source code publicly available at https://github.com/abdo-ui. Our dataset is generated using Python as random binary data and the fading coefficients are obtained from the STK simulator. To ensure a fair comparison, all benchmark models were trained with equal numbers of epochs and training samples. This ensures that any advantages observed in the proposed model can be attributed solely to its design. Also, note that the training is often done off-line and only once so that only the complexity during testing really matters. In addition, Table 5.1 provides a breakdown of the relevant parameters for each layer of the proposed AE(k, L) model.

The proposed AE$(7, 21)$ is compared with the uncoded IM/DD, LDPC with time sharing, the standard AE$(7, 21)$ presented in [19], and the sparse AE$(7, 21)$ presented in [33] based on the BER performance metric. The comparison is made under two channel conditions: AWGN MAC and Log-normal fading MAC. Figures 5.3 and

5.4 present the MAC channel results. The LDPC scheme, which utilizes IM/DD modulation with time sharing settings, is characterized by a block length of $b_l = 100$. The design of this LDPC configuration of code rate 1/3 follows the guidelines in [45]. In AWGN MAC channel illustrated in Fig. 5.3, our proposed model outperforms the standard AE by 2.1 dB at BER 10^{-6} and is ahead of both the sparse AE and uncoded IM/DD by 4 and 5.7 dB, respectively. As shown in Fig. 5.4, the proposed AE(7,21) is better than the standard AE by 1.5 dB at a BER of 10^{-6}. In addition, the proposed AE(7,21) with time sharing settings outperforms the LDPC with time sharing at low SNR regime and has 1.3 dB gain at BER 10^{-4}. Also, the proposed AE with time sharing has the same performance as the LDPC at BER 10^{-6}. In the Log-normal MAC and IC channel, the standard deviation σ is set to 0.2 [6]. The discrepancy in BER performance between the standard AE, sparse AE, and the proposed AE highlights the positive impact of integrating normalization layers and the min-max algorithm in the proposed model. In addition, we compare our system against the uncoded IM/DD, LDPC IM/DD, the sparse AE, and the AE system presented in [7] in the presence of two-user IC for Log-normal fading channel.

Fig. 5.5 illustrates the evaluation of the proposed AE in an IC channel-based AWGN setting. The results indicate that the proposed AE outperforms the sparse AE and the uncoded IM/DD by 3 dB and 5.5 dB, respectively. In addition, Fig. 5.6 shows a comparison of all models in a Log-normal two-user IC channel with $\sigma = 0.2$. The results demonstrate that the proposed AE achieves better performance than the standard AE proposed in [7], with an improvement of 1.8 dB at a BER of 10^{-6}. In addition, the

proposed AE with time sharing is just inferior with 0.3 compared to the LDPC with time sharing at BER 10^{-6}. However, the proposed AE with time sharing is better by 1.2 and 0.8 dB compared to the LDPC with time sharing at BER 10^{-3} and 10^{-4}, respectively.

Furthermore, Table 5.2 reveals that the proposed AE is more computationally efficient than the latest learning-based models. In the AWGN MAC channel, the standard AE$(7, 21)$ requires more 2.1 dB more in terms of SNR than the proposed AE$(7, 21)$ to achieve a BER of 10^{-6}, even though it has 23% more learnable parameters. Moreover, the sparse AE$(7, 21)$ requires a 4 dB higher SNR value than the proposed AE$(7, 21)$ to achieve 10^{-6} BER, despite having only 3% fewer learnable parameters. Fig. 5.7 visualizes the learned representations $\boldsymbol{x}$ of all messages as real constellation points for the proposed AE$(7, 21)$ in MAC channel. The histogram in Fig. 5.7 is a visualization of the learned constellations at the transmitter for all possible messages trained and tested for $A = 4$ over a two-user AWGN and log-normal fading channels. Obviously, the results in Fig. 5.7 are applied in the testing stage after achieving the best weights for minimizing the loss according to the min-max problem. The results in Fig. 5.7 are in symbol representations as the transmitter in the AE applies both channel coding and symbol mapping simultaneously considering both the peak and positivity constraints. The distribution of the encoded symbols is generated after the training stage. While Fig. 5.7 looks like an on-off-keying modulation and similar constellation points as represented in [6, 19], the scattering of the points is still different and not all points are exactly located in 0 and A. In particular Fig. 5.7, verifies that the AE output follows

Table 5.2: Computational complexity for each learning-based model.

Model	Number of learnable parameters
Sparse AE(7,21) [33]	60,298
Standard AE(7,21) [7, 19]	77,098
Proposed AE(7,21)	62,220

both positivity and peak intensity intensity constraints $\in [0, A]$. It is interesting to observe that the learned constellation points for the AE(7,21) are scattered in the interval $[0, A]$ with different relative frequencies. These findings indicate that the AE effectively acquired efficient coding, modulation techniques in a MAC scenario.

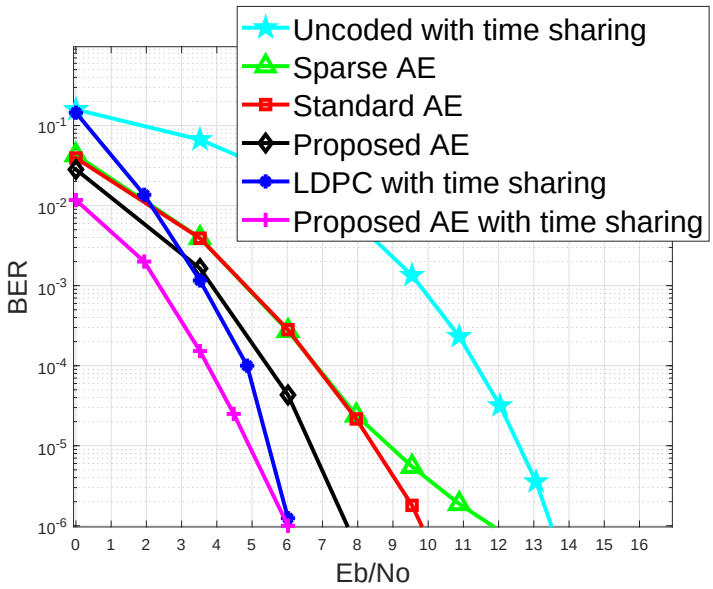


Figure 5.3: BER versus $\frac{E_b}{N_o}$ in a two-user MAC AWGN channel.

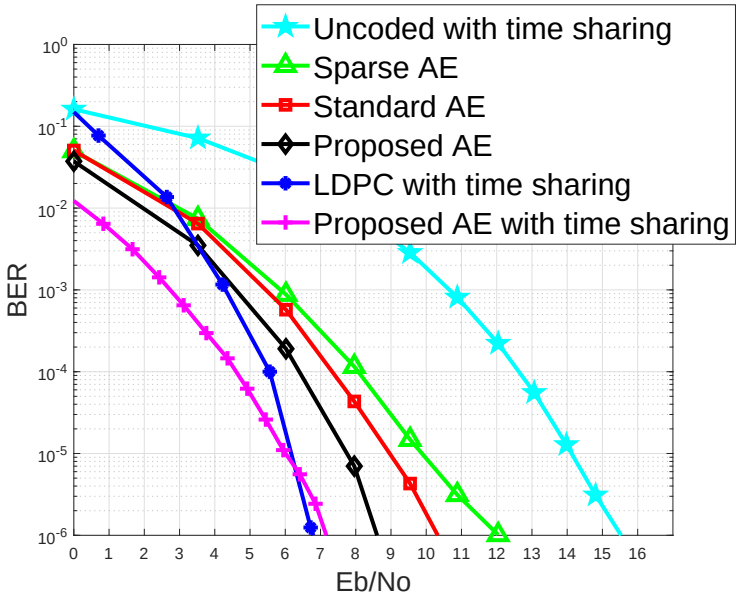


Figure 5.4: BER versus $\frac{E_b}{N_o}$ in a two-user MAC Log-normal channel.

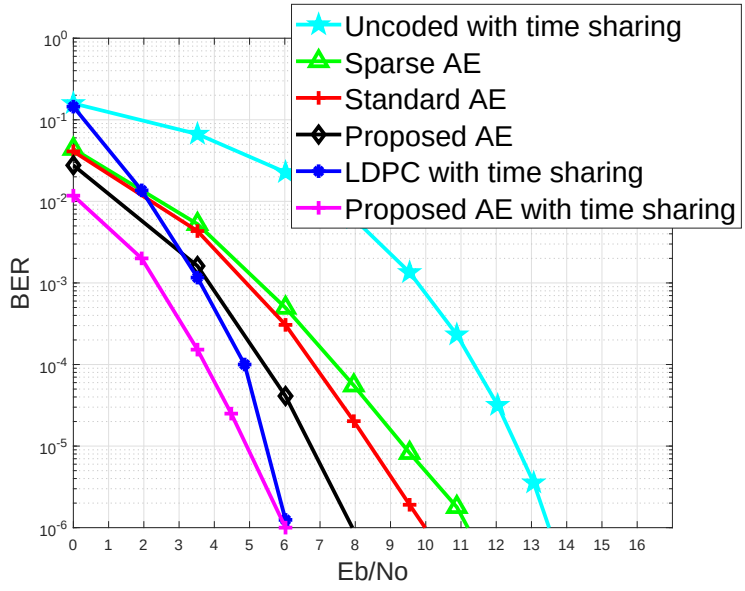


Figure 5.5: BER versus $\frac{E_b}{N_o}$ in a two-user IC AWGN channel.

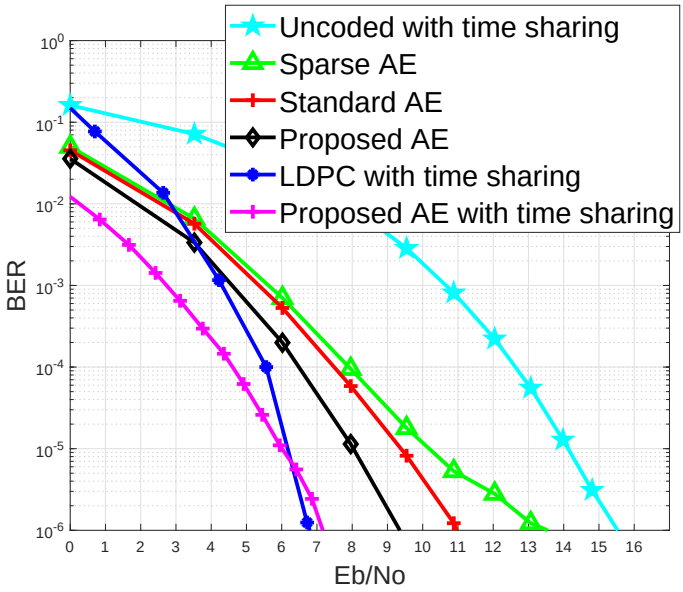


Figure 5.6: BER versus $\frac{E_b}{N_o}$ in a two-user IC Log-normal channel.

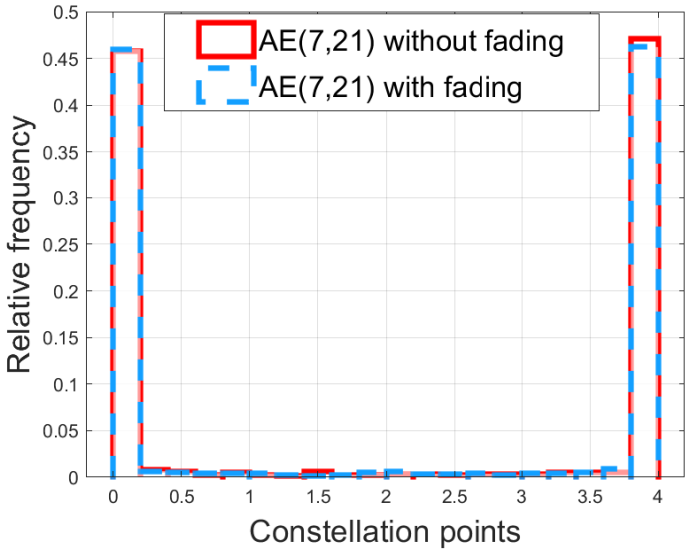


Figure 5.7: The constellation points against the relative frequency generated by the proposed AE$(7, 21)$ considering a peak intensity of $A = 4$.

5.5 Conclusion

In this chapter, we develop a DL AE model using a novel layered framework that incorporates BN in the encoder and LN in the decoders in SOC specifically tailored for multi-user environments. A realistic SOC channel model for fading channels is created using the STK simulator. Our model focuses not only on enhancing the BER performance but also on optimizing computational complexity making it a promising solution for improving communication reliability and efficiency in SOC systems. The proposed AE enables scalability for any number of users in a multi-user environment. The numerical results indicate that the proposed AE exhibits superior performance in terms of BER and computational complexity compared to the existing learning frameworks in both MAC and IC channels. The proposed AE results yields better results that all E2E learning frameworks currently considered in the state-of-the-art, to the best of our knowledge.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this dissertation, significant advancements were made in addressing the challenges of SOC through learning-based frameworks. Chapter 3 introduced a neural network-based channel estimator optimized for single-signal-to-noise ratio training, which eliminates the need for multiple estimators and reduces complexity. The proposed design achieved performance comparable to the MMSE estimator while demonstrating superior computational efficiency. Additionally, an AE-based symbol detection framework was developed, enabling joint optimization of transmitter and receiver components. The innovative multi-code rate AE model demonstrated adaptability across various fading channel conditions, achieving robust performance under both perfect and imperfect knowledge of the CSI.

In Chapter 4, the focus shifted to channels governed by Poisson statistics, a

critical challenge in SOC. Traditional approaches often rely on Gaussian approximations, which are limited in photon-constrained scenarios. This work proposed an AE framework integrated with the CMA-ES, achieving near-optimal error rate performance while bypassing the limitations of differentiable channel models. This contribution is particularly significant for photon-limited applications, where channel characteristics differ from conventional assumptions.

Chapter 5 expanded the scope to multi-user SOC environments, particularly addressing scenarios with multiple users transmitting over shared channels and interference between users. The proposed AE models incorporated layer normalization, significantly improving error rate performance while maintaining computational efficiency. These designs demonstrated the ability to handle multi-user interference effectively and adapt to diverse operational conditions. By achieving scalability and robust performance in these multi-user scenarios, the work set a new benchmark for practical applications of SOC in complex communication networks.

Collectively, these contributions advance the field by addressing key challenges in symbol detection, channel estimation, and multi-user environments. The proposed frameworks are not only computationally efficient but also highly adaptable, providing robust solutions to critical limitations in current SOC systems. This work lays a strong foundation for future developments in efficient and scalable communication frameworks for both near-Earth and deep-space applications.

6.2 Future Work

6.2.1 Secure AEs in Wiretap Channels

Physical layer security remains a cornerstone of modern communication systems, especially in scenarios where data confidentiality is critical. Our initial results show that leveraging AEs for wiretap channels significantly enhances secrecy by reducing information leakage to eavesdroppers while improving reliability for legitimate users. These findings indicate that our proposed AE-based models outperform state-of-the-art techniques and approach the performance of convolutional codes operating near channel capacity. This highlights their promise as a robust framework for secure communication. Future research will focus on fine-tuning these AE architectures to address diverse channel conditions and adversarial scenarios, with an aim to set new benchmarks for secure wireless communications.

6.2.2 AE in Relay SOC

Relay-based SOC systems are pivotal for extending communication reach and reliability in space and terrestrial networks. Our proposed AE models have shown superior performance in symbol detection and channel estimation compared to existing methods. Notably, these models demonstrate potential for achieving near-capacity performance similar to advanced convolutional codes. This establishes a strong foundation for using AEs to tackle the challenges of signal degradation and relaying delays. Future work will explore the integration of hybrid model-based and learning-driven approaches

to further optimize relay SOC systems for higher throughput and energy efficiency.

6.2.3 AE in Intelligent Reflecting Surface

Intelligent Reflecting Surfaces (IRS) have emerged as a transformative technology for enhancing communication by dynamically altering propagation environments. Our initial studies reveal that AE-based frameworks can effectively adapt IRS configurations to improve data fidelity and enhance channel reliability. These models, which outperform traditional approaches and approach the performance limits of capacity codes, provide a promising starting point for IRS-aided secure communication. Moving forward, we will develop adaptive AE algorithms that fully exploit IRS capabilities for secure, efficient, and dynamic communication in challenging scenarios, such as high-mobility or multi-user environments.

6.2.4 Full Hardware Implementation for Proposed AE Model

To determine the feasibility of deploying our deep learning AE model in low-power hardware, a necessity for satellites where power is an extremely valuable resource, we deployed our AE model on a Raspberry Pi 4, as shown in Fig. 6.1. A video presentation showing this project can be found at the website of Zouheir Rezki's lab.

To do this, we first pruned the weights of our DNN in order to preserve the scarce memory resources of the Raspberry Pi 4, reducing their amounts by 25%. Due to being able to effectively deploy our code on low power hardware, this means that it is feasible for deployment in power-constrained environments, such as on satellites.

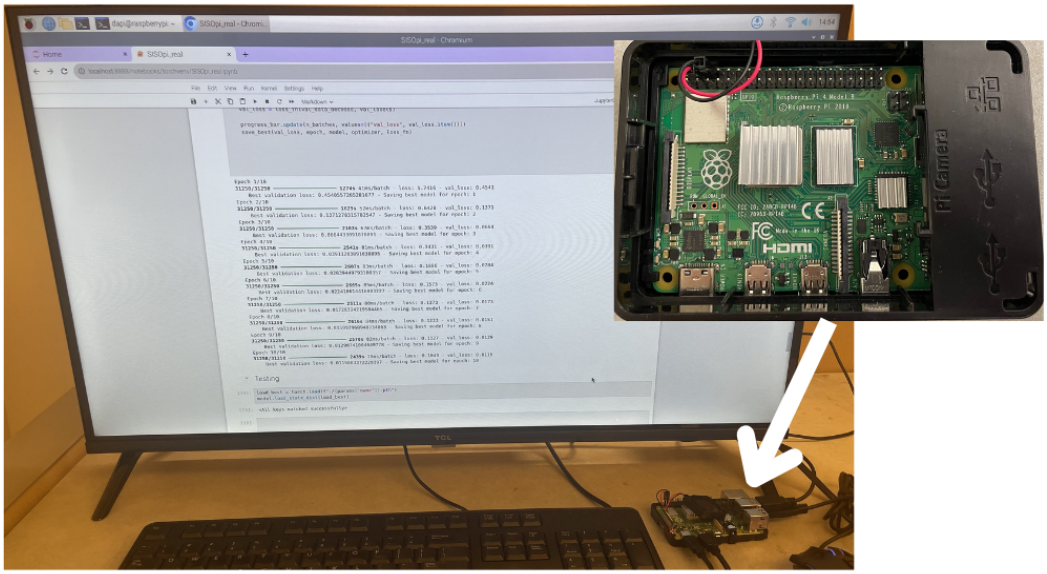


Figure 6.1: The proposed AE model running on Raspberry Pi 4 hardware.

After proving the feasibility of deployment, the next step in our research is to fully implement our model in hardware. This would consist of converting the data from the AE's encoder into analog values that would modulate the intensity of a laser diode at the transmitter using IM/DD, sending the laser beam across a channel of predetermined length, and receiving the laser beam with a photo-detector at the receiver.

Bibliography

[1] Abd El-Rahman A. El-Fikky and Zouheir Rezki. On the performance of autoencoder-based space optical communications. In *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pages 1466–1471, 2022.

[2] A. Biswas, D. Boroson, and B. Edwards. Mars laser communication demonstration: what it would have been. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6105 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 610502, February 2006.

[3] Hemani Kaushal and Georges Kaddoum. Optical communication in space: Challenges and mitigation techniques. *IEEE Commun. Surveys Tuts.*, 19(1):57–96, 2017.

[4] Hao Ye, Le Liang, Geoffrey Ye Li, JoonBeom Kim, Lu Lu, and May Wu. Machine learning for vehicular networks: Recent advances and application examples. *IEEE Vehicular Technology Magazine*, 13(2):94–101, 2018.

[5] Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. Deep reinforcement learn-

ing based resource allocation for v2v communications. *IEEE Transactions on Vehicular Technology*, 68(4):3163–3173, 2019.

[6] Zhao-Rui Zhu, Jian Zhang, Ru-Han Chen, and Hong-Yi Yu. Autoencoder-based transceiver design for owc systems in Log-normal fading channel. *IEEE Photon. J.*, 11(5):1–12, Aug. 2019.

[7] Timothy O'shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Trans. Cogn. Commun. Netw.*, 3(4):563–575, December 2017.

[8] Hao Ye, Geoffrey Ye Li, and Biing-Hwang Juang. Power of deep learning for channel estimation and signal detection in ofdm systems. *IEEE Wireless Communications Letters*, 7(1):114–117, 2018.

[9] Sebastian Dörner, Sebastian Cammerer, Jakob Hoydis, and Stephan Ten Brink. Deep learning based communication over the air. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):132–143, 2017.

[10] Hemani Kaushal and Georges Kaddoum. Optical communication in space: Challenges and mitigation techniques. *IEEE Communications Surveys Tutorials*, 19(1):57–96, 2017.

[11] Mohammad Ali Khalighi and Murat Uysal. Survey on free space optical communication: A communication theory perspective. *IEEE Communications Surveys Tutorials*, 16(4):2231–2258, 2014.

[12] Waseem Raza, Ethan Abele, John O'Hara, Behnaz Sadr, Peter LoPresti, Ali Imran, Wooyeol Choi, Ickhyun Song, Serhat Altunc, Obadiah Kegege, and Sabit Ekin. Toward a hybrid rf/optical lunar communication system (lunarcomm). *IEEE Net.*, 36(6):76–83, Nov. 2022.

[13] Jianing Wang, Jing Yi He, Yue Yang, Xiao Yu, and Xiaolong Ni. Stk and application in simulation of the space laser communication network. In *2012 International Conference on Optoelectronics and Microelectronics*, pages 308–312, 2012.

[14] Mohammed Elamassie and Murat Uysal. Incremental diversity order for characterization of fso communication systems over Log-normal fading channels. *IEEE Comm. Lett.*, 24(4):825–829, Apr. 2020.

[15] Haitham S. Khallaf, Kazutoshi Kato, Ehab Mahmoud Mohamed, Sadiq M. Sait, Halim Yanikomeroglu, and Murat Uysal. Composite fading model for aerial mimo fso links in the presence of atmospheric turbulence and pointing errors. *IEEE Wireless Comm. Lett.*, 10(6):1295–1299, Jun. 2021.

[16] Mohammadreza A. Kashani, Murat Uysal, and Mohsen Kavehrad. A novel statistical channel model for turbulence-induced fading in free-space optical systems. *Journal of Lightwave Technology*, 33(11):2303–2312, Jun. 2015.

[17] Chen Jiaxin and Han Junfeng. Research on initial pointing of inter-satellite laser communication. In *Proc. 12th International Conference 2020-2020 on Intelligent*

*Human-Machine Systems and Cybernetics (IHMSC),*Hangzhou, China, 2020, pp. 218-221.

[18] Anas Chaaban, Zouheir Rezki, and Mohamed-Slim Alouini. On the capacity of intensity-modulation direct-detection Gaussian optical wireless communication channels: A tutorial. *IEEE Commun. Surveys Tuts.*, 24(1):455–491, Firstquarter 2022.

[19] Morteza Soltani, Wael Fatnassi, Ahmed Aboutaleb, Zouheir Rezki, Arup Bhuyan, and Paul Titus. Autoencoder-based optical wireless communications systems. In *Proc. IEEE Globecom Workshops (GC Wkshps),* Abu Dhabi, United Arab Emirates, pages 1–6, Dec. 2018.

[20] Youngbin Na and Do-Kyeong Ko. Deep-learning-based high-resolution recognition of fractional-spatial-mode-encoded data for free-space optical communications. *Scientific Reports*, 11, 01 2021.

[21] Mohammad Karimzadeh and Mai Vu. Metrics and algorithms for designing convolutional codes with unequal error protection. *IEEE Transactions on Vehicular Technology*, 70(11):11169–11183, Nov. 2021.

[22] Dianxin Luan and John Thompson. Channelformer: Attention based neural solution for wireless channel estimation and effective online training. *IEEE Transactions on Wireless Communications*, pages 1–1, 2023.

[23] Kim Chia, Vishnu Monn Baskaran, KokSheik Wong, Moh Lim Sim, and

Chong Hin Chee. Recurrent network with attention for symbol detection in communication systems. In *2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–4, 2022.

[24] Mahdi Boloursaz Mashhadi and Deniz Gündüz. Pruning the pilots: Deep learning-based pilot design and channel estimation for mimo-ofdm systems. *IEEE Transactions on Wireless Communications*, 20(10):6315–6328, 2021.

[25] Abdul Karim Gizzini, Marwa Chafii, Ahmad Nimr, and Gerhard Fettweis. Enhancing least square channel estimation using deep learning. In *Proc. IEEE 91st Vehicular Technology Conference (VTC2020-Spring),* Antwerp, Belgium, pages 1–5, May 2020.

[26] David Neumann, Thomas Wiese, and Wolfgang Utschick. Learning the mmse channel estimator. *IEEE Trans. Signal Process.*, 66(11):2905–2917, Jun. 2018.

[27] Zhipeng Gao, Yuhao Wang, Xiaodong Liu, Fuhui Zhou, and Kat-Kit Wong. Ffdnet-based channel estimation for massive mimo visible light communication systems. *IEEE Comm. Lett.*, 9(3):340–343, Mar. 2020.

[28] Ahmed Aboutaleb, Wael Fatnassi, Morteza Soltani, and Zouheir Rezki. Symbol detection and channel estimation using neural networks in optical communication systems. In *Proc. IEEE Int. Conf. Commun. (ICC),* Shanghai, China, 2019,, pages 1–6.

[29] Zhaohui Huang, Dongxuan He, Jiaxuan Chen, Zhaocheng Wang, and Sheng Chen.

Autoencoder with fitting network for terahertz wireless communications: A deep learning approach. *China Commun.*, 19(3):172–180, Mar. 2022.

[30] Yihan Jiang, Hyeji Kim, Himanshu Asnani, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels. *Advances in neural information processing systems*, 32, 2019.

[31] Abdelrahman Elfikky, Morteza Soltani, and Zouheir Rezki. Learning-based autoencoder for multiple access and interference channels in space optical communications. *IEEE Communications Letters*, 27(10):2662–2666, 2023.

[32] Ling-Han Si-Ma, Zhao-Rui Zhu, and Hong-Yi Yu. Model-aware end-to-end learning for siso optical wireless communication over poisson channel. *IEEE Photonics Journal*, 12(6):1–15, 2020.

[33] Jinzhi Lin, Shengzhong Feng, Yun Zhang, Zhile Yang, and Yong Zhang. A novel deep neural network based approach for sparse code multiple access. *Neurocomputing*, 382:52–63, 2020.

[34] Abdelrahman Elfikky and Zouheir Rezki. Symbol detection and channel estimation for space optical communications using neural network and autoencoder. *IEEE Transactions on Machine Learning in Communications and Networking*, 2:110–128, 2024.

[35] Abdelrahman Elfikky, Morteza Soltani, and Zouheir Rezki. End-to-end learning

framework for space optical communications in non-differentiable poisson channel. *IEEE Wireless Communications Letters*, 13(8):2090–2094, 2024.

[36] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[37] AGI. System tool kit (STK) . `https://www.agi.com/products/stk`, 2020. [Online; accessed 20-April-2021].

[38] Mateusz Polnik, Luca Mazzarella, Marilena Di Carlo, Daniel K. L. Oi, Annalisa Riccardi, and Ashwin Arulselvan. Scheduling of space to ground quantum key distribution. *EPJ Quantum Technol.*, 7, Jan. 2020.

[39] Somia A Abd El-Mottaleb, Ahmed Métwalli, Mostafa Hassib, Abdelrahman A Alfikky, Heba A Fayed, and Moustafa H Aly. Sac-ocdma-fso communication system under different weather conditions: performance enhancement. *Opt. Quant. Electron.*, 53(11):1–18, Oct. 2021.

[40] Hoang D. Le, Vuong V. Mai, Chuyen T. Nguyen, and Anh T. Pham. Throughput analysis of incremental redundancy hybrid arq for fso-based satellite systems. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–5, 2019.

[41] Hoang D. Le and Anh T. Pham. On the design of fso-based satellite systems using incremental redundancy hybrid arq protocols with rate adaptation. *IEEE Transactions on Vehicular Technology*, 71(1):463–477, 2022.

[42] Savo G.. Glisic. *Advanced Wireless Communications: 4G Cognitive and Cooperative Broadband Technology.* Wiley, 2007.

[43] David JC MacKay. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[44] Chaity Banerjee, Tathagata Mukherjee, and Eduardo Pasiliao. Feature representations using the reflected rectified linear unit (rrelu) activation. *Big Data Mining Anal.*, 3(2):102–120, Jun. 2020.

[45] Bashar Tahir, Stefan Schwarz, and Markus Rupp. BER comparison between convolutional, Turbo, LDPC, and Polar codes. In *Proc. 24th International Conference on Telecommunications (ICT),* Limassol, Cyprus, pages 1–7, May 2017.

[46] Anas Chaaban, Zouheir Rezki, and Mohamed-Slim Alouini. On the capacity of intensity-modulation direct-detection Gaussian optical wireless communication channels: A tutorial. *IEEE Commun. Surveys Tuts.*, 24(1):455–491, Firstquarter 2022.

[47] Ilya Loshchilov. Cma-es with restarts for solving cec 2013 benchmark problems. In *2013 IEEE Congress on Evolutionary Computation*, pages 369–376. Ieee, 2013.

[48] Yapei Wu, Xingguang Peng, Handing Wang, Yaochu Jin, and Demin Xu. Cooperative coevolutionary cma-es with landscape-aware grouping in noisy environments. *IEEE Transactions on Evolutionary Computation*, 27(3):686–700, 2023.

[49] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry.

How does batch normalization help optimization? In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Proc. Conf. Neural Informat. Process. Syst, 2018*, volume 31, 2018.

[50] Shlomi Arnon, John Barry, George Karagiannidis, Robert Schober, and Murat Uysal. *Advanced optical wireless communication systems*. Cambridge university press, 2012.

[51] Amos Lapidoth, Stefan M Moser, and Michele A Wigger. On the capacity of free-space optical intensity channels. *IEEE Trans. Inf. Theory*, 55(10):4449–4461, October 2009.

[52] Anas Chaaban, Jean-Marie Morvan, and Mohamed-Slim Alouini. Free-space optical communications: Capacity bounds, approximations, and a new sphere-packing perspective. *IEEE Trans. Commun.*, 64(3):1176–1191, March 2016.

[53] Steve Hranilovic and Frank R Kschischang. Optical intensity-modulated direct detection channels: signal space and lattice codes. *IEEE Trans. Inf. Theor*, 49(6):1385–1399, June 2003.

[54] Hamid Hemmati. *Deep space optical communications*, volume 11. John Wiley & Sons, 2006.

[55] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[56] V. Cazaubiel, G. Planche, V. Chorvalli, L. Le Hors, B. Roy, E. Giraud, L. Vaillon, F. Carré, and E. Decourbey. LOLA: A 40000 km Optical Link between an Aircraft

and a Geostationary Satellite. In Andrew Wilson, editor, *ESA Special Publication*, volume 621 of *ESA Special Publication*, page 87, June 2006.

[57] James H. Churnside and Steven F. Clifford. Log-normal rician probability-density function of optical scintillations in the turbulent atmosphere. *J. Opt. Soc. Am. A*, 4(10):1923–1930, Oct 1987.

[58] Daniel J. Costello and G. David Forney. Channel coding: The road to channel capacity. *Proc. IEEE*, 95(6):1150–1177, Jun. 2007.

[59] Abd El-Rahman A. El-Fikky, Maram Essam Eldin, Heba A. Fayed, Ahmed Abd El Aziz, Hossam M. H. Shalaby, and Moustafa H. Aly. Nlos underwater vlc system performance: static and dynamic channel modeling. *Appl. Opt.*, 58(30):8272–8281, Oct. 2019.

[60] Syed Agha Hassnain Mohsan, Alireza Mazinani, Hassaan Sadiq, and Amjad Hussain. A survey of optical wireless technologies: practical considerations, impairments, security issues and future research directions. *Opt. Quant. Electron.*, 54, 03 2022.

[61] Kapila Wijaya Sri Palitharathna, Himal Suraweera, G M Roshan Godaliyadda, Vijitha Herath, and John Thompson. Neural network-based channel estimation and detection in spatial modulation vlc systems. *IEEE Communications Letters*, pages 1–1, 04 2022.

[62] Jacob Manning, David Langerman, Barath Ramesh, Evan Gretok, Christopher

Wilson, Alan George, James MacKinnon, and Gary Crum. Machine-learning space applications on smallsat platforms with tensorflow. 2018.

[63] Zabih Ghassemlooy, Murat Uysal, Mohammad Ali Khalighi, V Ribeiro, Florian Moll, Stanislav Zvanovec, and Aviceto Belmonte. An overview of optical wireless communications. *Optical Wireless Communications*, 2016.

[64] Xu Zhu and Ross D Murch. Performance analysis of maximum likelihood detection in a mimo antenna system. *IEEE Transactions on Communications*, 50(2):187–191, 2002.

[65] Joakim Jaldén, Björn Ottersten, and Wing-Kin Ma. Reducing the average complexity of ml detection using semidefinite relaxation. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 3, pages iii–1021. IEEE, 2005.

[66] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[67] Hyeji Kim, Yihan Jiang, Ranvir Rana, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Communication algorithms via deep learning. *arXiv preprint arXiv:1805.09317*, 2018.

[68] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Beery. Deep learning methods for improved decoding of linear codes. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):119–131, feb 2018.

[69] Mohannad Abu-Romoh, Ahmed Aboutaleb, and Zouheir Rezki. Automatic modulation classification using moments and likelihood maximization. *IEEE Comm. Lett.*, 22(5):938–941, 2018.

[70] Lubing Han, Feifei Gao, Zan Li, and Octavia A Dobre. Low complexity automatic modulation classification based on order-statistics. *IEEE Transactions on Wireless Communications*, 16(1):400–411, 2016.

[71] Haoran Sun, Xiangyi Chen, Qingjiang Shi, Mingyi Hong, Xiao Fu, and Nikos D Sidiropoulos. Learning to optimize: Training deep neural networks for wireless resource management. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–6. IEEE, 2017.

[72] Ahmed Alkhateeb, Sam Alex, Paul Varkey, Ying Li, Qi Qu, and Djordje Tujkovic. Deep learning coordinated beamforming for highly-mobile millimeter wave systems. *IEEE Access*, 6:37328–37348, 2018.

[73] Stephen Marsland. *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC, 2011.

[74] Samuel J Dolinar, Jon Hamkins, Bruce E Moision, Victor A Vilnrotter, and H Hemmati. *Optical modulation and coding*. Wiley-Interscience, 2006.

[75] Lijun Ge, Yue Zhang, Gaojie Chen, and Jun Tong. Compression-based lmmse

channel estimation with adaptive sparsity for massive mimo in 5g systems. *IEEE Systems Journal*, 13(4):3847–3857, 2019.

[76] Pei Liu, Kai Luo, Da Chen, and Tao Jiang. Spectral efficiency analysis of cell-free massive mimo systems with zero-forcing detector. *IEEE Transactions on Wireless Communications*, 19(2):795–807, 2019.

[77] Jakob Hoydis, Stephan Ten Brink, and Mérouane Debbah. Massive mimo in the ul/dl of cellular networks: How many antennas do we need? *IEEE Journal on selected Areas in Communications*, 31(2):160–171, 2013.

[78] Thomas L Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE transactions on wireless communications*, 9(11):3590–3600, 2010.

[79] Seung Joon Lee. On the training of mimo-ofdm channels with least square channel estimation and linear interpolation. *IEEE Communications Letters*, 12(2):100–102, 2008.

[80] Yusuf Acar, Hakan Doğan, and Erdal Panayırcı. Spline based channel estimation for stbc-sm systems over fast-varying rician fading channels. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, pages 188–191. IEEE, 2015.

[81] Weile Zhang, Feifei Gao, and Qinye Yin. Blind channel estimation for mimo-

ofdm systems with low order signal constellation. *IEEE Communications Letters*, 19(3):499–502, 2015.

[82] Abla Kammoun, Mérouane Debbah, Mohamed-Slim Alouini, et al. Asymptotic analysis of rzf over double scattering channels with mmse estimation. *IEEE Transactions on Wireless Communications*, 18(5):2509–2526, 2019.

[83] S Morteza Razavi and Tharmalingam Ratnarajah. Adaptive ls-and mmse-based beamformer design for multiuser mimo interference channels. *IEEE Trans. Veh. Technol.*, 65(1):132–144, 2015.

[84] Salil Kashyap, Christopher Mollén, Emil Björnson, and Erik G Larsson. Frequency-domain interpolation of the zero-forcing matrix in massive mimo-ofdm. In *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2016.

[85] Masumi Kuriyama and Tomoaki Ohtsuki. Low complexity and high accuracy channel interpolation with dividing ura into small uras for 3d massive mimo. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–5. IEEE, 2019.

[86] Alexander Osinsky, Andrey Ivanov, and Dmitry Yarotsky. Bayesian approach to channel interpolation in massive mimo receiver. *IEEE Communications Letters*, 24(12):2751–2755, 2020.

[87] Ling Zhang and Xianda Zhang. Mimo channel estimation and equalization us-

ing three-layer neural networks with feedback. *Tsinghua science and technology*, 12(6):658–662, 2007.

[88] Syed Junaid Nawaz, Sajjad Mohsin, and Ataul Aziz Ikaram. Neural network based mimo-ofdm channel equalizer using comb-type pilot arrangement. In *2009 International Conference on Future Computer and Communication*, pages 36–41. IEEE, 2009.

[89] Ling Yang, Binbin Xue, Mingming Nie, Changnian Liu, and Qiang Zhang. Semi-blind channel estimation of mimo-ofdm system based on extreme learning machine. In *2013 Sixth International Symposium on Computational Intelligence and Design*, volume 2, pages 164–168. IEEE, 2013.

[90] Tao Cui and Chintha Tellambura. Channel estimation for ofdm systems based on adaptive radial basis function networks. In *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, volume 1, pages 608–611. IEEE, 2004.

[91] Imad Barhumi, Geert Leus, and Marc Moonen. Optimal training design for mimo ofdm systems in mobile wireless channels. *IEEE Transactions on signal processing*, 51(6):1615–1624, 2003.

[92] Hongxiang Xie, Feifei Gao, and Shi Jin. An overview of low-rank channel estimation for massive mimo systems. *IEEE Access*, 4:7313–7321, 2016.

[93] Changqing Luo, Jinlong Ji, Qianlong Wang, Xuhui Chen, and Pan Li. Channel state information prediction for 5g wireless communications: A deep learning

approach. *IEEE Transactions on Network Science and Engineering*, 7(1):227–236, 2018.

[94] P Vimala and G Yamuna. Pilot design strategies for block sparse channel estimation in ofdm systems. *Indian Journal of Science and Technology*, 10(24), 2017.

[95] Ajay B Singh and Vivek K Gupta. Performance evaluation of mmse and ls channel estimation in ofdm system. *International Journal of Engineering Trends and Technology (IJETT)*, 15(1):39–43, 2014.

[96] Yawei Li, Lizuo Jin, A Kai Qin, Changyin Sun, Yew Soon Ong, and Tong Cui. Semi-supervised auto-encoder based on manifold learning. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 4032–4039. IEEE, 2016.

[97] Ling Yang, Ming Ming Nie, Zi Long Zhong, Bin Bin Xue, and Na Lv. Channel equalization of mimo-ofdm system based on extreme learning machine. In *Applied Mechanics and Materials*, volume 536, pages 1751–1757. Trans Tech Publ, 2014.

[98] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, and Shi Jin. Deep learning for wireless physical layer: Opportunities and challenges. *China Communications*, 14(11):92–111, 2017.

[99] Zhijin Qin, Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. Deep learning in physical layer communications. *IEEE Wireless Communications*, 26(2):93–99, 2019.

[100] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. Deep learning for massive mimo csi feedback. *IEEE Wireless Communications Letters*, 7(5):748–751, 2018.

[101] Tianqi Wang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Deep learning-based csi feedback approach for time-varying massive mimo channels. *IEEE Wireless Communications Letters*, 8(2):416–419, 2018.

[102] Shuichi Ohno and Georgios B Giannakis. Capacity maximizing mmse-optimal pilots for wireless ofdm over frequency-selective block rayleigh-fading channels. *IEEE Transactions on Information Theory*, 50(9):2138–2145, 2004.

[103] Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice. *Advances in neural information processing systems*, 30, 2017.

[104] Ranjitha Prasad, Chandra R Murthy, and Bhaskar D Rao. Joint channel estimation and data detection in mimo-ofdm systems: A sparse bayesian learning approach. *IEEE Transactions on signal processing*, 63(20):5369–5382, 2015.

[105] Zhou Zhou, Jun Fang, Linxiao Yang, Hongbin Li, Zhi Chen, and Rick S Blum. Low-rank tensor decomposition-aided channel estimation for millimeter wave mimo-ofdm systems. *IEEE Journal on Selected Areas in Communications*, 35(7):1524–1538, 2017.

[106] Wenbo Ding, Fang Yang, Wei Dai, and Jian Song. Time–frequency joint

sparse channel estimation for mimo-ofdm systems. *IEEE communications letters*, 19(1):58–61, 2014.

[107] Marco A. Fernandes, J. Leonardo Nascimento, Paulo P. Monteiro, and Fernando P. Guiomar. Highly reliable outdoor 400g fso transmission enabled by ann channel estimation. In *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2022.

[108] Laialy Darwesh and Natan S. Kopeika. Deep learning for improving performance of ook modulation over fso turbulent channels. *IEEE Access*, 8:155275–155284, 2020.

[109] Zhan Gao, Mark Eisen, and Alejandro Ribeiro. Resource allocation via model-free deep learning in free space optical communications. *IEEE Trans. Commun.*, 70(2):920–934, Feb. 2022.

[110] Mohammad Ali Amirabadi, Mohammad Hossein Kahaei, and S Alireza Nezamalhosseni. Low complexity deep learning algorithms for compensating atmospheric turbulence in the free space optical communication system. *IET Optoelectronics*, 16(3):93–105, 2022.

[111] Maged A. Esmail, Waddah S. Saif, Amr M. Ragheb, and Saleh A. Alshebeili. Free space optic channel monitoring using machine learning. *Opt. Express*, 29(7):10967–10981, Mar 2021.

[112] Mohamed Mahmoud, Ayman I. Boghdady, Abd El-Rahman A. El-Fikky, and

Moustafa H. Aly. Statistical studies using goodness-of-fit techniques with dynamic underwater visible light communication channel modeling. *IEEE Access*, 9:57716–57725, 2021.

[113] Hyeji Kim, Sewoong Oh, and Pramod Viswanath. Physical layer communication via deep learning. *IEEE J. Sel. Areas Inf. Theory*, 1(1):5–18, May. 2020.

[114] H. Hemmati, Abhijit Biswas, and Ivan Djordjevic. Deep-space optical communications: Future perspectives and applications. *Proceedings of the IEEE*, 99:2020 – 2039, 12 2011.

[115] Amos Lapidoth, Stefan M Moser, and Michele A Wigger. On the capacity of free-space optical intensity channels. *IEEE Trans. Inf. Theory*, 55(10):4449–4461, October 2009.

[116] Jingjing Xu, Xu Sun, Zhiyuan Zhang, Guangxiang Zhao, and Junyang Lin. Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels. In *Proc. Conf. Neural Informat. Process. Syst, 2019,* pp. 4381–4391.

[117] Larry C Andrews and Ronald L Phillips. *Laser Beam Propagation Through Random Media.* SPIE Pres, 2005.

[118] Jingjing Xu, Xu Sun, Zhiyuan Zhang, Guangxiang Zhao, and Junyang Lin. Understanding and improving layer normalization. volume 32, page 4381–4391, 2019.

[119] Mohammed Elamassie, Murat Uysal, Yahya Baykal, Mohamed Abdallah, and

Khalid Qaraqe. Effect of eddy diffusivity ratio on underwater optical scintillation index. *J. Opt. Soc. Am. A*, 34(11):1969–1973, Nov. 2017.

[120] Fayçal Ait Aoudia and Jakob Hoydis. Model-free training of end-to-end communication systems, 2019.

[121] Zabih Ghassemlooy, Stanislav Zvanovec, Mohammad-Ali Khalighi, Wasiu O. Popoola, and Joaquin Perez. Optical wireless communication systems. *Optik*, 151:1–6, 2017. Optical Wireless Communication Systems.

[122] Murat Uysal and Hatef Nouri. Optical wireless communications — an emerging technology. In *2014 16th International Conference on Transparent Optical Networks (ICTON)*, pages 1–7, 2014.

[123] Abd El-Rahman A. El-Fikky, Abdallah S. Ghazy, Haitham S. Khallaf, Ehab Mahmoud Mohamed, Hossam M. H. Shalaby, and Moustafa H. Aly. On the performance of adaptive hybrid MQAM–MPPM scheme over Nakagami and Log-normal dynamic visible light communication channels. *Appl. Opt.*, 59(7):1896–1906, Mar. 2020.

[124] Tobias A Eriksson, Henning Bülow, and Andreas Leven. Applying neural networks in optical communication systems: Possible pitfalls. *IEEE Photonics Technology Letters*, 29(23):2091–2094, 2017.

Appendix A

List of Acronyms

SOC Space Optical Communications

OWC Optical Wireless Communications

RF Radio Frequency

GEO Geostationary Earth Orbit

AE Autoencoder

NN Neural Network

CSI Channel State Information

BER Bit Error Rate

MSE Mean Squared Error

IM/DD Intensity Modulation Direct Detection

AWGN Additive White Gaussian Noise

MMSE Minimum Mean Square Error

SNR Signal-to-Noise Ratio

PDF Probability Density Function

GG Gamma-Gamma

STK System Tool Kit

LEO Low Earth Orbit

MAC Multiple Access Channel

IC Interference Channel

ReLU Rectified Linear Unit

DL Deep Learning

DNN Deep Neural Network

MLP Multilayer Perceptron

CSI Channel State Information

VL Visible Light

VLC Visible Light Communication

LS Least Square

OFDM Orthogonal Frequency Division Multiplexing

MIMO Multiple Input, Multiple Output

LMMSE Linear Minimum Mean Square Error

AWGN Additive White Gaussian Noise

CMA-ES Covariance Matrix Adaptation Evolution Strategy

FNN Feedforward Neural Network

Appendix B

Proof for Gradient Descent Calculations

The parameter $d^{[L]}$ of the single neuron output layer is defined as follows

$$d^{[L]} = \frac{\partial \mathcal{L}}{\partial z^{[L]}} = 2\left(h - \hat{h}\right). \tag{B.1}$$

The vector $\boldsymbol{d}^{[l]}$ in the l^{th} layer is given as

$$\boldsymbol{d}^{[l]} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{[l]}} = \left(\boldsymbol{\Theta}^{[l+1]^\top} \boldsymbol{d}^{[l+1]}\right) \odot \mathrm{ReLU}'\left(\boldsymbol{z}^{[l]}\right). \tag{B.2}$$

The gradient decent algorithm is employed in conjunction with backpropagation solving the optimization problem in (3.12) to reduce the loss function by updating the weights at the hidden and output layers.

Based on (10), (18) and (19), the gradient calculations are computed as follow:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Theta}^{[l+1]}} = \boldsymbol{d}^{[l+1]} \boldsymbol{a}^{[l]^\top}, \tag{B.3}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^{[l+1]}} = \boldsymbol{d}^{[l+1]}. \tag{B.4}$$

Moreover, the proposed channel estimator NN makes use of the Adaptive Moment Estimation (Adam) optimizer. Adam is a technique for computing adaptive learning rates for each weight parameter. In addition to storing a decaying average of past squared gradients v_t, we also keep track of them individually. We compute the exponentially decaying averages of past and past squared gradients as follows

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t, \tag{B.5}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2, \tag{B.6}$$

where the first and the second moment estimates are denoted by m_t and v_t, respectively. The decay rates for the first and second moment are defined as β_1 and β_2, respectively. Then the weight parameters are updated according to

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t} + \epsilon} m_t. \tag{B.7}$$

Finally, updating weights stop functioning whenever the difference in error between the two most recent times is negligible or the allocated number of epochs has been reached.

Appendix C

Proof for MMSE Channel Estimator in Log-normal Fading Channel

The received element y can be given by

$$y = hx + w, \tag{C.1}$$

where h is the true channel coefficients based on Log-normal fading channel and AWGN $w \sim N(0, 1)$. The criteria of MMSE estimator is based on

$$\widehat{h} = E[h \mid y] = \int_0^\infty h f(h/y) dh, \tag{C.2}$$

where $f(h \mid y)$ is defined as

$$f(h \mid y) = \frac{f(y \mid h) f(h)}{f(y)}, \tag{C.3}$$

where $f_h(h)$ is the PDF of the Log-normal distribution. Furthermore, the PDF of the received element y can be denoted as

$$f(y) = \int_0^\infty f(y \mid h) f(h) dh, \tag{C.4}$$

where $f(y \mid h)$ follows a Gaussian distribution with a mean $\mu = hA$ and unit variance, by substituting (27) and (28) in (26), $E[h/y]$ can be described as

$$E[h/y] = \int_0^\infty h \frac{f(y \mid h) f_h(h)}{\int_0^\infty f(y/z) f_z(z) dz} dh, \tag{C.5}$$

where $f_h(h)$ follows a Log-normal distribution and the PDF $f_h(h)$ is given by

$$\begin{aligned} &f_h(h) = \\ &\frac{1}{h\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\log h + \mu)^2}{2\sigma^2}\right\}, \text{for } h > 0. \end{aligned} \tag{C.6}$$

Afterwards, we deduce that $f(y \mid h)$ can be described as

$$f(y \mid h) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-hA)^2}{2}}. \tag{C.7}$$

Following along the same lines, by substituting (31) in (29), this yields to

$$\hat{h} = \int_0^\infty h \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-hA)^2}{2}} f_h(h)}{\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-hA)^2}{2}} f_z(z) dz} dh. \tag{C.8}$$

Now, based on (C.6) and (C.8), we finally obtain the channel estimate $\hat{h}$ based MMSE estimator

$$\hat{h} = \int_0^\infty \frac{e^{-\frac{(y-hA)^2}{2}} \exp\left\{\frac{-(\log h+\mu)^2}{2\sigma^2}\right\}}{\int_0^\infty e^{-\frac{(y-zA)^2}{2}} \frac{1}{z} \exp\left\{\frac{-(\log z+\mu)^2}{2\sigma^2}\right\} dz} dh. \tag{C.9}$$