Ph.D. DISSERTATION

Design of Wireless Communications Systems with Deep Learning and Optimization

딥러닝과 최적화를 활용한 무선 통신 시스템 설계

BY

KIM NAMIK

AUGUST 2021

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Design of Wireless Communications Systems with Deep Learning and Optimization

딥러닝과 최적화를 활용한 무선 통신 시스템 설계

지도교수 이 광 복
이 논문을 공학박사 학위논문으로 제출함

2021년 8월

서울대학교 대학원

전기 컴퓨터 공학부

김 남 익

김남익의 공학박사 학위 논문을 인준함

2021년 8월

위 원 장: _____
부위원장: _____
위 원: _____
위 원: _____
위 원: _____

Abstract

With the advent of 5G wireless systems, ultra reliable low-latency communications (URLLC) and massive machine-type communications (mMTC) have recently attracted growing attention. Applications in health care, connected cars, robotics, manufacturing, and free-viewpoint video are expected in low-latency communications, and they demand extremely short round-trip latency levels as low as 1 ms. On the other hand, mMTC mainly concerns the massive connectivity of a large number of devices (e.g. sensors, robots, vehicles, and machines) to the base station (BS). Since conventional communications systems (e.g. Long-Term Evolution (LTE)) are difficult to meet the requirements of low-latency communications or mMTC, novel techniques suitable for these communications environments are required. This dissertation proposes three techniques for mMTC or low-latency communications.

In the first part of the dissertation, we propose a deep learning-based spreading sequence design and active user detection (AUD) to support mMTC where a large number of devices access the base station using non-orthogonal spreading sequences. To design the whole communications system minimizing AUD error, we employ an end-to-end deep neural network (DNN) where the spreading network models the transmitter side and the AUD network estimates active devices. By using the AUD error as a loss function, network parameters including the spreading sequences are learned to minimize the AUD error. Numerical results reveal that the spreading sequences obtained from the proposed approach achieve higher AUD performance than the conventional spreading sequences in the compressive sensing-based AUD schemes, as well as in the proposed AUD scheme.

In the second part of the dissertation, a precoding scheme to reduce the root mean square (RMS) delay spread of precoded channels in a orthogonal frequency division multiplexing (OFDM) system is proposed. In order to reduce latency in OFDM sys-

tems while not increasing the overhead, it is of primary importance to reduce the effective delay spread of the channel and thus the length of the cyclic prefix (CP). We formulate an optimization problem with an upper bound of the RMS delay spread as the objective function and a signal-to-noise ratio for each subcarrier as constraints. Semi-definite relaxation (SDR) technique is used to convert the problem into a convex problem so as to find the optimal precoding vector. Numerical results confirm that the proposed precoding design provides a significant reduction in the RMS delay spread, especially when there are a large number of antennas at the base station.

In the last part of the dissertation, we addresses linear precoding design for sum rate maximization in low-latency OFDM systems. In order to mitigate the overhead of CP originating from shortened symbol duration for low-latency communications, 5G wireless systems need to adopt short CP lengths. As channel delay spread must be less than the CP length, we first derive the effective RMS delay spread and the achievable rate using the zero-forcing assumption. We construct a sum rate optimization problem for each user subject to delay spread constraints and then convert the problem into a solvable convex problem along with a SDR technique. The precoding matrix is finally obtained by solving optimization problems for all users. Numerical results reveal that the proposed scheme attains superior performance to the conventional sum rate optimization, as well as small RMS delay spread.

keywords: Massive machine-type communications, low-latency communications, deep neural network, non-orthogonal multiple access, precoding

student number: 2014-22545

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

With the advent of 5G wireless systems, ultra reliable low-latency communications (URLLC) and massive machine-type communications (mMTC) have recently attracted growing attention [1]–[4]. Applications in health care, connected cars, robotics, manufacturing, and free-viewpoint video are expected in low-latency communications, and they demand extremely short round-trip latency levels as low as 1 ms [1], [2]. On the other hand, mMTC mainly concerns the massive connectivity of a large number of devices (e.g. sensors, robots, vehicles, and machines) to the base station (BS) [4]. Since low-latency communications and mMTC have different requirements from conventional communications systems (e.g. Long-Term Evolution (LTE)), novel techniques to support these communications systems are essential.

Since low-latency communications require 1-ms round-trip latency levels [1]–[3], wider subcarrier spacing up to 480 kHz has been considered for an orthogonal frequency division multiplexing (OFDM) system design [3]. As the larger subcarrier spacing leads to the shorter symbol duration, shorter cyclic prefix (CP), i.e., 1/4 or 1/8 of the normal CP, which is 4.7 μs in LTE, needs to be employed not to increase the CP overhead due to shortened symbol duration. On the other hand, in order to avoid inter-symbol interference (ISI) and inter-carrier interference (ICI), the CP length must be larger than the channel delay spread [5]. Hence, it is crucial to reduce the effective

channel delay spread to realize low-latency communications with minimal CP overhead [6].

In mMTC, it is difficult to support mMTC in the conventional scheduling-based orthogonal multiple access (OMA) due to the lack of resources and the heavy signaling overhead so that grant-free non-orthogonal multiple access (GF-NOMA) has received special attention in recent years [7]. Using non-orthogonal spreading sequences, GF-NOMA allows devices to transmit data without complicated scheduling. In order to support the grant-free transmission, the BS needs to identify active devices, i.e., devices transmitting data, among all potential devices. This process, often called *active user detection* (AUD), is an essential step for the successful data detection in GF-NOMA [8].

This dissertation consists of three parts. In Chapter 2, we propose a deep learning-based spreading sequence design and AUD scheme for an mMTC system. In Chapter 3 and Chapter 4, a precoding scheme is proposed to reduce the root mean square (RMS) delay spread of precoded channels in a multiple-input single-output (MISO)-OFDM system and to maximize the sum rate subject to RMS delay spread constraints in a multi-user multiple-input multiple-output (MIMO)-OFDM system, respectively.

1.1 Deep Learning-based Spreading Sequence Design and Active User Detection for Massive Machine-Type Communications

Since only a small portion of machine-type devices is active at a time (see Figure 1.1), a transmit vector consisting of data symbols of all devices is readily modeled as a sparse vector [9]. In solving the AUD problem, compressive sensing (CS) techniques have been widely used [8], [10]. Over the years, various techniques based on the CS-based AUD schemes have been proposed [8], [10]. In these approaches, basically, the correlation between the received signal and the spreading sequence of each device is used

to identify active devices. That is, a device whose spreading sequence is maximally correlated with the received signal is chosen in each iteration. Since the detection performance depends heavily on the correlation among spreading sequences, selection of the spreading sequences with low-correlation structure is crucial to improve the AUD performance.

Recently, some efforts have been made to design the low-correlation spreading sequences. In [11] and [12], iterative algorithms to find out the collection of spreading sequences having small correlation structure have been proposed. While these approaches improve the detection performance to some extent, they do not consider the mMTC context (e.g., the number of devices, channel models, and activities of devices) so that the performance gain in practical mMTC environments is marginal.

An aim of this dissertation is to propose a deep learning-based communications system suited for mMTC. The key idea of the proposed scheme is to use an end-to-end deep neural network (DNN) minimizing the AUD error [13]. In a nutshell, we divide the proposed DNN into two parts: the *spreading network* (SN) at the transmitter side (devices) to model device activities, channels, and symbol spreading and the *AUD network* (AUDN) at the receiver side (BS) to estimate active devices. Due to the use of the end-to-end network from the transmitter all the way to the receiver, we can directly use the AUD error as a loss function in the entire network training. As a result of the training, we can directly obtain the spreading sequences (training parameters in the SN) minimizing the AUD error.

From the numerical evaluations in grant-free mMTC scenarios, we show that the spreading sequences obtained from the proposed scheme perform better than those of the conventional scheme minimizing mutual coherence in terms of the AUD error probability. We also show that the spreading sequences obtained from the proposed scheme can be applied to the conventional CS-based AUD schemes.

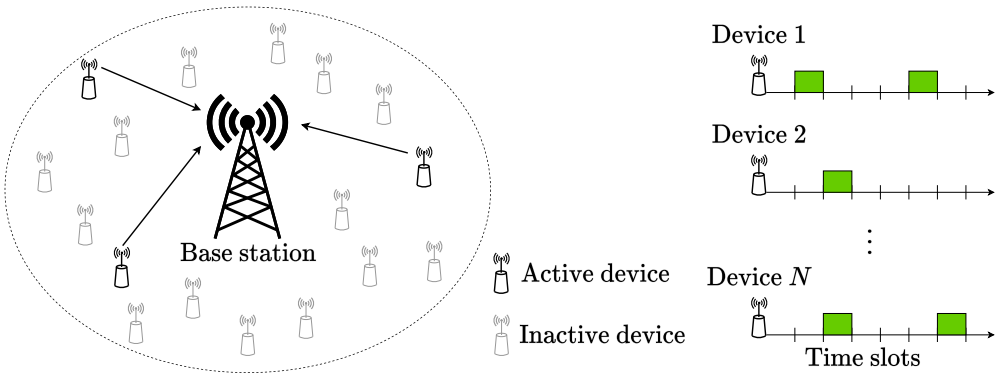

Figure 1.1: The illustration of the uplink mMTC scenario with N devices sporadically transmitting data symbols.

1.2 Precoding Design for Cyclic Prefix Overhead Reduction in a MISO-OFDM System

In recent years, a few preceding works have attempted to mitigate the effects of channel delay spread. In two studies [14] and [15], time-domain equalizers were designed to remove the residual ISI and ICI caused by the channel impulse response (CIR) exceeding the CP length by solving the Rayleigh quotient for the delay spread. In other work, to avoid the implementation of a complex equalizer on the receiver side, a precoding technique was employed [16], [17] as an alternative means of reducing the effective delay spread. In one such study in particular [17], a precoding matrix was designed to mitigate the effects of residual ISI and ICI for cases in which the channel delay spread is longer than the CP length. However, the performance of the precoding scheme in that case is degraded significantly as the excessive delay spread increases.

This dissertation presents a precoding design which reduces the RMS delay spread of precoded channels in a MISO-OFDM system. In contrast to previous approaches [16], [17] where precoding is used to mitigate the effect of an excessive delay spread for a fixed CP, the proposed precoding scheme is designed to proactively reduce the effective delay spread of the channel. As a result, the proposed design can reduce the CP length without causing the ISI and ICI. We derive the effective RMS delay spread and the signal-to-noise ratio (SNR) of each OFDM subcarrier as functions of the raw channel matrix and the precoding matrix. We then formulate a precoding optimization problem that minimizes the upper bound of the effective RMS delay spread while satisfying the SNR and power constraints. As the original problem is found to be a non-convex problem, we relax the problem using the semi-definite relaxation (SDR) technique presented in earlier work [18] into a convex problem to find the optimal solution.

1.3 Sum Rate Maximization with Shortened Cyclic Prefix in a MIMO-OFDM System

Although there are many works on sum rate maximization and CP overhead, respectively, both topics have not been simultaneously investigated for low-latency communications. Most papers on sum rate maximization have focused on linear precoding techniques owing to their practicability [19]–[22]. As the sum rate optimization generally demands the solution of non-convex problems, most of the preceding works attempted to find a local optimum through an alternative means of Karush-Kuhn-Tucker (KKT) conditions or a geometric program [19], [20]. Other precoding techniques imposed the zero-forcing condition to suppress multi-user interference and to decouple the multi-user channel into multiple independent subchannels [21], [22]. As the spectral efficiency decreases in proportion to the CP length for a given symbol duration, sum rate maximization must consider the CP overhead. To achieve both high spectral efficiency and low latency, several authors have pursued CP overhead reduction by shortening the effective channel delay spread or modifying OFDM frame structure [6], [23]. However, these works did not incorporate the CP overhead into the sum rate optimization problem.

This dissertation proposes a precoding scheme to maximize the sum rate subject to RMS delay spread constraints in a MIMO system. Contrary to earlier approaches [19]–[22] which solely concentrate on sum rate maximization, the proposed design incorporates the effective RMS delay spread into the optimization problem. In the same way as the previous works [21] and [22], we segregate the multi-user channel into multiple single-user channels by imposing zero-forcing conditions, which results in an optimization subproblem for each user. We examine the effective RMS delay spread and the rate as functions of the raw channel matrix and the precoding matrix. Analogous to other sum rate optimization in [19]–[22], the original optimization problem is found to be a non-convex problem. Using the SDR technique presented in [18], we

convert the problem into a convex problem and find the optimal solution for each user. Finally, the entire precoding matrix is constructed by composing precoding matrices for all users.

Chapter 2

DEEP LEARNING-BASED SPREADING SEQUENCE DESIGN AND ACTIVE USER DETECTION FOR MASSIVE MACHINE-TYPE COMMUNICATIONS

This work was in part presented at [24], where deep learning-aided spreading sequence design was proposed. In this dissertation, we refine the training and operation process and also supplement the performance analysis regarding AUD schemes. This work was accepted by the IEEE Wireless Communications Letters [25].

2.1 System Model

We consider the uplink transmission of the mMTC system where N devices access a single BS using spreading sequences of the length M (see Figure 1.1). We assume that the BS and all devices are equipped with one antenna and channels experience the flat fading. In most mMTC scenarios, the number of devices is larger than the length of spreading sequences (i.e., $M < N$) so that it is in general not possible to recover the transmit vector with the conventional recovery algorithm designed for overdetermined scenarios [26].

Let x_n and $\mathbf{s}_n \in \mathbb{R}^{M\times 1}$ be the symbol and spreading sequence of the n-th device,

respectively. Also, let $\boldsymbol{\delta} \in \mathbb{R}^{N\times 1}$ be the activity indicator vector (δ_n is 0 for an inactive device and 1 for an active device). The indicator δ_n of the n-th device follows the Bernoulli distribution:

$$\delta_n \sim Bern(p_n), \tag{2.1}$$

where p_n is the activity probability of the n-th device. Then, the received signal at the BS is given by

$$\begin{aligned}\mathbf{y} &= \sum_{n=1}^{N} \mathbf{s}_n h_n \delta_n x_n + \mathbf{w} \\ &= \mathbf{S}\mathbf{q} + \mathbf{w},\end{aligned} \tag{2.2}$$

where h_n is the channel between the n-th device and the BS, $\mathbf{S} = [\mathbf{s}_1 \ \cdots \ \mathbf{s}_N] \in \mathbb{R}^{M\times N}$ is the spreading matrix, $\mathbf{q} = [h_1\delta_1 x_1 \ \cdots \ h_N\delta_N x_N]$ is the composite vector of the symbols and the channels, and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is the additive white Gaussian noise (AWGN) vector. In this work, we consider the real-valued spreading sequences to make a fair comparison with the previous study using real values [12] but the extension to the complex scenarios is straightforward.

If the number of active devices, called *sparsity*, is K, then the AUD problem can be formulated as [27]

$$\tilde{\boldsymbol{\delta}} = \arg \min_{\|\boldsymbol{\delta}\|_0 = K} \|\mathbf{y} - \mathbf{S}\mathbf{q}\|_2^2 . \tag{2.3}$$

To solve this problem, greedy sparse recovery algorithm such as the orthogonal matching pursuit (OMP) has been employed [26], [28]. Since the AUD performance of the greedy algorithm relies heavily on the correlation among spreading sequences, previous studies focused on the minimization of the correlation between columns in the spreading matrix [11], [12]. In [12], for example, an approach to find out the quasi-orthogonal spreading sequences minimizing the mutual coherence has been proposed:

$$\min_{\mathbf{S}} \max_{1\le i\ne j\le N} \frac{|\langle \mathbf{s}_i, \mathbf{s}_j\rangle|}{\|\mathbf{s}_i\|_2 \|\mathbf{s}_j\|_2} \text{ and } \min_{\mathbf{S}} \left\|\mathbf{I} - \mathbf{S}^T\mathbf{S}\right\|_F^2 , \tag{2.4}$$

where $\|\cdot\|_F$ is the Frobenius norm. It has been shown that these approaches are effective in reducing the correlation among all spreading sequences [12]. However, since

the portion of active devices is very small (less than 10%) [29], it would be more effective to consider the device activities in the spreading sequence design. If we somehow minimize the correlation of spreading sequences for frequently active devices, multi-user interference can be reduced significantly, thereby achieving an improvement in the AUD performance.

2.2 DNN-based Spreading Sequence Design and Active User Detection

In order to design the end-to-end system (transmitter and receiver) minimizing AUD error, we employ the autoencoder-based DNN [13]. For the network training, we use the binary cross-entropy loss of AUD [30]:

$$\mathcal{L}(\boldsymbol{\delta}, \hat{\boldsymbol{\delta}}) = -\sum_{n=1}^{N} \Big(\delta_n \log(\hat{\delta}_n) + (1-\delta_n) \log(1-\hat{\delta}_n) \Big), \tag{2.5}$$

where $\hat{\boldsymbol{\delta}}$ is the estimated activity indicator vector and δ_n and $\hat{\delta}_n$ are the element of $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\delta}}$, respectively. Using the backpropagation mechanism, we can propagate the loss (the AUD error) all the way back to the transmitter and hence train the whole network parameters.

The basic structure of the proposed end-to-end network is illustrated in Figure 2.1. In essence, the proposed network consists of two subnetworks (SN and AUDN). The SN models the transmitter side (symbol spreading, device activities, and channels) and the AUDN estimates the activity indicator vector from the received signal. The estimated activity indicator vector generated from the end-to-end network is

$$\hat{\boldsymbol{\delta}} = g(\mathbf{x}; \boldsymbol{\Theta}), \tag{2.6}$$

where $\mathbf{x} = [x_1 \ \cdots \ x_N]$ is the input symbol vector, g is the mapping function between the input $\mathbf{x}$ and the output $\hat{\boldsymbol{\delta}}$ of the network, and $\boldsymbol{\Theta}$ is the set of all network parameters including the spreading sequences. Note that parameters in $\boldsymbol{\Theta}$ are updated during the

training phase using the stochastic gradient descent (SGD) algorithm:

$$\boldsymbol{\Theta}_j = \boldsymbol{\Theta}_{j-1} - \eta \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta}_{j-1}), \tag{2.7}$$

where $\boldsymbol{\Theta}_j$ is the parameters in the j-th training iteration, η is the learning rate, and $\nabla_{\boldsymbol{\Theta}} \mathcal{L}(\cdot)$ is the gradient of the loss function $\mathcal{L}$ [13]. After the training phase, the spreading sequences and AUD scheme are obtained from the trained end-to-end network (see Figure 2.1).

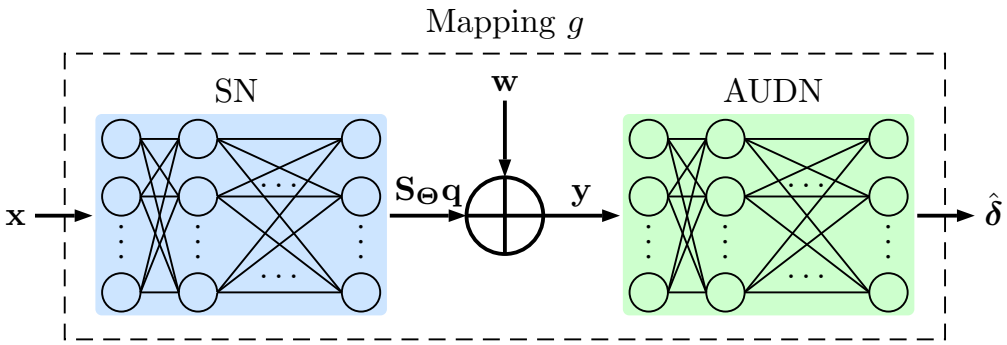

Figure 2.1: The structure of the end-to-end network.

2.2.1 SN Architecture

In Figure 2.2, we depict the SN structure in the transmitter. Instead of using the deterministic spreading sequences, we assign the trainable vector $\mathbf{s}'_n \in \mathbf{\Theta}$ to the n-th device ($n = 1, \cdots, N$). Hence, the spreading sequences are a part of the network parameters to be updated by the SGD algorithm. To model the device activities, the symbol x_n is multiplied by the activity indicator δ_n which follows the Bernoulli distribution, i.e. $\delta_n \sim Bern(p_n)$.

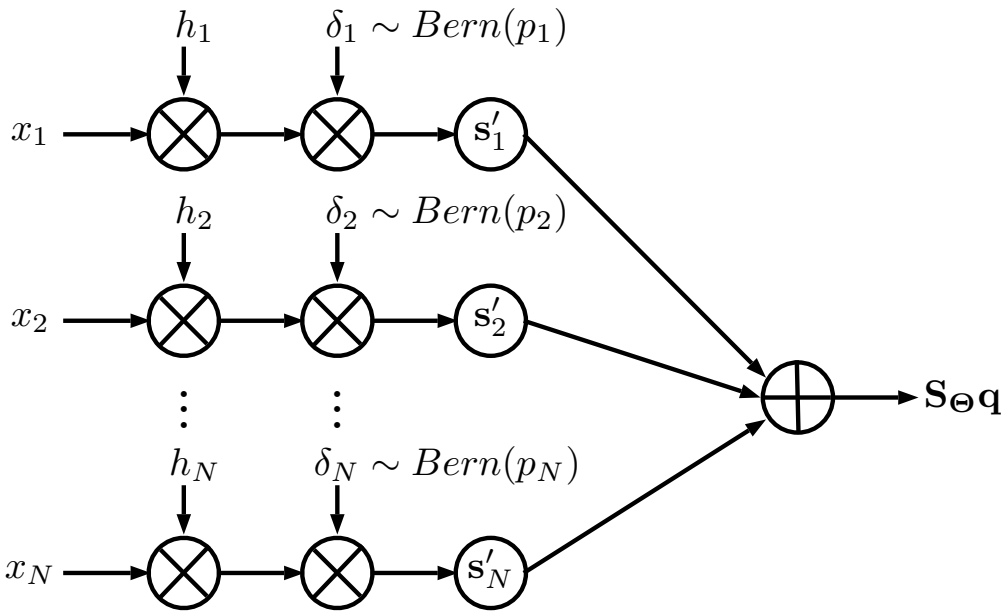

Figure 2.2: The detailed structure of the SN in the training phase.

2.2.2 AUDN Architecture

The main goal of the AUDN is to identify the activity indicator vector $\boldsymbol{\delta}$ from $\mathbf{y}$. Figure 2.3 depicts the structure of the AUDN. Because only a small number of machine-type devices are active at a time, the activity indicator vector can be well modeled as a sparse vector [9]. In solving this sparse recovery problem, we employ an iterative hard thresholding-based network (IHT-Net), a DNN-based sparse recovery algorithm based on iterative hard thresholding (IHT) [31]:

$$\min_{\|\mathbf{q}\|_0=K} \|\mathbf{y}-\mathbf{S}\mathbf{q}\|_2^2 . \tag{2.8}$$

In the IHT algorithm, the sparse vector $\mathbf{q}$ is iteratively updated as

$$\mathbf{q}^{(t+1)} = H_K[(\mathbf{I}_N - \mathbf{S}^T\mathbf{S})\mathbf{q}^{(t)} + \mathbf{S}^T\mathbf{y}], \tag{2.9}$$

where $\mathbf{q}^{(t)}$ is the estimate after t iterations and $H_K[\cdot]$ is the hard thresholding operator to enforce the sparsity K of the output vector. Similar to the IHT algorithm, one iteration in the IHT-Net is mapped to the neural layer of a DNN. Specifically, the t-th layer of the IHT-Net can be expressed as

$$\boldsymbol{\delta}^{(t+1)} = \mathrm{ReLU}[\boldsymbol{\Psi}^{(t)}\boldsymbol{\delta}^{(t)} + \boldsymbol{\beta}^{(t)}], \tag{2.10}$$

where $\boldsymbol{\delta}^{(t)}$ and $\boldsymbol{\delta}^{(t+1)}$ are the input and the output of the layer and $\boldsymbol{\Psi}^{(t)} \in \boldsymbol{\Theta}$ and $\boldsymbol{\beta}^{(t)} \in \boldsymbol{\Theta}$ are the trainable weight and bias of the layer. Note that the hard thresholding operator in the IHT is replaced by the rectified linear unit (ReLU).

In order to improve the accuracy of AUD when a massive number of devices are used, we adjust the number of nodes based on the number of devices. It is well-known from the *universal approximation theorem* that a DNN can approximate the desired function, provided that sufficiently many hidden nodes are available [32]. This implies that the AUDN with enough nodes can carry out the accurate AUD process even when the number of devices is large. In light of this, we set the width of layers in proportion to the number of devices N (e.g., $\boldsymbol{\delta}^{(t)} \in \mathbb{R}^{5N\times 1}$).

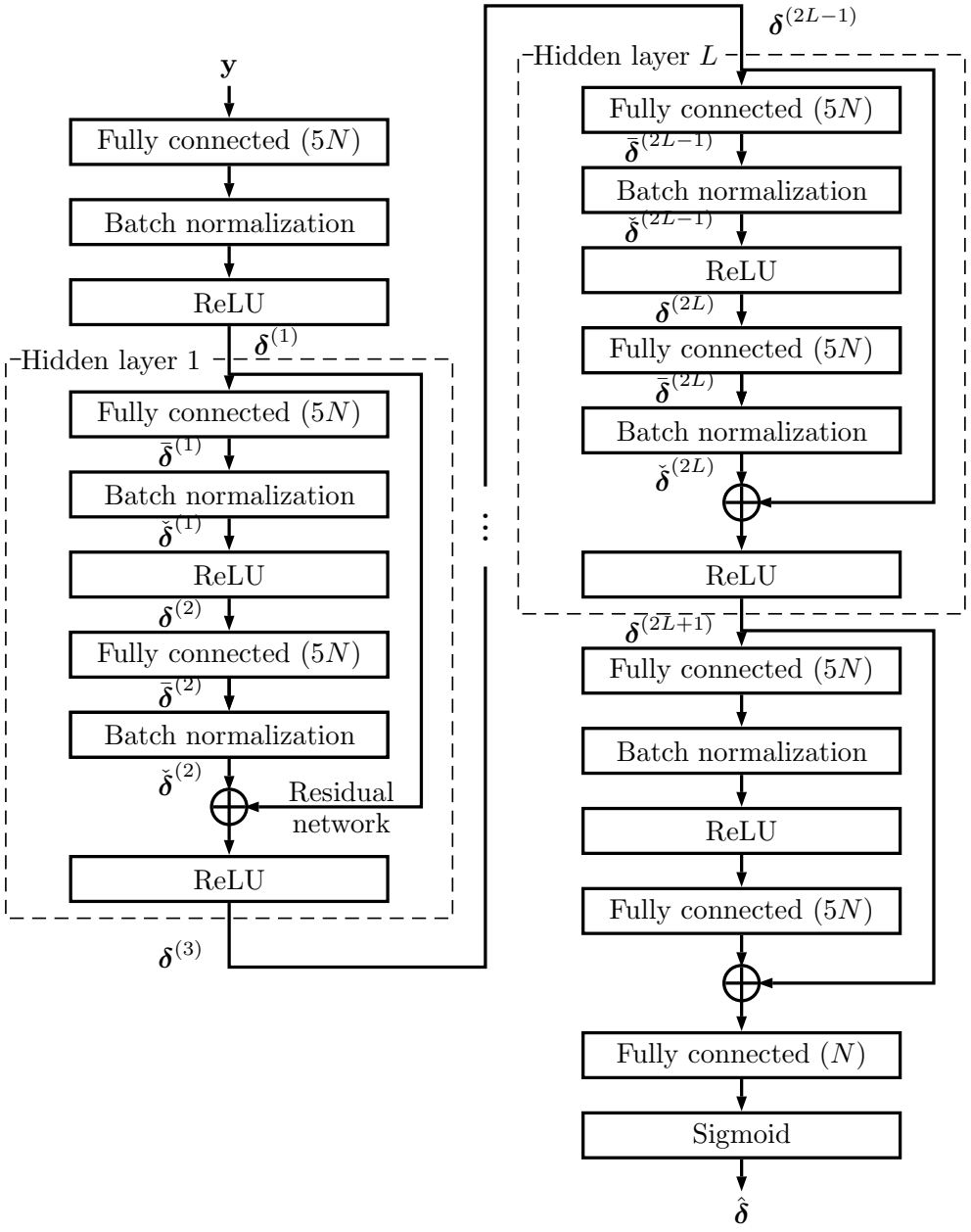

Figure 2.3: The detailed structure of the AUDN.

In this work, we exploit the batch normalization (BN) and residual networks (ResNets) to handle the problem occurring in the training phase [33], [34]. Since devices have different activities in mMTC, a variation of the received signal (the sum of all device signals) is large, which slows down the training speed and also hinders convergence because the network should handle a large variation of input data. In order to mitigate this, we add the BN layer to the data $\bar{\boldsymbol{\delta}}^{(t)} = \boldsymbol{\Psi}^{(t)}\boldsymbol{\delta}^{(t)} + \boldsymbol{\beta}^{(t)} = [\bar{\delta}_1^{(t)} \ \cdots \ \bar{\delta}_{5N}^{(t)}]$ so that the normalized data $\check{\delta}_1^{(t)}, \cdots, \check{\delta}_{5N}^{(t)}$ have zero means and unit variances for each batch size B:

$$\check{\delta}_i^{(t)} = \frac{\bar{\delta}_i^{(t)} - \mu_{B,i}}{\sigma_{B,i}}, \quad \text{for } i = 1, \cdots, 5N, \tag{2.11}$$

where $\check{\boldsymbol{\delta}}^{(t)} = [\check{\delta}_1^{(t)} \ \cdots \ \check{\delta}_{5N}^{(t)}]$ is the output of the BN layer and $\mu_{B,i} = \frac{1}{B}\sum_{b=1}^{B} \bar{\delta}_i^{(t)[b]}$ and $\sigma_{B,i} = \frac{1}{B}\sum_{b=1}^{B}(\bar{\delta}_i^{(t)[b]} - \mu_{B,i})^2$ are the mini-batch mean and variance. Since a large number of layers are used in the AUDN, the entire network suffers from the vanishing gradient problem. Note that as the gradient propagates backward all the way to the SN, new gradient obtained by multiplying local gradients gets smaller, making it difficult to update the weight of early layers (i.e., layers of the SN). In order to alleviate this so-called *vanishing gradient problem*, we exploit the ResNet architecture that adds the direct links between stacked layers. The t-th layer in the ResNet is given by

$$\boldsymbol{\delta}^{(t+1)} = \mathrm{ReLU}[\check{\boldsymbol{\delta}}^{(t)} + \boldsymbol{\delta}^{(t-1)}]. \tag{2.12}$$

Since the layers are directly connected by the ResNet (see Figure 2.3), the gradients can be transferred across the layers with much less distortion, resulting in an improvement in the training accuracy.

2.2.3 Operation

After the training phase, the trained parameters in the SN are used as the spreading sequences. As a result of the trained AUDN, the estimated activity indicator $\hat{\delta}_n$ is mapped to the unit interval $(0, 1)$ through the sigmoid activation ($f(x) = \frac{1}{1+e^{-x}}$). Finally, if $\hat{\delta}_n$ is greater than the threshold α, we declare the n-th device as an active

device. The threshold value α minimizing the AUD error is given by

$$\alpha^* = \arg\min_{\alpha}(|\{(b,n) \mid \hat{\delta}_n^{[b]} \leq \alpha \text{ if } \delta_n^{[b]} = 1\}| + |\{(b,n) \mid \hat{\delta}_n^{[b]} > \alpha \text{ if } \delta_n^{[b]} = 0\}|), \tag{2.13}$$

where b is the data index [35].

2.3 Numerical Results

2.3.1 Simulation Setup

We simulate the uplink mMTC system with 64 devices ($N = 64$). The length of spreading sequences is set to $M = 32$ and each spreading sequence $\mathbf{s}_n$ is normalized (i.e., $\|\mathbf{s}_n\|_2 = 1$). An AWGN channel model with the noise variance σ^2 is assumed for the sake of simplicity. The average symbol SNR is set to $1/\sigma^2$. We test two scenarios: activity probabilities are the same and different. We henceforth call these cases *homogeneous* and *heterogeneous activities*, respectively.

For comparison, we employ the conventional spreading sequences in [12] and the Gaussian random spreading sequences in which the elements follow the normal distribution $\mathcal{N}(0,1)$. To evaluate the AUD performance, AUDN is used at the receiver and is trained with the same training data for each set of spreading sequences. As an AUD performance measure, we use the activity error rate (AER) considering both missed detection and false alarms: $\text{AER} = 1 - \frac{|\boldsymbol{\Omega} \cap \hat{\boldsymbol{\Omega}}|}{|\boldsymbol{\Omega} \cup \hat{\boldsymbol{\Omega}}|}$ where $\boldsymbol{\Omega}$ and $\hat{\boldsymbol{\Omega}}$ are the support[1] of $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\delta}}$, respectively.

Data samples are constructed by combining symbols and activity indicators determined by the activity probabilities (see Figure 2.2). In the training phase, data samples with SNR in the range of 15 dB to 20 dB are used. We use 480,000 samples with 40 epochs for training, 60,000 samples for validation, and 60,000 samples for testing [31]. We employ an Adam optimizer for the SGD optimization. In the simulations,

[1] If $\boldsymbol{\delta} = [1\ 0\ 1\ 0]$, then the support is $\boldsymbol{\Omega} = \{1, 3\}$.

the number of hidden layers, the batch size, and the threshold value are set to $L = 10$, $B = 200$, and $\alpha = 0.4$, respectively. The learning rate η starts from 0.01 and is divided by 10 for every 10 epochs.

2.3.2 Homogeneous Activities

In the homogeneous activity scenario, the activity probabilities for all devices are the same. Figure 2.4 presents the average AER with different spreading sequences as a function of p_n at SNR $= 20$ dB. The average AER of the Gaussian random sequences is much larger due to the high correlation among the spreading sequences. On the other hand, the average AER of the proposed sequences is lower than that of the conventional sequences for each activity probability p_n.

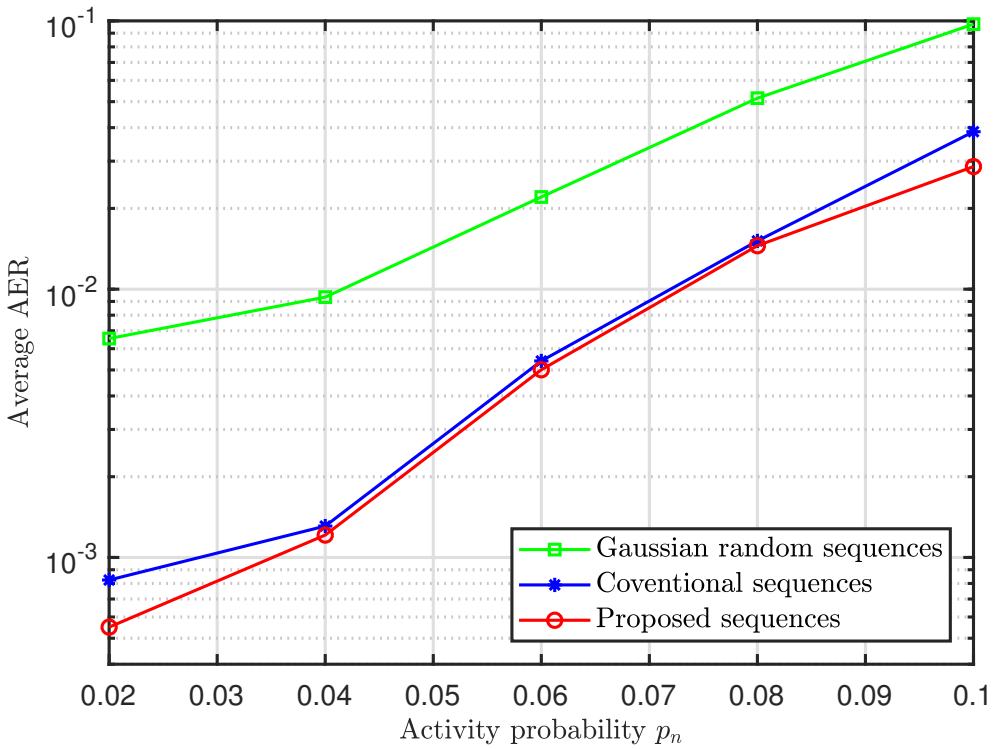

Figure 2.4: Average AER of the AUDN as a function of the activity probability at SNR $= 20$ dB

2.3.3 Heterogeneous Activities

In this subsection, we examine whether the proposed approach can generate activity-specific spreading sequences and thus improve the AUD performance when the activity probabilities are different.

In Figure 2.5, we evaluate the average AER of the AUDN as a function of SNR. The activity probability p_n is modeled by the uniform distribution on the interval of [0.01, 0.2] for scenario 1 and [0.01, 0.1] for scenario 2. We observe that the performance of the proposed sequences is significantly better than that of the conventional sequences. For example, the proposed sequences achieve about 3 dB gain in scenario 1 and 0.5 dB gain in scenario 2 over the conventional sequences at high SNR.

In Figure 2.6, we plot the average cross-correlation of each device as a function of p_n. The average cross-correlation of the i-th device is defined as

$$\mu_{i,avg} = \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} |\langle \mathbf{s}_i, \mathbf{s}_j \rangle|. \tag{2.14}$$

If $\mu_{i,avg}$ is small, the spreading sequence of the i-th device is less correlated with other spreading sequences. In terms of the AUD performance, it is desirable to have the spreading sequences such that the cross-correlation between any two active devices is as small as possible [8]. We observe that the proposed approach assigns the spreading sequences with lower $\mu_{i,avg}$ to the devices with higher activity probabilities. Since the average cross-correlation of devices with higher activity probabilities is reduced, active devices are less correlated, thereby improving the performance of the proposed approach in Figure 2.5.

Figure 2.7 depicts the average AER of the OMP and compressive sampling matching pursuit (CoSaMP) in the heterogeneous activity scenario [36]. When compared to Figure 2.5, we observe that the AUDN outperforms the greedy algorithms by a large margin (e.g. more than 4 dB gain for scenario 1). Further, the performance gap between the conventional and the proposed sequences becomes more severe in a wide range of SNR because the greedy algorithms depend heavily on the correlation among

spreading sequences. In the CoSaMP, for example, the proposed sequences achieve about 2.9 dB gain in scenario 1 and 1.3 dB gain in scenario 2 over the conventional sequences.

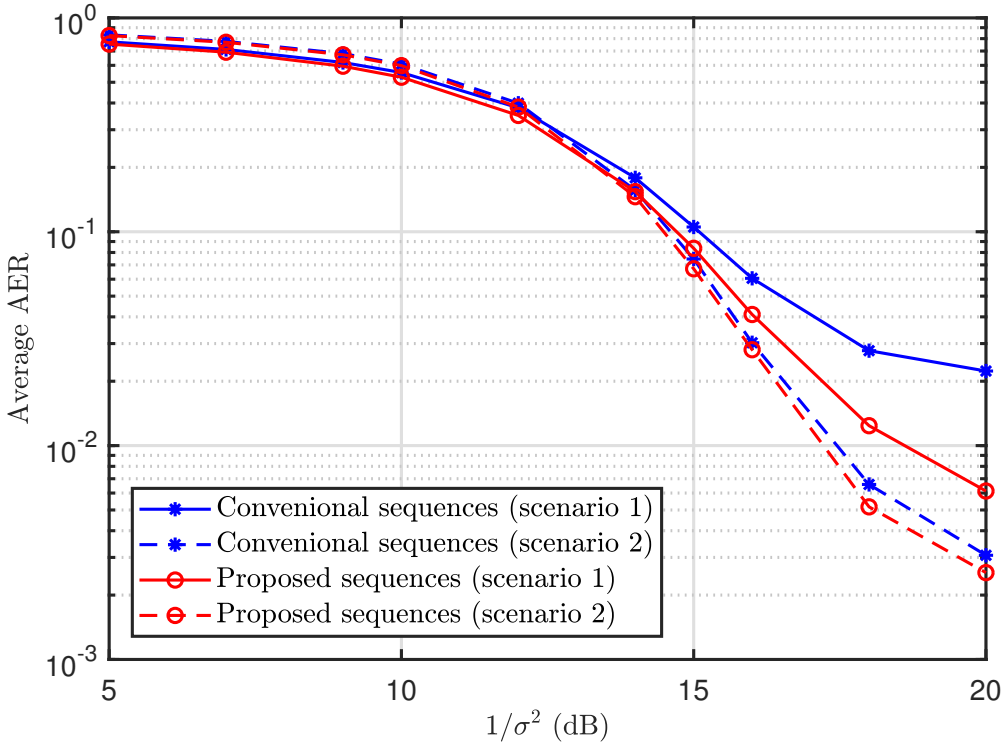

Figure 2.5: Average AER of the AUDN as a function of SNR in the heterogeneous activity scenario.

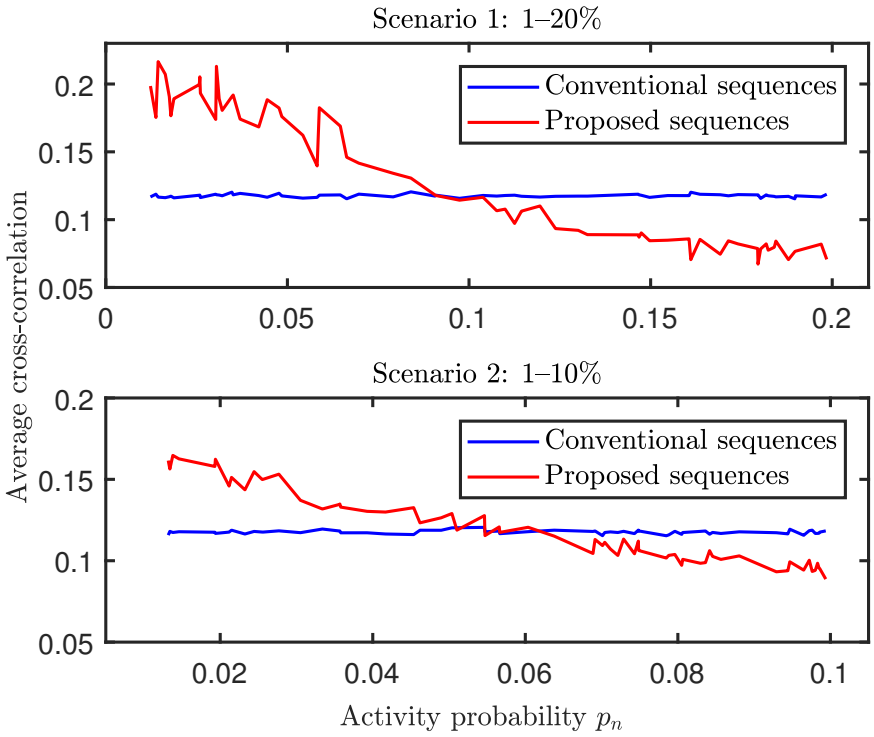

Figure 2.6: Average cross-correlation of each device as a function of the activity probability in the heterogeneous activity scenario.

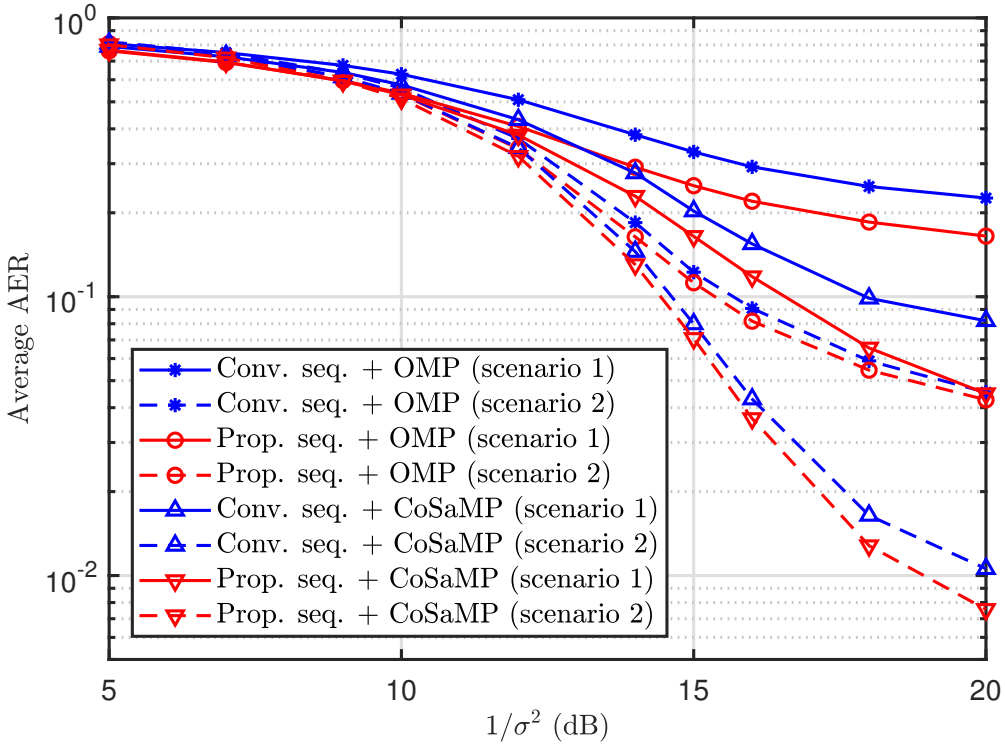

Figure 2.7: Average AER of the OMP and CoSaMP as a function of SNR in the heterogeneous activity scenario.

Chapter 3

PRECODING DESIGN FOR CYCLIC PREFIX OVERHEAD REDUCTION IN A MISO-OFDM SYSTEM

This work was published in the IEEE Wireless Communications Letters [6].

3.1 System Model

We consider the downlink transmission of a MISO-OFDM system with M antennas at the base station and N subcarriers. It is assumed that each user estimates the frequency-selective channel and feeds it back to the base station. The frequency response for the n-th subcarrier of the channel between the m-th transmit antenna and a certain user can be represented as

$$\begin{aligned} H_m(n) = \sum_{\ell=0}^{N-1} h_m(\ell) e^{-\frac{j2\pi n\ell}{N}} = \sum_{\ell=0}^{L-1} h_m(\ell) e^{-\frac{j2\pi n\ell}{N}}, \\ m = 1, 2, \cdots, M, \quad n = 0, 1, \cdots, N-1, \end{aligned} \tag{3.1}$$

where $h_m(\ell)$ is the CIR for the m-th antenna and L ($\leq N$) denotes the length of the CIR, i.e., $h_m(L) = h_m(L+1) = \cdots = h_m(N-1) = 0$. The aggregate channel vector for the n-th subcarrier can be defined as

$$\bar{H}(n) \triangleq \begin{bmatrix} H_1(n) & H_2(n) & \cdots & H_M(n) \end{bmatrix}^T. \tag{3.2}$$

Let $\bar{W}_m(n)$ denote the precoding vector associated with the channel vector $\bar{H}(n)$, which is defined as

$$\bar{W}(n) \triangleq \begin{bmatrix} W_1(n) & W_2(n) & \cdots & W_M(n) \end{bmatrix}^T. \tag{3.3}$$

Accordingly, the composite channel for the n-th subcarrier encompassing the precoding vector can be written as

$$H_{eq}(n) = \sum_{m=1}^{M} H_m(n) W_m(n) = \bar{H}(n)^T \bar{W}(n). \tag{3.4}$$

By aggregating the composite channels for all subcarriers, an equivalent channel matrix is formed as

$$\begin{aligned} \vec{H}_{eq} &\triangleq \begin{bmatrix} H_{eq}(0) \\ H_{eq}(1) \\ \cdots \\ H_{eq}(N-1) \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \bar{H}(0)^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{H}(1)^T & \mathbf{0} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{H}(N-1)^T \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \bar{W}(0) \\ \bar{W}(1) \\ \cdots \\ \bar{W}(N-1) \end{bmatrix}}_{\vec{W}}, \end{aligned} \tag{3.5}$$

where $\mathbf{0}$ is the $1 \times M$ all-zero vector.

3.2 Precoding Design

In this section, the proposed precoding design is presented. In Subsection 3.2.1, we express the effective RMS delay spread and the SNR of each subcarrier as functions of the raw channel matrix and the precoding matrix. In Subsection 3.2.2, we formulate a precoding optimization problem to minimize an upper bound of the effective RMS delay spread of the channel while satisfying an SNR constraint for each subcarrier. A method based on the SDR technique is also discussed to find the optimal solution.

3.2.1 Effective RMS Delay Spread and SNR

The effective RMS delay spread of the equivalent channel in the OFDM system is calculated as [14]

$$\tau_{RMS} = \sqrt{\frac{1}{E}\sum_{\ell=0}^{N-1}\left(\ell-\bar{\ell}\right)^2|h_{eq}\left(\ell\right)|^2}, \tag{3.6}$$

where $h_{eq}\left(\ell\right)$ is the equivalent CIR corresponding to $H_{eq}\left(n\right)$, and E and $\bar{\ell}$ are, respectively, given as

$$E = \sum_{\ell=0}^{N-1}|h_{eq}\left(\ell\right)|^2 \tag{3.7}$$

and

$$\bar{\ell} = \frac{1}{E}\sum_{\ell=0}^{N-1}\ell|h_{eq}\left(\ell\right)|^2. \tag{3.8}$$

Assuming that $\bar{\ell}$ is fixed, we represent the discrete Fourier transform (DFT) of $\ell-\bar{\ell}$ as

$$F\left(n\right) = \sum_{\ell=0}^{N-1}\left(\ell-\bar{\ell}\right)e^{-\frac{j2\pi n\ell}{N}}. \tag{3.9}$$

Using the convolution property and Parseval's relation [37], τ_{RMS}^2 can be expressed in terms of $H_{eq}\left(n\right)$ and $F\left(n\right)$ as

$$\tau_{RMS}^2 = \frac{\sum\limits_{n=0}^{N-1}\left|\frac{1}{N}\left(F\left(n\right)\circledast H_{eq}\left(n\right)\right)\right|^2}{\sum\limits_{n=0}^{N-1}\left|H_{eq}\left(n\right)\right|^2}, \tag{3.10}$$

where $\circledast$ denotes the circular convolution. Consequently, N convolutions can be represented with a matrix multiplication as

$$\frac{1}{N}\begin{bmatrix} F\left(0\right)\circledast H_{eq}\left(0\right) \\ F\left(1\right)\circledast H_{eq}\left(1\right) \\ \cdots \\ F\left(N-1\right)\circledast H_{eq}\left(N-1\right)\end{bmatrix} = \mathbf{F}\vec{H}_{eq}, \tag{3.11}$$

where

$$\mathbf{F} \triangleq \frac{1}{N}\begin{bmatrix} F(0) & F(N-1) & \cdots & F(1) \\ F(1) & F(0) & F(N-1) & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ F(N-1) & \cdots & F(1) & F(0) \end{bmatrix} \tag{3.12}$$

is an $N \times N$ matrix. Substituting (3.5) and (3.11) into (3.10), τ_{RMS}^2 can be rewritten as

$$\tau_{RMS}^2 = \frac{\vec{H}_{eq}^H \mathbf{F}^H \mathbf{F} \vec{H}_{eq}}{\vec{H}_{eq}^H \vec{H}_{eq}} = \frac{\vec{W}^H \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} \vec{W}}{\vec{W}^H \mathbf{H}^H \mathbf{H} \vec{W}}. \tag{3.13}$$

On the other hand, the SNR in the n-th subcarrier is defined with the composite channel $H_{eq}(n)$ and the noise variance σ^2 as [38]

$$\mathbf{SNR}(n) = \frac{|H_{eq}(n)|^2}{\sigma^2}. \tag{3.14}$$

The SNR in (3.14) can be transformed into a matrix form analogous to (3.13) as

$$\mathbf{SNR}(n) = \frac{\vec{H}_{eq}^H \mathbf{E}_n \vec{H}_{eq}}{\sigma^2} = \frac{\vec{W}^H \mathbf{H}^H \mathbf{E}_n \mathbf{H} \vec{W}}{\sigma^2}, \tag{3.15}$$

where the element on the i-th row and j-th column of the $N \times N$ matrix $\mathbf{E}_n$ is given as

$$(\mathbf{E}_n)_{ij} = \begin{cases} 1, & i = j = n+1, \\ 0, & \text{otherwise}. \end{cases} \tag{3.16}$$

3.2.2 Precoding Optimization

From (3.15), an SNR constraint for the n-th subcarrier can be imposed as

$$\mathbf{SNR}(n) = \frac{\vec{W}^H \mathbf{H}^H \mathbf{E}_n \mathbf{H} \vec{W}}{\sigma^2} \geq d_n, \tag{3.17}$$

where d_n denotes the minimum required SNR for the n-th subcarrier. Moreover, the denominator of τ_{RMS}^2 in (3.13) can be lower-bounded as

$$\vec{W}^H \mathbf{H}^H \mathbf{H} \vec{W} = \sigma^2 \sum_{n=0}^{N-1} \mathbf{SNR}(n) \geq \sigma^2 \sum_{n=0}^{N-1} d_n. \tag{3.18}$$

Equivalently, the upper bound of τ_{RMS}^2 can be derived as

$$\tau_{RMS}^2 \leq \frac{\vec{W}^H \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} \vec{W}}{\sigma^2 \sum_{n=0}^{N-1} d_n}. \tag{3.19}$$

To make the optimization problem tractable, we use the upper bound in (3.19) instead of τ_{RMS}^2 as the objective function. As a result, an optimization problem that minimizes the upper bound in (3.19) with the SNR constraints in (3.17) and a transmit power constraint can be established as

$$\begin{aligned} \min_{\vec{W}} \quad & \vec{W}^H \mathbf{S} \vec{W} \\ \text{s.t.} \quad & \mathbf{SNR}(n) = \frac{\vec{W}^H \mathbf{S}_n \vec{W}}{\sigma^2} \geq d_n, \ n = 0, 1, \cdots, N-1, \\ & \left\| \vec{W} \right\|^2 = P, \end{aligned} \tag{3.20}$$

where $\mathbf{S} \triangleq \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H}$, $\mathbf{S}_n \triangleq \mathbf{H}^H \mathbf{E}_n \mathbf{H}$, and P denotes the total transmit power at the base station. Note that only the numerator of the upper bound is used in (3.20), since the denominator is constant irrespectively of $\vec{W}$.

The problem in (3.20) is difficult to solve as it is a non-convex problem. To tackle this, we use an SDR technique [18]. From the property of trace, we know that $\vec{W}^H \mathbf{S} \vec{W} = \mathrm{tr}(\vec{W}^H \mathbf{S} \vec{W}) = \mathrm{tr}(\mathbf{S} \vec{W} \vec{W}^H)$. We also define $\mathbf{W} = \vec{W} \vec{W}^H$, which is a rank-one Hermitian positive semi-definite matrix. The optimization problem in (3.20) can then be rewritten as

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}(\mathbf{S}\mathbf{W}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{S}_n \mathbf{W}) \geq \sigma^2 d_n, \ n = 0, 1, \cdots, N-1, \\ & \mathrm{tr}(\mathbf{W}) = P, \\ & \mathbf{W} \succeq 0, \\ & \mathrm{rank}(\mathbf{W}) = 1. \end{aligned} \tag{3.21}$$

Note that the objective function and constraints apart from the rank constraint are convex in (3.21). By omitting the rank constraint, we obtain the following relaxed

optimization problem,

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}\left(\mathbf{S}\mathbf{W}\right) \\ \text{s.t.} \quad & \mathrm{tr}\left(\mathbf{S}_n\mathbf{W}\right) \geq \sigma^2 d_n,\ n = 0, 1, \cdots, N-1, \\ & \mathrm{tr}\left(\mathbf{W}\right) = P, \\ & \mathbf{W} \succeq 0, \end{aligned} \tag{3.22}$$

which becomes a convex optimization problem that can be solved using standard convex tools, such as CVX in MATLAB. Once we determine $\mathbf{W}$, the precoding vector $\vec{W}$ can be derived from the best rank-one approximation. Specifically, provided that the optimal solution of (3.22) is $\mathbf{W}_{opt}$, the eigen-decomposition of $\mathbf{W}_{opt}$ is given as

$$\mathbf{W}_{opt} = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^H, \tag{3.23}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are the eigenvalues of $\mathbf{W}_{opt}$, while $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_r$ are the corresponding eigenvectors, and $r = \mathrm{rank}\left(\mathbf{W}_{opt}\right)$. Given that the best rank-one approximation to $\mathbf{W}_{opt}$ is $\lambda_1 \mathbf{q}_1 \mathbf{q}_1^H$, $\sqrt{\lambda_1}\mathbf{q}_1$ is regarded as the solution of the problem (3.20) [18].

3.3 Numerical Results

We evaluate the performance of the proposed precoding design. The number of antennas at the base station, M, is set to 2, 4, 8, or 16. The number of OFDM subcarriers, N, is fixed at 16 and the length of the CIR, L, is fixed at 8. The CIR $\{h_m\left(0\right), h_m\left(1\right), \cdots, h_m\left(L-1\right)\}$ for the m-th antenna ($m = 1, 2, \cdots, M$) is assumed to follow a tapped delay line model, as in earlier research [16],

$$h_m\left(\ell\right) = \frac{a_{\ell,m}}{\sqrt{\tau}} \exp\left(-\frac{\ell}{2\tau}\right), \quad \ell = 0, 1, \cdots, L-1, \tag{3.24}$$

where $a_{\ell,m}$ is a complex normal random variable and τ follows a log-normal distribution with log standard deviation of 10 dB. In Figs. 1 and 2, the original channel and the

precoded channel represent the channel before and after precoding is applied, respectively. The minimum SNR requirement d_n is set according to the SNR corresponding to the maximum ratio transmission (MRT) precoding [39]; i.e., $W_m(n) = H_m^{\dagger}(n)$, where the superscript † denotes the complex conjugate. Accordingly, we set the SNR constraints as

$$d_n = \alpha \mathbf{SNR}_{MRT}(n), \tag{3.25}$$

where $\mathbf{SNR}_{MRT}(n)$ denotes the SNR corresponding to MRT precoding and the parameter $\alpha \in [0, 1]$ takes a real value. Note that less SNR loss is allowed with α closer to 1.

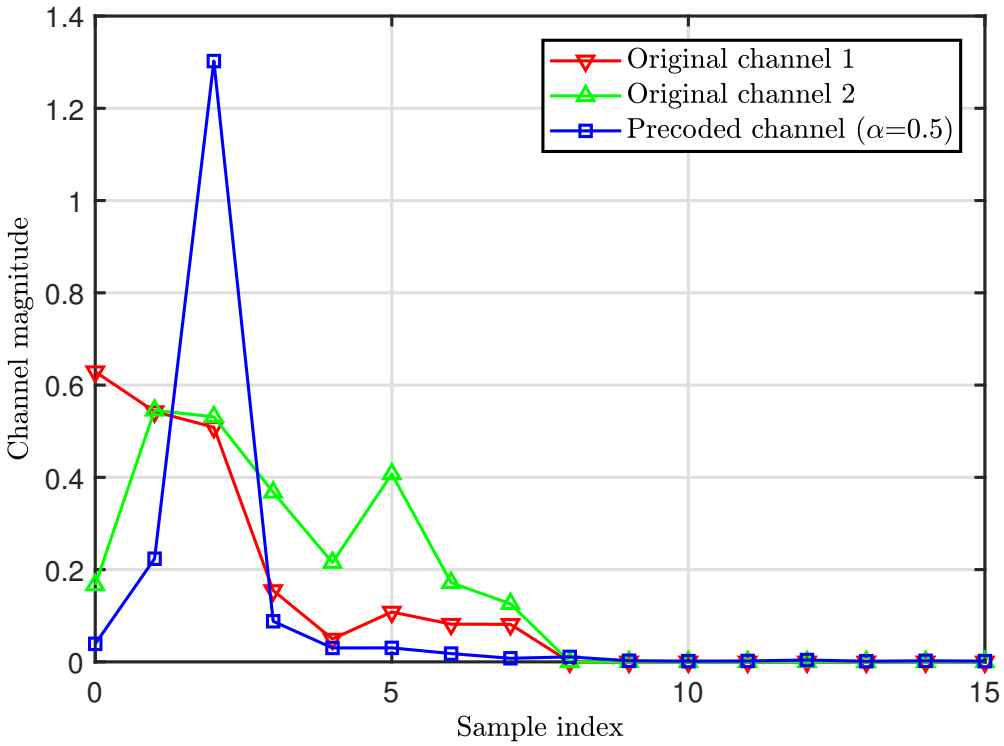

Figure 3.1: CIR of the original channels and precoded channel, when $M = 2$ and $\alpha = 0.5$.

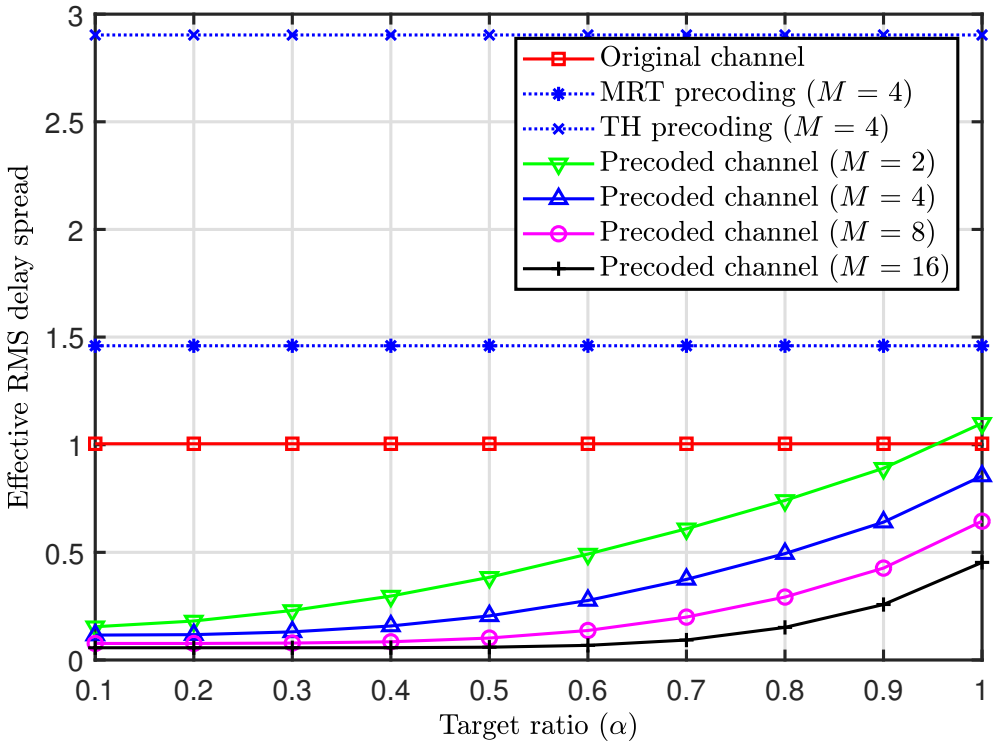

Figure 3.2: Average effective RMS delay spreads versus α for several values of M.

Table 3.1: Average Required CP Length and the Corresponding CP Overhead, when $M = 4$ and $\alpha = 0.5$

L	4	8	12
Original channel	3.73 (18.9%)	7.23 (31.1%)	9.20 (36.5%)
Precoded channel	0.79 (4.7%)	1.33 (7.7%)	1.50 (8.6%)

Figure 1 depicts the CIR for one specific channel realization with and without the proposed precoding scheme, when $M = 2$ and $\alpha = 0.5$. This figure shows that the CIR of the precoded channel has much less spread than the original channels for the two antennas. In fact, the RMS delay spreads of the original channels are 1.17 and 1.66 for the two antennas, whereas that of the precoded channel is 0.23.

Figure 2 shows how the effective RMS delay spreads averaged over 500 independent channel realizations vary with α for several values of M. For comparison purpose, we also present the results for MRT and Tomlinson-Harashima (TH) precoding in [40] which are pre-equalization techniques. As compared with the original channels, the proposed precoding provides less RMS delay spread in most cases. On the other hand, MRT and TH precoding yield larger delay spread than the others, and so a longer CP is required for those precoding schemes. When $M = 4$ and $\alpha = 0.5$, for instance, the average RMS delay spread of the proposed precoding is less than 25% of that of the original channels. It should be noted that this channel shortening is attained at the expense of allowing SNR loss as high as 3 dB, which is associated with $\alpha = 0.5$, as compared to MRT precoding. Nevertheless, the channel shortening gain still appears to be significant for a relatively large α. Moreover, the reduction in the effective RMS delay spread becomes more substantial with a larger value of M for an identical α, indicating that with the proposed precoding scheme, we can achieve a significant reduction in the effective RMS delay spread with negligible SNR loss once there are a sufficiently large number of antennas at the base station.

Table 3.1 tabulates the average required CP length and the corresponding CP overhead for different CIR lengths L when $M = 4$ and $\alpha = 0.5$. The CP lengths are simply calculated as 4 times the effective RMS delay spread, as in earlier work [41], and the CP overhead is defined as the ratio of the CP length to the overall OFDM symbol duration including the CP. Compared to the original channels, the precoded channel provides much less CP overhead as a result of channel shortening. Specifically, the proposed precoding reduces the CP overhead to 25% of the CP overhead of the original channels.

Chapter 4

SUM RATE MAXIMIZATION WITH SHORTENED CYCLIC PREFIX IN A MIMO-OFDM SYSTEM

This work was published in the IEEE Transactions on Vehicular Technology [42].

4.1 System Model

We consider the downlink transmission of a multi-user MIMO-OFDM system with N_t antennas at the transmitter and N_u users. Each user is assumed to be equipped with N_r antennas, and the base station estimates frequency-selective channels with N subcarriers. At the k-th user, the frequency response for the n-th subcarrier between the i-th transmit antenna and the j-th receive antenna is given as

$$\begin{aligned} H_{ij}^k(n) &= \sum_{\ell=0}^{N-1} h_{ij}^k(\ell) e^{-\frac{j2\pi n\ell}{N}} = \sum_{\ell=0}^{L-1} h_{ij}^k(\ell) e^{-\frac{j2\pi n\ell}{N}}, \\ & i = 1, 2, \cdots, N_t, \quad j = 1, 2, \cdots, N_r, \\ & n = 0, 1, \cdots, N-1, \end{aligned} \tag{4.1}$$

where $h_{ij}^k(\ell)$ is the corresponding CIR, and $L\ (\leq N)$ denotes the length of the CIR, i.e., $h_{ij}^k(L) = h_{ij}^k(L+1) = \cdots = h_{ij}^k(N-1) = 0$. Subsequently, the frequency response matrix of the k-th user can be represented as an $N_r \times N_t$ matrix, $\mathbf{H}_k(n) = \{H_{ij}^k(n)\}^T$.

Let $\mathbf{X}(n) \in \mathbb{C}^{N_u \times 1}$ denote the input vector consisting of each user's data $X_k(n)$, $k = 1, 2, \cdots, N_u$ and $\mathbf{W}(n) \in \mathbb{C}^{N_t \times N_u}$ be the corresponding precoding (pre-equalization) matrix for the n-th subcarrier. As long as the effective delay spread of the precoded channel is less than the CP length, we can represent the output vector of the channel at the k-th user as [5], [43]

$$\begin{aligned} \mathbf{Y}_k(n) &= \mathbf{H}_k(n)\mathbf{W}(n)\mathbf{X}(n) + \boldsymbol{\nu}_k(n) \\ &= \mathbf{H}_k(n) \sum_{i=1}^{N_u} \mathbf{W}_i(n) X_i(n) + \boldsymbol{\nu}_k(n), \end{aligned} \tag{4.2}$$

where $\mathbf{W}_i(n)$ is the i-th column vector of $\mathbf{W}(n)$, and $\boldsymbol{\nu}_k(n)$ denotes the $N_r \times 1$ noise vector.

4.2 Preliminaries for Precoding Design

In Subsection 4.2.1, we incorporate zero-forcing conditions into the system model. We then derive the effective RMS delay spread and the achievable rate for each user as functions of the precoding matrix in Subsection 4.2.2 and Subsection 4.2.3, respectively. The resulting equations constitute the objective function and constraints for the optimization problem in Section 4.3.

4.2.1 Zero-Forcing Conditions

In the same manner as [21] and [22], to eliminate other user interference at the k-th user in (4.2), the precoding vector $\mathbf{W}_k(n)$ must satisfy the following zero-forcing conditions:

$$\begin{aligned} \mathbf{H}_i(n)\mathbf{W}_k(n) = \mathbf{0}_{N_r \times 1} \quad & \text{if } i \neq k, \\ & i = 1, 2, \cdots, N_u. \end{aligned} \tag{4.3}$$

where $\mathbf{0}_{N_r \times 1}$ is the $N_r \times 1$ all-zero matrix. With the zero-forcing conditions imposed, the output signal in (4.2) can be rewritten as

$$\mathbf{Y}_k(n) = \mathbf{H}_k(n)\mathbf{W}_k(n)X_k(n) + \boldsymbol{\nu}_k(n). \tag{4.4}$$

Accordingly, at the j-th antenna of the k-th user, the scalar composite channel for the n-th subcarrier encompassing the precoding vector can be expressed as

$$H_{eq,jk}(n) = \mathbf{H}_{jk}(n)\mathbf{W}_k(n), \tag{4.5}$$

in which $\mathbf{H}_{jk}(n)$ is the j-th row vector of $\mathbf{H}_k(n)$. By aggregating the composite channels for all subcarriers, the equivalent channel matrix is formed as

$$\mathbf{H}_{eq,jk} \triangleq \begin{bmatrix} H_{eq,jk}(0) \\ H_{eq,jk}(1) \\ \cdots \\ H_{eq,jk}(N-1) \end{bmatrix} = \boldsymbol{\mathcal{H}}_{jk}\boldsymbol{\mathcal{W}}_k, \tag{4.6}$$

where $\boldsymbol{\mathcal{H}}_{jk} \in \mathbb{C}^{N \times NN_t}$ and $\boldsymbol{\mathcal{W}}_k \in \mathbb{C}^{NN_t \times 1}$ are, respectively, defined as

$$\boldsymbol{\mathcal{H}}_{jk} \triangleq \begin{bmatrix} \mathbf{H}_{jk}(0) & \mathbf{0}_{1\times N_t} & \cdots & \mathbf{0}_{1\times N_t} \\ \mathbf{0}_{1\times N_t} & \mathbf{H}_{jk}(1) & \mathbf{0}_{1\times N_t} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{1\times N_t} & \cdots & \mathbf{0}_{1\times N_t} & \mathbf{H}_{jk}(N-1) \end{bmatrix} \tag{4.7}$$

and

$$\boldsymbol{\mathcal{W}}_k \triangleq \begin{bmatrix} \mathbf{W}_k(0) \\ \mathbf{W}_k(1) \\ \cdots \\ \mathbf{W}_k(N-1) \end{bmatrix}. \tag{4.8}$$

Now let us define an $NN_r \times NN_t$ matrix,

$$\boldsymbol{\mathcal{H}}_i \triangleq \begin{bmatrix} \mathbf{H}_i(0) & \mathbf{0}_{N_r\times N_t} & \cdots & \mathbf{0}_{N_r\times N_t} \\ \mathbf{0}_{N_r\times N_t} & \mathbf{H}_i(1) & \mathbf{0}_{N_r\times N_t} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{N_r\times N_t} & \cdots & \mathbf{0}_{N_r\times N_t} & \mathbf{H}_i(N-1) \end{bmatrix}, \tag{4.9}$$

then the zero-forcing conditions in (4.3) can be rewritten as

$$\begin{aligned} \boldsymbol{\mathcal{H}}_i\boldsymbol{\mathcal{W}}_k = \mathbf{0}_{NN_r\times 1} \quad & \text{if } i \neq k, \\ & i = 1, 2, \cdots, N_u. \end{aligned} \tag{4.10}$$

From the fact that $\boldsymbol{\mathcal{H}}_i\boldsymbol{\mathcal{W}}_k = \mathbf{0}_{NN_r\times 1}$ holds if and only if $\|\boldsymbol{\mathcal{H}}_i\boldsymbol{\mathcal{W}}_k\|^2 = 0$, (4.10) can be represented in a quadratic form:

$$\begin{aligned} \|\boldsymbol{\mathcal{H}}_i\boldsymbol{\mathcal{W}}_k\|^2 = 0 \quad & \text{if } i \neq k, \\ & i = 1, 2, \cdots, N_u. \end{aligned} \tag{4.11}$$

4.2.2 Effective RMS Delay Spread

The effective RMS delay spread of the equivalent channel $H_{eq,jk}(n)$ in (4.6) is given as [14]

$$\tau_{RMS,jk} = \sqrt{\frac{1}{E_{jk}} \sum_{\ell=0}^{N-1} (\ell - \bar{\ell}_{jk})^2 |h_{eq,jk}(\ell)|^2}, \tag{4.12}$$

where $h_{eq,jk}(\ell)$ is the equivalent CIR corresponding to $H_{eq,jk}(n)$, and E_{jk} and $\bar{\ell}_{jk}$ are defined as

$$E_{jk} = \sum_{\ell=0}^{N-1} |h_{eq,jk}(\ell)|^2, \quad \bar{\ell}_{jk} = \frac{1}{E_{jk}} \sum_{\ell=0}^{N-1} \ell |h_{eq,jk}(\ell)|^2. \tag{4.13}$$

Assuming that $\bar{\ell}_{jk}$ is constant, we represent the DFT of $\ell - \bar{\ell}_{jk}$ as

$$F_{jk}(n) = \sum_{\ell=0}^{N-1} (\ell - \bar{\ell}_{jk}) e^{-\frac{j2\pi n\ell}{N}}. \tag{4.14}$$

From the preceding work [6], we can express the effective RMS delay spread in (3.6) as

$$\tau_{RMS,jk}^2 = \frac{\sum_{n=0}^{N-1} \left| \frac{1}{N} (F_{jk}(n) \circledast H_{eq,jk}(n)) \right|^2}{\sum_{n=0}^{N-1} |H_{eq,jk}(n)|^2}, \tag{4.15}$$

where $\circledast$ stands for the circular convolution. In addition, N convolutions can be represented with a matrix multiplication as

$$\frac{1}{N} \begin{bmatrix} F_{jk}(0) \circledast H_{eq,jk}(0) \\ F_{jk}(1) \circledast H_{eq,jk}(1) \\ \cdots \\ F_{jk}(N-1) \circledast H_{eq,jk}(N-1) \end{bmatrix} = \mathbf{F}_{jk}\mathbf{H}_{eq,jk}, \tag{4.16}$$

where

$$\mathbf{F}_{jk} \triangleq \frac{1}{N}\begin{bmatrix} F_{jk}(0) & F_{jk}(N-1) & \cdots & F_{jk}(1) \\ F_{jk}(1) & F_{jk}(0) & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ F_{jk}(N-1) & \cdots & F_{jk}(1) & F_{jk}(0) \end{bmatrix} \tag{4.17}$$

is an $N \times N$ matrix. Substituting (4.6) and (4.16) into (4.15), we finally have a matrix equation:

$$\begin{aligned} \tau^2_{RMS,jk} &= \frac{\mathbf{H}^H_{eq,jk}\mathbf{F}^H_{jk}\mathbf{F}_{jk}\mathbf{H}_{eq,jk}}{\mathbf{H}^H_{eq,jk}\mathbf{H}_{eq,jk}} \\ &= \frac{\boldsymbol{\mathcal{W}}^H_k\boldsymbol{\mathcal{H}}^H_{jk}\mathbf{F}^H_{jk}\mathbf{F}_{jk}\boldsymbol{\mathcal{H}}_{jk}\boldsymbol{\mathcal{W}}_k}{\boldsymbol{\mathcal{W}}^H_k\boldsymbol{\mathcal{H}}^H_{jk}\boldsymbol{\mathcal{H}}_{jk}\boldsymbol{\mathcal{W}}_k}. \end{aligned} \tag{4.18}$$

In OFDM systems, an RMS delay spread constraint is typically imposed by the CP length. Supposing that ρ is the maximum allowed RMS delay spread, the delay spread of the j-th antenna in (4.18) must satisfy

$$\tau^2_{RMS,jk} = \frac{\boldsymbol{\mathcal{W}}^H_k\boldsymbol{\mathcal{H}}^H_{jk}\mathbf{F}^H_{jk}\mathbf{F}_{jk}\boldsymbol{\mathcal{H}}_{jk}\boldsymbol{\mathcal{W}}_k}{\boldsymbol{\mathcal{W}}^H_k\boldsymbol{\mathcal{H}}^H_{jk}\boldsymbol{\mathcal{H}}_{jk}\boldsymbol{\mathcal{W}}_k} \leq \rho^2 \tag{4.19}$$

and equivalently

$$\boldsymbol{\mathcal{W}}^H_k(\boldsymbol{\mathcal{H}}^H_{jk}\mathbf{F}^H_{jk}\mathbf{F}_{jk}\boldsymbol{\mathcal{H}}_{jk} - \rho^2\boldsymbol{\mathcal{H}}^H_{jk}\boldsymbol{\mathcal{H}}_{jk})\boldsymbol{\mathcal{W}}_k \leq 0. \tag{4.20}$$

4.2.3 Achievable Rate

From (4.4), the achievable rate for the k-th user with all subcarriers is given as [22]

$$\begin{aligned} R_k &= \sum_{n=0}^{N-1} \log\left|\mathbf{I}_{N_r} + \frac{1}{\sigma^2}\mathbf{H}_k(n)\mathbf{W}_k(n)\mathbf{W}^H_k(n)\mathbf{H}^H_k(n)\right| \\ &= \sum_{n=0}^{N-1} \log\left|1 + \frac{1}{\sigma^2}\mathbf{W}^H_k(n)\mathbf{H}^H_k(n)\mathbf{H}_k(n)\mathbf{W}_k(n)\right|, \end{aligned} \tag{4.21}$$

where $\mathbf{I}_{N_r}$ is an $N_r \times N_r$ identity matrix and σ^2 is the noise variance. The second equality in (4.21) comes from Sylvester's determinant identity [44]. In order to represent the equation in terms of $\boldsymbol{\mathcal{W}}_k$, an $N_t \times NN_t$ matrix $\mathbf{E}_n$ is defined as

$$\mathbf{E}_n \triangleq \begin{bmatrix} \mathbf{0}_{N_t\times N_t} & \cdots & \mathbf{0}_{N_t\times N_t} & \mathbf{I}_{N_t} & \mathbf{0}_{N_t\times N_t} & \cdots \end{bmatrix} \tag{4.22}$$

such that $\mathbf{W}_k(n) = \mathbf{E}_n \boldsymbol{\mathcal{W}}_k$, $n = 0, 1, \cdots, N-1$. As a result, the achievable rate R_k for the k-th user can be expressed in terms of $\boldsymbol{\mathcal{W}}_k$ as

$$R_k = \sum_{n=0}^{N-1} \log(1 + \boldsymbol{\mathcal{W}}_k^H \mathbf{Q}_{kn} \boldsymbol{\mathcal{W}}_k), \tag{4.23}$$

where $\mathbf{Q}_{kn} \triangleq \frac{1}{\sigma^2} \mathbf{E}_n^H \mathbf{H}_k^H(n) \mathbf{H}_k(n) \mathbf{E}_n$ is an $NN_t \times NN_t$ positive semi-definite matrix.

4.3 Precoding Optimization

In this section, we formulate an optimization problem and present an SDR technique to find the solution. In particular, we aim to maximize the achievable rate of each user in (4.23) subject to the effective RMS delay spread constraint in (4.20) and zero-forcing constraints in (4.11). Correspondingly, a precoding optimization problem can be formulated as

$$\begin{aligned} \max_{\boldsymbol{\mathcal{W}}_k} \quad & \sum_{n=0}^{N-1} \log(1 + \boldsymbol{\mathcal{W}}_k^H \mathbf{Q}_{kn} \boldsymbol{\mathcal{W}}_k) \\ \text{s.t.} \quad & \boldsymbol{\mathcal{W}}_k^H \mathbf{S}_{jk} \boldsymbol{\mathcal{W}}_k \leq 0, \quad j = 1, 2, \cdots, N_r, \\ & \|\boldsymbol{\mathcal{H}}_i \boldsymbol{\mathcal{W}}_k\|^2 = 0, \quad i = 1, 2, \cdots, N_u, \quad i \neq k, \\ & \|\boldsymbol{\mathcal{W}}_k\|^2 = \frac{P}{N_u}, \end{aligned} \tag{4.24}$$

where $\mathbf{S}_{jk} \triangleq \boldsymbol{\mathcal{H}}_{jk}^H \mathbf{F}_{jk}^H \mathbf{F}_{jk} \boldsymbol{\mathcal{H}}_{jk} - \rho^2 \boldsymbol{\mathcal{H}}_{jk}^H \boldsymbol{\mathcal{H}}_{jk}$, and P denotes the total transmit power at the base station. Equal power allocation for all users is assumed in (4.24). Intuitively, the sum rate will decrease as the effective delay spread constraint becomes more stringent, i.e., as ρ becomes smaller. This is because more degree of freedom in the precoding should be exploited to limit the delay spread rather than to improve the sum rate. However, smaller ρ can support a shorter CP length, which provides longer time for data transmission. This tradeoff will be elaborated in Section 4.4.

Note that the problem in (4.24) is a non-convex quadratically constrained quadratic program (QCQP) because $\mathbf{S}_{jk}$ may not be a positive semi-definite matrix. In this regard, the SDR technique in [18] can be used to solve the problem. From properties of

trace operation, we observe that $\boldsymbol{\mathcal{W}}_k^H \mathbf{Q}_{kn} \boldsymbol{\mathcal{W}}_k = \mathrm{tr}(\boldsymbol{\mathcal{W}}_k^H \mathbf{Q}_{kn} \boldsymbol{\mathcal{W}}_k) = \mathrm{tr}(\mathbf{Q}_{kn} \boldsymbol{\mathcal{W}}_k \boldsymbol{\mathcal{W}}_k^H)$. If we define $\mathbf{X} \triangleq \boldsymbol{\mathcal{W}}_k \boldsymbol{\mathcal{W}}_k^H$, which is a rank-one positive semi-definite matrix, the optimization problem in (4.24) is equivalent to

$$\begin{aligned} \max_{\mathbf{X}} \quad & \sum_{n=0}^{N-1} \log(1 + \mathrm{tr}(\mathbf{Q}_{kn}\mathbf{X})) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{S}_{jk}\mathbf{X}) \leq 0, \quad j = 1, 2, \cdots, N_r, \\ & \mathrm{tr}(\boldsymbol{\mathcal{H}}_i^H \boldsymbol{\mathcal{H}}_i \mathbf{X}) = 0, \quad i = 1, 2, \cdots, N_u, \quad i \neq k, \\ & \mathrm{tr}(\mathbf{X}) = \frac{P}{N_u}, \\ & \mathbf{X} \succeq 0, \\ & \mathrm{rank}(\mathbf{X}) = 1. \end{aligned} \tag{4.25}$$

Notice that the constraints in (4.25) except for the last rank-one condition are convex. By omitting the rank-one constraint, we obtain a relaxed formulation which becomes a convex optimization problem. It can be solved by means of standard convex tools, such as CVX in MATLAB [45]. Using the best rank-one approximation, we extract the precoding vector $\boldsymbol{\mathcal{W}}_k$ from $\mathbf{X}$. Provided that the optimal solution of the relaxed problem is $\mathbf{X}_{opt}$, the eigen-decomposition of $\mathbf{X}_{opt}$ is given as

$$\mathbf{X}_{opt} = \sum_{i=1}^{r} \lambda_i \mathbf{x}_i \mathbf{x}_i^H, \tag{4.26}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are the eigenvalues of $\mathbf{X}_{opt}$, while $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_r$ are the corresponding eigenvectors, and $r = \mathrm{rank}(\mathbf{X}_{opt})$. Considering that the best rank-one approximation to $\mathbf{X}_{opt}$ is $\lambda_1 \mathbf{x}_1 \mathbf{x}_1^H$, $\sqrt{\lambda_1}\mathbf{x}_1$ is deemed the solution of the original problem in (4.24) [18]. To sum up, the precoding matrix $\mathbf{W}(n)$ over all subcarriers is drawn from N_u optimal solutions of (4.24) for all users.

4.4 Numerical Results

In this section, we evaluate the performance of the proposed precoding scheme in terms of the sum rate and effective RMS delay spread. It is assumed that the number

of transmit antennas at the base station is 4 or 8 and that each user is equipped with 1 or 2 antennas at the receiver. In all simulations, the number of OFDM subcarriers and the length of CIR are fixed to $N = 16$ and $L = 8$, respectively. The values of the CIR are assumed to follow independent identically distributed (i.i.d.) complex normal distributions $\mathcal{CN}(0, 1)$, and the noise covariance matrix is set to $\mathbf{I}_{N_r}$, i.e., $\sigma^2 = 1$. Correspondingly, the SNR is equal to the transmit power P. All the results are obtained by averaging 1,000 simulation runs with independent channel realizations. For comparison purpose, we also evaluate the performance of the zero-forcing beamforming (ZFBF) based method in [21]. Given that the problem formulation in the ZFBF based method resembles the proposed optimization problem except for delay spread constraints in (4.24), the sum rate of the proposed scheme is expected to be upper-bounded by that of the ZFBF based method.

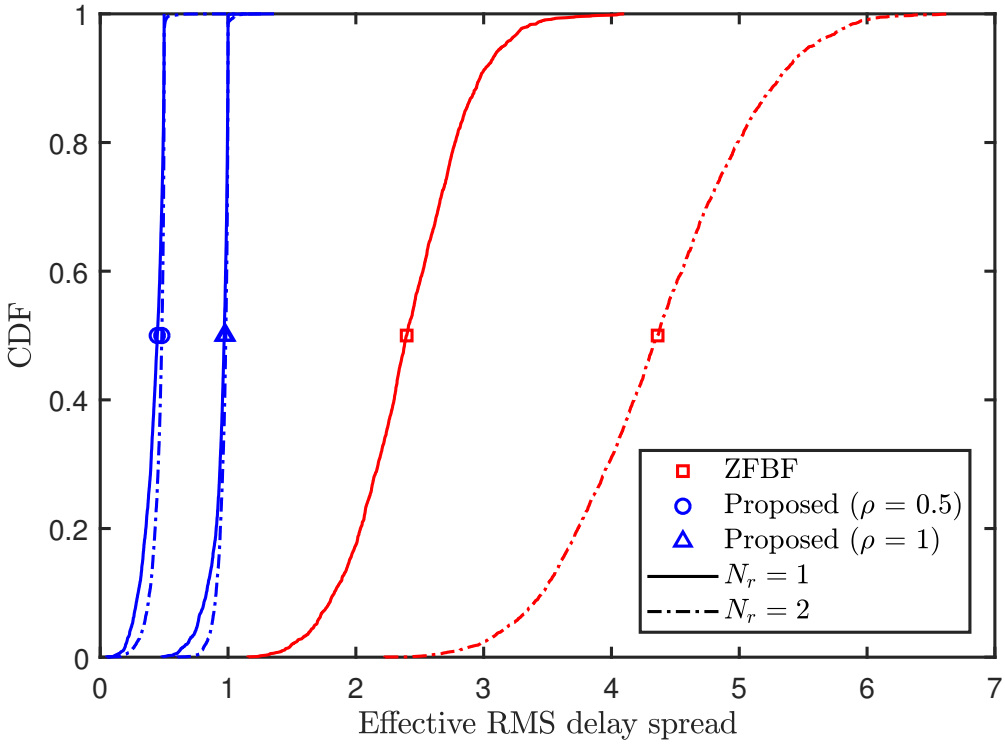

Figure 4.1: CDF of the effective RMS delay spread with $N_t = 4$, $N_u = 2$, and $N_r = \{1, 2\}$ at SNR $= 20$ dB.

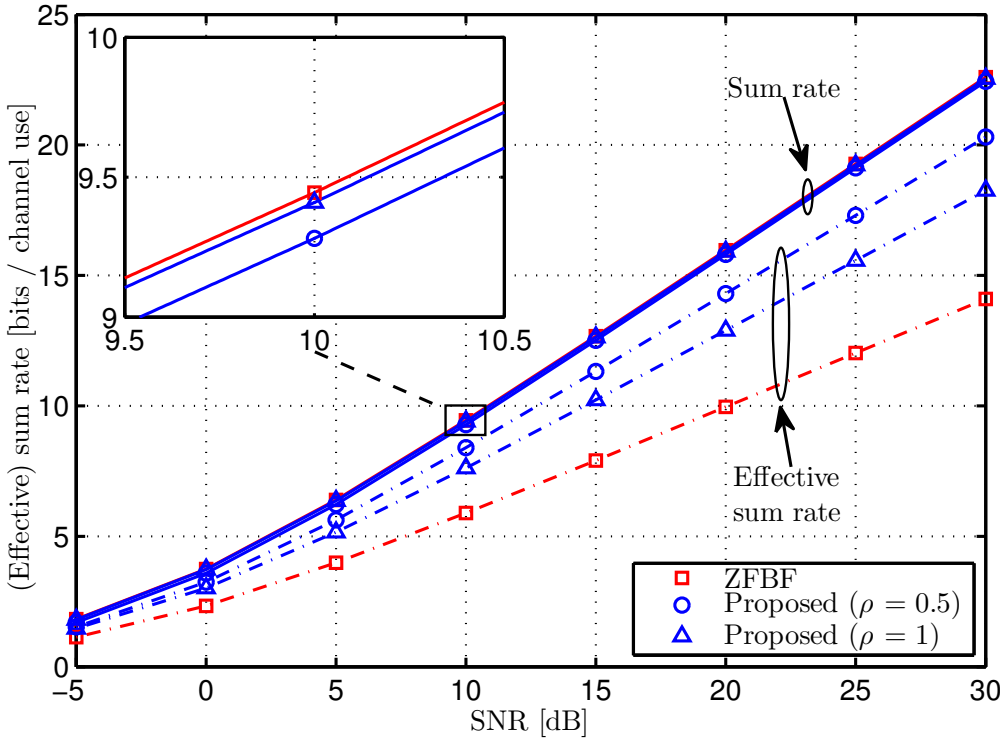

Figure 4.2: Sum rate and effective sum rate with $N_t = 4$, $N_u = 2$, and $N_r = 1$.

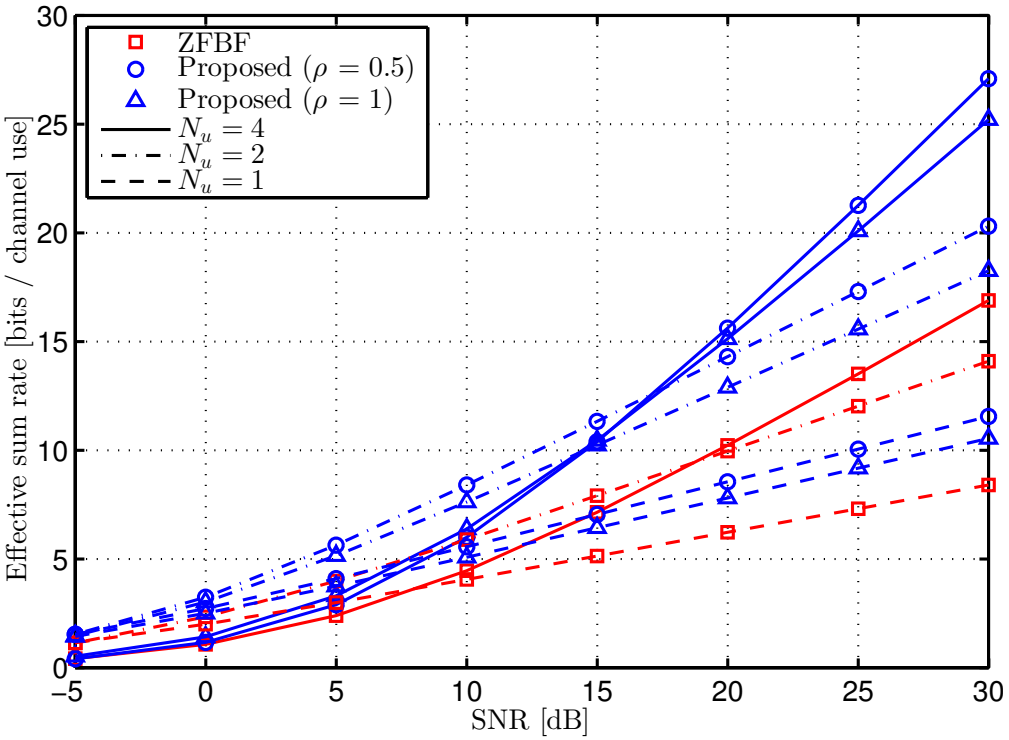

Figure 4.3: Effective sum rate for different numbers of users, $N_u = \{1, 2, 4\}$, when $N_t = 4$ and $N_r = 1$.

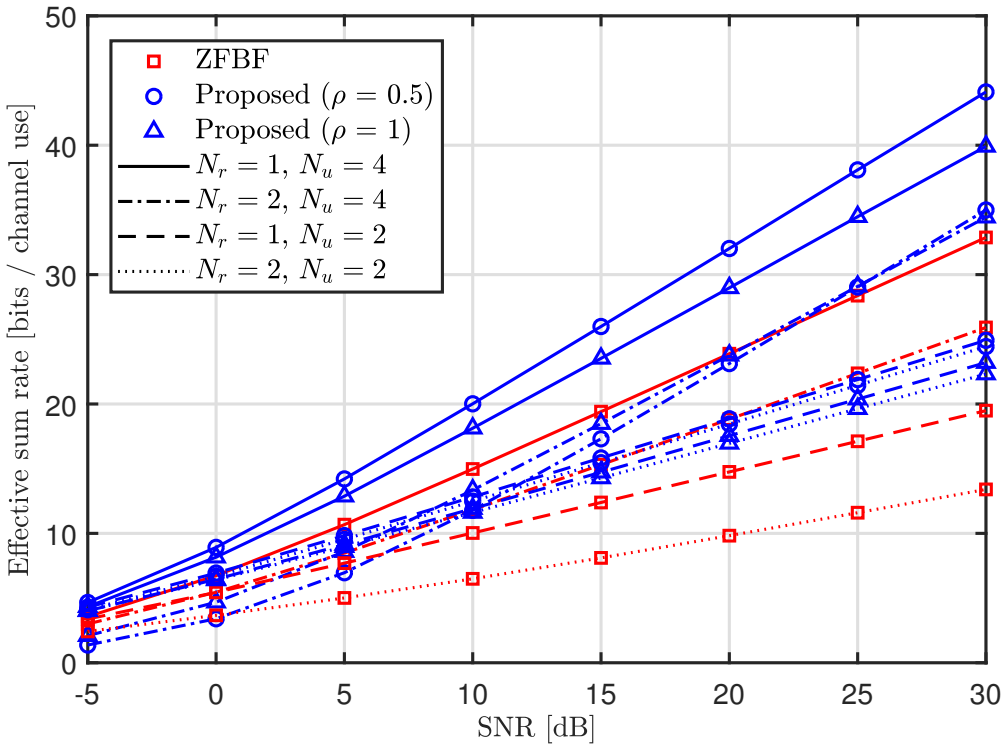

Figure 4.4: Effective sum rate for different numbers of antennas and users, $N_r = \{1, 2\}$ and $N_u = \{2, 4\}$, when $N_t = 8$.

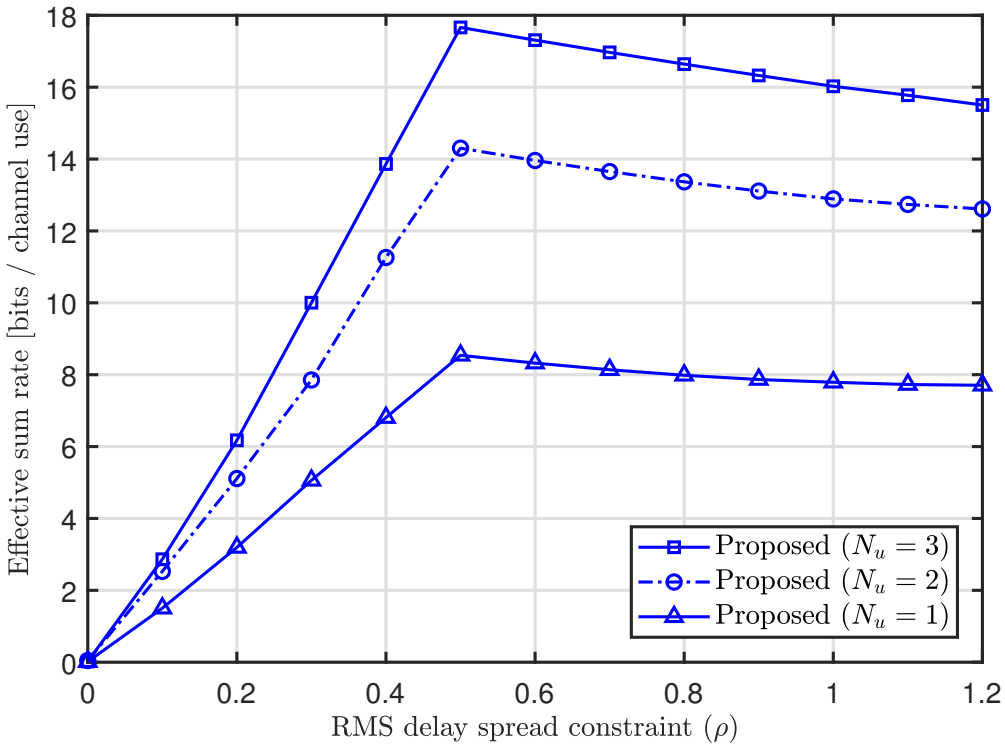

Figure 4.5: Effective sum rate versus the RMS delay spread constraint with $N_t = 4$ and $N_r = 1$ at SNR $= 20$ dB.

Figure 4.1 depicts the cumulative distribution function (CDF) of the effective RMS delay spread with $N_t = 4$, $N_u = 2$, and $N_r = \{1, 2\}$ at the SNR of 20 dB. It is demonstrated that the proposed scheme complies with the delay spread constraints, $\rho = 0.5$ and $\rho = 1$. The average values of the effective RMS delay spread for $N_r = 1$ amount to 0.42 and 0.94 in the proposed design, while it is 2.41 in the ZFBF based method. Given that the CP length is usually selected as several times the RMS delay spread [41], the ZFBF based method yields more than 5 times CP overhead as compared to the proposed design with $\rho = 0.5$. Hence, the proposed precoding scheme can accomplish low CP overhead via the delay spread constraint.

The achievable rate in (4.21) does not reflect the loss of data rate due to CP overhead. Considering the CP overhead, we can define *effective achievable rate* for the k-th user as

$$R_{eff,k} = \left(1 - \frac{L_{CP}}{L_{CP} + N}\right) R_k, \tag{4.27}$$

where L_{CP} denotes the CP length, which is assumed to be 4 times the mean value of the effective RMS delay spread in the simulation runs [41].

In Figure 4.2, we present the sum rate obtained from (4.21) and the effective sum rate obtained from (4.27). From the magnified image, it is verified that the original sum rate of the proposed scheme is upper-bounded by that of the ZFBF based method, although the performance gap between the two schemes is marginal. Moreover, the sum rate improvement with $\rho = 1$ is negligible as compared with $\rho = 0.5$, implying that constraints for the effective RMS delay spread scarcely affects the sum rate. The effective sum rate is larger with the proposed scheme than with the ZFBF based method, and the performance gap is mainly caused by the CP length or the effective RMS delay spread.

Figure 4.3 illustrates the effective sum rate for different numbers of users, $N_u = \{1, 2, 4\}$. The effective sum rate with the proposed design is greater than that of the ZFBF based method in all circumstances. When the number of users increases, the effective sum rate gap between the two schemes becomes more distinct because the

RMS delay spread of the ZFBF based method is larger. As the solution without the delay spread constraint yields marginal sum rate improvement but considerable CP overhead, the proposed precoding is appropriate for effective sum rate maximization. At high SNR, the effective sum rate increases as the number of users grows. However, even though the number of users changes from $N_u = 2$ to $N_u = 4$, the effective sum rate of both schemes decreases at low SNR. This indicates that, in the case of many users, efforts to eliminate multi-user interference do not have a significant impact on the sum rate maximization in low SNR regions where the noise dominates the interference.

Figure 4.4 depicts the effective sum rate with different numbers of receive antennas and users. Although the number of receive antennas increases, the sum rate gain is small because both precoding techniques do not consider the degrees of freedom for the receive antennas. Consequently, the reduction in the effective sum rate occurs by a longer CP length. In particular, the performance degradation of the ZFBF based method is severer due to the absence of the delay spread constraint. When $N_u = 2$, for example, the average RMS delay spread of the ZFBF based method is 1.61 with $N_r = 1$ and nearly doubles to 3.31 with $N_r = 2$. In the same situation, the average RMS delay spread of the proposed scheme with $\rho = 0.5$ amounts to 0.38 with $N_r = 1$ and 0.44 with $N_r = 2$.

In Figure 4.5, we present the effective sum rate versus the RMS delay spread constraint, ρ, for $N_t = 4$ and $N_r = 1$. When ρ is small, the sum rate optimization cannot achieve high rates because it is difficult to satisfy the tight constraint. For example, the sum rate performance with ρ under 0.5 is considerably small. Accordingly, the effective sum rate has a low value in spite of small CP overhead. For large values of ρ, the proposed design has the sum rate performance comparable to the ZFBF based method as shown in Figure 4.2, which leads to the improvement of the effective sum rate with reduced CP length. The optimal value of ρ may exist due to aforementioned tradeoff and appears to be constant regardless of the number of users. Hence, we can determine

the CP length from the optimal delay spread constraint in an OFDM system.

Chapter 5

CONCLUSION

This dissertation proposed a deep learning-based NOMA scheme and precoding design for 5G wireless communications.

5.1 Deep Learning-based Spreading Sequence Design and Active User Detection for Massive Machine-Type Communications

In this dissertation, we proposed a deep learning-based spreading sequence design and AUD scheme for an mMTC system. To design the communications system minimizing AUD error, we employed an end-to-end DNN. By properly training the whole network, we can obtain the spreading sequences and the AUD scheme optimized for mMTC environments. Numerical results demonstrated that the AUD performance of the proposed scheme is significantly better than that of the conventional schemes in the heterogeneous activity scenario. Further, we observed that the spreading sequences obtained from the proposed end-to-end DNN can improve the AUD performance even when we use the conventional greedy algorithms.

5.2 Precoding Design for Cyclic Prefix Overhead Reduction in a MISO-OFDM System

In this dissertation, a precoding design was proposed to reduce the effective RMS delay spread of the precoded channel and realize low-latency communications in a MISO-OFDM system. We formulated an optimization problem based on an upper bound of the effective RMS delay spread and the SNR for each subcarrier. To find the optimal precoding vector, the SDR technique is used to convert the optimization problem into a solvable convex problem. The proposed precoding scheme is found significantly to reduce the effective delay spread of the channel and thus the CP overhead of OFDM systems. Moreover, the advantage of the proposed approach becomes more distinct as the number of antennas at the base station increases. Hence, the proposed approach in combination with massive MIMO can provide a promising solution to realize low-latency communications with OFDM transmissions.

5.3 Sum Rate Maximization with Shortened Cyclic Prefix in a MIMO-OFDM System

We proposed a precoding scheme to maximize the sum rate under the RMS delay spread constraint for low-latency communications. Applying the zero-forcing conditions, we formulated an optimization problem for each user in terms of the achievable rate and the effective RMS delay spread. Subsequently, the optimal precoding matrix was obtained from the solutions of a convex problem relaxed by the SDR technique. Numerical results showed that the effective sum rate of the proposed design is significantly larger than that of the conventional zero-forcing method without delay constraints. Last but not least, the proposed precoding scheme can substantially reduce the effective delay spread of the channel, lowering CP overhead for low-latency communications.

Bibliography

[1] J. G. Andrews *et al.*, "What will 5G be?", *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] G. P. Fettweis, "The tactile Internet: Applications and challenges", *IEEE Veh. Technol. Mag*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[3] 3GPP, "Study on new radio access technology physical layer aspects", Tech. Rep. 38.802 v14.1.0, Jun. 2017.

[4] E. Dahlman *et al.*, "5G wireless access: Requirements and realization", *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, Dec. 2014.

[5] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2005.

[6] N. Kim, J. Ahn, O.-S. Shin, and K. B. Lee, "Precoding design for cyclic prefix overhead reduction in a MISO-OFDM system", *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 578–581, Oct. 2017.

[7] L. Dai *et al.*, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends", *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[8] B. K. Jeong, B. Shim, and K. B. Lee, "MAP-based active user and data detection for massive machine-type communications", *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8481–8494, Sep. 2018.

[9] C. Bockelmann *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks", *IEEE Access*, vol. 6, pp. 28 969–28 992, 2018.

[10] J. Ahn, B. Shim, and K. B. Lee, "Sparsity-aware ordered successive interference cancellation for massive machine-type communications", *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 134–137, Feb. 2018.

[11] M. Elad, "Optimized projections for compressed sensing", *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[12] G. Li *et al.*, "On projection matrix optimization for compressive sensing systems", *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.

[13] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer", *IEEE Trans. on Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[14] R. Schur and J. Speidel, "An efficient equalization method to minimize delay spread in OFDM/DMT systems", in *Proc. IEEE Inter. Conf. Commun.*, vol. 5, Helsinki, Finland, Jun. 2001, pp. 1481–1485.

[15] S. I. Husain, J. Yuan, and J. Zhang, "Impulse response shortening through limited time reversed channel in MB OFDM UWB systems", in *Proc. Inter. Symp. Commun. Inform. Technol.*, Oct. 2007, pp. 1269–1273.

[16] S. Payami and F. Tufvesson, "Delay spread properties in a measured massive MIMO system at 2.6 GHz", in *Proc. IEEE Inter. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2013, pp. 53–57.

[17] C. Kosanyakun and C. Kotchasarn, "A study of insufficient cyclic prefix by using precoding for MIMO-OFDM systems", in *Proc. IEEE Inter. Conf. Adv. Commun. Technol.*, Jul. 2015, pp. 727–732.

[18] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems", *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[19] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing", *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2338–2342, Sep. 2006.

[20] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing", *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Sep. 2008.

[21] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses", *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[22] C. Guthy, W. Utschick, R. Hunger, and M. Joham, "Efficient weighted sum rate maximization with linear precoding", *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2284–2297, Apr. 2010.

[23] J. Lorca, "Cyclic prefix overhead reduction for low-latency wireless communications in OFDM", in *Proc. IEEE Veh. Technol. Conf.*, Glasgow, U.K., May 2015.

[24] D. Kim, "Deep learning-aided spreading sequence design for massive machine-type communications", Master's thesis, Dept. of Electr. and Comput. Eng., Seoul National Univ., Seoul, South Korea, 2019.

[25] N. Kim, B. Shim, and K. B. Lee, "Deep learning-based spreading sequence design and active user detection for massive machine-type communications", *IEEE Wireless Commun. Lett.*,

[26] J. W. Choi *et al.*, "Compressed sensing for wireless communications: Useful tips and tricks", *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 2017.

[27] W. Kim, Y. Ahn, and B. Shim, "Deep neural network-based active user detection for grant-free noma systems", *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2143–2155, Apr. 2020.

[28] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit", *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[29] Y. Du *et al.*, "Efficient multi-user detection for uplink grant-free noma: Prior-information aided adaptive compressive sensing perspective", *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.

[30] J. Nam *et al.*, "Large-scale multi-label text classification — revisiting neural networks", in *Proc. Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, Nancy, France, Sep. 2014, pp. 437–452.

[31] B. Xin *et al.*, "Maximal sparsity with deep networks?", in *Proc. Adv. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016, pp. 4340–4348.

[32] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv:1502.03167*, Mar. 2015. [Online]. Available: https://arxiv.org/abs/1502.03167.

[34] K. He *et al.*, "Deep residual learning for image recognition", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[35] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization", *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1338–1351, Oct. 2006.

[36] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[37] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, Second. Upper Saddle River, NJ, USA: Prentice Hall, 1997.

[38] A. Ijaz, A. B. Awoseyila, and B. G. Evans, "Low-complexity time-domain SNR estimation for OFDM systems", *Electron. Lett.*, vol. 47, no. 20, pp. 1154–1156, Sep. 2011.

[39] T. K. Y. Lo, "Maximum ratio transmission", *IEEE Trans. Commun.*, vol. 47, pp. 1458–1461, Oct. 1999.

[40] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding", in *Proc. 4th ITG Conf. Source and Chnnel Coding*, Jan. 2002, pp. 139–147.

[41] K. Hassan, T. Rahman, M. Kamarudin, and F. Nor, "The mathematical relationship between maximum access delay and the RMS delay spread", in *Proc. Inter. Conf. Wireless Mobile Commun.*, Luxembourg City, Luxembourg, Jun. 2011, pp. 18–23.

[42] N. Kim, O.-S. Shin, and K. B. Lee, "Sum rate maximization with shortened cyclic prefix in a mimo-ofdm system", *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12 416–12 420, Dec. 2018.

[43] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[44] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*. New York, NY, USA: Springer, 1997.

[45] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming", 2017. [Online]. Available: http://cvxr.com/cvx/.

초 록

최근 5G 시스템의 등장으로 고신뢰 저지연 통신(ultra reliable low-latency communications, URLLC)과 대규모 사물 통신(massive machine-type communications, mMTC)이 주목을 받고 있다. 의료 서비스, 커넥티드 카, 로봇 공학, 제조업, 자유 시점 비디오 등 다양한 서비스들이 저지연 통신에서 예상되고, 이들은 1 ms 정도의 극도로 낮은 지연 시간을 요구한다. 한편, 대규모 사물 통신은 기지국에서 많은 기기(예를 들어 센서, 로봇, 자동차, 기계)의 방대한 연결성에 관한 것이다. 기존 통신 시스템(예를 들어 Long-Term Evolution (LTE))은 저지연 통신과 대규모 사물 통신의 요구 사항을 만족하기 어렵기에 이 통신 환경에 적합한 새로운 기술이 필요하다. 본 학위논문에서는 대규모 사물 통신과 저지연 통신을 위한 세 가지 기술을 제안한다.

논문의 첫 부분에서는 많은 기기가 비직교 확산 시퀀스를 사용해 기지국에 접속하는 대규모 사물 통신을 지원하는 딥러닝 기반의 확산 시퀀스 설계 및 활성 사용자 검출(active user detection, AUD) 방법을 제안한다. 검출 오류를 최소화하는 전체 통신 시스템을 설계하기 위해, 종단 간 심층 신경 네트워크(deep neural network, DNN)를 활용한다. 이 신경 네트워크에서 확산 네트워크는 송신기를 모델링하고 검출 네트워크는 활성 기기를 추정한다. 검출 오류를 손실 함수로 사용함으로써, 확산 시퀀스를 포함한 네트워크 변수들은 검출 오류를 최소화하도록 학습된다. 시뮬레이션 결과에서는 제안한 방법으로 얻어진 확산 시퀀스가 압축센싱 기반의 검출 기법과 제안한 검출 기법 모두에서 기존의 시퀀스보다 더 좋은 검출 성능을 달성하는 것을 보여준다.

논문의 두 번째 부분에서는 직교 주파수 분할 다중 방식(orthogonal frequency di-

vision multiplexing, OFDM) 시스템에서 프리코딩된 채널의 RMS (root mean square) 지연 확산을 줄이는 프리코딩 기법을 제안한다. OFDM 시스템에서 오버헤드를 증가시키지 않으면서 지연을 줄이기 위해서는 채널의 지연 확산과 그로 인한 CP (cyclic prefix)의 길이를 줄이는 것이 무엇보다 중요하다. 제안하는 기법에서는 RMS 지연 확산의 상한을 목적 함수로 하고 각 부반송파의 신호 대 잡음비를 제약조건으로 하는 최적화 문제를 설정한다. 최적화된 프리코딩을 찾을 수 있도록 원래 문제를 볼록 문제로 변환하기 위해 SDR (semi-definite relaxation) 기법을 사용한다. 시뮬레이션 결과에서는 제안한 프리코딩 설계가 특히 기지국에서 안테나의 수가 많을 때 RMS 지연 확산을 크게 줄이는 것을 보여준다.

논문의 마지막 부분에서는 저지연 OFDM 시스템에서 전송률 최대화를 위한 선형 프리코딩 설계를 다룬다. 저지연 통신에서 짧아지는 심볼 주기로 인한 CP의 오버헤드를 완화하기 위해 5G 무선 시스템은 짧은 CP를 사용할 필요가 있다. 채널의 지연 확산은 CP 길이보다 짧아야 하므로 먼저 실질적인 RMS 지연 확산과 달성 가능한 전송률을 제로 포싱 조건을 사용하여 유도한다. 다음으로 지연 확산 제약조건을 만족하는 전송률 최적화 문제를 사용자마다 정립하고 SDR 기법으로 해결 가능한 볼록 문제로 변환한다. 모든 사용자에 대해 최적화 문제를 푸는 것으로 전체 프리코딩 행렬을 얻는다. 시뮬레이션 결과에서는 제안한 기법이 작은 RMS 지연 확산과 함께 기존의 전송률 최적화보다 월등한 성능을 달성하는 것을 보여준다.

주요어: 대규모 사물 통신, 저지연 통신, 딥러닝, 비직교 다중 접속, 프리코딩

학번: 2014-22545

ACKNOWLEDGMENTS

이것으로 박사과정을 마치고 새로운 세상으로 한 발 내딛으려 합니다. 20대의 대부분의 시간을 박사과정으로 보냈기에 학위논문을 마무리하는 이 시점이 감회가 새롭습니다. 고군분투했던 지난 시간이 앞으로의 훌륭한 초석이 되어 줄 것을 믿어 의심치 않습니다. 본 학위논문을 완성하기까지 7년 간의 학위 과정 동안 학문적 · 정서적으로 저에게 도움을 주신 분들이 많습니다. 이 지면을 빌려 여러 분들께 감사 인사를 드리고자 합니다.

우선 7년 간 저의 연구와 대학원 생활을 지도해주신 이광복 교수님께 감사드립니다. 학사를 졸업하고 대학원을 입학해 여러 방면으로 미흡했던 저를 지도해주시고 조언해주신 덕분에 여기까지 올 수 있었습니다. 연구와 생활에 있어서 교수님께서 가지고 계신 높은 기준과 뛰어난 안목은 저의 견문을 넓히고 제가 발전할 수 있었던 원동력이 되었습니다.

저의 연구 지도를 위해 많은 시간을 할애해주시고 좋은 논문을 작성할 수 있도록 도움을 주셨던 심병효 교수님과 신오순 교수님께도 진심으로 감사드립니다. 교수님의 사려깊은 지도 덕분에 연구 성과를 내고 학업을 마무리할 수 있었습니다. 교수님들께 연구 및 논문 지도를 받게 된 것은 학위 과정동안 가장 감사하고 있는 일 중 하나입니다. 앞으로도 하시는 모든 일이 무탈하기를 바라겠습니다. 바쁘신 와중에 심사 위원으로 참석해주셔서 건설적인 조언을 해주신 최완 교수님과 문태섭 교수님께도 깊이 감사드립니다.

대학원 생활을 함께했던 연구실 선후배들에게도 감사를 전하고자 합니다. 처음 어색할 때 친근하게 다가와준 두희 형, 연구실의 중심을 잡아줬던 선익이 형, 똑똑하고 장난치는 것을 좋아하던 준수 형, 오랜 시간 동안 즐겁게 연구실 생활을 같이

하고 많은 것을 배울 수 있었던 진엽이 형, 엉뚱하고 수영을 좋아했던 준영이 형, 회사에서 오셔서도 저희와 잘 어울려주신 병국이 형 모두에게 감사드립니다. 정말 듬직하고 재미있는 승민이 형, 깐족거리는 것을 좋아하지만 마음은 따뜻한 우현이 형, 연구실 동기이자 연구실 생활을 풍성하게 만들어준 경원이 형, 친근하고 재미있는 병희 형, 잘생기고 잘된 의택이 형, 연구실에 마지막까지 남아 많은 시간을 보냈던 동우에게도 감사를 표합니다. 마지막으로 지금까지도 연구실에 도움을 주시는 여규랑 누나에게도 감사드립니다. 여러분 덕분에 연구실 생활을 정말 즐겁게 할 수 있었으며 힘들었던 시기도 견딜 수 있었습니다.

저의 벗들에게도 감사를 전하고자 합니다. 고등학교 시절부터 즐거운 시간을 함께한 본석이, 지환이, 용현이, 그리고 자주 보지 못하지만 모이면 재밌는 서울대 친구들 익성이 형, 형규 형, 구용이 형, 지석이, 형걸이 형, 현우 형, 형수 형, 민호, 지호 형에게 감사드립니다. 또한, 한성과학고 동문들과 서울대 전기공학부 동기들에게도 감사드립니다.

사랑하는 저의 가족들에게 감사를 전합니다. 모범적인 삶의 자세를 보여주시는 존경하는 아버지와 가족을 위해 많은 힘을 써주시는 어머니께 진심으로 감사드립니다. 두 분이 격려해주신 덕분에 제가 학위 과정을 잘 마칠 수 있었습니다. 그리고 집안에 활력을 불어넣어주는 동생 하리에게도 감사를 전합니다. 마지막으로 지면에 미처 담지 못했지만, 저를 격려해주신 모든 분들께도 진심으로 감사드립니다.