JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

SCHOOL OF GRADUATE STUDIES

Channel Estimation in Massive MIMO under Hardware Non-linearities using Stacked BLSTM

BY: DANIEL SIMENEH SHIFERAW

ADVISOR: KINDE ANLAY (ASSOCIATE PROF.)

CO-ADVISOR: MRS. SOFIA ALI (M.SC.)

A thesis submitted to School of Graduate Studies, Jimma University in fulfillment of the requirements for the degree of Masters of Science

in the field of

Communication Engineering

April 18, 2024

Jimma, Ethiopia

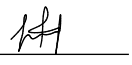
JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

Channel Estimation in Massive MIMO under Hardware Non-linearities using Stacked BLSTM

APPROVED BY THE BOARD OF EXAMINERS

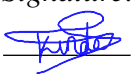
Chairperson:	*Signature:*	*Date:*
Mohammed Muntaz (Asst. Prof.)	_____	__ / __ / ____
Internal Examiner:	*Signature:*	*Date:*
Isayiyas Nigatu (Asst. Prof.)	_____	__ / __ / ____
External Examiner:	*Signature:*	*Date:*
Anteneh Wodajo (Asst. Prof.)		21 / 04 / 2024

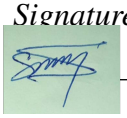
DECLARATION

I, DANIEL SIMENEH SHIFERAW, declare that this thesis titled, *"Channel Estimation in Massive MIMO under Hardware Non-linearities using Stacked BLSTM"* and the work presented herein are my own, except where explicitly stated otherwise in the text, and with the guidance of my advisor. I confirm that the work has not previously been submitted for a degree or any other qualification at this or any other University or institution, this has been clearly stated. Moreover, all sources of materials used in the thesis are acknowledged.

Student Name:	*Signature:*	*Date:*
DANIEL SIMENEH SHIFERAW	_____	__ / __ / ____

As Advisor, I hereby certify that I have read and evaluated this thesis paper prepared under my guidance, by DANIEL SIMENEH SHIFERAW entitled *"Channel Estimation in Massive MIMO under Hardware Non-linearities using Stacked BLSTM"* and recommend and would be accepted as a fulfilling requirement for the Masters of Science Degree in the field of Communication Engineering.

Main Advisor:	*Signature:*	*Date:*
KINDE ANLAY (ASSOCIATE PROF.)		22 / 04 / 2024

Co-Advisor:	*Signature:*	*Date:*
MRS. SOFIA ALI (M.SC.)		23 / 04/2024

ABSTRACT

The fifth and beyond generations of wireless communication technologies utilize an extension of Multiple-Input Multiple-Output (MIMO) principles called Massive MIMO. This technology improves communication system performance with a greater number of base station antennas sending signal beams to users' locations. However, if optimal channels are estimated without considering the distortion characteristics of the base station and user-side hardware, the overall performance can be reduced. Formerly, the non-linear hardware impairments impact channel estimation since this line of research is in its infancy. To overcome this issue, the proposed model is a recurrent neural network (RNN) particularly Stacked Bidirectional Long Short-Term Memory (BLSTM). In addition, this model can handle the sequential nature of channel data and temporal correlations in time-varying channels. Eventually, it is observed that the model can reduce the channel estimation error for Massive MIMO hardware non-linearity scenarios. The proposed model has been outperforming its benchmark, which has utilized Fully-Connected Deep Networks (FCDN), for robust channel estimation by handling the structure of the non-linear distortions. Further, it can be inferred that the BLSTM model can be integrated into physical layer design for future wireless systems within Massive MIMO networks.

Keywords: *Channel estimation, Deep learning, Hardware non-linearity, BLSTM, Massive MIMO wireless communication*

ACKNOWLEDGEMENTS

First of all, I would like to thank the Almighty GOD for everything. Next, my deepest gratitude goes to my advisor, Kinde Anlay (Associate Professor) for his active contribution to refining this research work. He was there to listen to my concerns, review my report, provide feedback, and show me the direction to steer me on the right path. Also, it is an honor to appreciate my co-advisor, Mrs. Sofia Ali, who consistently guides and supports this research work. Lastly, I want to convey my heartfelt appreciation to all of my family members for their unwavering help and encouragement over my years of study, as well as during the progress of this thesis. This accomplishment would not have been possible without them.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

3GPP	Third Generation Partnership Project
5G	Fifth Generation
ADAM	Adaptive Moment Estimation
ADC	Analog to Digital converter
AP	Access Point
AWGN	Additive White Gaussian Noise
BLSTM	Bidirectional Long Short-Term Memory
BRNN	Bidirectional Recurrent Neural Network
BS	Base Station
CDF	Cumulative Distribution Function
CEFnet	Channel Estimation and Feedback Network
CNN	Convolutional Neural Network
DAE	Denoising Auto Encoder
DA-MMSE	Distortion-Aware Minimum Mean Squared Error
DFT	Discrete Fourier Transform
DNN	Deep Neural Network
FCDN	Fully Connected Deep Network
FFNN	Feed Forward Neural Network
GaN	Gallium Nitride
GSCM	Geometry Based Stochastic Channel Model
HI	Hardware Impairment
i.i.d	Independent and Identically Distributed
LMMSE	Linear Minimum Mean Squared Error
LNA	Low Noise Amplifier
LOS	Line of Sight
LSTM	Long Short-Term Memory
MIMO	Multiple-Input Multiple-Output

MLP	Multilayer Perceptron
MRC	Maximal Ratio Combining
MSE	Mean Squared Error
NLOS	Non Line Of Sight
NMSE	Normalized Mean Squared Error
PFnet	Pilot compression and Feedback Network
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
ReLU	Rectifier Linear Unit
RNN	Recurrent Neural Network
SE	Spectral Efficiency
SIC	Successive Interference Cancellation
SNR	Signal to Noise Ratio
TNN	Two-layer Neural Network
UL	Uplink
UE	User Equipment
ZF	Zero Forcing

LIST OF SYMBOLS

$\mathbb{C}^{M\times K}$	The set of complex-valued $M \times K$ matrices.
$\mathbb{C}^{M}$, $\mathbb{C}^{K}$	Short forms $\mathbb{C}^{M\times 1}$ and $\mathbb{C}^{1\times K}$ for M and K complex vectors, respectively.
$\mathbf{D}^{*}$	The matrix $\mathbf{D}$ complex conjugate.
$\mathbf{D}^{\mathbf{T}}$	The matrix $\mathbf{D}$ transpose.
$\mathbf{D}^{\mathbf{H}}$	The matrix $\mathbf{D}$ conjugate transpose.
$tr(\mathbf{D})$	The trace of matrix $\mathbf{D}$.
$\Im\{\mathbf{d}\}$	The imaginary part of $\mathbf{d}$ as complex.
$\Re\{\mathbf{d}\}$	The real part of $\mathbf{d}$ as complex.
$\mathcal{N}_{\mathbb{C}}(\mathbf{0},\, \mathbf{D})$	The complex Gaussian distribution with zero-mean and $\mathbf{D}$ matrix.
$\lvert\mathbf{d}\rvert$	The magnitude (or absolute value) of a scalar $\mathbf{d}$ variable.
$\mathcal{N}(\mathbf{d},\, \mathbf{D})$	The real Gaussian distribution with $\mathbf{d}$ mean and $\mathbf{D}$ covariance matrix.
$\mathbf{0}_{M}$	The M vector with only zeros.
$\mathbf{I}_{M}$	The $M \times M$ identity matrix.
$\mathbb{E}_{\lvert\mathbf{D}}\{\mathbf{d}\}$	The conditional expectation of $\mathbf{d}$ given $\mathbf{D}$ matrix.
$\mathbb{E}\{\mathbf{d}\}$	The expectation of $\mathbf{d}$ variable.
$\log_{10}(\mathbf{d})$	The logarithm of $\mathbf{d}$ using the base 10.
j	The imaginary unit for the complex.
α	The path loss exponent in the large-scale fading.
κ	The hardware quality of transmitter or receiver.
σ^2	The noise variance.
μ	The additive noise.

Dedicated to my mom,
ALGANESH BIRHANU

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

AN enabling technology and a high-potential technology for 5G and beyond communication systems, and an extension of Multiple-Input Multiple-Output (MIMO), which utilizes multiple transmit antennas at the transmitter and multiple receive antennas at the receiver, is known as Massive MIMO. This technology can be utilized to send the signal beams toward the location of the users, as illustrated in Figure 1.1. Thus, it offers improved spectral efficiency, network capacity, and coverage with easily achievable data rates for the network [1], [2]. Massive MIMO is preferable due to its high capacity, increased energy efficiency, increased diversity, beam-forming, and directional signal transmission [3], [4].

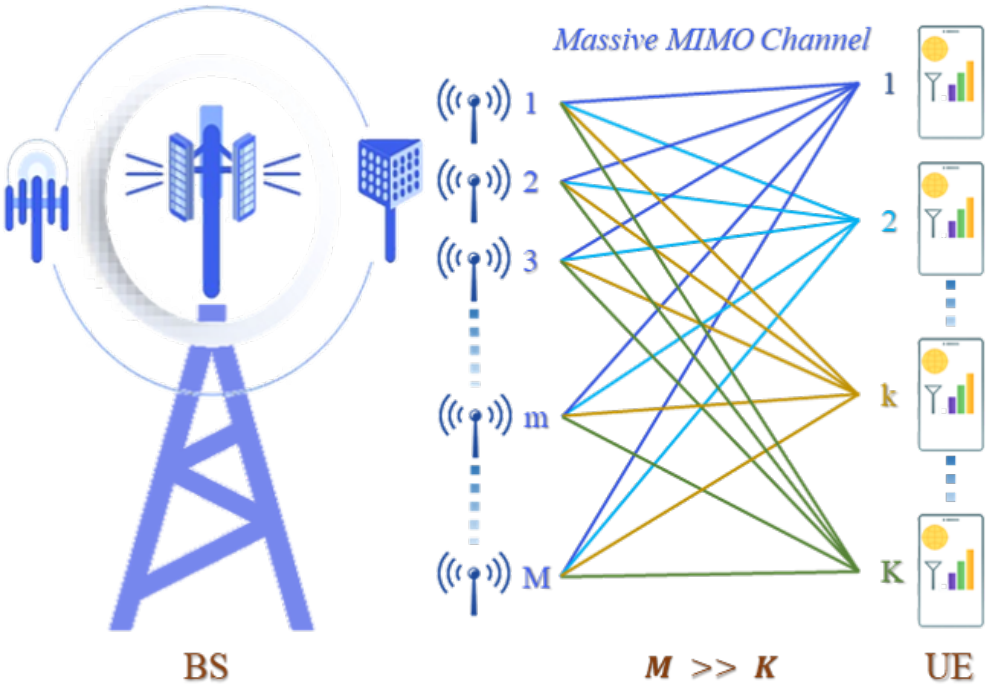


FIGURE 1.1: An illustration on Massive MIMO architecture.

Channel estimation is crucial for successful communication between transmitter and receiver, as the transmitter must know the channel for successful transmission [5]. Channel estimation using pilot sequences in the Uplink (UL) remains a challenge in ideal hardware at Base Station (BS) and User Equipment (UE), with transceiver impairments including amplifier non-linearities and quantization errors inevitable. Some literature has modeled continuous hardware impairments using a stochastic additive model. However, behavioral models that utilize some deterministic functions are expected to model the continuous non-linear distortion and are considered in many different research areas [6]–[8]. The non-linear system is treated by utilizing the Bussgang decomposition to find an equivalent linear system with uncorrelated distortion. This algorithm is a popular tool for analyzing the performance of systems that involve non-linear components. In a nutshell, the decomposition provides an exact probabilistic relationship between the output and the input of a non-linearity: the output is equal to a scaled version of the input plus uncorrelated distortion. The decomposition can either be used to compute exact performance results or lower bounds where the uncorrelated distortion is treated as independent noise [9], [10].

Traditional estimation methods, such as least squares (LS) and minimum mean square error (MMSE), often have high computational complexity, limited accuracy, and sensitivity to imperfect channel knowledge. The distortion, considered independent colored noise, is used to build a distortion-aware Bayesian linear minimum mean squared error (LMMSE) estimator. However, calculating the minimum mean squared error (MMSE) estimate for non-linear hardware impairments is complex, necessitating the development of deep learning models to outperform traditional Bayesian estimators. These models consider hardware non-linearities and the distortion's dependence on the desired signal [6], [11], [12].

Lately, Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) have been used in [1], [13], [14] for channel estimation under non-linear hardware impairments. Nevertheless, no RNN-based model has been applied in any of these research to tackle robust channel estimation problems. To further improve estimate quality, we chose the Recurrent Neural Network (RNN) among the most common deep learning algorithms through our experimental result. RNN performs better than other algorithms in terms of accuracy and speed since the architectures handle the sequential nature of the input data. It is determined that the proposed approach using the Bidirectional RNNs (BRNNs) model outperforms the conventional DNN-based models, which are outperforming old LMMSE estimators. Henceforth, the proposed approach can optimize and automate beamforming in Massive MIMO, which will lower complexity and can adapt to different channel environments: the BS receivers,

UEs, and transmission channel, and handle system configurations.

1.2 Statement of the Problem

Channel estimation is crucial for Massive MIMO to achieve its potential gains. However, because of the numerous antennas and intricate channel conditions, many research studies have modeled and analyzed the impact of hardware non-linearities on Massive MIMO using behavioral modeling. Recently, several studies have described distortion-aware receivers for UL signal detection in Massive MIMO. To apply these receivers, the BS must know the effective channels of the UE together with the received signal correlation matrix. If we don't consider the estimation of the effective channels by taking into account the BS and UE non-linear distortion characteristics and instead only the wireless channels, it can decrease or minimize the performance of the overall communication system. This has motivated us to consider the estimation of the effective channels, taking into account the BS and UE non-linear distortion characteristics, instead of only the wireless channels.

1.3 Research Question

In this thesis, we investigate how an RNN-based model, specifically Stacked Bidirectional Long Short-Term Memory (Stacked BLSTM), is used for more robust channel estimation under scenarios where hardware non-linearities are considered at both the BS and UE.

- What best neural network-based channel estimator can be applied to exploit the full structure of the hardware impairments and perform more effective channel estimation?
- How to tune and enhance the generalization and representational capacity of the neural network-based channel estimator?

1.4 Objectives

1.4.1 General Objective

To design the channel estimation under both the base station and user equipment non-linear distortions using the Stacked Bidirectional Long Short-Term Memory.

1.4.2 Specific Objectives

The specific objectives of the study are the following:

- To understand deeply the basic concept of channel estimation in Massive MIMO wireless network and related issues.
- To employ deep learning techniques, particularly stacked BLSTM for wireless channel estimation under hardware non-linearities.
- To optimize the performance of BLSTM-based channel estimation by using a state-of-the-art search algorithm.
- To analyze and compare the performance of deep learning and various Bayesian methods under hardware non-linearities.

1.5 Research Methodology

In this section, Figure 1.2 shows the overall diagram of the overall methodology.

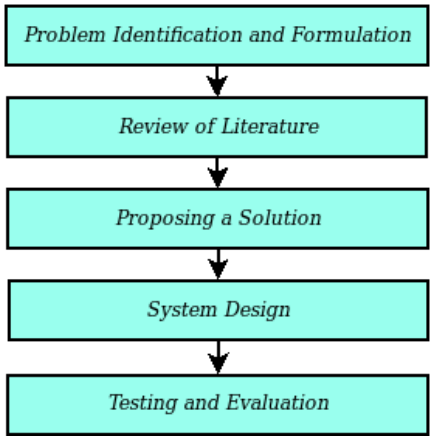


FIGURE 1.2: Overall methodology of the thesis

We first identified the issue with effective massive MIMO network channel estimation at the outset of our study. When we examined several research papers, journal publications, and both published and unpublished documents about the effectiveness of channel estimation for Massive MIMO under hardware non-linearities, we discovered this problem.

Second, to solve the problem with unrivaled performance, we studied the needed areas and chose the most feasible solution. For Massive MIMO, a deep learning technique was suggested for improving channel estimation quality under hardware non-linearities. Continuous-time channel responses are created and sampled using a worldwide standard wireless channel model to collect training and testing data.

1.6 Scope and Limitation of the Study

The basic finding of this study concentrated on performance improvement of channel estimation for Massive MIMO hardware non-linearities for reliable communication systems. We explored the effect of non-linearities in Radio Frequency (RF) hardware of the UE and BS and derived the effective channels for a given realization of the channels. However, we do not have any hardware-based implementation. However, similar to the related works, we have used the statistical distributions of the channels and tested the data at the software simulation level. Finally, we were successful in employing the RNN-based model and trained this model to perform channel estimation with samples according to the distribution.

1.7 Significance of the Study

The research topic of deep learning-based channel estimation in Massive MIMO, considering hardware impairments, has gained popularity in recent years. However, none of these studies have utilized RNN-based models to address robust channel estimation issues. This thesis demonstrates the Stacked BLSTM model with the hyper-tuned parameters attaining significant performance improvement over DNN-based models that were used in existing works. The estimated channels can be used for signal detection during UL data transmission. Therefore, we contributed to optimizing the quality of the channel estimation in Massive MIMO under hardware non-linearities that can be used for hybrid analog-digital beamforming architectures.

In summary, the proposed approach will optimize and automate Massive MIMO beamforming, minimizing complexity and adjusting according to various channel conditions, such as transmission channels, UEs, and BS receivers.

1.8 Thesis Organization

This thesis work has been done and organized as follows:

Chapter 1: Introduction to effective channel estimation in Massive MIMO communication system, also the current chapter. In this chapter, we give the background of the study, problem statement, objectives, scope, and significance in the practical applications of the system.

Chapter 2: Overviews of background theory, we demonstrate this in three parts: Massive MIMO, Channel estimation, Deep learning approaches, and Review of literature and research related to the study that usage model and the achieved results with a brief description of their strength and weaknesses.

Chapter 3: Data collection and dataset generation. Presents the approach or procedure followed to address the problem: this chapter goes into detail on the optimization, training, and evaluation of the neural network in deep learning architectures.

Chapter 4: The test results obtained were carried out. Two tests have been performed on the system to evaluate its performance in the effective channel estimation: Experimental setup, and Experimental result sections. In the following sections, we show an extract of the results obtained in these tests.

Chapter 5: Summary of the thesis work done, recommendation, and future works. We write the conclusions obtained and propose possible future related works, which either seek to solve the shortcomings and errors that throughout our approach we have been able to perform or propose alternative ways to tackle the same problem.

CHAPTER 2

LITERATURE REVIEW

This chapter presents three sections with their respective subsections to offer an overview of the basics of Massive MIMO, channel estimation approaches, deep learning, and related works.

2.1 Overview of Massive MIMO

In 5G and beyond wireless communication technology in 2006, Massive MIMO made a breakthrough in Multiple-Input Multiple-Output (MIMO) by extending cell coverage and accommodating a significantly higher number of users simultaneously. Massive MIMO equipped the base station with a large number of antennas, typically using configurations with around 64 to 256 antennas, thus improving the capacity and stability of communication channels. Figure 2.1 illustrates Massive MIMO technology utilizing 64 antenna array elements [15]–[17]. Massive MIMO offers advantages such as spatial multiplexing, expanding virtual channels, delivering multiple parallel data streams within the same resource block, and increasing capacity and data rates without additional towers and spectrum.

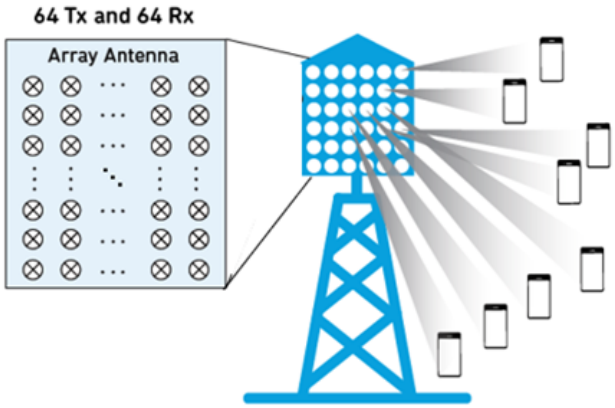


FIGURE 2.1: A base station with 64×64 antennas [17].

Nevertheless, there is still more work to be done in the areas of effective channel estimation and research and development on hardware non-linearities and their implications for the implementation of Massive MIMO [1], [10], [18]–[22].

2.2 Impact of Hardware Impairments

New system models are being developed to address hardware impairments and thus improve channel estimation accuracy and capacity in BS and single-antenna devices as the demand for higher data throughput in wireless systems increases. However, hardware non-linearities can cause signal impairments, affecting channel estimation accuracy and user capacity. Massive MIMO large-scale arrays can mitigate these issues by utilizing energy-efficient antenna elements and reducing transmit power.

Studies in [23]–[25] examined how hardware impairments affect BS antennas on Spectral Efficiency (SE) of UL in Massive MIMO. This distortion is correlated across the antennas but has often been approximated as uncorrelated to facilitate (tractable) SE analysis. The authors have also analyzed the distortion and correlation caused by third-order non-linearities. They emphasized the importance of considering distortion characteristics. A study in [26] also highlights the need for considering hardware imperfections and channel correlation in multi-user wireless systems because of their severe impacts on system performance while addressing channel estimation in modern wireless communication systems.

In essence, it is a known fact that hardware impairment impacts channel estimation. The impact of hardware impairments plays a crucial role in any practical communication system, yet such issues are so often omitted when investigating the performance of distributed massive multiple input multiple output (MIMO) systems [10]. To overcome this limitation, RNN-based effective channel estimation is used in this thesis, in which the residual hardware impairments are handled in the transmitter processing.

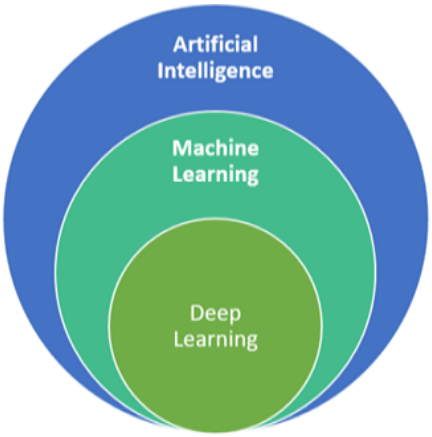
2.3 Deep Learning-Based Channel Estimation

In this section, we try to gain some insight into deep-learning concepts that are used for improved channel estimation in Massive MIMO. We are particularly interested in the state-of-the-art deep learning architectures: Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNNs). In this thesis, we chose RNN-type neural networks, especially Long Short-Term Memory (LSTM) as detailed in Section 2.3.3.

2.3.1 Overview of Deep Learning

Deep learning is about computers learning to think using structures modeled on the human brain. A machine using deep learning, which is one part of machine learning, can analyze images, videos, and unstructured data in ways machine learning can't

easily do. Deep learning is based on the idea that systems can learn from data, identify patterns, and make decisions.



Artificial Intelligence: *is which refers to the theory and development of computer systems able to perform tasks by normally requiring human intelligence.*

Machine Learning: *gives computers the capability to think and act with less human intervention and learn without being explicitly programmed.*

Deep Learning: *is a machine learning algorithm with a brain-like logical structure of models called artificial neural networks.*

FIGURE 2.2: Artificial intelligence, machine learning, and deep learning. Nadia Berchane (M2 IESCI, 2018).

Deep learning has its origins in *Warren S. McCulloch* and *Walter Pitts's* work in 1943, which was published as the first mathematical model network of neurons in the brain with computational circuits on Boolean quantities. Hence, the nodes in neural networks are considered "neurons," elsewhere in the literature. Successors to this work slightly refined this logical model by making mathematical "neurons" continuous functions that varied between zero and one. The neuron was excited or "fired" (took on the value one) if the inputs of these functions grew large enough; otherwise, it was quiescent. Figure 2.3 depicts a single neuron in deep learning constructed with input x, weight w, bias b, and activation function $f(.)$.

The deep learning model is constructed from multiple neurons and layers. It has an input layer, hidden layers, and an output layer. In general, deep learning algorithms are of two types, namely, Feed-Forward Neural Networks (FFNN) and Recurrent Neural Networks (RNN). If the connection of the neurons and propagation of the signals are only in a forward direction, it is categorized as FFNN. Whereas, if it has feedback from a previous neuron the network is categorized as RNN.

Mathematically, it can be represented as:

$$y = f\left(\sum_{k=1}^{3}(x_k w_k) + b\right) \tag{2.1}$$

2.3.2 Fully-Connected Deep Networks

A Fully-Connected Deep Network (FCDN) commonly be referred to as "neural networks," also known as a deep neural network (DNN) in data science. FCDN is a

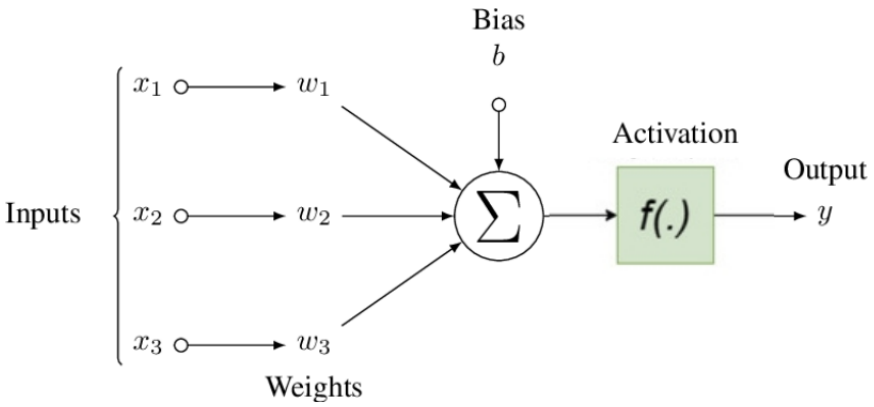


FIGURE 2.3: Structure of single neuron in deep learning.

neural network structure where every neuron in the network is connected to every other neuron in nearby layers, and every input vector affects every output vector due to all potential connections between layers.

FCDN has completely connected adjacent network layers, in which all nodes are interconnected. Also, it is often made up of several fully connected layers in such a way, that every input dimension influences every output dimension. FCDN with at least three layers—input, output, and one hidden layer—is known as a Multilayer Perceptron (MLP). For instance, as depicted in Figure 2.4 shows a fully connected MLP with multiple hidden layers.

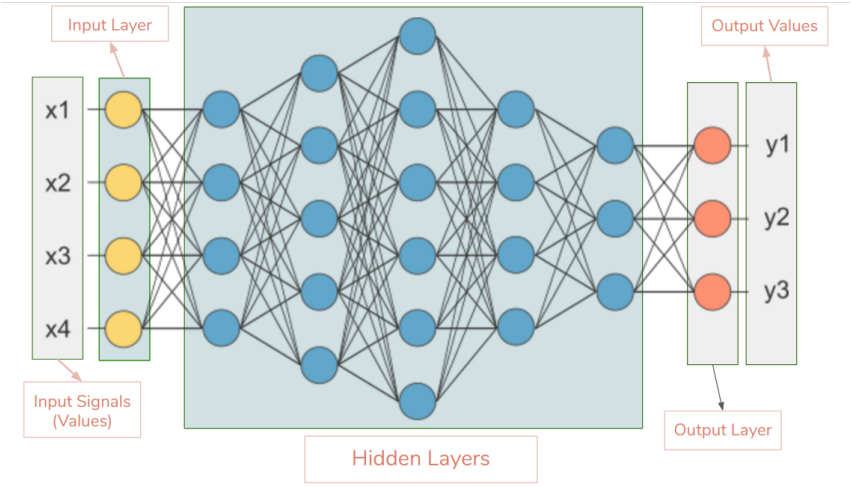


FIGURE 2.4: A fully connected MLP with multiple hidden layers[27].

2.3.3 Long Short-Term Memory

A bidirectional LSTM (BLSTM) is a variation of bidirectional recurrent neural network (BRNN) which is to connect two hidden layers of opposite directions to the

same output as proposed in [28]. The output layer can get information from past (backward) and future (forward) states simultaneously.

A stacked BLSTM model refers to the long short-term memory (LSTM) architecture comprised of multiple BLSTM layers, in the case of this thesis a pair of layers of BLSTM is chosen as recommended by many articles of literature such as [29]–[33], etc. More complex, recurrently connected subnets known as "memory cells" make up each LSTM. Each "memory cells" in an LSTM processes a single input sequence of time steps, and the result is a single value for the whole sequence output as a 2D array.

2.3.4 Related Works

Several papers explored effective channel estimation in Massive MIMO using Bayesian methods and deep learning models, with few considering hardware impairments. In this section, we reviewed the related works as follows:

In a 2019 study in [1], *Demir and Emil* presented a model that exploits the hardware impairment characteristics using feed-forward neural networks (FFNNs) for effective channel estimation of the UL data transmission in Massive MIMO. The authors made a remarkable change from traditional Linear Minimum Mean-Squared Error (LMMSE), which treats the distortion as independent colored noise and only uses its second-order statistics. Hence, they have analytically derived the impact of general hardware non-linearities on the UL systems of BS and UE. They have calculated the channel quality for distortion-aware receivers using finite-sized signal constellations. The training and validation data lengths were $3,000,000$ and $200,000$, respectively. They trained the DNN utilizing the hardware distortion characteristics and finally obtained the Mean Squared Error (MSE) of 5.22%. Their result was compared with traditional distortion-aware and unaware Bayesian LMMSE estimators, and they achieved better channel estimation quality. Nevertheless, the authors were unable to apply a memory-based or time-based, state-of-the-art model. They did not also apply any model optimization algorithms.

In [34], another study in 2019, *Chun, et al.* proposed a two-stage channel estimation scheme for Massive MIMO. These are (i) pilot-aided channel estimation using both a Two-layer Neural Network (TNN) and a DNN, and (ii) data-aided iterative channel estimation using only DNN. To enhance channel estimation performance and reduce the MSE, the authors carried out iterative data detection and channel estimation. They used $100,000$ channel samples to train the scheme and $10,000$ test samples to assess its performance. They achieved an MSE of nearly 15% for large pilot lengths,

which is smaller than that of the traditional LMMSE estimators. Notwithstanding, this MSE result can be further reduced by increasing the data size applying a modified architecture of DNN, and using state-of-the-art models.

In 2020, in [35] *Chen, et al.* constructed two deep learning-based joint channel estimation and feedback networks for the estimation, compression, and redesign of the downlink (DL) channels in the FDD Massive MIMO systems. These are the channel estimation and feedback network (CEFnet) and the pilot compression and feedback network (PFnet), which were used to achieve explicit and implicit channel estimation and feedback, respectively. The CEFnet was composed of both a lightweight CNN to obtain the effective channels and the Denoising Auto-Encoder (DAE) to compress and reconstruct the noisy channel matrices. Whereas, the PFnet compresses and sends back the pilot information directly to the BS without estimating the channels. The Geometry-based Stochastic Channel Model (GSCM) for MIMO systems generated the channels during their simulation. The numbers of samples in the training, validation, and testing sets are $100,000$, $30,000$, and $20,000$, respectively. Their simulation results compared the performance of the two networks (CEFnet and PFnet). These show that the CEFnet outperforms the PFnet and the MSE was around 10%. Still, this result is less satisfying since their dataset was smaller.

In 2021, [36] by *Rao, et al.* explored the UL Massive MIMO system, which used low-resolution Analog-to-Digital Converters (ADCs) and a spatial sigma-delta structure to reduce quantization noise at the BS. For channel estimation, they designed the optimal LMMSE estimator, which was based on the Bussgang decomposition that reformulates the nonlinear quantizer model using an equivalent linear model plus quantization noise. Also, they examined the achievable rate in the UL domain using Maximal Ratio Combining (MRC), Zero-Forcing (ZF), and LMMSE receivers; they also presented a lower bound for the achievable rate with the MRC receiver. Their results demonstrated that the spatial approach achieved better channel estimates and efficiency than systems using first-order one- and two-bit ADCs for direct quantization when the users are confined to a specific angular sector or when the array elements are closer than one-half wavelength; however, the performance declined because they did not use state-of-the-art models.

In 2021, in [37] *Ghazanfari et al.* studied both a model-aided approach and a deep-learning-based approach for the downlink effective channel estimation in multi-cell Massive MIMO systems operating in a time-division duplex. The first approach was based on a closed-form expression that was obtained using asymptotic analysis, whereas the second approach was used for mapping between the available information and effective channel gain. For deep learning, the entire data set consists of

$1,000,000$ input-output vector pairs for a typical user K randomly located in cell l for 1000 realizations of large-scale fading and 1000 small-scale fading. We selected $400,000$ for training, $100,000$ for validation, and the rest of $500,000$ for the testing phase. They found that the deep learning-based approach was the best of the considered estimators.

In 2022, in [38] the same authors proposed a DNN-based channel estimator with hardware impairments, with a focus on UL multi-user MIMO systems. They trained DNN to utilize the hardware distortion characteristics and to learn the structure of the quasi-memoryless polynomials that were used for modeling the BS hardware impairments. The authors assumed the QPSK modulation scheme is adopted for UL data transmission, which determines the effective channel. The pilot length is the sequence of the columns of the matrix of Discrete Fourier Transform (DFT). They utilized two hidden layers, each with 300 neurons in the DNN. They assessed the loss function as the MSE. Both inputs and outputs of the neural network are scaled using the Standard Scaler and MinMax Scaler. They selected the training set of $3,000,000$ samples and $200,000$ of the validation set. Their model performance is compared with the distortion-aware and unaware Bayesian LMMSE estimators. They demonstrated that the DNN model improves the estimation quality by exploiting impairment characteristics, while LMMSE methods treat distortion as noise. Although they have successfully trained the DNN using Rayleigh fading, the expressions of effective channels can be applied to evaluate the performance of models handling different channels and hardware impairments.

In 2022, a study by Atzeni and Tölli [39] presented an analytical framework for channel estimation and data detection in Massive MIMO UL systems with 1-bit ADCs and tractable independent and identically distributed (i.i.d.) Rayleigh fading. The authors addressed the class of scaled least-squares estimators and the linear minimal MSE estimator while providing closed-form equations for the MSE of the channel estimation. For the data detection, they provided closed-form expressions of the expected value and the variance of the estimated symbols when maximum ratio combining is adopted, which can be exploited to efficiently implement minimum distance detection. The authors' analytical findings explicitly depend on key system parameters such as the SNR, the number of UE, and the pilot length, thus enabling a precise characterization of the performance of channel estimation and data detection with 1-bit ADCs.The proposed analysis gives important practical insights into the design and implementation of 1-bit quantized systems. Their analysis highlighted a fundamental SNR trade-off, according to which operating at the right noise level significantly enhances the system's performance.

Even though some related works have been done so far by considering the hardware impairment to estimate the channel, they did not use state-of-the-art algorithms to hyper-tune the DNN model; consequently, this results in unsatisfactory performance. Therefore, to handle such an issue, we employed the stacked BLSTM with a grid search algorithm for effective channel estimation. To the best of our knowledge, this thesis is the first work that uses an RNN-type deep learning model for channel estimation under both BS and UE non-linear distortions.

CHAPTER 3

SYSTEM DESIGN

This chapter presents the system design for the proposed system. Figure 3.1 depicts a block diagram of the designed system, which incorporates three typical design elements: data collection, modeling and hyper-tuning, and evaluation (training and testing). The data has been collected from an online generation platform. To model our system, we have employed Stacked BLSTM. Finally, we have used accuracy, *MSE* and *NMSE* to evaluate the performance of the system. The details are given below with the respective subsections.

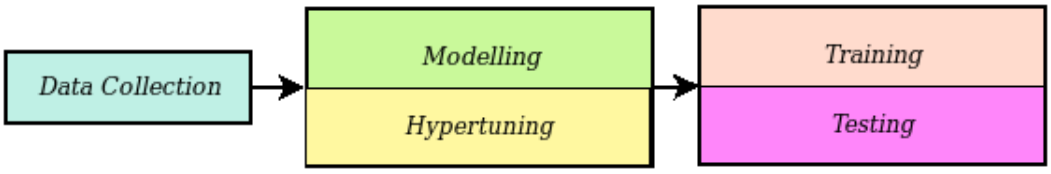


FIGURE 3.1: Block diagram of the proposed system.

3.1 Data Collection

This section explains how the dataset is extracted from Massive MIMO and is used to estimate channels at the Base Station (BS) using Uplink (UL) data transmission. The UL system modeled in this thesis is according to [8], [40], [41]. The process of UL data transmission mainly consists of two phases. The first phase involves all UE transmitting pilot sequences simultaneously, and BS estimates channels based on these pilots. The second phase involves all UE transmitting data symbols; BS decodes these data using channel estimates from the first phase [40]. To assess the performance of Massive MIMO systems, a channel model must consider the main characteristics of large antenna arrays, including array geometry, channel response correlations, and the physical location and orientation of BS and UE [8].

The effective channels and distortion variances are calculated as per Equation 3.12, 3.13 and 3.14 using the input generated for the proposed model as per Equation 3.4, 3.6 and 3.9 as third-order non-linear behaviors of both the UE and BS hardware.

3.1.1 Uplink System Modeling

In this section, we analyze a single-cell Massive MIMO UL system. Figure 3.2 depicts a single hexagonal cell with a BS at the central position that's serving K single-antenna UE by equipping M antennas. In a circularly symmetric Gaussian distribution, let any UE be linked to one of the BS's M antennas, and the *matrix of channel gain*[1] is denoted by $\mathbf{G} = [g_1, g_2, \ldots, g_K] \in \mathbb{C}^{M \times K}$.

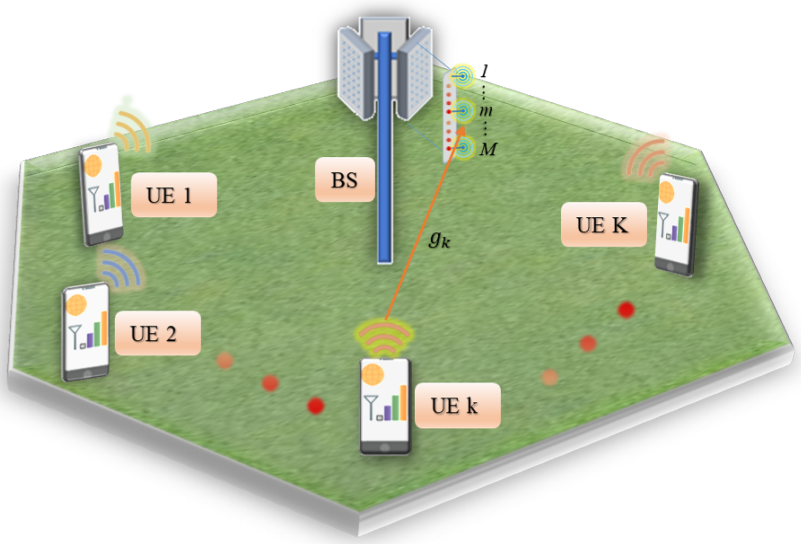


FIGURE 3.2: An illustration of single-cell UL in Massive MIMO.

The channel matrix $\mathbf{H} = [h_1, h_2, \ldots, h_K] \in \mathbb{C}^{M \times K}$ for $h_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{1})$ and the small-scale *i.i.d Rayleigh fading* $h_k = [h_{k,1}, h_{k,2}, \ldots, h_{k,M}]^T \in \mathbb{C}^M$. Whereas the large-scale fading coefficient, $\beta_k = \frac{1}{M} tr(\mathbf{R}_k)$, characterizes long-term channel effects such as *pathloss exponent*[2] α, *shadow fading*[3] $F_k \sim \mathcal{N}(0, \sigma_{sf}^2)$, and $\mathbf{R}_k \in \mathbb{C}^{M \times M}$ is *positive semi-definite* spatial correlation matrix[4] with an eigenstructure, then the *channel gain* g_k is the *channel response* h_k times square root of *channel attenuation* β_k as follows.

$$g_k = h_k \sqrt{\beta_k} \tag{3.1}$$

[1]Matrix of channel gain is modeled as

$$\mathbf{G} = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,K} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M,1} & g_{M,2} & \cdots & g_{M,K} \end{bmatrix}$$

[2]α is a measure of the rate at which signal power decays with the transmitter-receiver distance.
[3]A variance σ_{sf}^2 of the shadow fading is responsible for determining the size of random variations.
[4]Hermitian matrices are *positive semi-definite* if all non-negative eigenvalues are present.

In any arbitrary coherence block, the received signal at the BS without any noise or distortion is $\mathbf{u} = [u_1,\, u_2,\, \ldots,\, u_M]^T \in \mathbb{C}^M$. This is expressed as Equation 3.2.

$$\mathbf{u} = \sum_{i=1}^{K} g_i s_i = \mathbf{G}\mathbf{s} \tag{3.2}$$

where $\mathbf{s} = [s_1,\, s_2,\, \ldots,\, s_K]^T \in \mathbb{C}^K$ is an information-bearing signal during UL data transmission and $\mathbf{s} \sim \mathcal{N}_{\mathbb{C}}(0,\, p\mathbf{I}_K)$ is also independent. The UL signal from UE k is $s_k \in \mathbb{C}$ with transmission power $p_k = \mathbb{E}\{|s_k|^2\}$, irrespective of whether it is a random $s_k \sim \mathcal{N}_{\mathbb{C}}(0,\, p_k)$ or a deterministic pilot signal with $p_k = |s_k|^2$.

I) Modeling UE Hardware Non-linearities

In [42], it has been assumed that the stochastic additive or multiplicative model when analyzing the BS hardware distortion at each UE. However, considering UE hardware as perfect is the performance-limiting factor since it is not averaged out over the BS antennas. A deterministic *third-order behavioral hardware non-linearities modeling* is typically proposed for more practical communication systems while considering potential symbol-dependent distortion. In this thesis, the hardware non-linearities on the UE side are effectively modeled using a *third-order quasi-memoryless polynomial model*. Equation 3.3 reveals the impact of k^{th} UE hardware impairment on information-bearing signals. $\varsigma = [\varsigma_1,\, \varsigma_2,\, \ldots,\, \varsigma_K]^T \in \mathbb{C}^K$ uses complex Gaussian functions to maximize differential entropy while evaluating SE.

$$s_k = \sqrt{\eta_k}\left(\tilde{b}_{0k}\varsigma_k + \tilde{b}_{1k}\varsigma_k^3\right) \qquad \text{for } k = 1,\, 2,\, ...,\, K \tag{3.3}$$

where ς_k is the actual *desired signal* to be transmitted with $\mathbb{E}\{|\varsigma_k|^2\} = 1$, but the mean is zero. $\sqrt{\eta_k}$ is the scaling factor and from $p_k = \mathbb{E}\{|s_k|^2\}$ as the average power of s_k, we have[5] $\mathbb{E}\{|s_k|^2\} = \chi_2\eta_k$ thus $\eta_k = \frac{p_k}{\chi_2}$. Since UEs can have unequal *path-losses* and/or *transmit powers*, we assume variable-gain power amplifiers are used at the UEs. The normalized complex scalar coefficients $\{\tilde{b}_{0k}, \tilde{b}_{1k}\}$ are as follows:

$$\tilde{b}_{lk} = \frac{b_{lk}}{\left(b_{\text{UE}}^{\text{off}}\,\mathbb{E}\{|\varsigma_k|^2\}\right)^l} \qquad \text{for } l = 0,\, 1, \tag{3.4}$$

where back-off $b_{\text{UE}}^{\text{off}}$ is input while adjusting the operating point of the power amplifier can be consistent across all UEs. This UE modeling uses a *normalized distortion*.

[5] $\chi_l \triangleq \mathbb{E}\{|\tilde{b}_{0k}\varsigma_k + \tilde{b}_{1k}\varsigma_k^3|^l\}$ for $l = 2,\, 4,\, ...$ *is the even order moments of i.i.d distorted signals.*

II) Modeling BS Hardware Non-linearities

The non-ideal BS receiver hardware is modeled as a *non-linear quasi-memory-less function* with distorted amplitude and phase of the received signal. The signal detection method utilizes the received signal $\mathbf{y} = [y_1, y_2, \ldots, y_M]^T \in \mathbb{C}^M$ as per Equation 3.5, which is available in *digital baseband*.

$$\mathbf{y} = \mathbf{z} + \mathbf{n} \tag{3.5}$$

where $\mathbf{z} = [z_1, z_2, \ldots, z_M]^T \in \mathbb{C}^M$ is the distorted signal due to hardware non-linearties, and $n \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \sigma^2\mathbf{I}_M)$ is independent Additive White Gaussian Noise (*AWGN*) at the BS with a zero mean and variance of σ^2.

The third-order non-linearities in the BS hardware—oscillators, amplifiers, mixers, etc. The effects of the distorted baseband are modeled as:

$$z_m = \tilde{a}_{0m}u_m + \tilde{a}_{1m}u_m^3 \qquad \text{for } m = 1,\ 2,\ ...,\ M \tag{3.6}$$

where z_m is *distorted noise-free signal* received at the m^{th} BS antenna and $\{\tilde{a}_{0m}, \tilde{a}_{1m}\}$ are complex coefficients in distorted *AM/AM* and *AM/PM* of Gallium Nitride (*GaN*).

Non-linear amplifications in the *Low-Noise Amplifier* (*LNA*) and quantization errors in the *Analog-to-Digital Converter* (*ADC*) are typical characteristics of BS hardware. However, the BS back-off $b_{\text{BS}}^{\text{off}}$ is applied in the *LNA* to eliminate clipping and reduce the non-linear amplification effects. The complex scalar coefficient of the m^{th} BS antenna $\tilde{a}_{lm}$ is expressed as

$$\tilde{a}_{lm} = \frac{a_{lm}}{\left(b_{\text{BS}}^{\text{off}}\,\mathbb{E}\{|u_{lm}|^2\}\right)^l} = \frac{\alpha}{\left(b_{\text{BS}}^{\text{off}}\,\sum_{k=1}^{K}\beta_k p_k\right)^l} \qquad \text{for } l = 0,\ 1, \tag{3.7}$$

3.1.2 Problem Formulation

The BS must know the *pilot sequence* of a specific UE to estimate its channel. Hence, pilots are deterministic sequences assigned when the UE establishes a connection with the BS using an arbitrary allocation method, and *channel gains* are constant inside a coherence block but the signals and noise undergo new realizations with each sample. Assume τ_{p_k} samples are reserved for UL pilot signaling in each coherence block. Each UE transmits a *pilot sequence* that spans these τ_{p_k} samples. The *pilot sequence* of UE k is deemed as $\varphi_k \in \mathbb{C}^{\tau_{p_k}}$. Also, let the unit-magnitude of elements provide a constant power level, which means that $||\varphi_k||^2 = \varphi_k^H\varphi_k = \tau_{p_k}$. Each φ_k

element is scaled by the UL transmission power as $\sqrt{p_k}$ and then transmitted as the signal $s_k = \sqrt{p_k}\varphi_k^T$ using τ_{p_k} samples [8], this is the received signal $\mathbf{y}_m^p \in \mathbb{C}^{M\times\tau_{p_k}}$ at the BS. Thus, $\mathbf{y}_m^p$ can be utilized to estimate the *channel gains* as an observation from the BS, which is given by $\mathbf{y}_m^p = \mathbf{z}_m^p + \mathbf{n}_m^p$ become as follows

$$\mathbf{y}_k^p = \sum_{k=1}^{K} g_k \left(\sqrt{p_k}\varsigma_k + \omega_k\right) + \mathbf{n}_k^p \tag{3.8}$$

where $\mathbf{n}_m^p \in \mathbb{C}^{M\times\tau_{p_k}}$ is the additive uncorrelated thermal noise of BS receiver with $\mathbf{n}_m^p \sim \mathcal{N}_{\mathbb{C}}(0_{\tau_{p_k}}, \sigma_{BS}^2\mathbf{I}_{\tau_{p_k}})$ and ω_k is circularly symmetric distortion with each i.i.d. element distributed as variance of $(1-\kappa_k)p_k$ for $\kappa_k \in [0,\ 1]$ is level of UE k hardware non-linearities, and the transmit power of UE k can be $p_k = \frac{\rho_k\sigma_{BS}^2}{\beta_k}$ for ρ_k is the SNR.

The distorted information-bearing signal in the absence of *power scaling* parameters denoted as $\mathbf{v} = [v_1,\ v_2,\ \ldots,\ v_K]^T \in \mathbb{C}^K$ thus $v_k = \tilde{b}_{0k}\varsigma_k + \tilde{b}_{1k}\varsigma_k^3$ and according to Equation 3.3, we can have $s_k = \sqrt{\eta_k}v_k$. The *channel gain with power scaling* $\tilde{g}_k \in \mathbb{C}^K$ can be defined as

$$\tilde{g}_k \triangleq g_k\sqrt{\eta_k} \tag{3.9}$$

In a nutshell, the non-linear function of the *channel matrix* $\mathbf{G}$ and the *digital base-band signals* $\mathbf{y}$, which can be analyzed as *the effective channel matrix* $\mathbf{C}_{y\varsigma} \in \mathbb{C}^{M\times K}$ based on methods of Bussgang decomposition in [1], [11], [41]. So, the effective channel $[\mathbf{C}_{y\varsigma}]_k$ between UE k and one of the BS antenna m with the third-order hardware non-linearities can be defined as

$$[\mathbf{C}_{y\varsigma}]_k = \mathbb{E}_{|\mathbf{G}}\{z_m\varsigma_k^*\} = \tilde{a}_{0m}\mathbb{E}_{|\mathbf{G}}\{u_m\varsigma_k^*\} + \tilde{a}_{1m}\mathbb{E}_{|\mathbf{G}}\{u_m^3\varsigma_k^*\} \tag{3.10}$$

$$= \tilde{a}_{0m}\sum_{i=1}^{K} g_i\mathbb{E}\{s_i\varsigma_k^*\} + \tilde{a}_{1m}\sum_{i_1=1}^{K} g_{i_1}\sum_{i_2=1}^{K} g_{i_2}^*\sum_{i_3=1}^{K} g_{i_3}\mathbb{E}\{s_{i_1}s_{i_2}^*s_{i_3}\varsigma_k^*\} \tag{3.11}$$

$$= \tilde{a}_{0m}\,\tilde{g}_k\mathbb{E}\{v_k\varsigma_k^*\} + \tilde{a}_{1m}\sum_{i_1=1}^{K} \tilde{g}_{i_1}\sum_{i_2=1}^{K} \tilde{g}_{i_2}^*\sum_{i_3=1}^{K} \tilde{g}_{i_3}\mathbb{E}\{v_{i_1}v_{i_2}^*v_{i_3}\varsigma_k^*\} \tag{3.12}$$

where $\mathbb{E}_{|\mathbf{G}}$ is the conditional expectation given $\mathbf{G}$. For the effective channel, each element's constellation[6] with *the even-order moments* of the UL data signal ς_k given by $\zeta_l = \mathbb{E}\{\varsigma_k\}^l$ for $l = 2,\ 4,\ \ldots$, and as per Equation C.5 we can have Equation 3.13.

[6]Let symbol constellations have the circular shift symmetry of 90°.

$$
\begin{aligned}
[\mathbf{C}_{y\varsigma}]_k = \; & \tilde{a}_{0m}\,\tilde{g}_k\left(\tilde{b}_{0k}+\zeta_4\tilde{b}_{1k}\right)+\tilde{a}_{1m}(\tilde{g})_k^3\left(\zeta_{10}B_{1,1,1}+2\zeta_8B_{1,1,0}+\zeta_8B_{1,0,1}\right.\\
& \left.+2\zeta_6B_{0,0,1}+\zeta_6B_{0,1,0}+\zeta_4B_{0,0,0}\right)\\
& +2\tilde{a}_{1m}\,\tilde{g}_k\left(\tilde{b}_{0k}+\zeta_4\tilde{b}_{1k}\right)\left(\zeta_6B_{1,1}+\zeta_4B_{1,0}+\zeta_4B_{0,1}+B_{0,0}\right)\sum_{\substack{i=1\\ i\neq k}}^{K}|\tilde{g}_i|^2
\end{aligned}
\tag{3.13}
$$

Also, for analyzing the SE and maximizing the differential entropy we considered the data symbols as $\varsigma \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_K,\, \mathbf{I}_K)$ and the received signal at the BS in terms of this can be $\mathbf{y} = \mathbf{C}_{y\varsigma}\varsigma + \mu$ where ς and μ are uncorrelated *actual data signals* and *additive noise*, respectively. While quantifying the performance of receivers, the distortion correlation matrix $\mathbf{C}_{\mu\mu}$ can be obtained as follows

$$
\mathbf{C}_{\mu\mu} = \mathbb{E}_{|\mathbf{G}}\{\mu\mu^H\} = \left(\mathbf{C}_{zz}+\sigma^2\mathbf{I}_M\right) - \mathbf{C}_{y\varsigma}\mathbf{C}_{y\varsigma}^H \text{ herein } \mathbf{C}_{zz} = \mathbb{E}_{|\mathbf{G}}\{\mathbf{z}\mathbf{z}^H\} \tag{3.14}
$$

where $(\mathbf{C}_{zz}+\sigma^2\mathbf{I}_M)$ is the *correlated received data symbols* at the BS M antennas.

3.1.3 Dataset Preparation

To train and test the proposed system, we have used an online-generated dataset from the derived functions. We utilized distortions and channel gains of UL data transmission for training the proposed model. The target output is obtained as effective channels and distortion correlation matrix, with antenna characteristics considered for training and testing datasets. Table 3.1 reveals the overall characteristics of the prepared dataset.

TABLE 3.1: Description of the prepared dataset

PARAMETERS	CHARACTERISTICS/VALUES
Input Data Type	Channel gain matrix
Data Format	NumPY arrays (.npy)
Means of Collection	Online generation
Means of Generation	Using analytically derived functions
Generation Protocol	3GPP specification [43]
Generation Platform	Google COLAB
Number of antennas of the BS (M)	100
Number of UE in the cell (K)	20

We generated the *channel gains* and saved as *NumPY* arrays (.npy) data format. The data has been generated according to the 3rd Generation Partnership Project Urban Microcell Model (*3GPP* [43]) on Google COLAB. The data-generating algorithm for the models of channel gains and both UE and BS hardware non-linearities is presented in Algorithm 1 in Appendix B.

3.2 Model Designing and Hypertuning

3.2.1 Model Designing

As we discussed in Section 2.3.4, [1] has used the stacked layers of Fully-Connected Deep Network (FCDN). Using this related work as a benchmark, we have carried out our thesis work with a more advanced modeling approach, the Stacked Bidirectional LSTM (BLSTM). Two BLSTMs each with $30K$ neurons are stacked and used as a model of the proposed system. Because of the reason stated in Section 2.3.3, at the input layer, we reshaped the data from $2D$ to $3D$ by appending the timestep variable as a value of 1.

As illustrated in Figure 3.3, the proposed model architecture has four layers in total; the input layer $\{x_1, x_2, \ldots, x_k\}$, two layers of BLSTM and the output layer $\{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_k\}$. Hence, the two BLSTM layers are stacked, each of which is linking two hidden layers pointing in opposing directions (the backward $\{\mathrm{b}_1, \mathrm{b}_2\}$ and forward $\{\mathrm{f}_1, \mathrm{f}_2\}$) and the output layer can concurrently retrieve information (in the form of weights $\{\mathrm{U}, \mathrm{W}\}$) from both directions of these layers.

3.2.2 Model Hyper-tuning

The Stacked BLSTM has a lot of hyper-parameters that need to be fine-tuned. These include the number of epochs, batch size, number of neurons in each layer, optimizer functions, activation functions, etc. Hyper-tuning gives us the power to optimize model performance, but tuning a large number of hyper-parameters presents a major challenge regarding computation time [44]. Hence, we chose the following hyper-parameters and evaluated their values for training the Stacked BLSTM model:

- The number of epochs, the number of times the entire training dataset is shown to the network during training.
- The batch size also needs optimization in the training of the network, defining how many patterns to read at a time and keep in memory. Iterative gradient

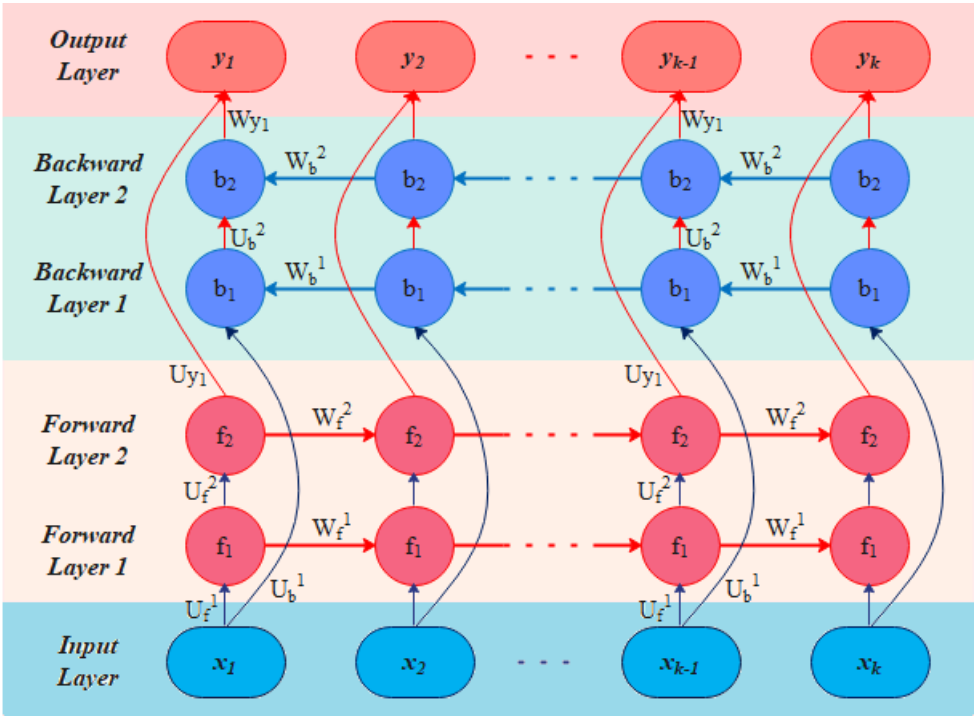


FIGURE 3.3: The Stacked BLSTM architecture.

descent is the number of patterns shown to the network before the weights are updated. Also, BLSTM requires more optimal batch sizes and epochs.

- The number of neurons in a layer is another important parameter to set up. In general, the network's capacity is determined by the number of neurons in a layer, at least in that particular architecture.
- An optimization algorithm is selected to train a model and tune its parameters, such as the learning rate, momentum, etc. The learning rate controls how much to update the weight at the end of each batch, and the momentum controls how much to let the previous update influence the current weight update.
- The activation function, which controls the non-linearity of individual neurons and when to fire, can be selected. It used to be either the sigmoid, relu, linear, tanh, etc. functions, which enable the model to learn more complex things and one among these can be more suitable for this particular problem. These activation functions are given in Table A.1 in terms of mathematical model and graph. Herewith, we hyper-tuned also the selection of the weight initialization method.

3.3 Model Training and Testing

3.3.1 Model Training

For the m^{th} antenna's channels, the first $2K$ inputs of the Stacked BLSTM model are the real $\Re$ and imaginary $\Im$ parts of the processed *received signals* in the UL training phase by correlating them with *pilot sequences* as $\varphi_k^H y_m^p$ for $k = 1, \ldots, K$, which represents a *naive estimate* of the *channel gain* g_k excluding the additional distortion terms, and which are different processed signals for different users. Thus, the UEs have independent small-scale fading coefficients. Unlike this, large-scale fading coefficients are the same among all UEs. The last K inputs depend on the square roots of the large-scale fading coefficient times the corresponding power level for every UE and can be expressed as: $\sqrt{[\beta_k\, p_k]}$ for $k = 1, \ldots, K$, known at the BS. To handle any set of UE locations without the need for repeatedly training the model.

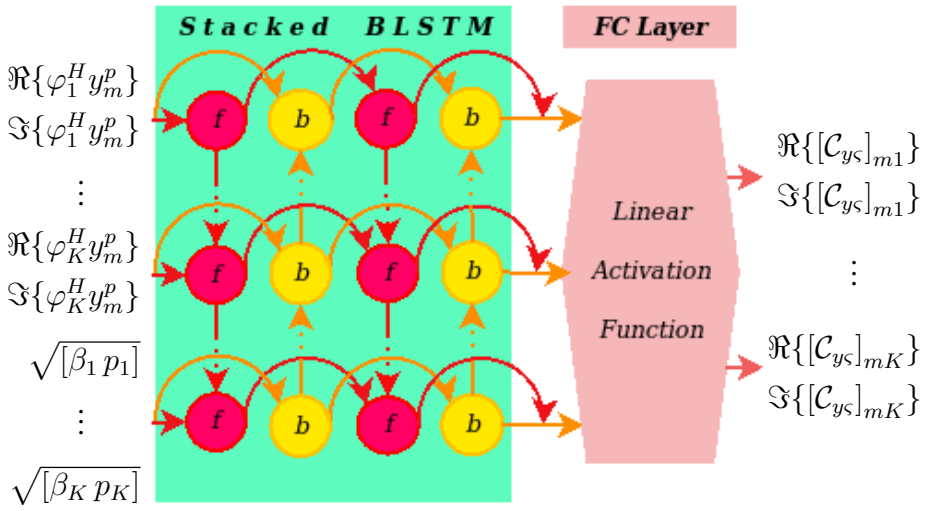


FIGURE 3.4: The Stacked BLSTM activated by *linear* function at the output layer is used for effective channel estimation.

The outputs of the Stacked BLSTM model in Figure 3.4 are with the real $\Re$ and imaginary $\Im$ parts of the effective channel elements. The output layer of this model used a fully connected (*FC*) layer with a *linear* activation since the outputs can take both positive and negative values. Whereas a similar model is used for an estimation of the element-wise distortion correlation in Figure 3.5, the *ReLU* activation is used at the output layer. Thus, we can get the logarithm of the diagonal elements of the distortion correlation matrix after normalizing it with the noise variance, and since $\log_{10}\left([\mathcal{C}_{\mu\mu}]_m / \sigma^2\right)$ is always greater than or equal to 1, we can also exploit the knowledge that the output is always non-negative. The reason behind taking the logarithm is to provide more uniformity to the output, which improves the capability of the model to learn faster.

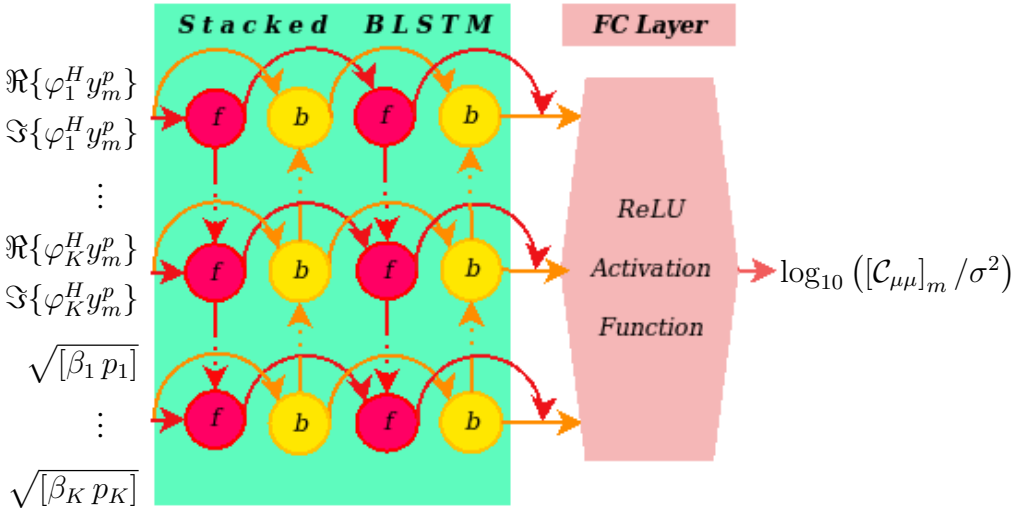


FIGURE 3.5: The Stacked BLSTM excited by *ReLU* activation function at the output layer is used for effective channel estimation.

Even though we successfully trained the models illustrated in Figures 3.4 and 3.5, the amount of fluctuating *SNR* was still one of the major difficulties. But to make learning easier and achieve improved learning performance, we considered lexicographical rearrangement of inputs' and outputs' indices in ascending order. In addition, the data corresponding to the outliers were excluded from the training set. Thus, the channel gains, which are the last K inputs of the model empirically result in improved learning. In addition, the outputs of the model in Figure 3.4 are scaled using the *MinMax Scaler*, which improves learning.

3.3.2 Model Testing

We tested the Stacked BLSTM model to investigate if it has learned how the desired signal could be affected due to the hardware non-linearities during UL data transmission. Hence, we have used a separate testing set containing the set of input-output vector pairs (given as $\{\mathbf{r}_0^t,\, \tilde{\mathbf{r}}_p^t\}_{t=1}^T$ where T is the testing set size, $\tilde{\mathbf{r}}_p^t$ is the desired output, and $\mathbf{r}_0^t$ is input vector), a loss function is applied to improve the parameters for the p^{th} layer, the weight and bias pair $\{W_p,\, b_p\}_{p=1}^P$ as Equation 3.15.

$$L\left(\{W_p,\, b_p\}_{p=1}^P\right) = \frac{1}{T}\sum_{t=1}^{T} l(\tilde{\mathbf{r}}_p^t,\, \mathbf{r}_p^t) \tag{3.15}$$

3.4 Evaluation Metrics

Accuracy and error rates are the standard and most popular evaluation metrics in machine learning while assessing the performance of estimators. Table 3.2 shows

a *confusion matrix*, which is a tabular organization of the estimations made by the proposed model.

	POSITIVE ESTIMATION	NEGATIVE ESTIMATION
POSITIVE	$True\ Positive\ (TP)$	$False\ Negative\ (FN)$
NEGATIVE	$False\ Positive\ (FP)$	$True\ Negative\ (TN)$

TABLE 3.2: Confusion matrix table.

According to Table 3.2, the accuracy is calculated as per Equation 3.16. Thereby, the accuracy and error rates are mutually complementary, i.e. $accuracy\ (A) = 1 - error\ rate\ (\mathbf{e})$.

$$A = \frac{\#\,Correct\ Estimation}{\#\,Total\ Estimation} = \frac{TP + TN}{TP + TN + FN + FP} \tag{3.16}$$

During UL signal transmission, from UE k and the effective Signal-to-Noise Ratio (*SNR*) denoted as ρ_k, which is also known as the *transmission power throughout* and given by

$$\rho_k = \frac{p_k \beta_k}{\sigma^2} \tag{3.17}$$

Statistics also defined the mean squared error (*MSE*), which is popularly used in regression problems, measures the accuracy of estimation using the data set, with a minimal *MSE* indicating high estimation quality [8]. Hence, the *MSE* formula as a *loss function* for the channel estimate of UE k, given the estimator error vector is $e_k = g_k - \hat{g}_k$ with the model estimate $\hat{g}_k$ can be defined as

$$MSE = \mathbb{E}||e_k||^2 = \mathbb{E}||tr(e_k(e_k)^H)||^2 = tr(\mathbf{C}_k) \tag{3.18}$$

where e_k is a complex Gaussian vector with the element distributed in the range of $\sigma^2_{e_k} = \frac{\beta_k}{1+\tau_{p_k}\rho_k}$ within the *pilot length* is τ_{p_k} for the sequences, which are the columns in the matrix of Discrete Fourier Transform (*DFT*).

Also, the normalized MSE (*NMSE*) measures the relative estimate error per antenna, making it an appropriate metric. We compared the estimation quality of different estimation schemes using the *NMSE*. This is expressed as:

$$NMSE_k = \frac{tr(\mathbf{C}_k)}{tr(\mathbf{R}_k)} \tag{3.19}$$

CHAPTER 4

RESULT AND DISCUSSION

This chapter presents experimental results of the model proposed for the channel and distortion variance estimation tasks. The hyper-parameters best settings are achieved by using the grid search method. We also graphically analyzed the effects of different estimator models over the *NMSE* vs. *CDF*, with their corresponding details.

4.1 Experimental Setup

To perform the training-testing operation, we have started with the data splitting task. Table 4.3 reveals a description of the parameters used for our experiment.

TABLE 4.1: Description of parameters used for our experiment

PARAMETER	VALUE/TECHNIQUE
Dataset splitting approach	Random
Training dataset size	$3\,Million$
Validation dataset size	$0.2\,Million$
Carrier Frequency	$2\,GHz$
Bandwidth	$20\,MHz$
GaN amplifier operating at	$2.1\,GHz$
Coverage area of BS antenna	$(250 \times 250)\,m^2$
Height of antenna for BS	$10\,m$
Height of antenna for UE	$1.5\,m$
Max. transmission power of UE antenna	$200\,mW$
$b_{\text{BS}}^{\text{off}}$ and $b_{\text{UE}}^{\text{off}}$	$7\,dB$
Noise variance (σ^2)	$-96\,dBm$
Channel fading type	Rayleigh

The data splitting has been done using a random splitting approach. The total data was split into two sets: 3×10^6 for the training set and 2×10^5 for the testing set. The training set was generated by using large-scale fading parameters according to the 3rd Generation Partnership Project (3GPP) Urban Microcell model [43] with a $2\,GHz$ carrier frequency and $20\,MHz$ bandwidth. For each training sample, the users are dropped randomly in a cell of $250\,m \times 250\,m$. The number of BS antennas is $M = 100$, and $K = 20$ users are uniformly distributed in the cell. The BS antenna array and the UE heights from the ground are $10\,m$ and $1.5\,m$, respectively. The noise variance is $\sigma^2 = -96dBm$ and the maximum received power difference $\Delta = 20\,dB$ is used to set the transmission power $\{p_k\}$.

The large-scale fading coefficients, shadowing parameters, probability of *LOS* and *NLOS*, and the Rayleigh factors are simulated based on [43]. The first $2K$ inputs of the proposed model are scaled using the *Standard Scaler*, and the others using the *MinMax Scaler*. The scaling is needed for proper training, and the motivation for these two types of scaling is as follows: the first $2K$ inputs can have both positive and negative values; hence, a *Standard Scaler* that removes the mean and normalizes the input data such that it has unit variance is used for these inputs. Whereas, the other K inputs represent the square root of the channel gain over noise, which is always positive. Moreover, to prevent the large deviation between channel gains, these inputs are scaled between 0.1 and 0.9 using *MinMax Scaler*.

Next, we selected a state-of-the-art optimization algorithm for setting values of the Stacked BLSTM model hyper-parameters, namely grid search, which is designed to conduct hyper-tuning systematically by going through each of the sets of hyper-parameter values automatically during the model training process [45]. The merit of this algorithm is that it attempts all possible combinations of values before arriving at the best setting. But it was computationally expensive. Consequently, we reduced the amount of training-validation set to 1×10^5 and 1×10^4, respectively, only during this experiment to overcome the issue.

4.2 Experimental Result

4.2.1 Hyper-tuning Result

The performance of the proposed stacked BLSTM model was mainly determined by the setups in hyper-parameters, including activation functions, neurons per layer, number of layers, etc. Table 4.2 reveals a description of the hyper-tuning parameters with their respective values and techniques.

TABLE 4.2: Description of the hyper-tuning parameters

PARAMETER	VALUE/TECHNIQUE
Modelling approach (algorithm)	Grid search
Batch size	1000
Number of epochs	50
Hidden unit	1200
Learning rate	0.001
Kernel activation function	ReLU
Kernel initialization mode	Normal
Dropout rate	0.5
Optimizer function	Adamax
Output activation function	Linear
Number of layers	2×BLSTM
Evaluation metrics	MSE and Accuracy

A grid search algorithm was employed to discover the best performance of the model hyper-parameter. After comprehensive evaluation of several hyper-parameter values, the following tabular results are recorded as shown in Table 4.3, 4.4, and 4.5. The model with 1200 hidden layer neurons and each with a dropout rate of 0.5 outperformed. Table 4.6 shows that *Adamax* optimizer with a learning rate of $1e-3$ (i.e. 0.001) for training and the batch size and the maximum number of epochs is set as 1000 and 50, respectively. For every experiment, we employed the early stopping by setting the patience value of 5 indicating the frequency of epochs on which stop improvement in the *MSE* and need no further iteration.

TABLE 4.3: Stacked BLSTM model architectures.
None: supports any size of data.

LAYER TYPE	DESCRIPTION	OUTPUT SIZE
Input	2D array numeric data	None × 60
Reshape	3D array (1 as timestep)	None × 60 × 1
BLSTM + relu	1200 hidden cells	None × 1200 × 1
BLSTM + relu	1200 hidden cells	None × 1200 × 1
Output	project onto 40 classes	None × 40

Among the softmax, relu, linear, tanh, etc. output activation functions, the linear function achieved the best results with an accuracy of about 92.02%.

TABLE 4.4: Grid search algorithm on activation functions.

Hyperparameter values	Accuracy (%)	Loss	Best
softmax	60.63	0.0707	
softplus	91.93	0.0045	
softsign	73.40	0.0117	
relu	77.85	0.0740	
linear	92.02	0.0056	✓

We can see that the best results were achieved with a "normal" weight initialization scheme achieving a performance of about 92.17%.

TABLE 4.5: Grid search on kernel initialization mode.

Hyper-parameter values	Accuracy (%)	Loss	Best
uniform	60.63	0.0707	
lecun uniform	91.93	0.0045	
normal	92.17	0.0043	✓
zero	77.85	0.0740	
glorot normal	12.05	0.0707	
he normal	12.05	0.0707	
he uniform	12.05	0.0707	
glorot uniform	92.02	0.0056	

Therefore, the proposed model contained two BLSTM layers each with $30K$ neurons corresponding to the antennas per coherence block. In addition, for minimization of over-fitting and improving the model's ability to learn and generalize, tuning the dropout rate for regularization involves fitting both the dropout percentage and the weight constraint. The dropout rates were between 0.0 to 0.9 and the *MaxNorm* weight constraint values were between 0 to 5. The algorithm also selected the dropout rate of 50% and *MaxNorm* weight constraint of 3 resulting in the best accuracy of 93.87%.

TABLE 4.6: Grid search on the learning rates.

Learning rates	Accuracy (%)	Loss	Best
0.1	70.30	0.0707	
0.001	93.87	0.0045	✓
0.0001	92.93	0.0046	
0.00001	77.85	0.0740	

4.2.2 Comparison Results

Both Figure 4.1 and Figure 4.2 depict the $NMSE$ of effective channel and distortion variance estimates for $K = 20$ UE over the 3rd-order non-linear distortion. After a careful analysis of the evaluation results, the *MSE* is 5.22% for the benchmark. In contrast, it's 0.89% for the proposed method when an early stopping of $patience = 5$ is applied to both the benchmark and our proposed model. Figure 4.1 reveals the comparison result between the benchmark and our proposed model for the *QPSK* modulation technique. To demonstrate the resilience of the proposed approach to modulation variations, we repeated a similar experiment for *16-QAM*. At the UE transmitter in the practical communication systems, *16-QAM* experiences amplitude distortions of constellation symbols in addition to phase distortions. Figure 4.2 reveals the comparison result between the benchmark and our proposed model for the *16-QAM* modulation technique.

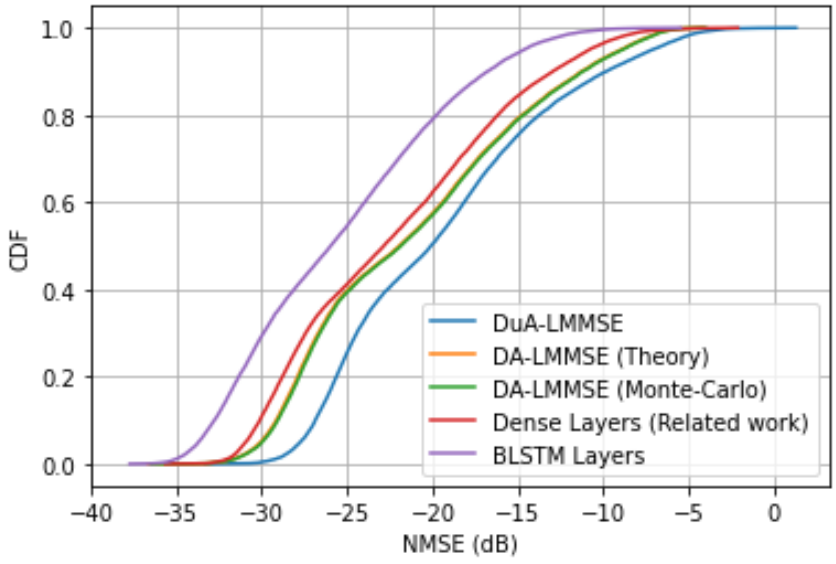


FIGURE 4.1: The third-order non-linear distortion with *QPSK* based comparison between the benchmark and proposed models.

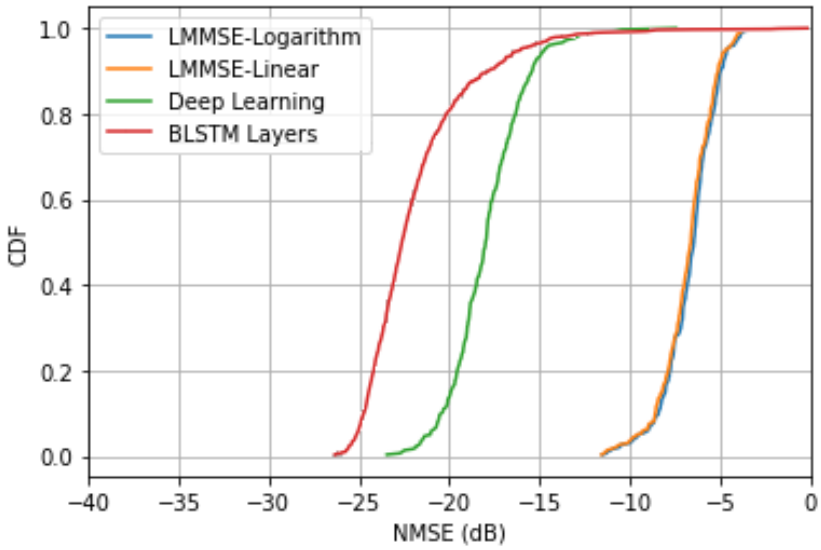


FIGURE 4.2: The third-order non-linear distortion with *16-QAM* based comparison between the benchmark and proposed models.

4.3 Discussion of the Results

The proposed and benchmark models were trained to use the hardware distortion characteristics of the channel gain information. In such an equal *SNR* scenario, the performance gap between the BLSTM and benchmark estimator increases with the *SNR* value. Thus, the Stacked BLSTM architecture-based model is showing an outperforming accuracy after estimating the effective channels and distortion variances rather than the model in the FCDN-based benchmark model within the related work.

Generally, the observed results implied that the proposed model is more convenient to be trained to handle varying users' *SNR* effectively and implicitly learn the *SNR* distribution of the considered propagation environment. Hence, we conclude that the proposed model handles the structure of the hardware distortion more than the benchmark estimators in the related work.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In this thesis, we presented the stacked BLSTM model for channel and distortion variance estimation in the system of Massive MIMO networks, known for employing antenna numbering in the tens or even hundreds at the base station. Formerly, we presented related work as a benchmark model and reviewed older literature on the traditional Bayesian LMMSE estimators. Similar to the benchmark, the proposed model inputs are the channel gains, which are the only required large-scale fading parameters while model training, thus it is used without any additional complexity.

During the experiment, the proposed model hyper-parameters were tuned to obtain a higher estimation quality than the benchmark. We have finally proven that the model could be trained and utilized to provide significantly better estimation in practical Rayleigh fading channel setups with varying SNRs. Unlike Rician, while using Rayleigh, we have observed NLOS waves as there is no dominant LOS component. After training the joint effect of non-linear distortions in the BS and UE hardware, we have demonstrated effective channel estimation and detection in the Massive MIMO.

In conclusion, the proposed model could experimentally demonstrate outperforming results to exploit the structure of transceiver hardware compared to the performance of channel estimation using the benchmark model and previous sub-optimal model-based methods. This implies that the proposed system can be suitable for automating and optimizing beamforming in Massive MIMO, reducing complexity, and adapting dynamic channels.

5.2 Recommendation

The proposed model with stacked BLSTM architecture has been outperforming the benchmark model with FCDN architecture, which was in turn outperforming the old LMMSE estimators. Hence, we can incorporate this model to automate and optimize beamforming in Massive MIMO, reducing complexity and improving adaptation to time-varying channels. In general, we recommend the following as future work to achieve further performance improvement:

- *Pilot contamination:* Future works should consider the impact of pilot contamination on the UE, where the UE from adjacent cells cooperatively share pilot signals to mitigate this issue in regards [46], [47].
- *Dynamic timesteps:* While providing the BLSTM input, we have considered the timestep as 1 only due to the complexity of the data, and we recommend a dynamic timestep.
- *Transfer learning:* Leveraging knowledge from other domains could reduce training data requirements and improve generalization.
- *Federated learning:* Collaborative learning across multiple devices without sharing raw data could address privacy concerns and data scarcity.
- In addition, we have empirically comprehended that increasing the number of neurons per BLSTM results in better performance when compared to the addition of the stack.

BIBLIOGRAPHY

[1] Özlem Tugfe Demir and Emil Björnson. "Channel Estimation in Massive MIMO Under Hardware Non-Linearities: Bayesian Methods Versus Deep Learning". In: *IEEE Open Journal of the Communications Society* 1 (Dec. 2019), pp. 109–124. DOI: 10.1109/OJCOMS.2019.2959913.

[2] Xiao Yu, Wenqian Shen, Rui Zhang, et al. "Channel Estimation for XL-RIS-Aided Millimeter-Wave Systems". In: *IEEE Transactions on Communications* (2023), pp. 1–1. DOI: 10.1109/TCOMM.2023.3286450.

[3] Mahmoud Albreem, Markku Juntti, and Shahriar Shahabuddin. "Massive MIMO Detection Techniques: A Survey". In: *IEEE Communications Surveys and Tutorials* PP (Aug. 2019), pp. 1–1. DOI: 10.1109/COMST.2019.2935810.

[4] Nikolaos I. Miridakis, Theodoros A. Tsiftsis, and Corbett Rowell. "Distributed Spatial Multiplexing Systems with Hardware Impairments and Imperfect Channel Estimation under Rank-1 Rician Fading Channels". In: *CoRR* abs/1610.08073 (2016). arXiv: 1610.08073. URL: http://arxiv.org/abs/1610.08073.

[5] Jiakang Zheng, Jiayi Zhang, Luming Zhang, et al. "Efficient Receiver Design for Uplink Cell-Free Massive MIMO With Hardware Impairments". In: *IEEE Transactions on Vehicular Technology* PP (Feb. 2020), pp. 1–1. DOI: 10.1109/TVT.2020.2975354.

[6] Jayesh Kotecha and Akbar Sayeed. "Transmit Signal Design for Optimal Estimation of Correlated MIMO Channels". In: *Signal Processing, IEEE Transactions on* 52 (Mar. 2004), pp. 546 –557. DOI: 10.1109/TSP.2003.821104.

[7] David Neumann, Thomas Wiese, and Wolfgang Utschick. "Learning the MMSE Channel Estimator". In: *IEEE Transactions on Signal Processing* 66.11 (June 2018), pp. 2905–2917. DOI: 10.1109/tsp.2018.2799164. URL: https://doi.org/10.1109%2Ftsp.2018.2799164.

[8] Emil Björnson, Jakob Hoydis, and Luca Sanguinetti. "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency". In: *Foundations and Trends® in Signal Processing* 11.3-4 (2017), pp. 154–655. ISSN: 1932-8346. DOI: 10.1561/2000000093. URL: http://dx.doi.org/10.1561/2000000093.

[9] Jamal Zafar, Usman Malik, Imran Rashid, et al. "Effects of transmit non-linearity on performance of massive MIMO systems". In: Oct. 2016, pp. 1–4. DOI: 10.1109/ICET.2016.7813205.

[10] Ulf Gustavsson, César Sánchéz-Perez, Thomas Eriksson, et al. "On the Impact of Hardware Impairments on Massive MIMO". In: *CoRR* abs/1411.7197 (2014). arXiv: 1411.7197. URL: http://arxiv.org/abs/1411.7197.

[11] Emil Björnson, Luca Sanguinetti, and Jakob Hoydis. *Can Hardware Distortion Correlation be Neglected When Analyzing Uplink SE in Massive MIMO?* 2018. arXiv: 1805.07958 [cs.IT].

[12] Daniel Rönnow and Peter Handel. "Nonlinear Distortion Noise and Linear Attenuation in MIMO Systems—Theory and Application to Multiband Transmitters". In: *IEEE Transactions on Signal Processing* PP (Aug. 2019), pp. 1–1. DOI: 10.1109/TSP.2019.2935896.

[13] Sai Subramanyam Thoota, Chandra Murthy, and Ramesh Annavajjala. "Quantized Variational Bayesian Joint Channel Estimation and Data Detection for Uplink Massive MIMO Systems with Low-resolution ADCS". In: Oct. 2019, pp. 1–6. DOI: 10.1109/MLSP.2019.8918721.

[14] Christopher Mollén, Ulf Gustavsson, Thomas Eriksson, et al. *Impact of Spatial Filtering on Distortion from Low-Noise Amplifiers in Massive MIMO Base Stations*. 2018. arXiv: 1712.09612 [cs.IT].

[15] Steve Taranovich. *Creating 5G Massive MIMO: Part 1*. 2022. URL: https://www.5gtechnologyworld.com/creating-5g-massive-mimo-part-1/.

[16] *Massive MIMO*. 2021. URL: https://futurenetworks.ieee.org/topics/massive-mimo.

[17] Kapilavai Suma. *Best practices to accelerate 5G base station deployment: Your RF front-end massive MIMO Primer*. 2019. URL: https://www.qorvo.com/design-hub/blog/best-practices-to-accelerate-5g-base-station-deployment.

[18] Emil Björnson, Erik G. Larsson, and Thomas L. Marzetta. "Massive MIMO: ten myths and one critical question". In: *IEEE Communications Magazine* 54.2 (2016), pp. 114–123. DOI: 10.1109/MCOM.2016.7402270.

[19] Junting Chen and Vincent K. N. Lau. "Two-Tier Precoding for FDD Multi-Cell Massive MIMO Time-Varying Interference Networks". In: *IEEE Journal on Selected Areas in Communications* 32.6 (2014), pp. 1230–1238. DOI: 10.1109/JSAC.2014.2328391.

[20] Ansuman Adhikary, Junyoung Nam, Jae-Young Ahn, et al. "Joint Spatial Division and Multiplexing—The Large-Scale Array Regime". In: *IEEE Transactions on Information Theory* 59.10 (2013), pp. 6441–6463. DOI: 10.1109/TIT.2013.2269476.

[21] C.B. Papadias and A.J. Paulraj. "Space-time signal processing for wireless communications: a survey". In: *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*. 1997, pp. 285–288. DOI: 10.1109/SPAWC.1997.630370.

[22] D. Gerlach and A. Paulraj. "Adaptive transmitting antenna arrays with feedback". In: *IEEE Signal Processing Letters* 1.10 (1994), pp. 150–152. DOI: 10.1109/97.329842.

[23] R. Raich and G.T. Zhou. "On the modeling of memory nonlinear effects of power amplifiers for communication applications". In: *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.* 2002, pp. 7–10. DOI: 10.1109/DSPWS.2002.1231065.

[24] Christopher Mollén, Ulf Gustavsson, Thomas Eriksson, et al. "Impact of Spatial Filtering on Distortion From Low-Noise Amplifiers in Massive MIMO Base Stations". In: *IEEE Transactions on Communications* 66.12 (2018), pp. 6050–6067. DOI: 10.1109/TCOMM.2018.2850331.

[25] Jiakang Zheng, Jiayi Zhang, Luming Zhang, et al. "Efficient Receiver Design for Uplink Cell-Free Massive MIMO With Hardware Impairments". In: *IEEE Transactions on Vehicular Technology* 69.4 (2020), pp. 4537–4541. DOI: 10.1109/TVT.2020.2975354.

[26] Emil Bjornson, Luca Sanguinetti, and Jakob Hoydis. "Can Hardware Distortion Correlation be Neglected when Analyzing Uplink SE in Massive MIMO?" In: *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2018, pp. 1–5. DOI: 10.1109/SPAWC.2018.8446008.

[27] Muhammad Uzair and Noreen Jamil. "Effects of Hidden Layers on the Efficiency of Neural Networks". In: *2020 IEEE 23rd International Multitopic Conference (INMIC)*. 2020, pp. 1–6. DOI: 10.1109/INMIC50486.2020.9318195.

[28] Mike Schuster and Kuldip Paliwal. "Bidirectional recurrent neural networks". In: *Signal Processing, IEEE Transactions on* 45 (Dec. 1997), pp. 2673 –2681. DOI: 10.1109/78.650093.

[29] Chong Bian, Huoliang He, and Shunkun Yang. "Stacked bidirectional long short-term memory networks for state-of-charge estimation of lithium-ion batteries". In: *Energy* 191 (2020), p. 116538.

[30] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, et al. "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values". In: *Transportation Research Part C: Emerging Technologies* 118 (2020), p. 102674.

[31] Khaled A Althelaya, El-Sayed M El-Alfy, and Salahadin Mohammed. "Evaluation of bidirectional LSTM for short-and long-term stock market prediction". In: *2018 9th international conference on information and communication systems (ICICS)*. IEEE. 2018, pp. 151–156.

[32] Ziyong Ran, Desheng Zheng, Yanling Lai, et al. "Applying stack bidirectional LSTM model to intrusion detection". In: *CMC-Comput. Mater. Continua* 65.1 (2020), pp. 309–320.

[33] DK Thara, BG PremaSudha, and Fan Xiong. "Epileptic seizure detection and prediction using stacked bidirectional long short-term memory". In: *Pattern Recognition Letters* 128 (2019), pp. 529–535.

[34] Chang Jae Chun, Jae Mo Kang, and Il Min Kim. "Deep Learning-Based Channel Estimation for Massive MIMO Systems". English. In: *IEEE Wireless Communications Letters* 8.4 (Aug. 2019), pp. 1228–1231. ISSN: 2162-2337. DOI: 10.1109/LWC.2019.2912378.

[35] Tong Chen, Jiajia Guo, Chao-Kai Wen, et al. "Deep Learning for Joint Channel Estimation and Feedback in Massive MIMO Systems". In: *CoRR* abs/2011.07242 (2020). arXiv: 2011.07242. URL: https://arxiv.org/abs/2011.07242.

[36] Shilpa Rao, Gonzalo Seco-Granados, Hessam Pirzadeh, et al. "Massive MIMO Channel Estimation With Low-Resolution Spatial Sigma-Delta ADCs". In: *IEEE Access* 9 (2021), pp. 109320–109334. DOI: 10.1109/ACCESS.2021.3101159.

[37] Amin Ghazanfari, Trinh Van Chien, Emil Björnson, et al. "Learning to Perform Downlink Channel Estimation in Massive MIMO Systems". In: *CoRR* abs/2109.02463 (2021). arXiv: 2109.02463. URL: https://arxiv.org/abs/2109.02463.

[38] Amin Ghazanfari, Trinh Van Chien, Emil Björnson, et al. "Model-Based and Data-Driven Approaches for Downlink Massive MIMO Channel Estimation". In: *IEEE Transactions on Communications* 70.3 (2022), pp. 2085–2101. DOI: 10.1109/TCOMM.2021.3133939.

[39] Italo Atzeni and Antti Tölli. "Channel Estimation and Data Detection Analysis of Massive MIMO With 1-Bit ADCs". In: *IEEE Transactions on Wireless Communications* 21.6 (2022), pp. 3850–3867. DOI: 10.1109/TWC.2021.3124709.

[40] Pengxiang Li, Yuehong Gao, Zhidu Li, et al. "A Pilot Assignment Scheme for Single-Cell Massive MIMO Circumstances". In: *Journal of Communications* 12 (Feb. 2017), pp. 91–97. DOI: 10.12720/jcm.12.2.91-97.

[41] Emil Björnson, Luca Sanguinetti, and Jakob Hoydis. "Hardware Distortion Correlation Has Negligible Impact on UL Massive MIMO Spectral Efficiency". In: *IEEE Transactions on Communications* 67.2 (Feb. 2019), 1085–1098. ISSN: 1558-0857. DOI: 10.1109/tcomm.2018.2877331. URL: http://dx.doi.org/10.1109/TCOMM.2018.2877331.

[42] Anastasios Papazafeiropoulos, Bruno Clerckx, and Tharmalingam Ratnarajah. "Rate-Splitting to Mitigate Residual Transceiver Hardware Impairments in Massive MIMO Systems". In: *IEEE Transactions on Vehicular Technology* 66.9 (2017), pp. 8196–8211. DOI: 10.1109/TVT.2017.2691014.

[43] "Further Advancements for E-UTRA Physical Layer Aspects (Release 9)". In: *3GPP Standard TS 36.814* (July 2017).

[44] Hao Shen. "Towards a Mathematical Understanding of the Difficulty in Learning With Feedforward Neural Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[45] Petro Liashchynskyi and Pavlo Liashchynskyi. "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS". In: *CoRR* abs/1912.06059 (2019). arXiv: 1912.06059. URL: http://arxiv.org/abs/1912.06059.

[46] Rao Asif, Mustafa Shakir, Ateeq Rehman, et al. "Performance Evaluation of Spectral Efficiency for Uplink and Downlink Multi-Cell Massive MIMO Systems". In: *Journal of Sensors* 2022 (June 2022). DOI: 10.1155/2022/7205687.

[47] Mohamed, Abdelfettah Belhabib, and Abdelhamid Riadi. "Encryption Based Strategy to Overcome the Problem of Pilot Contamination Within Multi-cellular Massive MIMO Systems". In: *Wireless Personal Communications* 119 (Oct. 2021), 2639–2655. DOI: 10.1007/s11277-021-08349-8.

[48] Minsoo Go. *E 15. [Recap] Activation Functions - EN — wikidocs.net*. https://wikidocs.net/196911. [Accessed 01-01-2024]. May 2023.

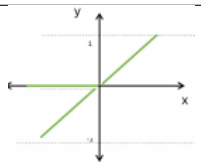
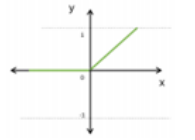
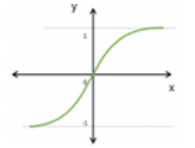
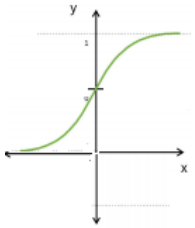
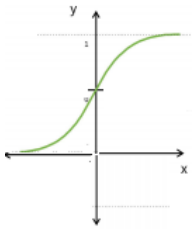
APPENDIX A

ACTIVATION FUNCTIONS SUMMARY

A.1 Deep Learning Activation Functions

The most commonly used activation functions in deep neural networks are *linear*, *relu*, *tanh*, *sigmoid*, and *softmax*.

TABLE A.1: Some activation functions and their characteristics for deep neural networks [48].

Activation Function	Mathematical model	Graph	Range
Linear	$f(x) = a \times x$		$(-\infty, \infty)$
Rectifier Linear Unit (ReLU)	$f(x) = max(0, x)$		$(0, \infty)$
Hyperbolic tangent (tanh)	$f(x) = tanh(x)$		$(-1, 1)$
Sigmoid (Logistic)	$f(x) = \frac{1}{(1+e^{-x})}$		$(0, 1)$
Softmax	$f(x) = \frac{e^{xi}}{\sum_{j=1}^{k}(e^{xj})}$		$(0, 1)$

APPENDIX B

DATASET GENERATION

Algorithm 1: Training dataset generation

Result: input training set, output training set

M = 100, K = 20, backOff = $10^{0.7}$, $\varphi = DFT$(K), training data size = $3,000,000$;

polyCoef = np.array([[$0.9798 - 0.0075j,\ -0.2419 + 0.0374j$]]);

$\tilde{b}$ = polyCoef/np.array([[1, backOff]]), ($\varsigma,\ \chi$) = moments($3, \tilde{b}$);

$p_{max} = 0.2$;

for *variable* r_1 *in range of (2)* **do**

  for *variable* r_2 *in range of (2)* **do**

    $B2_{r_1,r_2} = \tilde{b}_{0,r_1} \times (\tilde{b}_{0,r_2})^*$;

    for *variable* r_3 *in range of (2)* **do**

      $B3_{r_1,r_2,r_3} = \tilde{b}_{0,r_1} \times (\tilde{b}_{0,r_2})^* \times \tilde{b}_{0,r_3}$;

    end

  end

end

pilott2 = $\tilde{b}_{0,0} \times \varphi + \tilde{b}_{0,1} \times \varphi^3$;

for *variable* t *in range of (training data size)* **do**

  (G, $channelGONoise$, betaNLOS, GMean, powerCoef) = channelGen(M, K, 1, p_{max});

  η = (powerCoef/$\chi_{0,0}$).reshape(K,);

  β_k = (powerCoef/$||\tilde{b}_{0,0} + \tilde{b}_{0,1}||^2$).reshape(K,);

  $normalized\,gain = \beta_k \times$ powerCoef;

  $\tilde{a}$ = polyCoef/np.array([[1, backOff]]);

  $\tilde{a}_{:,1} = \tilde{a}_{:,1}$/np.sum($normalized\,gain \times powerCoef$);

  $effective\,channel$ = effectiveChannel3rd(M,K,$\tilde{a}$,$\tilde{b}$, np.matmul(G[:,:,0], np.diag($\sqrt{\eta}$)), ς);

  distortion = distCorr($\tilde{a}$, np.matmul(G[:,:,0], np.diag($\sqrt{\eta}$)), χ, 1, $effective\,channel$);

  U = np.matmul(pilott2, np.matmul(G[:,:,0], np.diag($((\beta_k)^{0.5})^T$);

  $Z = \tilde{a}_{:,0}\times$ U + $\tilde{a}_{:,1}\times$U^3;

  $Y = Z + \sqrt{0.5}\times$ (np.random(K,1) + $1j\times$np.random(K,1));

  inputs = np.matmul($\{\varphi\}^H, Y$);

  input training set [$t : t+1,\ 0$:K] = $\Re\{$(inputs)$^T\}$;

  input training set [$t : t+1,$ K: 2×K] = $\Im\{$(inputs)$^T\}$;

  input training set [$t : t+1,\ 2\times$K: 3×K] = $\{(normalized\,gain)^{0.5}\}$;

  output training set [$t : t+1,\ 0$:K] = $\Re\{effective\,channel\}$;

  output training set [$t : t+1,$ K: 2×K] = $\Im\{effective\,channel\}$;

  output training set [$t : t+1,\ 2\times$K] = $\log(\Re\{distortion\})$;

end

APPENDIX C

SYMBOL MOMENTS FOR CONSTELLATION

On the basis of the standard, let's consider circular shift and there exists the 90° for symmetric finite-sized constellation data signals.

$$B_{r_1,r_2} = \tilde{b}_{r_1}\tilde{b}^*_{r_2} \tag{C.1}$$

$$B_{r_1,r_2,r_3} = \tilde{b}_{r_1}\tilde{b}^*_{r_2}\tilde{b}_{r_3} \tag{C.2}$$

This works also for Gaussian data signals under symmetry circular shift of the 90°.

$$\mathbb{E}\{v_{i_1}v^*_{i_2}v_{i_3}\varsigma^*_k\} = \mathbb{E}\left\{|v_k|^2 v_k\varsigma^*_k\right\} \qquad \text{for} \quad i_1 = i_2 = i_3 = k \tag{C.3}$$

$$= \mathbb{E}\left\{\left|\tilde{b}_0\varsigma_k + \tilde{b}_1|\varsigma_k|^2\varsigma_k\right|^2\left(\tilde{b}_0|\varsigma_k|^2 + \tilde{b}_1|\varsigma_k|^4\right)\right\} \tag{C.4}$$

Let's combine Equations from C.2 to C.4 in accordance with the vector B and symbol moments.

$$\mathbb{E}\{v_{i_1}v^*_{i_2}v_{i_3}\varsigma^*_k\} = \begin{cases} \zeta_{10}B_{1,1,1} + 2\zeta_8 B_{1,1,0} + \zeta_8 B_{1,0,1} & \\ \quad + 2\zeta_6 B_{0,0,1} + \zeta_6 B_{0,1,0} + \zeta_4 B_{0,0,0}, & \text{if } i_1 = i_2 = i_3 = k \\ \zeta_6 B_{1,1} + \zeta_4 B_{1,0} + \zeta_4 B_{0,1} + B_{0,0}, & \text{if } i_1 = k \neq i_2 = i_3 \\ \zeta_6 B_{1,1} + \zeta_4 B_{1,0} + \zeta_4 B_{0,1} + B_{0,0}, & \text{if } i_1 = i_2 \neq i_3 = k \\ 0, & \text{otherwise.} \end{cases} \tag{C.5}$$

In Equation C.5, the middle two are the same because of the reason that the actual data symbols are independent corresponding to the UEs.