Massive MIMO Technology for Next Generation of Wireless Networks

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Parna Sabeti

Electrical Engineering Department

The University of Dublin
Trinity College

This dissertation is submitted for the degree of

Doctor of Philosophy

Trinity College Dublin June 2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent / do not consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed:

_____

Parna Sabeti

Parna Sabeti
June 2020

Summary

Large scale antenna or massive multiple input multiple output (MIMO) systems are one of the key enabling technologies for fifth generation (5G) of wireless communications networks and beyond. This technology offers huge advantages in terms of energy efficiency, spectral efficiency, robustness and reliability. However, there are some challenges that prevent the realization of full potential of massive MIMO technology. For instance, performance of massive MIMO systems heavily relies on accurate synchronization. While orthogonal frequency division multiplexing (OFDM) is commonly used as a multicarrier modulation technique in massive MIMO systems, and it is very sensitive to frequency synchronization errors. In contrast, filter-bank multicarrier modulation (FBMC) based waveforms are more robust against frequency offset. Thus, the application of FBMC pulse amplitude modulation (FBMC-PAM) to massive MIMO is proposed in this thesis as an alternative to OFDM. It is also demonstrated that due to the absence of cycle prefix (CP), FBMC-PAM can provide a better bit error rate (BER) performance than OFDM. In addition, it is observed that as the number of base station (BS) antennas increases, the performance of the massive MIMO system with FBMC-PAM saturates. In fact, in an asymptotic regime, the noise effect and multi-user interference (MUI) are averaged out, but some residual effects of multipath channel will remain. This saturation level, which is the upper bound for the system performance, is mathematically calculated and confirmed with simulations. Moreover, it is shown that by increasing the number of subcarriers, a higher upper bound can be achieved. On the other hand, the CP in OFDM systems effectively mitigates the effects of multipath channel, and there will be no saturation in an asymptotic regime. Following this results, we focus on OFDM-based massive MIMO systems for the rest of this study. Hence, to meet the requirement of accurate synchronization in these systems, a low-complexity frequency synchronization technique is proposed. It is shown that the phase information of the covariance of the received signals at the BS antennas is a function of carrier frequency offset (CFO), and if real-valued pilots are utilized, CFO can be simply calculated from the phase information. It should be noted that due to the spatial multiplexing in massive MIMO systems, all the users can simultaneously share the entire available bandwidth, and channel state information (CSI) of users is used to distinguish their signals. However, at the CFO estimation stage, the CSI knowledge is unknown. Thus, a set of rectangular-shaped real-valued pilots are

designed to preserve the orthogonality of the users, and a closed form formula is calculated for CFO estimation. Since this technique has strict limitations on pilot design, another CFO estimation technique is proposed which is more general and can work with any pilots. In this technique, the desired user signal is separated from the received signal by using a matrix orthogonal to the space spanned by the pilot of that user. It is proved that the objective function of the proposed optimization problem is unimodal and can be simply solved by Golden search algorithm. Furthermore, it is noticed that the massive MIMO system with the time domain CFO compensation requires a separate receiver for each user which imposes a huge amount of computational complexity to the system. Therefore, a frequency domain CFO compensation technique is proposed which takes place after combining the received signals at the BS. Thus, one receiver is sufficient for all the users in the network, and the complexity of the receiver is considerably reduced. In addition, it is proved that by applying this CFO compensation technique, even in the presence of CFO estimation error, the scattering effect of CFO is removed and only a phase shift remains. Hence, two iterative error correction algorithms are proposed to improve the synchronization accuracy. Simulation results demonstrate that the BER of the system with this synchronization technique matches that of the perfect synchronous system.

We then move on to the challenge of CSI acquisition in massive MIMO systems. This is important because the CSI estimation becomes a bottleneck as the scale of the antenna array increases or the number of users in the network grows large. To avoid the pilot overhead, a deep learning (DL) aided blind channel estimation technique is proposed. First, the MUI is canceled by calculating the orthogonal complement space matrix of the MUI. Then, based on the asymptotic orthogonality of the massive MIMO channels, the first OFDM symbol of all the users is extracted as a virtual pilots. However, in practice, noise and MUI are not completely removed, and the remaining part is also intensified after this process. Therefore, denoising convolutional neural network (DnCNN) is deployed as a denoiser to deal with the remaining interference. Another DL-based algorithm called U-Net is also proposed to be used as a denoiser that can outperform DnCNN when the noise level is high. Moreover, a ResNet architecture followed by a feedforward neural network is proposed to force the network to converge to the expected values. This further enhances the performance of the virtual pilot detection. Finally, maximum likelihood (ML) estimator is employed to estimate the CSIs.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Nicola Marchetti, my co-supervisor Prof. Linda Doyle, and my adviser Dr. Arman Farhang for their guidance, patience, encouragement and support. Their insightful feedback greatly impacted my work and improved my research skills during the last four years. I would also like to thank Dr Irene Macaluso for the great discussions that helped me to broaden my knowledge in the context of deep learning algorithms.

Moreover, I must express my very profound gratitude to my husband, Hamidreza, and to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. Indeed, this accomplishment would not have been possible without them.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Large scale multiple input multiple output (MIMO) or massive MIMO technology can boost system performance in many aspects, which makes it one of the key technologies for fifth generation (5G) of wireless networks and beyond [1–4]. It can tremendously enhance network capacity by enabling users to simultaneously utilize the entire available bandwidth [5–7]. This leads to improved spectral efficiency [8]. In addition, by increasing the number of base station (BS) antennas (array gain), the required power to achieve a desired information rate can be reduced [8]. Moreover, large-scale antenna arrays are crucial for achieving highly directional beamforming. Hence, emergence of millimeter wave (mmWave) systems, where beamforming plays a vital role, has further highlighted the importance of massive MIMO, [9]. However, massive MIMO technology has uncovered entirely new problems that are not the concern of conventional small scale MIMO systems. Therefore, this thesis is dedicated to study and address some of the main challenges of the massive MIMO technology to better prepare it for the next generation of the mobile networks.

1.1 Massive MIMO challenges

Despite all the aforementioned benefits of massive MIMO, performance of these systems heavily relies on an accurate frequency synchronization among multiple users. In fact, orthogonal frequency division multiplexing (OFDM) used in multicarrier massive MIMO systems is highly sensitive to carrier frequency offset (CFO) due to the large side lobes of its waveform. Moreover, CFO estimation is quite challenging in massive MIMO systems because of the coexistence of multiple frequency synchronization errors corresponding to different users. More precisely, as all the users can simultaneously share the entire available bandwidth, and their orthogonality is granted by spatial diversity, the channel state information (CSI) of the users are required to distinguish their signals. However, at the stage of CFO estimation, such

information is not available. In addition, since different users have different CFO values, their CFO should be compensated separately. Thus, after compensating the CFOs in the time domain, the users need their individual receivers. This can lead to a huge computational burden for the system as the number of BS antennas and/or the number of users in the network increases.

Moreover, as the scale of the antenna array increases or the number of users grows, the CSI acquisition becomes a bottleneck for massive MIMO systems. Channel estimation can be data-aided in which the estimation is based on the knowledge of pilot signals. However, such techniques waste considerable amount of resources and energy for transmitting pilots. In addition, if a pilot signal is reused by a user in a neighboring cell, the pilot contamination degrades the performance of the channel estimation. There are some blind channel estimation techniques in the literature, which either have some constants, for instance on sparsity level of the channels, that limit their applications, or impose a large amount of computational complexity to the system.

1.2 Contributions

To address the above-mentioned challenges and enable the realization of full potential of massive MIMO systems, the following contributions have been made in this thesis:

1. Filter bank multicarrier with pulse amplitude modulation (FBMC-PAM) is applied to massive MIMO systems as an alternative modulation technique to OFDM. FBMC-PAM waveform has smaller side-lobes compared to OFDM, which means it is more robust against CFO. Moreover, unlike OFDM, FBMC-PAM dose not need cyclic prefix (CP), and consequently, offers a higher spectral efficiency. It can transmit at the same rate as OFDM opposed to the conventional FBMC-based systems. FBMC-PAM has a minimum overlapping factor leading to a lower computational complexity than other FBMC-based waveforms. Contributions of this part can be summarized in the following.

    - Performance of FBMC-PAM is studied in massive multiple-input multiple-output (MIMO) systems in terms of signal to interference ratio (SIR). It is discussed that although in massive MIMO channels, multiuser interference as well as channel noise average out, there remains some residual interference due to the channel distortions even for an infinite number of BS antennas. Consequently, the SIR saturates after the number of BS antennas reaches a certain value.

- To quantify this saturation level, an upper bound for the SIR performance of FBMC-PAM is calculated in such channels.
- Moreover, it is shown that by choosing an appropriate number of subcarriers, although the SIR saturates, FBMC-PAM can equalize the channel efficiently and the error floor in the bit error rate (BER) can be completely removed.
- It is also shown that though the small overlapping factor of FBMC-PAM brings shortcomings in terms of equalization, in massive MIMO systems, FBMC-PAM provides a better BER performance than OFDM.

2. Thanks to the properties of massive MIMO in the asymptotic regime, a low-complexity frequency synchronization technique is developed for multi-user OFDM-based massive MIMO systems. The contributions related to the proposed synchronization technique are as follows:

   - It is shown that the phase information of the covariance matrix of the received signals is solely dependent on the users' CFOs. Hence, if a real-valued pilot is chosen, CFO values can be straightforwardly calculated from the received signal covariance matrix. There is then no need to deal with a complex optimization problem, and CFO can be estimated by the calculated closed-form equation.
   - Another CFO estimation technique is proposed which is more general and is not limited to a real-valued pilot. The computational complexity of this technique increases only linearly with respect to the number of BS antennas and is independent of the number of users.
   - A frequency domain CFO compensation technique is proposed that takes place after combining the received signals at the BS antennas. As a result, the compensation process is independent of the number of BS antennas. Moreover, one combiner can be used for all the users, unlike the conventional CFO compensation performed in the time domain which requires separate receivers for different users. Therefore, the computational complexity of the receiver is substantially reduced.
   - The effect of CFO estimation error on the channel estimation and CFO compensation is studied. It is proved that by applying our CFO compensation technique, estimation error is canceled through CFO compensation process and only a constant phase shift remains. Based on this result, a CFO estimation error correction algorithm is proposed. As it is illustrated through simulation, mean square error (MSE) performance is

improved by three orders of magnitude after error correction. According to the bit error rate (BER) curves in our simulations, we demonstrate that an almost perfect synchronization can be achieved.

3. A deep learning-aided blind channel estimation technique is proposed for multi-user OFDM-based massive MIMO systems. This proposed technique consists of three steps: multiuser interference (MUI) cancellation, virtual pilot detection, and channel estimation. In the first step, we assign a few null subcarriers to each user signal, and find the orthogonal complement space matrices of the users based on the null subcarriers. We further utilize these matrices to cancel the MUI. Next, we exploit the asymptotic orthogonality of the channel vectors in large scale MIMO systems to average out the channel impulse responses and extract the transmitted data of the first transmitted OFDM symbol of each user as a virtual pilot. In the last step, we use the detected data for channel estimation. The main contributions of this part can be summarized as follows:

   - It is proved that by exploiting asymptotic orthogonality of the massive MIMO channel vectors and without the knowledge of channel impulse responses, the transmitted data can be detected.
   - Since in this proposed virtual pilot detection technique, several noise affected observations of each transmitted symbol are obtained, DnCNN algorithms can be employed as a denoiser to improve the performance of the proposed technique.
   - Another DL-based algorithm named U-Net is utilized as a denoiser that can outperform DnCNN in low SNRs.
   - A residual neural network (ResNet) followed by a fully connected feed forward neural network is used to jointly denoise and detect the data symbols of the virtual pilots. Since this technique learns both additive and multiplicative noise, it can enhance the virtual pilot estimation, and as a result, a performance close to the one of data-aided channel estimation can be achieved.

1.3 Structure of the thesis

The thesis is laid out as follows. Chapter 2 provides the basics of massive MIMO systems as well as related literature review of synchronization and channel estimation techniques for these systems. Moreover, the technologies that are utilized in this thesis such as advanced waveforms and deep learning algorithms are explained

in this chapter. Chapter 3 looks at the application of FBMC-PAM as an alternative to OFDM modulation in massive MIMO systems and presents the performance analysis of the massive MIMO systems with FBMC-PAM modulation in an asymptotic regime. Chapter 4 focuses on the synchronization problem in massive MIMO systems. Low-complexity CFO estimation and CFO compensation techniques are proposed in this chapter. Chapter 5 moves on to the CSI acquisition concern in massive MIMO systems, and a DL-aided blind channel estimation technique is proposed. Finally, Chapter 6 concludes the thesis and suggests possible future studies.

1.4 Publication record

The following is the list of publications from this thesis:

- Sabeti, Parna, Arman Farhang, Nicola Marchetti, and Linda Doyle. "Performance analysis of FBMC-PAM in massive MIMO." In 2016 IEEE Globecom Workshops (GC Wkshps), pp. 1-7. IEEE, 2016.
- Sabeti, Parna, Arman Farhang, Nicola Marchetti, and Linda Doyle. "Low-complexity CFO compensation for OFDM-based massive MIMO systems." In 2017 IEEE Globecom Workshops (GC Wkshps), pp. 1-6. IEEE, 2017.
- Sabeti, Parna, Arman Farhang, Nicola Marchetti, and Linda Doyle. "CFO Estimation for OFDM-Based Massive MIMO Systems in Asymptotic Regime." In 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), pp. 99-103. IEEE, 2018.
- Sabeti, Parna, Arman Farhang, Nicola Marchetti, and Linda Doyle. "Frequency Synchronization for OFDM-Based Massive MIMO Systems." IEEE Transactions on Signal Processing 67, no. 11 (2019): 2973-2986.
- Sabeti, Parna, Arman Farhang, Irene Macaluso, Nicola Marchetti, and Linda Doyle. "Blind Channel Estimation for Massive MIMO: A Deep Learning Assisted Approach." In 2020 IEEE International Conference on Communications (ICC), IEEE, 2020.
- Sabeti, Parna, Arman Farhang, Irene Macaluso, Nicola Marchetti, and Linda Doyle. "DL-aided blind channel estimation for OFDM-Based Massive MIMO Systems." Submitted to IEEE Transactions on Wireless Communications.

Chapter 2

Background

Massive MIMO is a multiuser MIMO system where a base station (BS) with multiple transceiver antennas simultaneously serves multiple mobile terminals through spatial multiplexing. The number of BS antennas in this system is assumed to be around an order of magnitude larger than in conventional wireless systems. As the number of BS antennas tends to infinity, the processing gain of the system tends to infinity too, and as a result, the effect of both noise and multiuser interference are removed. Therefore in theory, the network capacity can be increased unboundedly by growing the number of BS antennas.

2.1 Basics of massive MIMO systems

The Massive MIMO systems can operate either in time division duplex (TDD) mode or frequency division duplex (FDD) mode. In TDD, the uplink and the downlink transmissions are separated in time, but they take place in the same frequency resource. While in FDD, the uplink and the downlink transmissions simultaneously occupy different frequency resources. There are several reasons that make TDD mode more desirable in massive MIMO systems. For instance, in TDD, only the BS needs to know the channel impulse responses. This is because relying on reciprocity of the transmission channel, it can be assumed that the uplink channels is identical to the downlink channels and do not change during the uplink and the downlink transmission. Therefore, the uplink channel estimated by the BS can be used for the precoding in the downlink transmission. Instead, in case of FDD, the uplink and downlink have different channels, and the users should estimate the downlink channels and send this information back to the BS to be used in precoding. This complicates the control process and adds a considerable overhead. In addition, FDD uses a large amount of frequency spectrum, at least twice the spectrum needed

by TDD. There must also be adequate guard bands between the uplink and the downlink channels, which leads to the waste of resources.

Since there is plenty of FDD spectrum currently in use, some studies have been devoted to the area of massive MIMO in FDD. In this thesis, we consider the TDD mode.

2.1.1 Uplink data transmission

Consider a multiuser OFDM-based MIMO setup similar to the one discussed in [6], where there are P single antenna users communicating to a BS equipped with $M \gg P$ antennas in a cell. We assume that different users' wireless channels are statistically independent and time invariant during one OFDM packet. Hence, they can simultaneously share all the available subcarriers as their signals can be distinguished through their respective channel gains. Having N active subcarriers, the $N \times 1$ vector of the κ^{th} OFDM symbol for user p before CP assignment can be obtained as

$$\mathbf{x}_p^{\kappa} = \mathbf{F}_N^{\text{H}} \mathbf{d}_p^{\kappa}, \tag{2.1}$$

where the $N \times 1$ vector $\mathbf{d}_p^{\kappa}$ contains the κ^{th} data symbol of user p, which is normalized to have a power of unity. Also, $\mathbf{F}_N$ is the normalized N-point DFT matrix with the elements $F_N[i,k] = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}ik}$ for $i,k = 0, \cdots, N-1$. The CP length, N_{CP}, is considered to be longer than the channel impulse response (CIR) length, L. Here, we consider perfect time and frequency synchronization. After CP removal, the OFDM symbol received at the m^{th} BS antenna can be written as

$$\mathbf{r}_m^{\kappa} = \sum_{p=0}^{P-1} \mathbf{x}_p^{\kappa} \otimes \mathbf{h}_{m,p} + \mathbf{n}_m^{\kappa}, \tag{2.2}$$

where $\otimes$ denotes circular convolution, $\mathbf{n}_m \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_N)$ is the complex additive white Gaussian noise (AWGN) with the variance of σ_n^2 at the m^{th} BS antenna and $\mathbf{h}_{m,p}$ is the $L \times 1$ CIR vector between user p and BS antenna m. We assume the channel taps to be a set of independent and identically distributed (i.i.d.) random variables that follow the complex normal distribution $\mathcal{CN}(0, \boldsymbol{\rho}_p)$, where $\boldsymbol{\rho}_p$ is an $L \times 1$ vector representing the p^{th} user's channel power delay profile (PDP).

Representing the circular convolution of the transmit data with the channel as $\mathbf{H}_{m,p}\mathbf{x}_p^{\kappa}$, the received signal at BS antenna m in the frequency domain can be written as

$$\bar{\mathbf{r}}_m^{\kappa} = \sum_{p=0}^{P-1} \mathbf{F}_N \mathbf{H}_{m,p} \mathbf{x}_p^{\kappa} + \bar{\mathbf{n}}_m^{\kappa}, \tag{2.3}$$

where $\mathbf{H}_{m,p}$ is a circulant matrix with the first column $\mathbf{h}_{m,p}$ which is zero-padded to have the length of N, and $\bar{\mathbf{n}}_m = \mathbf{F}_N \mathbf{n}_m$ is the frequency domain noise vector. By defining $\widetilde{\mathbf{H}}_{m,p} \triangleq \mathbf{F}_N \mathbf{H}_{m,p} \mathbf{F}_N^{\mathrm{H}}$, we have

$$\bar{\mathbf{r}}_m^{\kappa} = \sum_{p=0}^{P-1} \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p^{\kappa} + \bar{\mathbf{n}}_m^{\kappa}. \tag{2.4}$$

It is worth noting that due to the circulant property of $\mathbf{H}_{m,p}$, $\widetilde{\mathbf{H}}_{m,p}$ becomes a diagonal matrix, [10], with the main diagonal equal to $\bar{\mathbf{h}}_{m,p} = \mathbf{F}_N \mathbf{h}_{m,p}$, i.e. $\widetilde{\mathbf{H}}_{m,p} = \mathrm{diag}(\bar{\mathbf{h}}_{m,p})$.

Let us define the vector $\bar{\mathbf{r}}^{\kappa}[k] = [\bar{r}_0^{\kappa}[k], \bar{r}_1^{\kappa}[k], \cdots, \bar{r}_{M-1}^{\kappa}[k]]^{\mathrm{T}}$ which contains the k^{th} samples of the received signals at all the BS antennas in the frequency domain for $k = 0, 1, \cdots, N-1$. Then, the combiner output is given by

$$\bar{\mathbf{y}}^{\kappa}[k] = \mathbf{Z}[k] \bar{\mathbf{r}}^{\kappa}[k], \tag{2.5}$$

where $\bar{\mathbf{y}}^{\kappa}[k] = [\bar{y}_0^{\kappa}[k], \bar{y}_1^{\kappa}[k], \cdots, \bar{y}_{P-1}^{\kappa}[k]]^{\mathrm{T}}$ is a $P \times 1$ vector containing the k^{th} samples of the users' signals, and $\mathbf{Z}[k]$ is a $P \times M$ linear combiner on subcarrier k. The combiner can be maximum ratio combining (MRC), zero-forcing (ZF), or minimum mean square error (MMSE). For each of them, $\mathbf{Z}[k]$ can be calculated as

$$\mathbf{Z}[k] = \begin{cases} \mathbf{D}_k^{-1} \boldsymbol{\mathcal{H}}^{\mathrm{H}}[k] & \text{for MRC,} \\ (\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}} \boldsymbol{\mathcal{H}}[k])^{-1} \boldsymbol{\mathcal{H}}^{\mathrm{H}}[k] & \text{for ZF,} \\ (\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}} \boldsymbol{\mathcal{H}}[k] + \sigma_n^2 \mathbf{I}_P)^{-1} \boldsymbol{\mathcal{H}}^{\mathrm{H}}[k] & \text{for MMSE,} \end{cases} \tag{2.6}$$

where $\boldsymbol{\mathcal{H}}[k] = [\bar{\mathbf{h}}_0[k], \bar{\mathbf{h}}_1[k], \cdots, \bar{\mathbf{h}}_{P-1}[k]]$ is an $M \times P$ matrix whose p^{th} column is $\bar{\mathbf{h}}_p[k] = [\bar{h}_{0,p}[k], \bar{h}_{1,p}[k], \cdots, \bar{h}_{M-1,p}[k]]^{\mathrm{T}}$, and $\bar{\mathbf{h}}_{m,p} = \mathbf{F}_N \mathbf{h}_{m,p}$. In addition, $\mathbf{D}_k$ is a diagonal matrix as

$$\mathbf{D}_k = \mathrm{diag}\left\{ \left\|\bar{\mathbf{h}}_0[k]\right\|^2, \left\|\bar{\mathbf{h}}_1[k]\right\|^2, \cdots, \left\|\bar{\mathbf{h}}_{P-1}[k]\right\|^2 \right\}. \tag{2.7}$$

Note that in the asymptotic regime, $P \times P$ normalization matrix $\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}} \boldsymbol{\mathcal{H}}[k]$ also becomes a diagonal matrix equal to $\mathbf{D}_k$. Thus, when M is large MRC and ZF are the same. Moreover, when M tends to infinity, according to the law of large numbers, $\mathbf{D}_k \to M \mathbf{I}_P$. The law of large numbers (LLN) states that the average of the results obtained from a large number of experiments should be close to the expected value and will become closer to the expected value as we have more experiments. There are two types of LLN, weak law (WLLN) and strong law (SLLN). WLLN says that the sample average converges in probability towards the expected value, while

SLLN indicates that the sample average converges almost surely, or with probability 1, towards the expected value. Therefore, the difference of these two types is their convergence types. Here, $\mathbf{D}_k$ convergence is almost surely. Therefore, the combiner output for user q can be written as

$$\bar{y}_q^{\kappa}[k] = \frac{1}{M}\bar{\mathbf{h}}_q^{\mathrm{H}}[k]\bar{\mathbf{r}}^{\kappa}[k], \tag{2.8}$$

for $k = 0, 1, \cdots, N-1$. Substituting $\bar{\mathbf{r}}$ from (2.4) into (2.8), we can rewrite the q^{th} user's signal at the combiner output as

$$\bar{\mathbf{y}}_q^{\kappa} = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\sum_{p=0}^{P-1}\widetilde{\mathbf{H}}_{m,p}\mathbf{d}_p^{\kappa} + \tilde{\mathbf{n}}^{\kappa}, \tag{2.9}$$

where $\tilde{\mathbf{n}}^{\kappa} = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\bar{\mathbf{n}}_m^{\kappa}$. Then, defining an $N \times N$ interference matrix as

$$\mathbf{\Omega}_{q,p} = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\widetilde{\mathbf{H}}_{m,p}, \tag{2.10}$$

we have

$$\bar{\mathbf{y}}_q = \mathbf{\Omega}_{q,q}\mathbf{d}_q + \sum_{\substack{p=0\\p\neq q}}^{P-1}\mathbf{\Omega}_{q,p}\mathbf{d}_p^{\kappa} + \tilde{\mathbf{n}}^{\kappa}. \tag{2.11}$$

Since we can assume that different users' channels are uncorrelated, when the number of BS antennas, M, is large, $\mathbf{\Omega}_{q,q}$ becomes a diagonal matrix with the main diagonal equal to the PDP of the channel in the frequency domain, i.e., $\bar{\boldsymbol{\rho}}_q = \boldsymbol{\rho}_q$. This phenomena is called the channel hardening [11]. Moreover, if the pair $\mathbf{h}_{m,p}$ and $\mathbf{h}_{m,q}$ for $p \neq q$ provide asymptotically favorable propagation, the elements of $\mathbf{\Omega}_{q,p}$ tend to zero for $p \neq q$ as M increases, and multi-user interference will be averaged out [11]. In addition, for large values of M, $\tilde{\mathbf{n}}^{\kappa}$ tends to the mean of the variable $\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\bar{\mathbf{n}}_m^{\kappa}$, which is equal to the mean of noise and is zero. Therefore, in that case the output of the combiner is equal to the transmitted data, i.e., $\hat{\mathbf{d}}_q^{\kappa} = \bar{\mathbf{y}}_q^{\kappa}$.

2.1.2 Downlink data transmission

For downlink transmission, BS applies a precoding on each user's signal to shape it toward a specific location of that user. The value of the precoding is set based on the knowledge of the uplink channel $\mathbf{h}_{m,p}$. If matched filter (MF) precoding is used, the $M \times P$ precoding matrix on the k^{th} subcarrier is

$$\mathbf{W}[k] = \boldsymbol{\mathcal{H}}[k]\mathbf{D}_k^{-1}. \tag{2.12}$$

Since for large values of M, $\mathbf{D}_k \to N\mathbf{I}_P$, the $(m,p)^{\text{th}}$ entry of matrix $\mathbf{W}[k]$ is equal to $\bar{h}_{m,p}[k]/N$. Then, adding all the user signals together, the $N \times 1$ downlink signal from the BS antenna m at OFDM symbol κ before CP assignment can be written as

$$\mathbf{x}_m^{\kappa,\text{dl}} = \frac{1}{M}\sum_{p=0}^{P-1} \mathbf{F}_N^{\text{H}} \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p^{\kappa,\text{dl}}, \tag{2.13}$$

where $\mathbf{d}_p^{\kappa,\text{dl}}$ is the data of user p for downlink transmission on OFDM symbol κ. It is assumed that the CP length is sufficient to mitigate the effect of the multipath channels, and also, to have perfect time synchronization. Thus, after CP removal, the received signal at the user q is obtained as

$$\mathbf{r}_q^{\kappa,\text{dl}} = \sum_{m=0}^{M-1} \mathbf{H}_{m,q}^{\text{dl}} \mathbf{x}_m^{\kappa,\text{dl}}, \tag{2.14}$$

where $\mathbf{H}_{m,q}^{\text{dl}}$ is a circulant matrix whose first column $\mathbf{h}_{m,q}^{\text{dl}}$ is the vector of the downlink channel impulse response between the BS antenna m and the user q. Then, the received signal in the frequency domain is given by

$$\begin{aligned}
\bar{\mathbf{r}}_q^{\kappa,\text{dl}} &= \sum_{m=0}^{M-1} \mathbf{F}_N \mathbf{H}_{m,q}^{\text{dl}} \mathbf{x}_m^{\kappa,\text{dl}} \\
&= \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{F}_N \mathbf{H}_{m,q}^{\text{dl}} \sum_{p=0}^{P-1} \mathbf{F}_N^{\text{H}} \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p^{\kappa,\text{dl}} \\
&= \frac{1}{M}\sum_{p=0}^{P-1}\sum_{m=0}^{M-1} (\mathbf{F}_N \mathbf{H}_{m,q}^{\text{dl}} \mathbf{F}_N^{\text{H}}) \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p^{\kappa,\text{dl}} \\
&= \frac{1}{M}\sum_{p=0}^{P-1}\sum_{m=0}^{M-1} \widetilde{\mathbf{H}}_{m,q}^{\text{dl}} \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p^{\kappa,\text{dl}},
\end{aligned} \tag{2.15}$$

where $\widetilde{\mathbf{H}}_{m,q}^{\text{dl}} = \text{diag}(\bar{\mathbf{h}}_{m,q}^{\text{dl}})$, and $\bar{\mathbf{h}}_{m,q}^{\text{dl}} = \mathbf{F}_N \mathbf{h}_{m,q}^{\text{dl}}$. Based on the reciprocity of the transmission channel, the downlink channels in the frequency domain are equal to the Hermitian transpose of the frequency domain uplink channels, i.e., $\bar{\mathbf{h}}_{m,q}^{\text{dl}} = \bar{\mathbf{h}}_{m,q}^{*}$. Therefore, the received signal at user q can be rewritten as

$$\begin{aligned}
\bar{\mathbf{r}}_q^{\kappa,\text{dl}} &= \left\{ \frac{1}{M}\sum_{m=0}^{M-1} \widetilde{\mathbf{H}}_{m,q}^{\text{H}} \widetilde{\mathbf{H}}_{m,q} \right\} \mathbf{d}_q^{\kappa,\text{dl}} + \sum_{\substack{p=0 \\ p\neq q}}^{P-1} \left\{ \frac{1}{M}\sum_{m=0}^{M-1} \widetilde{\mathbf{H}}_{m,q}^{\text{H}} \widetilde{\mathbf{H}}_{m,p} \right\} \mathbf{d}_p^{\kappa,\text{dl}} \\
&= \mathbf{\Omega}_{q,q} \mathbf{d}_q^{\kappa,\text{dl}} + \sum_{\substack{p=0 \\ p\neq q}}^{P-1} \mathbf{\Omega}_{q,p} \mathbf{d}_p^{\kappa,\text{dl}},
\end{aligned} \tag{2.16}$$

As in uplink transmission, $\mathbf{\Omega}_{q,p}$ corresponds to the multiuser interference which tends to zero when the number of BS antennas is large. In addition, $\mathbf{\Omega}_{q,q}$ is also a diagonal matrix with the main diagonal equal to the PDP of the channel in the frequency domain, $\bar{\boldsymbol{\rho}}_q$. Thus, if the user knows the PDP of its channel, it can recover its signal from the received signal as

$$\hat{\mathbf{d}}_p^{\kappa,\mathrm{dl}} = \mathbf{\Omega}_{q,q}^{-1}\bar{\mathbf{r}}_q^{\kappa,\mathrm{dl}}. \tag{2.17}$$

2.2 Synchronization in massive MIMO

All the advantages of massive MIMO systems rely on accurate synchronization. Massive MIMO systems heavily suffer from synchronization errors, such as, timing offset (TO) and carrier frequency offset (CFO), [12–14]. In OFDM, the effects of TO can be absorbed into the cyclic prefix (CP) provided that an adequately long CP is utilized, [15]. However, frequency synchronization is still a challenging problem. Although there exists a substantial amount of work reported on the frequency synchronization for conventional multi-user OFDM systems, [16–18], not all of them are applicable to massive MIMO systems. In other words, when the number of BS antennas grows, the complexity of the conventional synchronization algorithms increases to an extent that may not be practical.

The frequency synchronization technique consists of two steps: CFO estimation and CFO compensation. In the following, we present the problem associated to each step, and then, we mention the existing solutions in the literature and their shortcomings.

2.2.1 CFO estimation

In massive MIMO systems, all the BS antennas can be co-located at one BS at the center of a cell called co-located antenna (CA) layout, or they can be grouped into geographically distributed clusters in each cell and connected to a central processing unit (CPU) by fiber which is known as distributed antenna (DA) layout. In CA layout, the transmit power can be reduced, a larger number of BS antennas can be integrated, and it can be simply upgraded by installing large scale antenna arrays on the existing BS sites. On the other hand, by distributing the BS antennas, the minimum access distance from each user to the BS antennas can be significantly reduced, on average resulting in a much improved channel condition. However, its implementation is very expensive, especially when the number of BS antennas is large.

Moreover, in CA massive MIMO systems, the BS is equipped with coherent oscillators at all the antennas. Therefore, the synchronization errors of the received signals at different BS antennas are the same. While, in case of DA layout, the synchronization errors are different at each BS antenna. However, current massive MIMO testbeds, e.g. [19], use coaxial cables to distribute synchronization signals from a master clock source in CPU to slave devices and, subsequently, to the RF front-ends at each BS antenna with DA layout. Therefore, all the BS antennas can be synchronized, and as in CA massive MIMO systems, they can experience the same synchronization error.

It is worth noticing that the users' signal have different amount of CFO, and CFO estimation should be performed separately for each user. On the other hand, in massive MIMO systems, the users share all the time and frequency resources and their orthogonality is provided by spatial diversity. In order to exploit the spatial diversity, the channel state information (CSI) of the users are required, while such an information is not available at the CFO estimation stage. This makes the CFO estimation a challenging task in massive MIMO systems.

In the past decade, a number of studies have been conducted to address the synchronization problem in OFDM-based MIMO systems. The work in [20] derived a joint CFO and channel estimator based on the maximum likelihood (ML) criterion for small-scale MIMO OFDM system, whose computational complexity exponentially increases with the number of users. A sub-optimal estimation algorithm using constant amplitude zero auto-correlation (CAZAC) training sequences was proposed in [21]. Another ML-based estimator with iterative interference cancellation was proposed in [22] for coordinated multi-point (CoMP) MIMO OFDM systems. The authors in [23] developed an algorithm for joint CFO compensation and multi-user detection in small-scale multi-user MIMO systems. In another work [24], the authors proposed a pilot-based two stage CFO and phase noise (PN) compensation for point to point high frequency MIMO OFDM, where both stages use a conventional least squares (LS) method and need to calculate the inverse of a full rank matrix. In [25], authors proposed a joint ML-based CFO and channel estimation method which requires a multidimensional grid search. To solve this issue, the authors in [26] converted the ML CFO estimator into a set of line search problems. However, this algorithm still requires per antenna CFO estimation for all the users leading to a substantial amount of computational burden.

To reduce the complexity of frequency synchronization for massive MIMO systems, some works combine the received signals at all the BS antennas and extract the users' CFOs from the resulting signal. The authors in [27] proposed a joint spatial-frequency alignment technique in which users' pilots can be distinguished

using their estimated angles of arrival. However, when the users are not spatially separated, their angles of arrival are very close, if not the same. As a result, the users' signals cannot be accurately distinguished from each other. In [28], an improved user grouping scheme has been designed to deal with this issue, where the CFO estimation and data detection are jointly performed for the users that are close to one another. Despite the imposed computational burden due to the multi-user interference (MUI) cancellation, this technique still cannot accurately compensate the CFO effects. In [29], the authors proposed a scattered pilot-based frequency synchronization technique which can estimate CFO for each user individually by exploiting the spatial dimensions offered by the large number of BS antennas. The authors also designed a beamforming matrix for MUI cancellation. The amount of complexity of this technique is also large and goes up rapidly by increasing the number of BS antennas. In [30], the authors proposed a blind CFO estimation technique based on the covariance matrix of the received signals over all BS antennas in the frequency domain, where different null subcarriers are assigned to different users. This approach is followed by a blind MUI cancellation and channel estimation. A more computationally efficient blind CFO estimation was proposed in [31] which cancels the MUI by using the eigendecomposition of the covariance matrix of the received signals over all the BS antennas in frequency domain. After MUI cancellation, the system is like a set of parallel single user signals, and the conventional single user CFO estimation proposed in [32] is performed. In addition, an iterative CFO estimation procedure was proposed to enhance the MUI cancellation and CFO estimation accuracy in [31]. In each iteration of the algorithm proposed in [32], all received signals are taken to the frequency domain and the eigendecomposition of their covariance matrix is calculated which adds a large amount of computational complexity to the system.

A low complexity CFO estimation was proposed in [33], where the authors used series of impulses for the users' pilots to provide time-domain orthogonality among different users. Hence, the received signals on each sample carry the information of only one user. Therefore, CFO of a certain user can be estimated from the combination of the received signals on the respective samples over all the BS antennas. The complexity of this approach increases only linearly with the number of BS antennas and is also independent of the number of users. However, long pilot sequences are required for an accurate estimation, especially when we need to serve many users. This technique is originally proposed for single carrier systems, while it can be straightforwardly extended to multicarrier systems. The performance of this technique is studied in [34]. Since the proposed technique in [33] suffers from high peak to average power ratio (PAPR), the spatially averaged periodogram based CFO

estimation was proposed which uses a constant envelope pilot sequence to preserve users' orthogonality through different phases. This technique still needs a long pilot sequence, and also its computational complexity is higher than for our proposed techniques.

In summary, the aforementioned techniques either suffer from high computational complexity or performance loss.

2.2.2 CFO compensation

In the conventional massive MIMO systems, the CFO compensation takes place in time domain. In fact, once the CFO is estimated, its effect is removed by multiplying the Hermitian transpose of the CFO matrix to the received signal at each BS antenna. It is important to notice that since the CFO is different for different users, it should be compensated for each user individually. After CFO compensation, each user requires a separate receiver which includes M data streams, where M is the number of BS antennas. Considering the large number of BS antennas in massive MIMO systems, this leads to a tremendous amount of computational complexity. It should also be noted that this computational burden increases linearly with the number of BS antennas and/or the number of users in the network.

2.3 Channel estimation for massive MIMO

As the scale of the antenna array increases or the number of users in the network grows, the acquisition of channel state information (CSI) becomes a bottleneck for massive MIMO performance [35]. Many studies have been conducted to design efficient and reliable techniques for massive MIMO channel estimation. Channel estimation can be either data-aided or blind. In data-aided techniques, a known data sequence called pilot is transmitted to help the receiver to estimate the transmission channel. On the other hand, in blind channel estimation technique, the receiver only relies on some properties of the signal, e.g. cyclostationary of the signal or the sparsity of the channel. Although the data-aided estimation is more accurate than the blind estimation, considerable amount of resources is wasted on transmitting known signal which reduces the spectral efficiency of the system.

2.3.1 Data-aided channel estimation

Considering massive MIMO system in TDD mode, channel estimation should be performed once per coherence time-frequency block of T_c seconds and W_c Hz as illustrated in Fig 2.1. It is assumed that T_c and W_c are smaller or equal to the

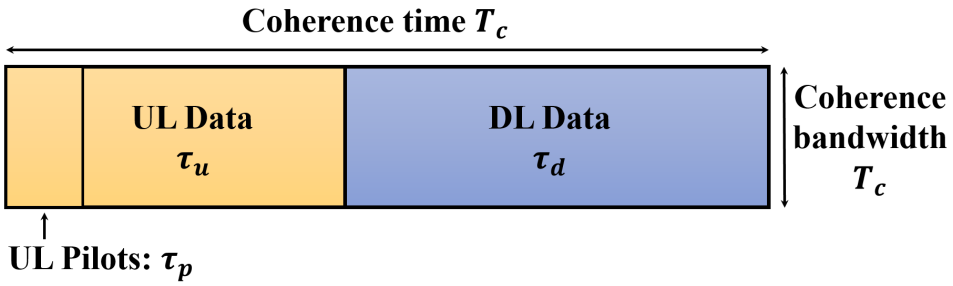

Figure 2.1 A coherent time-frequency block [36] .

coherence time and the coherence bandwidth of all users, respectively. In case of data-aided channel estimation, this coherence block is divided into three phases; uplink pilot transmission, uplink data transmission, and downlink data transmission, [36]. In the estimation phase, users transmit a sequence of known pilot symbols to the BS, and the BS estimates the channel coefficients based on the knowledge of the pilot signals. In the uplink data transmission phase, the BS can detect transmitted data from users based on the estimated channels. Due to the reciprocity of the channel, the downlink channel, which is the conjugate transpose of the uplink channel, is known to the BS. Therefore, there is no need for downlink pilot or feedback of CSI. In fact, the BS can perform a pre-coding on the data symbol based on the estimated channels before transmitting the downlink signals, and as it thoroughly explained in section 2.1.2, the users can detect the transmitted data without any knowledge of the channel impulse responses.

It is worth mentioning that the number of pilot symbols should be equal or greater than the number of users to assure a vanishing channel estimation error, which means a substantial portion of resources is lost [35, 37]. In [38], it is shown that up to half of the coherence time may be dedicated to pilot transmission.

In order to determine the number of required pilots in OFDM-based massive MIMO system, we can define the so called frequency smoothness interval, [6], as

$$N_{smoth} = \frac{1}{T_g \Delta f}, \tag{2.18}$$

where Δf denotes the subcarrier spacing and T_g is the guard or CP interval as $T_g = T_s - T_u$ with T_s and T_u equal to the OFDM symbol time interval and the useful symbol duration, respectively. In fact, the frequency smoothness interval is the reciprocal of the guard interval based on subcarrier spacings. In other words, channels can be assumed to be piecewise-constant over N_{smoth} successive subcarriers, and consequently, only one pilot symbol per smoothness interval is sufficient per user.

Moreover, the maximum number of users a network can have is

$$P_{max} = \tau_p N_{smooth}, \tag{2.19}$$

where τ_p is the number of OFDM symbols dedicated to pilot transmission. For pilot allocation, one can assign each user one unique time-frequency slot for its pilot (e.g., one subcarrier within each smoothness interval and within one OFDM symbol). More generally, a set of orthogonal pilot sequences with the length of $\tau_p N_{smooth}$ can be assigned to the users to assure their orthogonality [6]. Then, channel can be estimated by least-square (LS) channel estimator on pilots subcarriers, and by interpolation for other subcarriers.

Since transmission of pilots wastes a considerable amount of resources, some channel estimation techniques have been proposed to reduce the number of required pilots. The work in [39] proposed a technique based on compressed sensing. In fact, since signals are usually sent through a limited range of angles from massive antenna array, massive MIMO channel exhibits a sparse structure in angular domain [39]. By exploiting this sparsity, compressed sensing is used estimate the channel with small number of pilot signals [40, 41]. In [42], a semi-blind channel estimation based on message-passing algorithm is proposed. It uses the knowledge of a short pilot sequence inserted in user packet for noisy matrix factorization. In [43], a joint channel-and-data estimation is proposed that use the partially detected data as soft pilots to enhance the channel estimation techniques in an iterative algorithm.

Beside the length of the pilot sequence, there is another issue in data-aided channel estimation techniques for multi-cell systems called pilot contamination. In fact, acquisition of accurate CSI is a challenging task since transmission resources are reused among different cells. The same band of frequencies can be reused with factor one, three, or seven among the neighbor cells. This means same pilot sequences can be reused by multiple users in different cells. Thus, if their signals have enough power to reach to a BS in other cell, they will interfere with the user's pilot in that desired cell, and corrupt the pilot-based channel estimation. There are several studies devoted to address the pilot contamination problem [44–46]. It is worth noticing that pilot contamination problem is not a concern in blind channel estimation techniques as they do not relies on pilot signal for channel estimation.

2.3.2 Blind channel estimation

To further reduce the channel acquisition overhead, another line of research is dedicated to blind channel estimation which does not use any prior knowledge of the signals from transmitter [47, 48]. In particular, these works also explore massive

MIMO channel sparsity in angular domain. In fact, once the covariance matrix of the received signals is calculated, the channel can be estimated by factorizing the covariance matrix based on the channel sparsity. However, in these works, the number of BS antennas should be smaller than the length of the coherence time of the channel in discrete domain, which is unlikely in a massive MIMO scenario. In [43], another blind estimation technique is proposed that can work with a large number of BS antennas, but it depends on the sparsity level of the channel. Moreover, the sparse matrix factorization problem is non-convex. Thus, the authors proposed a projection-based bilinear generalized approximate message passing (P-BiG-AMP) algorithm as an approximate solution which imposes a considerable amount of computational complexity to the system. In [42], the authors addressed the aforementioned issues in [43] by assigning a very short pilot sequence into each users' symbol.

2.4 Deep learning

In 1996, machine learning (ML) was defined as a branch of artificial intelligence (AI) to improve performance of a task by experience [49]. In other words, a computer program is said to learn from an experience with respect to some class of tasks and performance measure if its performance improved by that experience at the given task. After a long study from 20th century, several algorithms have been proposed for ML, such as logistic regression, support vector machine (SVM), and artificial neural networks (ANN). Deep learning (DL) is a class of ML algorithms which are based on neural network (NN) architecture. In fact, DL adopts an ANN with several layers called deep neural network (DNN) to progressively learn some features in the input data and model the relationship between the input and the output. The term deep refers to the number of layers in the network, so the more layers, the deeper the network. Traditional ANN contains two or three layers, while DL can have hundreds.

In this section, first, we explain the basics of DL and introduce some DL algorithms that are used in this thesis. Then, we review the applications of DL for physical layer, and more specifically, channel estimation. At the end, we present the DL algorithms as a denoiser in image processing, which is applied to the field of communications in this thesis.

2.4.1 Basics of deep learning

A basic DL model is fully connected feed-forward DNN as it is shown in Fig 2.2, where each neuron (circle) is connected to all the neurons of the adjacent layers. It consists of an input layer, several hidden layers, and an output layer. At each

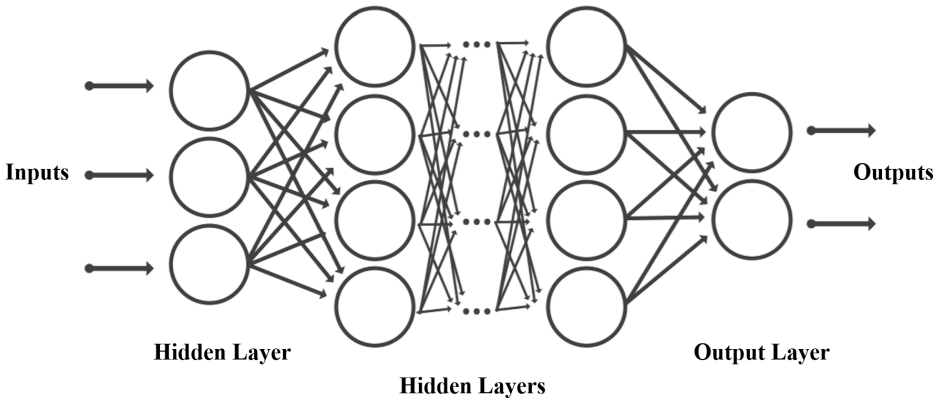

Figure 2.2 Fully connected feedforward NN

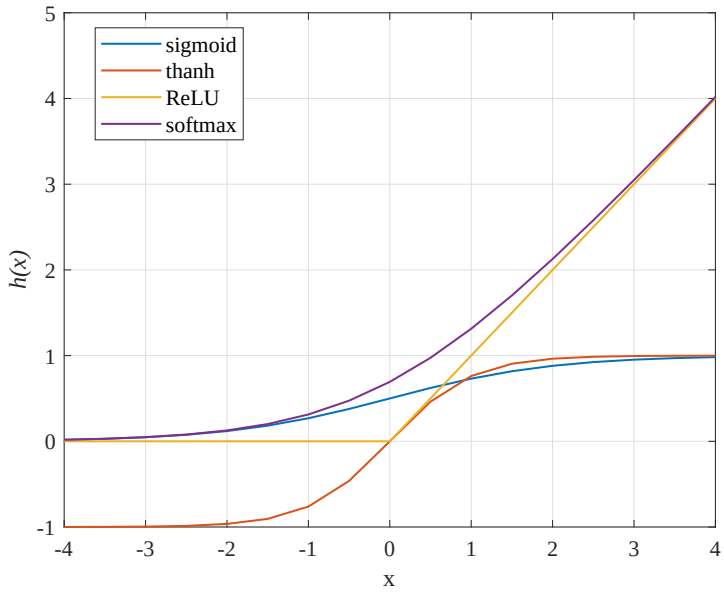

Figure 2.3 Comparison of different activation functions used for DL

layer, the values on the neurons are multiplied by some weights, added together, and passed an activation function to form the values on the next layer. Therefore, if $a_n^{(l)}$ is the value of n^{th} neuron of l^{th} layer of the DNN, then the $a_m^{(l+1)}$ can be calculated as

$$a_m^{(l+1)} = h(\sum_{n=0}^{N_l} \theta_{n,m}^{(l)} a_n^{(l)}) \tag{2.20}$$

where N_l is the number of neurons in layer l, $h(.)$ is an activation function, and $\theta_{n,m}^{(l)}$ is the weight between the neuron n of layer l and the neuron m of layer $l+1$. Different activation function has been introduced for DNN. Such as rectified linear units (ReLu), sigmoid, tanh, and softplus which their functions are presented in equation (2.21).

$$h(x) = \begin{cases} \frac{1}{1+\mathrm{e}^{-\mathrm{x}}}, & \text{sigmoid}, \\ \tanh(\mathrm{x}), & \tanh, \\ \max(0, \mathrm{x}), & \text{ReLu}, \\ \ln(1+\mathrm{e}^{\mathrm{x}}), & \text{softplus}. \end{cases} \tag{2.21}$$

In addition, Fig 2.3 compares these activation functions. ReLU is the most common one which has a zeros value for negative inputs and return the input value as an output for positive inputs.

Depending on the application, a DL algorithm can be trained in supervised, unsupervised or reinforcement manner. In supervised DL algorithms, we need to have a massive set of training data. A training data set contains a large number of labeled data which are the input of the network and their associated output (what we expect the network to give as an output). While in unsupervised algorithms, the network is supposed to learn a pattern in the training data set and without any associated output. The reinforcement learning tries to learn the rules based on the rewards it get after each action. The focus of this thesis is only on supervised learning. In this case, training a DNN means calculating the parameters in the network that minimize the loss function for the given training data set. The loss function should be established to measure the difference between the actual output of the network and the expected value in the output data set. The loss function should be chosen according to the task given to the neural network. Commonly, loss function is mean square error (MSE) as

$$l(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|^2, \tag{2.22}$$

or categorical cross entropy as

$$l(\mathbf{u}, \mathbf{v}) = -\sum_j u_j \log(v_j), \tag{2.23}$$

where $\mathbf{u}$ is the vector of the desired output, and $\mathbf{v}$ is the vector of the actual output of the network. u_j and v_j are the j^{th} element of the output vector, which is the j^{th} neuron of the final layer.

Gradient descent (GD) algorithm is commonly used to optimize the network weights by minimizing the loss function. The gradient of weights required in the optimization are calculated using back-propagation algorithm. However, for a very deep neural network, GD might have a slow convergence. So, to achieve faster convergence and also decrease the computational complexity, classic GD is updated into stochastic GD (SGD). In addition, to avoid severe fluctuation in the training with SGD, mini-batch SGD was introduced, which is a trad-off between classic GD and SGD in terms of complexity and speed. These algorithms may still converge to the local optimal solution. To deal with this issue and further increase the training speed, other algorithms have been proposed such as Adagrad, RMSProp, Momentum, and Adam [50]. Moreover, back propagation (BP) algorithm has been proposed as an efficient method for calculating the gradient of weights through the optimization process.

Another issue that one may encounter is that a trained network, which works well with the training data set, has a poor performance on the test set. It might be due to under-fitting (high-bias) or over-fitting (high-variance). In the case of under-fitting, increasing the training set can solve the problem, and in the other case, adding regularization helps. In order to choose a proper regularization coefficient, we can divide the available data set into the two groups, a training data set and a cross-validation data set. There is no exact formula for selecting this parameter. The best approach is to try a few and see how well they work.

There are various libraries and frameworks established to accelerate the deployment of DL architectures. To name a few, MXNet [51], Caffe [52], TensorFlow [53], Theano [54], Keras [55], and Torch [56] are some frameworks in which a DNN can be designed just in a few lines of code. In these frameworks, the architecture of a DL algorithm should be determined along with all its details, such as the number of layers, neurons, and all their connections. The type of the activation function and the loss function should also be indicated. Depending on the requirement of each scenario, different DNN architectures can be defined. In the following, we introduce some architectures that are used in this thesis.

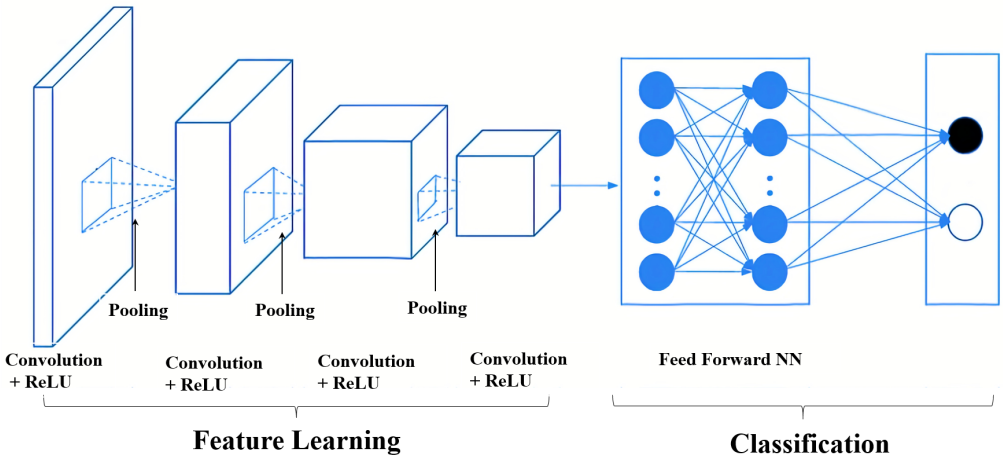

Figure 2.4 CNN architecture [59].

2.4.1.1 CNN architecture

A convolutional neural network (CNN, or ConvNet) is one of the most popular algorithms for deep learning with images and videos. In general, CNN layers can be divide in two groups, feature detection and classification, as it is shown in Fig 2.4. Feature detection layers perform one of the three types of operations on the data: convolution, pooling, or ReLU. Convolution puts the input images through a set of convolutional filters, each of which activates certain features from the images. Pooling simplifies the output by performing nonlinear down-sampling, reducing the number of parameters that the network needs to learn about. ReLU allows for faster and more effective training by mapping negative values to zero and maintaining positive values. These three operations are repeated over tens or hundreds of layers, with each layer learning to detect different features.

Classification layers, like a classic fully connected feed-forward neural network, are composed of an input layer, an output layer, and many hidden layers in between. The number of neurons at the output layer is equal to the number of classes the network can predict.

2.4.1.2 ResNet architecture

As the number of hidden layers and neurons increase the training becomes more difficult. Vanishing/exploding gradients is one of the problem that appear when the network grows [57]. This problem has been addressed to some extent by normalized initiation and intermediate normalization [58]. However, with these solutions, when the network depth increases, its accuracy gets saturated and then degrades rapidly. As it is reported in [59], this degradation is not due to the oversampling, and adding more layers also leads to a higher training error. In [60], a deep residual learning

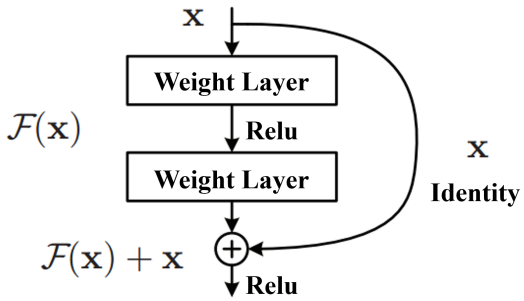

Figure 2.5 A building block of a ResNet architecture [60] .

is proposed as a solution for the mentioned degradation. The authors reformulated the layers as learning residual functions with reference to the layer input, instead of learning unreferenced function. In other words, if $\mathcal{G}(\mathbf{x})$ is the mapping that is supposed to be fit by a few layers (not necessary the entire network), we can let the layers learn the residual function $\mathcal{F}(\mathbf{x}) = \mathcal{G}(\mathbf{x}) - \mathbf{x}$ instead. Then the original function becomes $\mathcal{F}(\mathbf{x}) + \mathbf{x}$. Fig 2.5 shows that by adding a feed-forward shortcut connection the formulation $\mathcal{F}(\mathbf{x})+\mathbf{x}$ can be realized. Since the shortcut is an identity mapping, it does not add any extra parameter or complexity, and the result network can still be trained by SGD and back-propagation, and can be implemented using common libraries (e.g. Cafe) without modifying the solvers. It is shown in [60] that better training (lower training and cross validation error) can be achieved with ResNet network compared to the classical CNN. Moreover, ResNet can add extra accuracy by increasing the number of layers without any performance degradation.

2.4.1.3 U-net architecture

The main idea of U-net is to add some successive layer to a usual convolutional network in which the pooling operator is replaced by upsampling operators. Thus, the U-net architecture can be considered as a cascade of two networks as it is illustrated in Fig 2.6, a contracting network (left side) and expansive network (right side). Since these two networks are symmetric, they yield a u-shaped architecture.

The contracting path follows the typical architecture of convolutional network with ReLU and max pooling operation. As one can see in Fig 2.6, there are five sets of layers in contracting path. Each set has two convolutional layers with a ReLU function as an activation function, and followed by a 2×2 max pooling. After each max pooling, the matrix sizes are halved and the number of channels are doubled.

The expansive path is symmetric to the contracting network, the difference is that it contains upsampling operators instead of the max pooling. Moreover, after each up-conversion in expansive path, a copy of the output of the respective convolutional layer in contracting path (before the max pooling operator) is attached

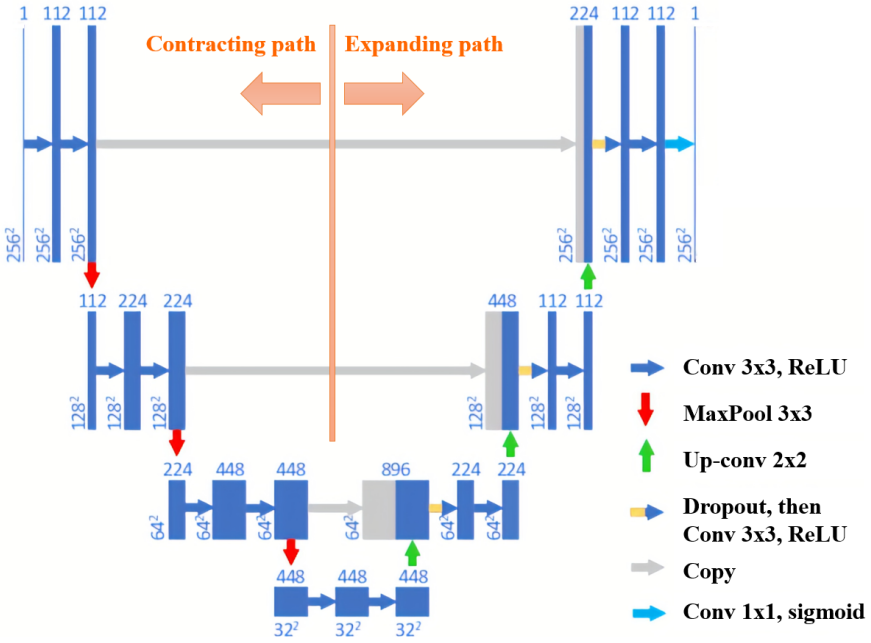

Figure 2.6 U-net architecture for 32 by 32 pixels in the lowest resolution [61] .

to the up-sampled output as the input of the next convolutional layer. This feed-forwarded information allows the network to propagate context information to the higher resolution layers.

In this architecture, the contracting path captures the context in the image, and then, expanding path enables precise localization. It has been shown that this network can outperform the classic convolutional neural network in image segmentation [61]. Note that since in image segmentation, the full context of image is required at the output, this network does not have any fully connected layers at the end, and only uses the valid part of each convolution.

2.4.2 Application of deep learning in physical layer

Limited learning ability of conventional ML algorithms has restricted its application for so many years. From 2012, advanced tools and techniques have dramatically improved DL algorithms. DL architectures such as deep belief networks, recurrent neural networks (RNN) and aforementioned neural networks have been applied to fields including computer vision, speech recognition, natural language processing, social network filtering, medical image analysis, bio-informatics, and drug design, where in some cases, they have produced results superior to human experts. Moreover, the access to the open-source DL libraries increased the potential of using learning algorithms for different fields [62].

Unlike the aforementioned DL applications, where it is difficult to find a mathematical model for the feature representation, various models have been developed

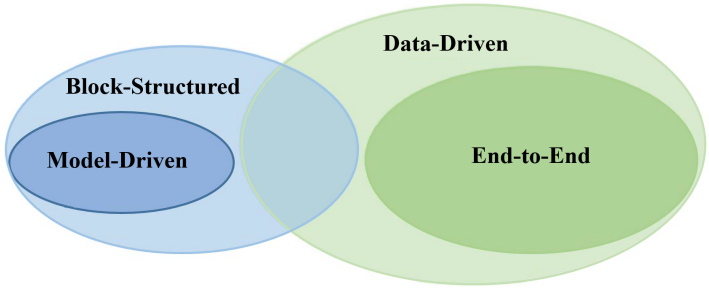

Figure 2.7 Different categories of DL for communication systems [63] .

to describe communication systems. However, there are still some cases that may contain unknown effects which are difficult to be modeled like underwater or molecular communications. Moreover, traditional communication systems consist of several cascaded block, such as modulation, channel encoding, channel estimation, and data detection. Each of these blocks are designed and optimized individually that may not guarantee the global optimality of the system. Thus, the gap between the theory and practice has motivated many researchers to work on intelligent communications.

There are two main approaches of applying DL to the physical layer. First, the entire system can be completely replaced by DNN to achieve a DL-based end-to-end communications. Second, a part of the existing Block-structured communication can be improved/augmented by DL algorithm. A DL-based communication system can be data-driven where the whole system or few blocks are considered as a black box which can be tuned by using labeled data. This learning method requires a large training data set and sufficient computing resources to be accurately trained. On the other hand, there are model-driven techniques in which some prior knowledge of communication system is exploited to reduce the number of trainable parameters in the system, and consequently, shorten the training time. However, the data-driven techniques can be robust under variant circumstance because they have less presumption. While model-driven methods may suffer when the prior assumption is not accurate in the real scenario.

Fig 2.7 demonstrates these different categories of DL application for communication systems as it is suggested in [55]. The end-to-end DL-based communication system is in the data-driven category since the whole system should be replaced with a DNN, and it is trained only by using labeled data and without any presumption. While DL-based block-structured communication systems can be either data-driven or model driven. For data-driven block-structure method, a given block is replaced by a DNN. But in the case of model-driven, the network is simplified by employing the conventional mathematical model or prior knowledge, and as a result, the number of trainable parameters is reduced. For physical layer communications, block-structured model-driven DL method is more attractive since the mathematical

models are usually available and it is possible to model a rough and broad solution with DNN for one or several cascade functions to enhance the system performance.

2.4.2.1 End-to-end communications

As it is shown in Fig 2.8, in DL-based end-to-end communication systems, transmitter is replaced by a DNN which decode the source data. The channel is considered as one unknown layer of the end-to-end system, and the receiver is another DNN which decodes the received signal to recover the source data [63]. The loss function is defined based on an end-to-end loss on the recovery accuracy. The weights of the model are optimized in a supervised learning manner using SGD with the gradients of the loss function back-propagated from the output layer to the input layer. However, the gradient cannot be back-propagated through an unknown channel, and any assumption of channel transfer function will bias the model. Since an accurate channel transfer function is difficult to be obtained in real communication system, the performance of the trained model will be dropped in practice.

There are some techniques proposed to solve the missing gradient problem. In [64], the end-to-end system is trained based on the reinforcement learning in which the transmitter is considered as an agent and the channel plus receiver is regarded as an environment. Then, the agent learns to take actions to maximize the cumulative rewards received from environment. The end-to-end loss is fed back from the receiver to transmitter as the reward from environment. Moreover, in [65] and [66] an approach based on conditional generative adversarial net (GAN) is proposed which circumvent the channel information requirement by modeling the conditional distribution of the channel. In this technique, there are three DNNs for transmitter, receiver and channel generator which are trained iteratively. The conditional GAN needs received pilot signals to learn to mimic the channel effect. Then, the gradient can back-propagate through the trained conditional GAN. Since the received pilot information along with the received signal is fed to this network, the output is specific to the instantaneous channel, and thus, it can be applied to a more realistic time-varying channels. The work in [67] also trains a DL-based end-to-end system in real channel without any assumption on channel models by utilizing stochastic perturbation. [68] extend the idea of end-to-end communication for OFDM-based systems. The works in [69] and [70] consider such a system for single user MIMO scenario.

In these works, the communication system is considered as a black-box modeled by DL architecture and no explicit CSI is obtained. Therefore, the proposed techniques in these papers are not effective in massive MIMO systems where the CSI is required to exploit the spatial diversity. Moreover, these papers studied single

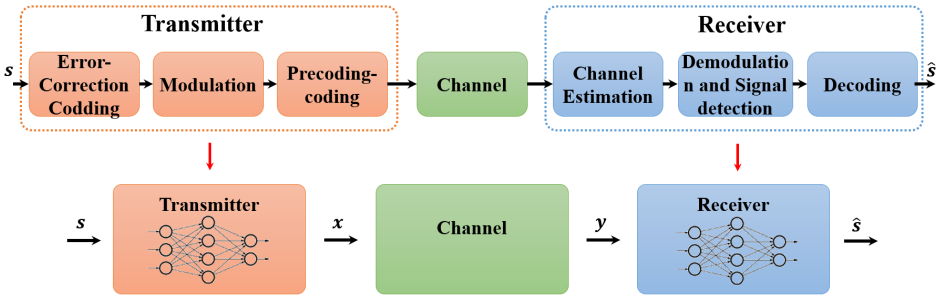

Figure 2.8 The DL-based end-to-end communication system [69] .

antenna or small scale MIMO systems, and their techniques are not compatible with massive MIMO systems.

2.4.2.2 Block-structured communications

DL can be applied to block structured communication systems for different tasks, such as channel coding [71], or CSI compression and reconstruction in FDD massive MIMO [72, 73]. Regarding channel estimation, the work in [74] proposed a joint channel estimation and data detection algorithm for OFDM receiver based on DNN. In fact, DNN is trained to recover transmitted data symbol by feeding the received data symbol and pilots as inputs to the network. Another work in [75], proposed a DL-based channel estimation techniques in which received pilot values are considered as a low-resolution image, and the channel is estimated using image super-resolution network cascaded with a image restoration network. Since these works are for single antenna systems, their application to massive MIMO system will impose huge amount of computational complexity to the receiver.

In [76], a low-complexity channel estimator with embedded NN is proposed for massive MIMO systems. However, the author assumed a flat fading channel and a single user case. The proposed technique also requires pilot signals to estimate the channel. In [77], a framework that integrates massive MIMO into deep learning is proposed to estimate direction of arrival (DOA) and channel impulse responses. In fact, a DNN covers the whole massive MIMO receiver and is trained to estimate DOA. The channels are calculated based on the estimated DOA, plus the transmitted pilot sequence. Moreover, if the number of BS antennas increases, the size of DNN layers should grow accordingly, which leads to a considerable computational burden.

2.4.3 Deep neural network denoiser

Image denoising is a classical topic in the field of image processing, since it is an essential step in many practical applications. The goal of image denoising is to

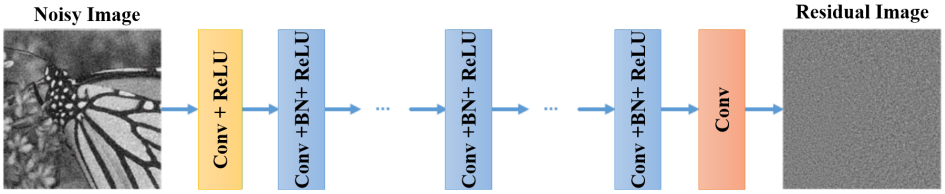

Figure 2.9 DnCNN denoiser [83] .

recover a clean image from a noisy observation. Various models have been developed image denoising, such as non-local self-similarity (NSS) models [78], sparse models [79], gradient models [80] and Markov random field (MRF) models [81]. Despite their high denoising quality, most of them involve a complex optimization problem in the testing stage which usually is a non-convex problem. This make the the denoising process time consuming and very complex. To overcome this drawback, some learning methods have been designed that do not need the optimization process in the test phase.

Authors in [82] proposed an image denoiser technique based on CNN architecture named denoising convolutional neural networks (DnCNNs). DnCNN uses the residual learning, [60], and batch normalization, [58], to improve the denoising performance and to speed up the training process. Fig 2.9 presents the architecture of DnCNN, which consists of three sets of layers. The first one, has only convolutional layer and ReLU, the middle layers have convolutional layer plus batch normalization and ReLU, and the final layer consists only one convolutional layer. As it is suggested in [82], this algorithm learns the residual noise instead of the original image. Then, this detected noise should be subtracted from the input noisy observation to achieve the clean image. Moreover, the DnCNN has been applied in the field of communications for denoising signals. In [83], a denoising-based approximate message passing (LDAMP) neural network is proposed where DnCNN is applied in an iterative manner to estimate the channel for mm-Wave massive MIMO systems. However, this work has studied only a single user scenario, and its performance is the same as the compressive sensing based CSI estimation.

2.5 Advanced waveforms

Media-rich bandwidth hungry applications in the next generation of wireless networks need more localized signals in time and frequency than achievable by orthogonal frequency division multiplexing (OFDM) [84]. This paves the way towards more efficient carrier aggregation and more relaxed synchronization requirements than OFDM. Therefore, in the journey towards finding an appropriate waveform for the physical layer (PHY) of next generation wireless networks, a number of waveforms

have been revisited as evolutions of OFDM, [85], or new filter bank multicarrier (FBMC) based waveforms have appeared in the literature. In the following, we review some of these waveforms.

2.5.1 Evolutions of OFDM

OFDM has been the most popular signaling method in both wired, [86], and wireless [87] systems. It offers a minimum complexity compared to other waveforms. Moreover, multipath effect of channel impulse responses can be simply eliminated by adding a cyclic prefix (CP) to each OFDM symbol. However, using CP reduce the bandwidth efficiency of OFDM systems. OFDM is also very sensitive to frequency synchronization error. Moreover, the subcarriers filter in OFDM have large amount of out of band (OOB) emission, which means users can not effectively localized their signals. Thus, in some use cases such as cognitive radio, either it causes considerable amount of interference to the adjacent users, or more bandwidth should be wasted for longer guard band.

Inspired by OFDM modulation and to address its drawbacks, some waveforms have been introduced [88]. In generalized frequency division multiplexing (GFDM), filtering is imposed on each subcarrier band to minimize the overlapping among subcarriers [89]. Therefore, in case of inaccurate synchronization, GFDM has less inter-user interference compared to OFDM system. Moreover, in GFDM, one CP can be assigned to many GFDM symbols or even one packet (provided that the channel variation over each packet is negligible). This can significantly improve the bandwidth efficiency. However, in GFDM, adjacent subcarriers get more interference from each other which its compensation adds more complexity to the system and causes some performance loss. Later, the concept of GFDM is extended to FBMC, and circular-FBMC (C-FBMC) is introduced to address the aforementioned issues in GFDM [90]. Another waveform following same principle as GFDM is cyclic-block filtered multitone (CB-FMT) [91]. Unlike GFDM and C-FBMS, the subcarrier filters have no overlapping in CB-FMT which leads to bandwidth efficiency loss.

Moreover, filtered OFDM (f-OFDM) [92] has been proposed to reduce the spectral leakage in OFDM. Single carrier frequency division multiple access (SC-FDMA) has been introduced to reduce the peak to average power ratio (PAPR) of OFDMA. Unique word OFDM (UW-OFDM) encodes reduced set of QAM symbols to an OFDM symbol instead of adding CP [93]. The CP in the conventional OFDM is replaced by a set of zero samples in zero-padded OFDM (ZP-OFDM) [94], and by a known sequence in known symbol padded OFDM (KSP-OFDM) [95]. Universal filtered multicarrier (UFMC) is another alternative to OFDM which is a combination of f-OFDM and ZP-OFDM [96].

2.5.2 FBMC-based waveform

An alternative waveform to OFDM, which is resilient against synchronization errors, has a very low level of adjacent carrier leakage ratio (ACLR) and higher bandwidth efficiency due to the absence of cyclic prefix (CP), is filter bank multicarrier with offset-quadrature amplitude modulation (FBMC-OQAM) [97],[98]. FBMC-OQAM is a modulation technique that was initially proposed around 50 years ago [99]. It is considered a strong candidate for the PHY layer of 5G networks [100]. Recently, frequency spreading FBMC (FS-FBMC), as a new implementation of FBMC-OQAM has been proposed that is simpler and more efficient than the conventional implementation in terms of equalization and synchronization [97]. However, OQAM is used in FBMC-OQAM to preserve orthogonality in real domain and avoid interference among adjacent subcarriers. Utilization of OQAM in FBMC-OQAM requires doubling the processing rate and hence a higher computational complexity compared with OFDM [101]. Moreover, OQAM is unable to efficiently support multiple-input multiple-output (MIMO) capabilities introduced in the current wireless standards. Therefore, an interesting FBMC technique named FBMC-QAM is recently introduced that can transmit quadrature amplitude modulation (QAM) rather than OQAM and support MIMO [102, 103]. To preserve orthogonality among all the subcarriers, FBMC-QAM uses different filters for the even and odd subchannels. However, the authors in [103] do not consider the time-frequency localization in their derived orthogonality conditions. As a result, this system suffers from a large amount of out-of-band emissions. To overcome this problem, further enhancements can be achieved through optimizing the trade off between the spectral confinement and self-interference [104]. Moreover, this system is still under investigation and its sensitivity to synchronization errors especially in multiuser scenarios is yet to be studied.

The main focus of our work is on filter bank multicarrier with pulse amplitude modulation (FBMC-PAM), i.e., a new multicarrier modulation technique based on lapped transform (LT) [102]. FBMC-PAM is capable of transmitting QAM data symbols at the same rate as OFDM without the need for doubling the processing rate. The sensitivity of this system to synchronization error is analyzed and its superiority to OFDM is shown [102, 105]. FBMC-PAM is easy to implement in contrast to FBMC-QAM as the filtering can be implemented in the frequency domain. Consequently, it can take advantage of frequency domain equalization leading to an enhanced performance, thanks to its higher frequency resolution, compared to the single-tap equalizer utilized in the polyphase-based implementations [97]. The overlapping factor in LT-based systems, K, has a fixed value of $K = 2$. This leads to a shorter ramp-up and ramp-down period than the existing FBMC-based systems

in the literature with overlapping factors larger than 2. On the other hand, FBMC-PAM has a lower frequency resolution than the conventional FS-FBMC systems, with overlapping factors larger than two. It is worth mentioning that opposed to the conventional FS-FBMC systems having $2K-1$ equalization taps per subcarriers, FBMC-PAM has only 2 taps per subcarrier. This leads to a poorer equalization than FS-FBMC systems. To the best of our knowledge, LT-based multicarrier is only investigated in single antenna systems. In this thesis, we study these systems in massive MIMO to take advantage of the self-equalization property that is present in such channels, i.e., the linear combining of the received signals at different antennas averages out the channel distortions, [100], and enhances the equalization accuracy.

2.5.2.1 FBMC-PAM modulation

Lapped Transform was introduced several decades ago in order to enhance the spectral characteristics of the discrete Fourier transform (DFT) and discrete cosine transform (DCT) [106]. As a matter of fact, LT is a second class of cosine modulated filter bank (CMFB) systems, which have been developed independently to the first class, pseudo quadrature mirror filter (QMF). Their main difference is that the pseudo QMF systems lead to a nearly perfect reconstruction, while LT can achieve a perfect reconstruction [107].

The lapped orthogonal transform (LOT) is a particular type of LT in which two time-domain symbols overlap, i.e., overlapping factor is $K=2$ [108]. This technique has been used for speech compression in the last three decades [106]. Recently, in [109], it is proposed to use LOT for data transmission. In order to take advantage of frequency domain equalization as underlined in [110], an implementation using a $2N$-point DFT is introduced for LOT, named lapped-OFDM. Thus, for N subcarriers, the i^{th} output sample of lapped-OFDM transmitter can be shown as

$$x(i)=\sum_{n=-\infty}^{+\infty}\sum_{k=0}^{N-1} d_{n,k}\,g(i-nN)\cos\Big[(n-\frac{1}{2}+\frac{N}{2})(k-\frac{1}{2})\frac{\pi}{N}\Big], \tag{2.24}$$

where $d_{n,k}$ is the data symbol at the k^{th} subcarrier and n^{th} multicarrier symbol, and

$$g(i)=-\sin\Big[(i-\frac{1}{2})\frac{\pi}{2N}\Big], \tag{2.25}$$

is a pulse shaping filter.

In LOT, the transform coefficients are real. Hence, there is a spectral constraint, due to the hermitian symmetry generated by real coefficients [109]. Therefore, in order to avoid this constraint, a complex lapped transform (CLT), which was introduced in [111] as a complex extension of LOT, is utilized to build a multicarrier

transmission system [105]. It can transmit $2N$ real data symbols at rate $1/N$, and the i^{th} sample of its transmitted symbol can be obtained as

$$x(i) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{KN-1} d_{n,k} g(i - nN) e^{j(\frac{\pi}{N})(n-\frac{1}{2}+\frac{M}{2})(k-\frac{1}{2})}. \tag{2.26}$$

Another type of LT with $K = 2$ is modulated lapped transform (MLT) [112]. The MLT is similar to the LOT, but they have two main differences. First, the MLT basis function is not symmetrical. Second, MLT is basically a pseudo QMF system that meets the orthogonality condition of LT, and consequently, has a perfect reconstruction. Similar to CLT, the modulated complex lapped transform (MCLT) can be obtained from the MLT transform. By using MCLT as a multicarrier modulation technique, it is shown that the system can transmit $2N$ PAM data symbols at the rate $1/N$ [102]. Therefore, this system is called FBMC-PAM, and the i^{th} sample of its symbol can be shown as

$$x(i) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{KN-1} d_{n,k} g(i - nN) e^{j(\frac{\pi}{N})(n+\frac{a}{2}+\frac{N}{2})(k+\frac{1}{2})}, \tag{2.27}$$

where

$$g(i) = \sin[(i + \frac{a}{2})\frac{\pi}{2N}]. \tag{2.28}$$

The value of a depends on the type of symmetry chosen for the prototype filter. For $a = 0$ and $a = 1$, the prototype filter has odd and even symmetry, respectively. For a=0, MCLT and cosine-modulated multitone (CMT) are the same [113]. Moreover, in [114], authors mathematically prove that the CMT is similar to the staggered modulated multitone (SMT), or OFDM/OQAM. In the following section, we focus on FBMC-PAM.

Chapter 3

FBMC-PAM for massive MIMO systems

As the standardization activities started to form the foundation of 5G wireless networks, a wave of research begun to find an alternative waveform for OFDM that can better serve the needs of 5G. To this end, some new waveforms have been introduced in the literature. Among them, FBMC-PAM, which is introduced in section 2.5.2.1, looks like a strong candidate for 5G. It has the same symbol rate as OFDM, and lower computational complexity than existing FBMC-based systems. FBMC-PAM also has the smaller overlapping factor than other FBMC-based systems which means shorter ramp-up and ramp-down periods. In this chapter, we studied the application of FBMC-PAM to massive MIMO systems, and investigated its performance in asymptotic regime when the number of BS antennas goes to infinity.

3.1 FBMC-PAM for massive MIMO

We consider the multiuser massive MIMO setup discussed in [6] where P users communicate with the BS in time division duplex (TDD) manner, and BS is equipped with M transmit/receive antennas. All the users in a cell can simultaneously use all the subcarriers. Due to the fact that the channel gain vectors for different users are statistically independent, as the number of BS antennas grows large, the users' signals can be distinguished through their respective channel gain vectors [6]. We also consider the FBMC-PAM structure proposed in [102]. Since the overlapping factor of this system has a fixed amount $K = 2$, we replace this parameter with its value in the rest of the chapter. At the transmitter, data symbols are multiplied by the phase shift matrix, $\mathbf{T} = \text{diag}\{\boldsymbol{\phi}\}$, that is a $2N \times 2N$ diagonal matrix whose elements are

$$[\boldsymbol{\phi}]_k = \frac{1+j}{\sqrt{2}} j^k \mathrm{e}^{j\frac{\pi}{2N}(k+\frac{1}{2})}, 0 \leq k \leq 2N-1. \tag{3.1}$$

Then, the pulse shaping filter $\mathbf{G} = \text{circ}\{\mathbf{g}\}$ is applied, which is a $2N \times 2N$ circulant matrix. Its first column is $\mathbf{g} = [g_0, g_1, \mathbf{0}_{1\times 2N-2}]^{\text{T}}$, where $g_0 = -\frac{1}{2j}\text{e}^{-j\frac{\pi}{4N}}$, $g_1 = \frac{1}{2j}\text{e}^{j\frac{\pi}{4N}}$ and $\mathbf{0}_{N_1\times N_2}$ is an $N_1 \times N_2$ zero matrix. Hence, after IDFT operation, the $2N \times 1$ vector of the n^{th} multicarrier symbol of the p^{th} user can be obtained as

$$\mathbf{s}_{n,p} = \mathbf{F}_{2N}^{\text{H}}\mathbf{G}\mathbf{T}\mathbf{d}_{n,p}, \tag{3.2}$$

where $\mathbf{d}_{n,p}$ is the $2N \times 1$ data vector and $\mathbf{F}_{2N}$ is the $2N$-point normalized DFT matrix with the elements kn

$$[\mathbf{F}_{2N}]_{kn} = \frac{1}{\sqrt{2N}}\text{e}^{-j\frac{2\pi}{KN}nk}. \tag{3.3}$$

Let us define the $N \times 2N$ overlapping matrices

$$\mathbf{\Gamma}_{-1} = [\mathbf{I}_{\text{N}}, \mathbf{0}_{\text{N}\times\text{N}}], \tag{3.4}$$

and

$$\mathbf{\Gamma}_0 = [\mathbf{0}_{N\times N}, \mathbf{I}_{\text{N}}], \tag{3.5}$$

where $\mathbf{I}_N$ is an $N \times N$ identity matrix. Then, after overlap and add, the $N \times 1$ vector of the transmit symbol can be obtained as

$$\mathbf{x}_{n,p} = \mathbf{\Gamma}_{-1}\mathbf{s}_{n-1,p} + \mathbf{\Gamma}_0\mathbf{s}_{n,p}. \tag{3.6}$$

At the receiver side, the signal received at the m^{th} BS antenna can be shown as

$$\mathbf{r}_m = \sum_{p=0}^{P-1} \mathbf{h}_{m,p} * \mathbf{x}_p + \mathbf{v}_m, \tag{3.7}$$

where $*$ indicates the linear convolution operator, $\mathbf{x}_p = [\mathbf{x}_{0,p}^{\text{T}}, \cdots, \mathbf{x}_{Q-1,p}^{\text{T}}]^{\text{T}}$ is the $QN \times 1$ vector that includes Q concatenated FBMC-PAM symbols of the user p, $\mathbf{v}_m$ is the complex additive white Gaussian noise (AWGN), $\mathbf{v}_m \sim \mathcal{CN}(0, \sigma_v^2\mathbf{I}_{NQ})$, where σ_v^2 is a noise variance, and $\mathbf{h}_{m,p}$ is the channel impulse response between the p^{th} user and the m^{th} BS antenna. It is assumed that the entries of the vectors $\mathbf{h}_{m,p}$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance equal to the elements of $\boldsymbol{\rho}$, i.e, $\mathbf{h}_{m,p} \sim \mathcal{CN}(0, \text{diag}(\boldsymbol{\rho}))$. The vector $\boldsymbol{\rho} = [\rho(0), \rho(1), \cdots, \rho(L-1)]^T$ is the channel power delay profile (PDP) and L is the length of the channel impulse response. Considering perfect timing and frequency synchronization, the first block at the receiver is a sliding window of size $2N$, which picks $2N$ samples of the received signal in the position of the symbol n.

Hence, the output of the n^{th} window can be written as

$$\mathbf{r}_{n,m} = \sum_{p=0}^{P-1} \left(\mathbf{H}_{m,p}^{(1)} \mathbf{x}_{n,p}^{(1)} + \mathbf{H}_{m,p}^{(0)} \mathbf{x}_{n,p}^{(0)} \right) + \tilde{\mathbf{v}}_m. \tag{3.8}$$

where $\mathbf{x}_{n,p}^{(1)} = [\mathbf{x}_{n,p}, \mathbf{x}_{n+1,p}]^{\mathrm{T}}$ and $\mathbf{x}_{n,p}^{(0)} = [\mathbf{x}_{n-2,p}, \mathbf{x}_{n-1,p}]^{\mathrm{T}}$ are $2N \times 1$ vectors, and $\tilde{\mathbf{v}}_m$ is $\mathbf{v}_m$ that is windowed. Considering (3.6), they can be obtained as

$$\mathbf{x}_{n,p}^{(0)} = \mathbf{\Pi}_{-1} \mathbf{s}_{n-1,p} + \mathbf{\Pi}_0 \mathbf{s}_{n,p} + \mathbf{\Pi}_{+1} \mathbf{s}_{n+1,p}, \tag{3.9}$$

and

$$\mathbf{x}_{n,p}^{(1)} = \mathbf{\Pi}_{-1} \mathbf{s}_{n-3,p} + \mathbf{\Pi}_0 \mathbf{s}_{n-2,p} + \mathbf{\Pi}_{+1} \mathbf{s}_{n-1,p}, \tag{3.10}$$

where

$$\mathbf{\Pi}_{-1} \triangleq \begin{bmatrix} \mathbf{\Gamma}_{-1} \\ \mathbf{0}_{2N \times N} \end{bmatrix}, \mathbf{\Pi}_0 \triangleq \begin{bmatrix} \mathbf{\Gamma}_{-1} \\ \mathbf{\Gamma}_0 \end{bmatrix}, \mathbf{\Pi}_{+1} \triangleq \begin{bmatrix} \mathbf{0}_{2N \times N} \\ \mathbf{\Gamma}_0 \end{bmatrix}. \tag{3.11}$$

$\mathbf{H}_{m,p}^{(1)}$ and $\mathbf{H}_{m,p}^{(0)}$, are $2N \times 2N$ Toeplitz matrices that create the channel affected symbol n and the tail of the adjacent block, respectively, when they are multiplied to the vectors $\mathbf{x}_{n,p}^{(1)}$ and $\mathbf{x}_{n,p}^{(0)}$. They are defined as

$$\mathbf{H}_{m,p}^{(1)} = \begin{bmatrix} h_{m,p}(0) & 0 & 0 & \cdots & 0 \\ h_{m,p}(1) & h_{m,p}(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{m,p}(L-1) & h_{m,p}(L-2) & h_{m,p}(L-3) & \cdots & 0 \\ 0 & h_{m,p}(L-1) & h_{m,p}(L-2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_{m,p}(0) \end{bmatrix}, \tag{3.12}$$

and

$$\mathbf{H}_{m,p}^{(0)} = \begin{bmatrix} 0 \cdots & h_{m,p}(L-1) & h_{m,p}(L-2) \cdots & h_{m,p}(1) \\ 0 \cdots & 0 & h_{m,p}(L-1) \cdots & h_{m,p}(2) \\ \vdots \ddots & \vdots & \vdots \ddots & \vdots \\ 0 \cdots & 0 & 0 \cdots & h_{m,p}(L-1) \\ \vdots \ddots & \vdots & \vdots \ddots & \vdots \\ 0 \cdots & \cdots & 0 \cdots & 0 \end{bmatrix}. \tag{3.13}$$

The received symbol, n, at the m^{th} antenna, after application of DFT can be obtained as

$$\mathbf{y}_{n,m} = \mathbf{F}_{2N} \mathbf{r}_{n,m}. \tag{3.14}$$

Then, the equalization is performed for each user separately. For the sake of simplicity, the matched filter (MF) detector is chosen for equalization. Therefore, if the vector $\overline{\mathbf{h}}_{k,p} = [\overline{h}_{0,p}(k), \overline{h}_{1,p}(k), \ldots, \overline{h}_{M,p}(k)]^{\mathrm{T}}$ contains $2N$ point DFT coefficients of the CIR between the p^{th} user and the BS antennas at subcarrier k, the combiner vector is $\boldsymbol{\psi}_{k,p} = \overline{\mathbf{h}}_{k,p} / \left\|\overline{\mathbf{h}}_{k,p}\right\|^2$, where $\|.\|$ indicates 2-norm or Euclidean norm. Thus, the equalized signal at this subcarrier is obtained as

$$\mathbf{z}_{n,p}(k) = \boldsymbol{\psi}_{k,p}^{\mathrm{H}} \tilde{\mathbf{y}}_{n,k}, \tag{3.15}$$

where $\tilde{\mathbf{y}}_{n,k} = [y_{n,0}(k), y_{n,1}(k), \ldots, y_{n,M}(k)]^{\mathrm{T}}$. Finally, the detected data symbols are given by

$$\widehat{\mathbf{d}}_{n,p} = \Re\left\{\mathbf{T}^H \mathbf{G}^H \mathbf{z}_{n,p}\right\}. \tag{3.16}$$

Note that performance of MF is asymptotically the same as zero forcing (ZF) equalization when M tends to infinity. In fact, the ZF equalizer is obtained as $\boldsymbol{\psi}_k = \bar{\mathbf{H}}_k (\bar{\mathbf{H}}_k^{\mathrm{H}} \bar{\mathbf{H}}_k)^{-1}$, where $\bar{\mathbf{H}}_k = [\overline{\mathbf{h}}_{k,1}, \overline{\mathbf{h}}_{k,2}, \ldots, \overline{\mathbf{h}}_{k,P}]$ is an $M \times P$ matrix. When M tends to infinity, $\check{\mathbf{H}} \triangleq \bar{\mathbf{H}}_k^{\mathrm{H}} \bar{\mathbf{H}}_k$ become a diagonal matrix, and its elements are

$$[\check{\mathbf{H}}]_{p,q} = \overline{\mathbf{h}}_{k,p}^{\mathrm{H}} \overline{\mathbf{h}}_{k,q} = \begin{cases} \left\|\overline{\mathbf{h}}_{k,p}\right\|^2, & p = q, \\ 0, & p \neq q, \end{cases} \tag{3.17}$$

Thus, the ZF equalizer is asymptotically equal to the MF detector. Hence, for the sake of simplicity, we limit our SIR analysis in the next section to the MF detector.

3.2 Asymptotic SIR analysis

In this section, we analytically derive the asymptotic SIR of the massive MIMO detector discussed in the previous section for FBMC-PAM. Since the channel vectors of the users become orthogonal in the large antenna regime, no multiuser interference and additive noise remain. Therefore, without loss of generality, in this section, we consider one active user to study the behavior of the ISI and ICI imposed by the multipath channel. We also investigate if the channel distortions are completely removed when there are an infinite number of antennas due to the self-equalization property. To keep the equations simple, we drop the user indices from the user's channel and data matrices and vectors, respectively. Moreover, we do not consider the additive noise in the formulation in this section. To this end, if the tail of the n^{th} multicarrier symbol, i.e., the part of the symbol that is placed out of the respective

sliding window, is added and subtracted in (3.8), then, the received signal at the DFT input can be written as

$$\mathbf{r}_{n,m} = \boldsymbol{\mathcal{H}}\mathbf{x}_n^{(1)} - \mathbf{H}_m^{(0)}\mathbf{x}_n^{(1)} + \mathbf{H}_m^{(0)}\mathbf{x}_n^{(0)}. \tag{3.18}$$

where $\boldsymbol{\mathcal{H}} = \mathbf{H}_m^{(1)} + \mathbf{H}_m^{(0)}$ is a circulant matrix, and its multiplication by vector $\mathbf{x}_n^{(1)}$ is equivalent to the circular convolution of the channel impulse response and the vector $\mathbf{x}_n^{(1)}$. Therefore, the effect of this matrix can be eliminated by frequency domain equalization of the received signal with the reciprocal of the channel gain in a given subcarrier. Thus, the first term on the right hand side of (3.18) does not lead to any interference, and after applying the filtering matrix $\mathbf{G}$ and the phase shift matrix $\mathbf{T}$, the n^{th} symbol can be extracted from this term thoroughly. Accordingly, The other two terms cause interference.

By substituting $\mathbf{x}_n^{(1)}$ and $\mathbf{x}_n^{(0)}$ from equations (3.10) and (3.9) into (3.18), the interference part of the equation (3.18) can be shown as

$$\mathbf{q}_{n,m} = \mathbf{q}_{n,m}^{(1)} + \mathbf{q}_{n,m}^{(0)}, \tag{3.19}$$

where

$$\mathbf{q}_{n,m}^{(1)} = -\mathbf{H}_m^{(0)}\boldsymbol{\Pi}_{-1}\mathbf{s}_{n-1,p} - \mathbf{H}_m^{(0)}\boldsymbol{\Pi}_0\mathbf{s}_{n,p} - \mathbf{H}_m^{(0)}\boldsymbol{\Pi}_{+1}\mathbf{s}_{n+1,p}, \tag{3.20}$$

and

$$\mathbf{q}_{n,m}^{(0)} = \mathbf{H}_m^{(0)}\boldsymbol{\Pi}_{-1}\mathbf{s}_{n-3,p} + \mathbf{H}_m^{(0)}\boldsymbol{\Pi}_0\mathbf{s}_{n-2,p} + \mathbf{H}_m^{(0)}\boldsymbol{\Pi}_{+1}\mathbf{s}_{n-1,p}. \tag{3.21}$$

Assuming that the data symbols are independent, the powers of these terms can be separately calculated and added to each other.

As the number of antennas goes to infinity, according to the law of large numbers, the equalized signal can be obtained as

$$\mathbf{z}_n(k) = \boldsymbol{\psi}_k^{\mathrm{H}}\tilde{\mathbf{y}}_k \to \mathbb{E}\left\{\overline{\mathbf{h}}_k^{\mathrm{H}}(m)\tilde{\mathbf{y}}_k(m)\right\}. \tag{3.22}$$

Thus, considering that data symbols, $s_{n,p}$, have variance of unity, the power of the third term of the right hand side of (3.20) is given by

$$P_3^{(1)} = \frac{1}{(2N)^2}\left\|\Re\left\{\mathbf{T}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{B}_{+1,m}\mathbf{T}\mathbf{G}\right\}\right\|_{\mathrm{F}}^2, \tag{3.23}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes Frobenius norm and

$$\mathbf{B}_{+1,m} \triangleq \mathbb{E}\left\{\overline{\mathbf{H}}_m^{\mathrm{H}}\mathbf{F}_{2N}\mathbf{H}_m^{(0)}\boldsymbol{\Pi}_{+1}\mathbf{F}_{2N}\right\}. \tag{3.24}$$

$\overline{\mathbf{H}}_m$ is a $2N \times 2N$ diagonal matrix with diagonal elements $\left[\overline{\mathbf{H}}_m\right]_{k,k} = \overline{\mathbf{h}}_k(m)$. Expanding $\mathbf{B}_{+1,m}$, its elements can be calculated as

$$[\mathbf{B}_{+1,m}]_{p,q} = \frac{1}{2N}\mathbb{E}\Bigg\{\sum_{n=0}^{2N-1}\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} h_m^*(l')h_m(l)\mathbf{\Pi}_{+1}(n,n') \times \mathrm{e}^{-j\frac{2\pi}{2N}\left(lq-nq-l'p+n'p\right)}\omega(n-l+2N)\Bigg\}, \tag{3.25}$$

where

$$\omega(n) = \begin{cases} 1, & 0 \le n \le 2N-1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.26}$$

Note that $[\mathbf{\Pi}_{+1}]_{nn'} = 1$ only if $n' = n - N$. Thus, after some manipulations, we get

$$\begin{aligned}[\mathbf{B}_{+1,m}]_{p,q} &= \frac{1}{2N}\sum_{n=0}^{2N-1}\sum_{l=0}^{L-1}\rho(l)\mathrm{e}^{-j\frac{2\pi}{2N}(lq-nq-lp+np-pN)} \\ &\quad\times\omega(n-l+2N) \\ &= \frac{(-1)^p}{2N}\sum_{l=0}^{L-1}\rho(l)\mathrm{e}^{-j\frac{2\pi}{2N}(q-p)l}\sum_{n=0}^{l-1}\mathrm{e}^{-j\frac{2\pi}{2N}(q-p)n}.\end{aligned} \tag{3.27}$$

For $p \neq q$,

$$\begin{aligned}[\mathbf{B}_{+1,m}]_{p,q} &= \frac{(-1)^p}{2N}\sum_{l=0}^{L-1}\rho(l)\mathrm{e}^{-j\frac{2\pi}{2N}(q-p)l}\frac{1-\mathrm{e}^{j\frac{2\pi}{2N}(q-p)l}}{1-\mathrm{e}^{j\frac{2\pi}{2N}(q-p)}} \\ &= \frac{(-1)^p}{2N(1-\mathrm{e}^{j\frac{2\pi}{2M}(q-p)})}\left(1-\bar{\boldsymbol{\rho}}(q-p)\right),\end{aligned} \tag{3.28}$$

where $\bar{\boldsymbol{\rho}} = \mathbf{F}_{2N}\tilde{\boldsymbol{\rho}}$ contains $2N$-point DFT samples of zero padded version of the channel PDP, i.e., $\tilde{\boldsymbol{\rho}} = \left[\boldsymbol{\rho}^{\mathrm{T}}, \mathbf{0}_{1\times(2N-L)}\right]^{\mathrm{T}}$. It is worth mentioning that in this chapter, we consider a normalized PDP, i.e., $\sum_{l=0}^{L-1}\rho(l) = 1$.

For $p = q$,

$$[\mathbf{B}_{+1,m}]_{p,q} = \frac{(-1)^p}{2N}\sum_{l=0}^{L-1} l\rho(l) = \frac{(-1)^p}{2N}\bar{\tau}, \tag{3.29}$$

where

$$\bar{\tau} = \sum_{l=0}^{L-1} l\rho(l), \tag{3.30}$$

is the average delay spread of channel. One can easily realize that the power of the third term of (3.21) is equal to the power of the third term in (3.20), i.e. $P_3^{(0)} = P_3^{(1)}$.

Besides, since $\mathbf{H}_m^{(0)}\mathbf{\Pi}_{-1} = [\mathbf{0}_{2N\times 2N}]$, the power of the first terms of (3.20) and (3.21) are $P_1^{(1)} = P_1^{(0)} = 0$. Moreover, with the same approach, the power of the second terms of (3.20) and (3.21), $P_2^{(0)} = P_2^{(1)}$, can be calculated as

$$P_2^{(1)} = \frac{1}{(2N)^2} \left\| \Re \left\{ \mathbf{T}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{B}_{0,m}\mathbf{T}\mathbf{G} \right\} \right\|_{\mathrm{F}}^2, \tag{3.31}$$

where

$$\mathbf{B}_{0,m} \triangleq \mathbb{E}\left\{ \overline{\mathbf{H}}_m^{\mathrm{H}}\mathbf{F}_{2N}\mathbf{H}_m^{(0)}\mathbf{\Pi}_0\mathbf{F}_{2N} \right\}. \tag{3.32}$$

Noting that $\mathbf{\Pi}_0 = \mathbf{I}_{2N}$, we obtain

$$[\mathbf{B}_{0,m}]_{p,q} = \frac{1}{2N(1 - \mathrm{e}^{j\frac{2\pi}{2N}(q-p)})} \left(1 - \bar{\boldsymbol{\rho}}(q-p)\right), \tag{3.33}$$

for $p \neq q$, and

$$[\mathbf{B}_{0,m}]_{p,q} = \frac{1}{2N}\bar{\tau}, \tag{3.34}$$

for $p = q$. Finally, the asymptotic SIR can be derived as

$$\mathrm{SIR} = \frac{1}{2P_3^{(1)} + 2P_2^{(1)}}, \tag{3.35}$$

From (3.35), one can realize that when the number of BS antennas tends to infinity, the SIR saturates at the calculated level. Furthermore, with the same approach, an equation similar to (3.35) can be achieved for any FS-FBMC systems with $K \geq 2$ in order to calculate and compare their saturation level in massive MIMO channels.

In addition, through the SIR relationship derived in this section, the upper bound for SIR performance of the system can be calculated according to the number of subcarriers for a given channel PDP. As a result, the number of subcarriers in a certain band, which in turn defines the subcarrier spacing, can be set to achieve a target SIR.

3.2.1 Comparison with OFDM

As it is proved in section 3.2, the last two terms on the right hand side of equation (3.18) are the source of remaining interference for FBMC-PAM modulation in the asymptotic regime of massive MIMO systems. These terms are associated with the tail of the transmitted symbols after passing through the multipath channel. In OFDM modulation, the tail of the adjacent symbol is removed from the desired symbol when we drop the CP at the receiver. Therefore, the received signal at the

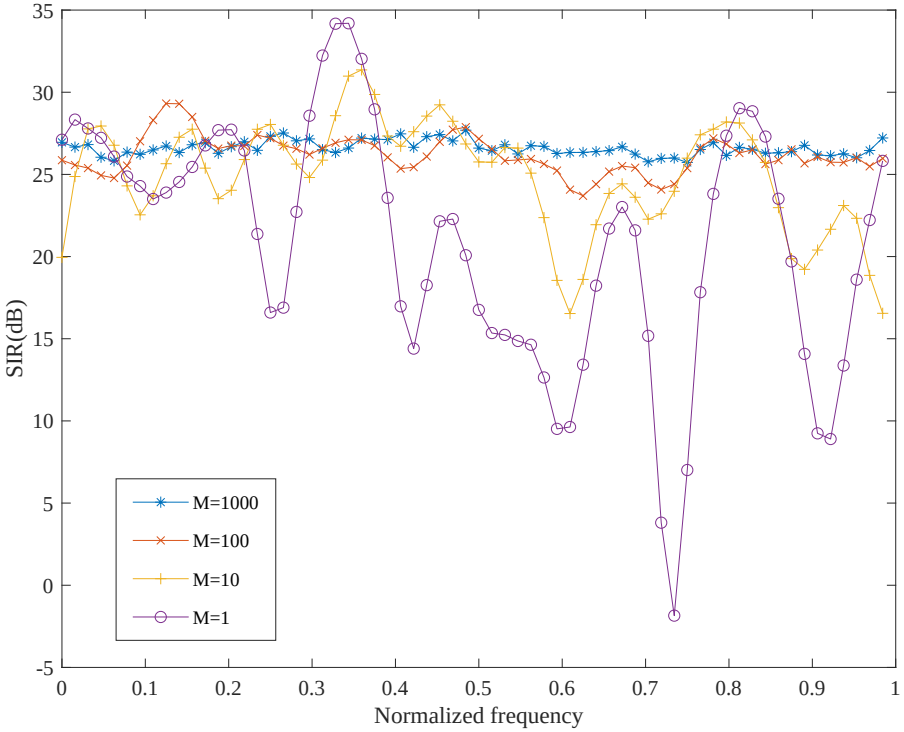

Figure 3.1 SIR performance for $N = 32$ subcarriers and different number of BS antennas.

DFT input can be written as

$$\begin{aligned}\mathbf{r}_{n,m} &= \mathbf{H}_m^{(1)}\mathbf{x}_n^{(1)} + \mathbf{H}_m^{(0)}\mathbf{x}_n^{(1)} \\ &= \boldsymbol{\mathcal{H}}\mathbf{x}_n^{(1)}.\end{aligned} \tag{3.36}$$

where $\boldsymbol{\mathcal{H}} = \mathbf{H}_m^{(1)} + \mathbf{H}_m^{(0)}$ is a circulant matrix, its multiplication by vector $\mathbf{x}_n^{(1)}$ is equivalent to the circular convolution of the channel impulse response and the vector $\mathbf{x}_n^{(1)}$, and its effect can be eliminated by a channel equalizer. Since the interference from multipath channel is removed, and the MUI and noise effects are averaged out in asymptotic regime, the upper bound for the SIR of the OFDM-based massive MIMO systems is infinity. In other words, in OFDM-based massive MIMO systems, the SIR performance unboundedly increases as the number of BS antennas tends to infinity.

3.3 Simulation results

In this section, a broad set of numerical results is presented to confirm the theoretical developments of this chapter. It is worth noting that our simulations are based on

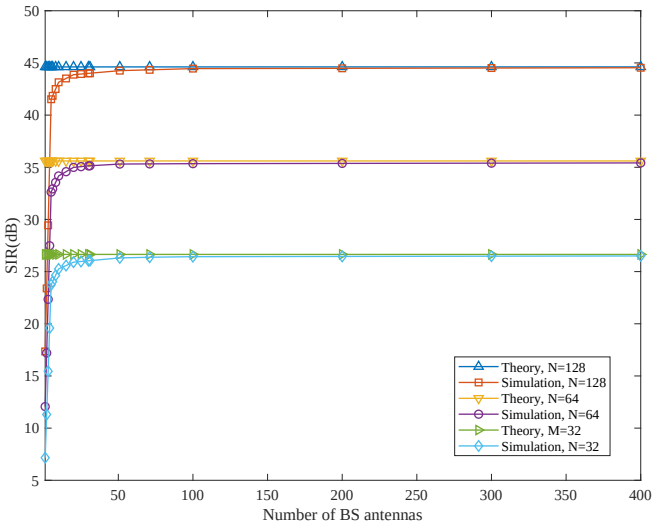

Figure 3.2 SIR as a function of the number of BS antennas, M, for different number of subcarriers, N.

a sample set of channel responses generated according to the SUI-4 channel model proposed by the IEEE802.16 broadband wireless access working group [115].

As it is discussed in [100], FBMC has a self-equalization property in massive MIMO. This means it can approximately flatten the channel over each subcarrier band, so wider subcarriers can be used. Fig. 3.1 provides evidence of this property for FBMC-PAM in massive MIMO. In this figure, the effect of increasing the number of BS antennas on SIR is depicted in a single user scenario. For these simulations, noise free channel is considered in order to study the impact of increasing the number of BS antennas on the equalization accuracy. Note that the number of points along the normalized frequency axis depends on the frequency spacing $(1/2N)$. As it is shown, when the number of antennas grows large, the channel is flattened over all the subcarrier bands.

Then, the analytical SIR relationship derived in equation (3.35) is assessed in Fig. 3.2. The SIR curves are depicted for $N = 32$, 64, and 128 subcarriers in a single user scenario. It is shown that when the number of the BS antennas increases, the SIR is enhanced by about 20dB compared to the single antenna case. As the figure illustrates, the SIR reaches the saturation level when the number of BS antennas is higher than 50. It is worth mentioning that the first line of massive MIMO products including Ericsson AIR 6468, Huawei AAU and Nokia Airscale have 64 transceiver antennas [116–118]. These products are designed for 4G LTE and considered as a pre-5G product, and they are already deployed by many operators such as Sprint [119]. Since the individual antenna element size is smaller in mmWave frequencies, a large number of elements, even more than 100, can be deployed. But, according to Fig. 3.2, we can not gain any performance improvement by increasing the number

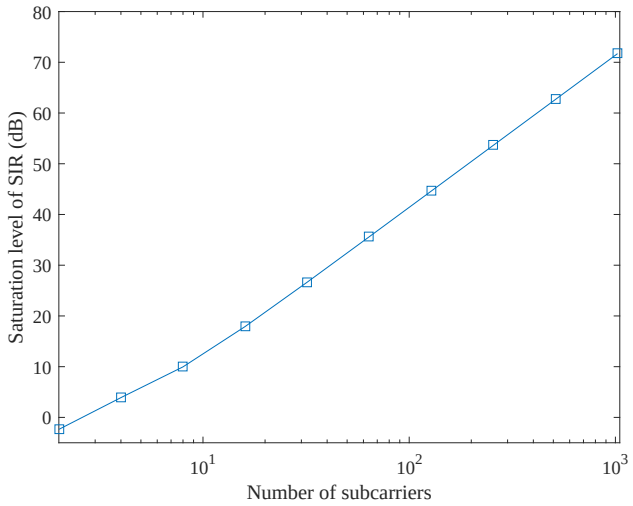

Figure 3.3 SIR saturation level versus number of subcarriers, N.

of BS antennas to more than 50 in the case of FBMC-PAM based massive MIMO system. However, this saturation level is increased by the number of subcarriers. The level of saturation is calculated using equation (3.35), and depicted in Fig. 3.2 where it is shown that the theoretical results perfectly match the simulations, provided that the number of antennas is large enough. Hence, it can be concluded that for a given number of subcarriers, the system performance has an upper bound, and it is impossible to improve it by applying more antennas at the BS.

Fig. 3.3 depicts the SIR saturation level as a function of the number of subcarriers, N. Therefore, in order to hit a given target SIR, N should be properly set. In Fig. 3.4, the BER performance with respect to $\mathrm{E_b/N_0}$ for $N = 64$ is investigated. The term $\mathrm{E_b/N_0}$ is the ratio of signal energy associated with each user data bit to noise power spectral density. Then, $\mathrm{E_s/N_0}$, signal energy associated with each user data symbol to noise power spectral density ratio, or simply SNR, signal to noise ratio, can be calculated as

$$\mathrm{SNR} = \frac{\mathrm{E_b}}{\mathrm{N_0}} + 10\log_{10}(\mathrm{M_{qam}}), \tag{3.37}$$

where M_{qam} is the number of bits per symbol. The curves for different constellation sizes are obtained in the multiuser case, i.e., $P = 4$ users. Note that FBMC-PAM does not have a powerful equalization property due the fact that it uses 2 equalizer taps per subcarrier. Thus, in single antenna systems, the channel distortions is not equalized effectively, and its BER curve has an error floor [102]. As it is shown, by applying FBMC-PAM to massive MIMO, despite the SIR is saturated, it is high enough so that the error floor can be completely removed without any complex equalization. To evaluate the system performance, OFDM is chosen as a

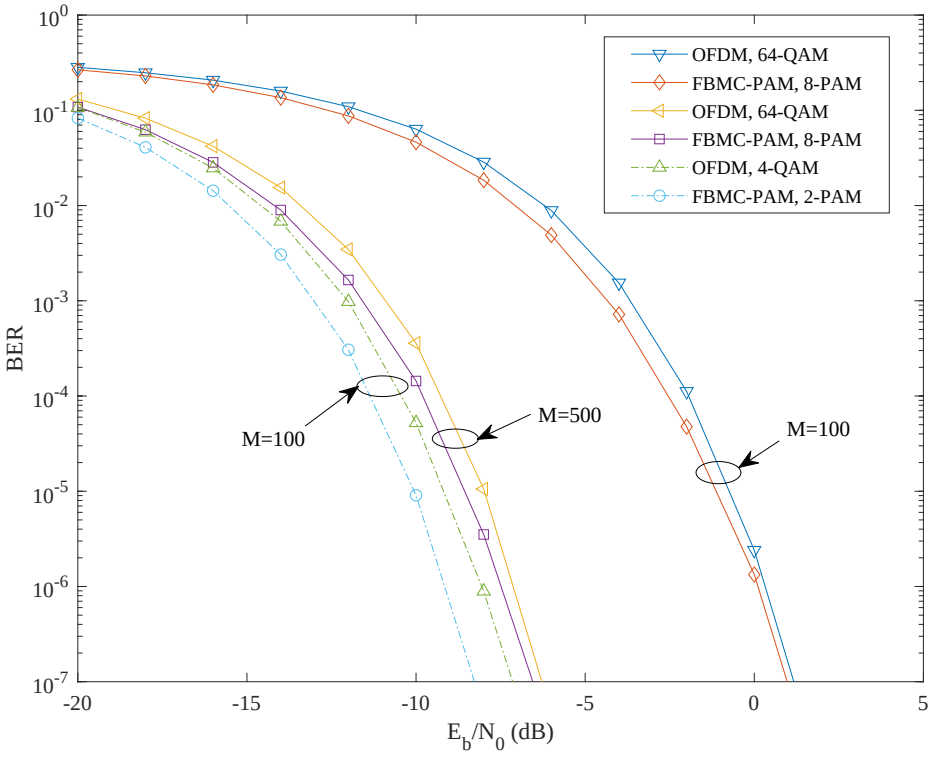

Figure 3.4 BER versus E_b/N_0 for $N = 64$ subcarriers, $P = 4$ users and the multipath channel.

benchmark, because it has an infinite SIR as it is discussed in section 3.2.1. In fact, it can cope with the multipath channel effects by using a CP longer than the length of the channel. We consider the CP length of $N/4 = 16$. The curves are obtained for 2-PAM and 8-PAM for FBMC-PAM, which are equivalent to 4-QAM and 64-QAM for OFDM, respectively. This is because the real and imaginary parts of symbols in FBMC-PAM are modulated separately. Since the absence of CP makes FBMC-PAM more bandwidth efficient, FBMC-PAM has a superior BER performance than OFDM. Furthermore, it is shown that by increasing the number of antennas, the noise effect can be considerably reduced, leading to a substantial performance improvement.

3.4 Conclusion

FBMC-PAM, a new candidate waveform for 5G, was studied in massive MIMO. It was shown that although in massive MIMO channels, multiuser interference as well as channel noise average out, the system performance is limited due to the residual channel distortions even for a very large number of BS antennas. In other words, when the number of BS antennas tends to infinity, the SIR performance

saturates. The level of this saturation, which constitutes an upper bound to the system performance, was analytically calculated in terms of channel parameters and the number of subcarriers, N. Then it was discussed that N should be set appropriately to achieve a target SIR. Simulation results were provided to confirm the theoretical developments. Besides, FBMC-PAM has an equalization problem in single antenna systems leading to an error floor in BER performance. This is due to fact that it only uses two equalizer taps per subcarrier. However, as we have shown in this chapter, FBMC-PAM can sufficiently equalize the multipath channel in massive MIMO systems and get rid of the BER error floor, given that the number of subcarriers, N, is properly chosen. As we have shown, FBMC-PAM has a better BER performance compared to OFDM due to its higher bandwidth efficiency.

Chapter 4

Frequency synchronization for massive MIMO

As discussed in the previous chapter, OFDM seems a reasonable waveform choice for massive MIMO systems. However, the most important drawback of OFDM, i.e. its sensitivity to frequency offset can hinder massive MIMO performance. In fact, due to the large side-lobes of OFDM waveform, even a small amount of frequency offset can lead to a considerable amount of inter carrier interference (ICI), and consequently, tremendously degrade the massive MIMO performance. Therefore, a precise frequency synchronization is crucial for OFDM-based massive MIMO systems. On the other hand, as the number of BS antennas in massive MIMO systems is large, in the order of hundreds, the computational complexity of synchronization can become a burden for the system. In this chapter, we propose an accurate low-complexity frequency synchronization technique for OFDM-based massive MIMO systems which consists of three parts; CFO estimation, CFO compensation, estimation error correction.

The rest of this chapter is organized as follow. In section 4.1, we propose two different techniques for CFO estimation, and a low complexity CFO compensation technique in section 4.2. In section 4.3, the performance of the proposed CFO estimation and compensation techniques in the presence of CFO estimation error is analysed, and then, an iterative algorithm is proposed to correct the error and boost the performance. In section 4.4, the computational complexity of the proposed synchronization technique is calculated. The simulation results and conclusion are presented in sections 4.5 and 4.6, respectively.

4.1 Carrier frequency offset estimation

In multiuser massive MIMO systems, CFO estimation is challenging since all the users simultaneously share all the available bandwidth. In fact, their orthogonality is guaranteed by spacial diversity, while in CFO estimation stage we do not have the knowledge of CSI to separate the users' signals. There are some techniques in the literature that try to address this issue. As it is discussed in section 2.2.1, these techniques have high computational complexity for large scale antenna systems. In this section, we propose two low complexity CFO estimation techniques, the first one is based on a closed-form formula with a set of rectangular shape pilots, and the second one has a unimodal cost function.

4.1.1 CFO estimation in close form

Representing the circular convolution of the transmit data with the channel in equation (2.2) as $\mathbf{X}_p^{\kappa}\mathbf{h}_{m,p}$, the OFDM symbol received at the m^{th} BS antenna after CP removal in the presence of CFO can be written as

$$\mathbf{r}_m^{\kappa} = \sum_{p=0}^{P-1} \mathbf{\Phi}_p^{\kappa}\mathbf{X}_p^{\kappa}\mathbf{h}_{m,p} + \mathbf{n}_m, \tag{4.1}$$

where $\mathbf{X}_p^{\kappa}$ is an $N \times L$ matrix including the first L columns of the circulant matrix $\text{circ}(\mathbf{x}_p^{\kappa})$ and $\mathbf{\Phi}_p^{\kappa}$ is an $N \times N$ diagonal CFO matrix with the diagonal elements $\Phi_p^{\kappa}[l,l] = e^{j\frac{2\pi}{N}\epsilon_p(l+(N+N_{\text{CP}})\kappa+N_{\text{CP}})}$ for $l = 0, \cdots, N-1$, and ϵ_p is the CFO normalized to subcarrier spacing.

Let us consider the first OFDM symbol within each data packet as users' pilots. To keep the formulation simple, without any loss of generality, we omit the symbol index in our derivation, i.e. the superscript κ. If we multiply the first received OFDM symbol at each antenna in equation (4.1) to its Hermitian and average them over all the BS antennas, we have

$$\begin{aligned}
\mathbf{R} &= \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{r}_m\mathbf{r}_m^{\text{H}} \\
&= \sum_{p=0}^{P-1}\sum_{q=0}^{P-1} \mathbf{\Lambda}_p\mathbf{C}_{p,q}\mathbf{\Lambda}_q^{\text{H}} + \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{n}_m\mathbf{n}_m^{\text{H}} \\
&+ \frac{1}{M}\sum_{m=0}^{M-1}(\sum_{p=0}^{P-1} \mathbf{\Lambda}_p\mathbf{h}_{m,p})\mathbf{n}_m^{\text{H}} + \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{n}_m(\sum_{q=0}^{P-1} \mathbf{\Lambda}_q\mathbf{h}_{m,q})^{\text{H}},
\end{aligned} \tag{4.2}$$

where $\mathbf{\Lambda}_p = \mathbf{\Phi}_p \mathbf{X}_p$ and $\mathbf{C}_{p,q} \triangleq \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{h}_{m,p} \mathbf{h}_{m,q}^{\mathrm{H}}$. In the asymptotic regime, i.e. $M \to \infty$, the two last terms in (4.2) tend to zero and the second one becomes a diagonal matrix equal to $\sigma_n^2 \mathbf{I}_N$. Hence, for large values of M, $\mathbf{R}$ can be written as

$$\mathbf{R} = \sum_{p=0}^{P-1} \sum_{q=0}^{P-1} \mathbf{\Lambda}_p \mathbf{C}_{p,q} \mathbf{\Lambda}_q^{\mathrm{H}} + \sigma_n^2 \mathbf{I}_N. \tag{4.3}$$

Moreover, according to the law of large numbers, as M grows large, $C_{p,q}[i,j] \to \mathbb{E}\left\{h_{m,p}[i] h_{m,q}^*[j]\right\}$. In massive MIMO systems, the distance between BS antennas is set to be more than half of the wavelength of the transmitted signal, $\lambda/2$. In that case, the channels between a given user and different BS antennas are a set of i.i.d. random variables. Therefore, $\mathbb{E}\left\{h_{m,p}[i] h_{m,q}^*[j]\right\} = 0$ for $i \neq j$ and $p \neq q$. Thus, $\mathbf{C}_{p,q}$ is a zero matrix for $p \neq q$, and $\mathbf{C}_{p,p}$ is a real-valued diagonal matrix with the main diagonal $\boldsymbol{\rho}_p$. Therefore, based on equation (4.3), each entry of $\mathbf{R}$ is given by

$$\begin{aligned} R[i,j] &= \sum_{p=0}^{P-1} \Phi_p[i,i] \sum_{l=0}^{N-1} X_p[i,l] C_{p,p}[l,l] X_p^{\mathrm{H}}[l,j] \Phi_p^{\mathrm{H}}[j,j] \\ &= \sum_{p=0}^{P-1} e^{j\frac{2\pi}{N}\epsilon_p(i-j)} \sum_{l=0}^{N-1} X_p[i,l] C_{p,p}[l,l] X_p^{\mathrm{H}}[l,j]. \end{aligned} \tag{4.4}$$

Accordingly, if we choose $\mathbf{X}_p$ to be real, the angles of the elements of matrix $\mathbf{R}$ will be a linear function of the users' CFOs. Note that due to the summation over p in equation (4.4), in order to be able to estimate different users' CFOs, we need to choose orthogonal pilot sequences for different users in the time domain. To this end, a pilot should take zero value where other users' pilots are non-zero. In addition, due to the channel delay spread, users' pilots spread over L samples. Thus, an additional guard of L zero samples is required between users' pilot elements. Considering these two points, we propose a pilot signal that has L non-zero elements every $2LP$ samples in the time-domain, which can be given by $\mathbf{d}_p = \mathbf{F}_N \boldsymbol{\gamma}_p$ in the frequency domain where $\boldsymbol{\gamma}_p$ is an $N \times 1$ vector. The elements of γ_p can be shown as

$$\gamma_p[i] = \sum_{a=0}^{A-1} \mathrm{Rect}_L[i - 2L(p + aP) - L/2)], \tag{4.5}$$

where $i = 0, 1, \cdots, N-1$, $p = 0, 1, \cdots, P-1$, and $A = \lfloor N/(2PL) \rfloor$. From equation (4.4), one may realize that the angles of the elements of $\mathbf{R}$ relevant to user p are a

linear function of ϵ_p. Hence, ϵ_p can be estimated as

$$\epsilon_p = \frac{N}{A(A-1)L^2} \sum_{a=0}^{A-1} \sum_{\substack{a'=0 \\ a' \neq a}}^{A-1} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{\angle R[i+i_{p,a}, j+i_{p,a'}]}{2\pi(i-j+2LP(a-a'))} \tag{4.6}$$

where $i_{p,a} = 2L(p+aP)$. It is worth mentioning that we calculate $\mathbf{R}$ once for all the users. Moreover, since there is no optimization problem to solve in this technique, it has a light computational burden. In fact, the computational complexity with respect to the number of complex multiplications is in the order of $\mathcal{O}(MN^2)$, which increases linearly by increasing the number of BS antennas, M. However, the proposed pilot for this technique can lead to high peak to average power ratio (PAPR). Hence, we propose an alternative CFO estimation technique that is generic and does not limit us to a specific pilot. Thus, the computational complexity with respect to the number of complex multiplications is $\mathcal{O}(MN^2)$, which increases linearly by increasing the number of BS antennas, M.

4.1.2 CFO estimation with unimodal cost function

In this section, instead of using orthogonal pilots, we use the orthogonal projection of each user's pilot to separate user's signals. We denote $\mathbf{P}_{X_q} = \mathbf{I}_N - \mathbf{X}_q(\mathbf{X}_q^{\mathrm{H}}\mathbf{X}_q)^{-1}\mathbf{X}_q^{\mathrm{H}}$ as the orthogonal projection onto the space spanned by the columns of $\mathbf{X}_q$. Then, if we consider the training sequence designed such that $\mathbf{\Lambda}_q^{\perp} = \mathbf{P}_{X_q}\mathbf{\Phi}_q^{\mathrm{H}}$, then, $\mathbf{\Lambda}_q^{\perp}\mathbf{\Lambda}_p = 0$ holds for $q = p$. Therefore, if we estimate CFO, $\hat{\epsilon}_q$, calculate $\hat{\mathbf{\Lambda}}_q^{\perp}$, and multiply it to $\mathbf{R}$ in equation (4.2), we have

$$\hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{R} = \hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{\Lambda}_q\mathbf{C}_{q,q}\mathbf{\Lambda}_q^{\mathrm{H}} + \hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{V}_q, \tag{4.7}$$

where

$$\mathbf{V}_q = \sum_{\substack{p=0 \\ p \neq q}}^{P-1} \mathbf{\Lambda}_p\mathbf{C}_{p,p}\mathbf{\Lambda}_p^{\mathrm{H}} + \sigma_n^2\mathbf{I}_N, \tag{4.8}$$

denotes the noise plus multi-user interference with respect to user q. Then, if $\hat{\epsilon}_q = \epsilon_q$, the first term on the right-hand side of equation (4.7) is zero, and the Frobenius norm of the resulting matrix gets its minimum value. Therefore,

$$\hat{\epsilon}_q = \underset{\epsilon}{\operatorname{argmin}} \quad \left\| \hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{R} \right\|^2. \tag{4.9}$$

The cost function of this optimization problem is a unimodal function, i.e. it is monotonically decreasing for $\epsilon \leq \epsilon_q$ and monotonically increasing for $\epsilon \geq \epsilon_q$. The

proof is provided in Section 4.1.2.1. Therefore, this optimization problem can be solved by applying the Golden section search algorithm [120]. This approach is generic, and therefore, there is no restriction on choosing the pilot sequence.

4.1.2.1 Proof of unimodality of the cost function

The objective function of our optimization problem is the square of Frobenius norm of matrix $\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{R}$ as it is shown in equation (4.9). By substituting this matrix from (4.7), we have

$$\left\|\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{R}\right\|^2 = \left\|\hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{\Lambda}_q\mathbf{C}_{q,q}\mathbf{\Lambda}_q^{\mathrm{H}} + \hat{\mathbf{\Lambda}}_q^{\perp}\mathbf{V}_q\right\|^2, \tag{4.10}$$

where the first term corresponds to the desired user q, and the second one is due to the noise plus multi-user interference with respect to user q. Equation (4.10) can be expanded as

$$\begin{aligned}\left\|\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{R}\right\|^2 = {} & \left\|\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{\Lambda}_q\mathbf{W}_q\right\|^2 + \left\|\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{V}_q\right\|^2 \\ & + 2\Re\left\{\left\langle\check{\mathbf{\Lambda}}_q^{\perp}\mathbf{\Lambda}_q\mathbf{W}_q, \check{\mathbf{\Lambda}}_q^{\perp}\mathbf{V}_q\right\rangle\right\},\end{aligned} \tag{4.11}$$

where $\mathbf{W}_q = \mathbf{C}_{q,q}\mathbf{\Lambda}_q^{\mathrm{H}}$, $\check{\mathbf{\Lambda}}_q^{\perp} = \mathbf{P}_{X_q}\check{\mathbf{\Phi}}_q^{\mathrm{H}}$ for the trial CFO matrix $\check{\mathbf{\Phi}}_q$, and $\langle .,. \rangle$ is the Frobenius inner product. In the following, first we analyze these three terms separately, and then, we conclude that our objective function is unimodal.

For the sake of simplicity and without loss of generality, we drop the index q from all of the vectors and matrices for the rest of this section. In addition, let us define $\mathbf{\Xi}_1 = \check{\mathbf{\Lambda}}^{\perp}\mathbf{\Lambda}\mathbf{W}$ and $\mathbf{\Xi}_2 = \check{\mathbf{\Lambda}}^{\perp}\mathbf{V}$, and rewrite the equation (4.11) as

$$\begin{aligned}\left\|\check{\mathbf{\Lambda}}^{\perp}\mathbf{R}\right\|^2 = {} & \|\mathbf{\Xi}_1\|^2 + \|\mathbf{\Xi}_2\|^2 \\ & + 2\Re\left\{\langle\mathbf{\Xi}_1, \mathbf{\Xi}_2\rangle\right\}.\end{aligned} \tag{4.12}$$

The first term of the right hand side of equation (4.12) can be calculated as

$$\|\mathbf{\Xi}_1\|^2 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left|\Xi_1[i,j]\right|^2. \tag{4.13}$$

Then by defining $\mathbf{G} = \check{\mathbf{\Lambda}}^{\perp}\mathbf{\Lambda}$, we have

$$\begin{aligned}
|\Xi_1[i,j]|^2 &= \left|\sum_{l=0}^{N-1} G[i,l]W[l,j]\right|^2 \\
&= \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} G[i,l]W[l,j]G^*[i,k]W^*[k,j] \\
&= \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} |G[i,l]|\,|G^*[i,k]|\,\mathrm{e}^{j(\angle G[i,l]-\angle G[i,k])} \\
&\quad \times |W[l,j]|\,|W^*[k,j]|\,\mathrm{e}^{j(\angle W[i,l]-\angle W[i,k])}. \qquad (4.14)
\end{aligned}$$

The elements of the matrix $\mathbf{W}$ are independent of the trial CFO values, and the ij^{th} entry of $\mathbf{G}$ is given by

$$\begin{aligned}
G[i,j] &= \sum_{l=0}^{N-1} P_X[i,l]\check{\Phi}^*[l,l]\Phi[l,l]X[l,j] \\
&= \sum_{l=0}^{N-1} P_X[i,l]\tilde{\Phi}^*[l,l]X[l,j] \\
&= \left\langle \tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j \right\rangle \\
&= \left\| \tilde{\boldsymbol{\lambda}}_i^{\perp} \right\| \|\mathbf{x}_j\| \cos(\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j)), \qquad (4.15)
\end{aligned}$$

for $0 \le i,j \le N-1$ where $\tilde{\Phi}[l,l] = e^{j\frac{2\pi}{N}\tilde{\epsilon}(l+N_{\mathrm{CP}})}$, and $\tilde{\epsilon}$ is the difference between the desired CFO, ϵ, and the trial CFO value, $\check{\epsilon}$. In addition, $\tilde{\boldsymbol{\lambda}}_i^{\perp}$ is the i^{th} row of the matrix $\tilde{\mathbf{\Lambda}}^{\perp} = \mathbf{P}_X\tilde{\mathbf{\Phi}}^{\mathrm{H}}$, $\mathbf{x}_j$ is the j^{th} column of the matrix $\mathbf{X}$, and $\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j)$ is the angle between the two complex vectors $\tilde{\boldsymbol{\lambda}}_i^{\perp}$ and $\mathbf{x}_j$. Since $\mathbf{P}_X$ is defined to be the orthogonal projection onto the space spanned by the columns of the matrix $\mathbf{X}$, the i^{th} row of the matrix $\mathbf{P}_X$, and the vector $\mathbf{x}_j$ are orthogonal for any $i,j \in [0, N-1]$. Hence, if $\check{\epsilon} = \epsilon$, $\tilde{\mathbf{\Phi}} = \mathbf{I}_N$ and we have

$$\left\langle \tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j \right\rangle = 0, \qquad (4.16)$$

this means $\cos(\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j)) = 0$ and $\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j) = \pm\pi/2$ for any i and j. Thus, $\mathbf{G}$ becomes a zero matrix, and $\|\mathbf{\Xi}_1\|^2 = 0$.

If $\check{\epsilon} \neq \epsilon$ the orthogonality between the two vectors $\tilde{\boldsymbol{\lambda}}_i^{\perp}$ and $\mathbf{x}_j$ is destroyed. Consider the singular value decomposition (SVD) of the matrix $\mathbf{P}_X$ as $\mathbf{P}_X = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$. Then, by multiplying $\mathbf{P}_X$ to $\tilde{\mathbf{\Phi}}$, the right singular vectors of the matrix $\mathbf{P}_X$, i.e. the columns of $\mathbf{V}$, are multiplied to the matrix $\tilde{\mathbf{\Phi}}$. This means that the orthogonal basis of the matrix $\mathbf{P}_X$ is rotated by $\tilde{\epsilon}$, which is the difference between the desired CFO,

ϵ, and the trial CFO, $\check{\epsilon}$. Therefore, the resulting matrix is not orthogonal to $\mathbf{X}$ anymore, and the length of the projection of $\tilde{\boldsymbol{\lambda}}_i^{\perp}$ on $\mathbf{x}_j$, i.e. $\left\langle \tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j \right\rangle$, is a function of $\tilde{\epsilon}$. Note that $\cos(\theta_{\mathrm{C}}(\mathbf{a}, \mathbf{b}))$ has a complex value which can be represented as

$$\cos(\theta_C(\mathbf{a}, \mathbf{b})) = \gamma \mathrm{e}^{i\varphi}, \tag{4.17}$$

where

$$\gamma = \cos(\theta_{\mathrm{H}}(\mathbf{a}, \mathbf{b})) = |\cos(\theta_{\mathrm{C}}(\mathbf{a}, \mathbf{b}))|, \tag{4.18}$$

$0 \leq \theta_{\mathrm{H}} \leq \pi/2$ is called the Hermitian angle between the vectors $\mathbf{a}$ and $\mathbf{b}$ of the complex vector space, and $-\pi \leq \varphi \leq \pi$ is called their (Kasner's) pseudo-angle [121]. Since $\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j) = \pm\pi/2$ for $\tilde{\epsilon} = 0$, any rotation less than $\pi/2$ will increase the value of $\cos(\theta_{\mathrm{H}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j)) = |\cos(\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j))|$, and hence the larger the rotation, the larger the value of $\cos(\theta_{\mathrm{H}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j))$. If $|\tilde{\epsilon}| < 1$, the rotation cannot be greater than $\pi/2$. Therefore, as $|\tilde{\epsilon}|$ grows, the absolute value of the elements in the matrix $\mathbf{G}$ increase.

Moreover, since all the orthogonal basis vectors of the matrix $\mathbf{P}_X$ are rotated by the value of $\tilde{\epsilon}$, $\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j)$ is the same for every $i, j \in [0, N-1]$. Hence, considering $\angle G[i,j] = \angle \cos(\theta_{\mathrm{C}}(\tilde{\boldsymbol{\lambda}}_i^{\perp}, \mathbf{x}_j))$, $\angle G[i,l] = \angle G[i,k]$ for any amount of $\tilde{\epsilon}$, and we have

$$\begin{aligned} |\Xi_1[i,j]|^2 = &\sum_{l=0}^{N-1} |G[i,l]|^2 \, |W[l,j]|^2 \\ &+ 2 \sum_{l=0}^{N-2} \sum_{k=l+1}^{N-1} |G[i,l]| \, |G^*[i,k]| \\ &\times |W[l,j]| \, |W^*[k,j]| \cos(\angle W[i,l] - \angle W[i,k])). \end{aligned} \tag{4.19}$$

Then, because $|G[i,j]|$ increases as $|\tilde{\epsilon}|$ grows, $|\Xi_1[i,j]|^2$ will also increase. Hence, if $|\check{\epsilon}_2 - \epsilon| > |\check{\epsilon}_1 - \epsilon|$, we have $\|\boldsymbol{\Xi}_1\|^2|_{\check{\epsilon}_2} > \|\boldsymbol{\Xi}_1\|^2|_{\check{\epsilon}_1}$ for any $\check{\epsilon}$, and $\|\boldsymbol{\Xi}_1\|^2 = 0$ for $\check{\epsilon} = \epsilon$. As a result, the first term of the right hand side of equation (4.12) is a unimodal function with respect to the trial CFO value.

The second term of the right hand side of equation (4.12) can be calculated as

$$\|\boldsymbol{\Xi}_2\|^2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |\Xi_2[i,j]|^2, \tag{4.20}$$

where

$$|\Xi_2[i,j]|^2 = \left|\sum_{l=0}^{N-1} \check{\Lambda}^{\perp}[i,l]V[l,j]\right|^2 = \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} \check{\Lambda}^{\perp}[i,l]V[l,j](\check{\Lambda}^{\perp})^*[i,k]V^*[k,j]. \tag{4.21}$$

Then, by substituting $\check{\mathbf{\Lambda}}^{\perp} = \mathbf{P}_X\check{\mathbf{\Phi}}^{\mathrm{H}}$ in equation (4.21), we have

$$|\Xi_2[i,j]|^2 = \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} P_X[i,l]P_X^*[i,k] \times V[l,j]V^*[k,j]\mathrm{e}^{j\frac{2\pi}{N}\check{\epsilon}(k-l)}. \tag{4.22}$$

By substituting (4.22) in (4.20), $\|\mathbf{\Xi}_2\|^2$ is given by

$$\|\mathbf{\Xi}_2\|^2 = \zeta + \sum_{l=0}^{N-1}\sum_{\substack{k=0\\k\neq l}}^{N-1} T[l,k]\mathrm{e}^{j\frac{2\pi}{N}\check{\epsilon}(k-l)}, \tag{4.23}$$

where

$$\zeta = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{l=0}^{N-1} |P_X[i,l]|^2\,|V[l,j]|^2\,, \tag{4.24}$$

and

$$T[l,k] = \sum_{i=0}^{N-1} P_X[i,l]P_X^*[i,k]\sum_{j=0}^{N-1} V[l,j]V^*[k,j]. \tag{4.25}$$

Note that ζ is constant with respect to $\check{\epsilon}$. Moreover, since $T[l,k] = T^*[k,l]$, $\|\mathbf{\Xi}_2\|^2$ can be rewritten as

$$\|\mathbf{\Xi}_2\|^2 = \zeta + 2\sum_{l=0}^{N-2}\sum_{\substack{k=l+1\\k\neq l}}^{N-1} |T[l,k]| \times \cos(\angle T[l,k] + \frac{2\pi}{N}\check{\epsilon}(k-l)). \tag{4.26}$$

The second term in equation (4.26) is a sum of $N(N-1)/2$ number of cosine functions of $\check{\epsilon}$ with different periods in the range of $1 < N/(k-l) \leq N$. Thus, the value of these cosine functions vary slowly. In addition, since $l \neq k$ for $T[l,k]$ in equation (4.23), the elements of the matrices $\mathbf{P}_X$ and $\mathbf{V}$ in equation (4.25) are

statistically independent. Consequently, the elements $T[l,k]$ are very small and tend to zero. Thus, the second term on the right hand side of equation (4.12) can be approximated to a constant value with respect to the trial CFO.

Finally, considering the third term on the right hand side of equation (4.12), we have

$$\Re\left\{\langle \boldsymbol{\Xi}_1, \boldsymbol{\Xi}_2 \rangle\right\} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} |\Xi_1[i,j]|\,|\Xi_2[i,j]|\cos(\theta_\Xi), \tag{4.27}$$

where $\theta_\Xi = \angle\Xi_2[i,j] - \angle\Xi_1[i,j]$. Following similar line of derivations as above, it can be shown that $|\Xi_1[i,j]|$ is also unimodal with respect to the trial CFO values, and $|\Xi_2[i,j]|$ remains constant as $\check{\epsilon}$ changes. Since $|\Xi_2[i,j]|$ is always positive, $|\Xi_1[i,j]|\,|\Xi_2[i,j]|$ is just a scaled version of $|\Xi_1[i,j]|$ which is unimodal with the same minimum point as $|\Xi_1[i,j]|$ for any $i,j \in [0, N-1]$. Then, for given i and j, if $\cos(\theta_\Xi) > 0$, it only multiplies to the scale of $|\Xi_1[i,j]|$. Else if $\cos(\theta_\Xi) < 0$, beside changing the scale, it also turns it to a downward unimodal function.

In addition, since adding a constant value does not disturb the unimodality, the sum of the first two terms in equation (4.12) is an upward unimodal function that gets its minimum at the trial CFO value that minimizes $|\Xi_1[i,j]|$. Therefore, as long as the value of the third term is smaller than the sum of the two first terms for any i,j, and $\check{\epsilon}$, the overall cost function in (4.12) will be an upward unimodal function. To this end, the following inequality should be fulfilled for any i and j value.

$$|\Xi_1[i,j]|^2 + |\Xi_2[i,j]|^2 > |\Xi_1[i,j]|\,|\Xi_2[i,j]|\cos(\theta_\Xi). \tag{4.28}$$

Since $\cos(\theta_\Xi) \in [-1,1]$, the sufficient condition to guarantee unimodality is

$$|\Xi_1[i,j]|^2 + |\Xi_2[i,j]|^2 > |\Xi_1[i,j]|\,|\Xi_2[i,j]|\,. \tag{4.29}$$

We know that the inequality

$$(|\Xi_1[i,j]| - |\Xi_2[i,j]|)^2 > 0, \tag{4.30}$$

is always true. Hence,

$$\begin{aligned} |\Xi_1[i,j]|^2 + |\Xi_2[i,j]|^2 &> 2\,|\Xi_1[i,j]|\,|\Xi_2[i,j]| \\ &> |\Xi_1[i,j]|\,|\Xi_2[i,j]|\,, \end{aligned} \tag{4.31}$$

is also true, and the sufficient condition for unimodality is fulfilled. Therefore, even when $\cos(\theta_{\Xi}) < 0$, the final expression of the cost function for given i and j values is an upward unimodal function with respect to the trial CFO values.

As a conclusion, we have proved that the first term on the right hand side of the equation (4.12) is unimodal, the second term is constant which means the sum of the first two terms is also unimodal. Moreover, for any i, j and ϵ, the third term of this equation preserves the unimodality of the whole objective function. Therefore, our objective function is upward unimodal and its minimum can be found using search algorithms such as Golden section search.

4.2 Carrier frequency offset compensation

As it is discussed in section 2.2.2, the CFO compensation technique for massive MIMO systems in the literature is performed in the time domain. Since users' signals have different amount of CFO, after CFO compensation, each user has a separate signal stream, and consequently, needs a separate receiver. Therefore, as the number of BS antennas grows large, CFO compensation can impose a considerable amount of complexity to the system. To address this problem, we propose to compensate the users' CFOs in the frequency domain and after combining the received signals at the BS antennas.

4.2.1 CFO compensation in frequency domain

Representing the circular convolution of the transmit data with the channel in equation (2.2) as $\mathbf{H}_{m,p}\mathbf{x}_p$, the received signal at BS antenna m in the frequency domain can be written as

$$\bar{\mathbf{r}}_m = \sum_{p=0}^{P-1} \mathbf{F}_N \mathbf{\Phi}_p \mathbf{H}_{m,p} \mathbf{x}_p + \mathbf{F}_N \mathbf{n}_m, \tag{4.32}$$

where $\mathbf{H}_{m,p}$ is a circulant matrix with the first column $\mathbf{h}_{m,p}$ which is zero-padded to have the length of N. Then, considering $\mathbf{F}_N^{\mathrm{H}}\mathbf{F}_N = \mathbf{I}_N$, equation (4.32) can be rewritten as

$$\bar{\mathbf{r}}_m = \sum_{p=0}^{P-1} \mathbf{F}_N \mathbf{\Phi}_p \mathbf{F}_N^{\mathrm{H}} \mathbf{F}_N \mathbf{H}_{m,p} \mathbf{F}_N^{\mathrm{H}} \mathbf{d}_p + \bar{\mathbf{n}}_m, \tag{4.33}$$

where $\bar{\mathbf{n}}_m = \mathbf{F}_N \mathbf{n}_m$ is the frequency domain noise vector. By defining $\mathbf{E}_p \triangleq \mathbf{F}_N \mathbf{\Phi}_p \mathbf{F}_N^{\mathrm{H}}$ and $\widetilde{\mathbf{H}}_{m,p} \triangleq \mathbf{F}_N \mathbf{H}_{m,p} \mathbf{F}_N^{\mathrm{H}}$, we have

$$\bar{\mathbf{r}}_m = \sum_{p=0}^{P-1} \mathbf{E}_p \widetilde{\mathbf{H}}_{m,p} \mathbf{d}_p + \bar{\mathbf{n}}_m. \tag{4.34}$$

It is worth noting that due to the circulant property of $\mathbf{H}_{m,p}$, $\widetilde{\mathbf{H}}_{m,p}$ becomes a diagonal matrix, [10].

Let us define the vector $\bar{\mathbf{r}}[k] = [\bar{r}_0[k], \bar{r}_1[k], \cdots, \bar{r}_{M-1}[k]]^{\mathrm{T}}$ which contains the k^{th} samples of the received signals at all the BS antennas in the frequency domain for $k = 0, 1, \cdots, N-1$. Then, the combiner output is given by

$$\bar{\mathbf{y}}[k] = \mathbf{Z}[k]\bar{\mathbf{r}}[k], \tag{4.35}$$

where $\bar{\mathbf{y}}[k] = [\bar{y}_0[k], \bar{y}_1[k], \cdots, \bar{y}_{P-1}[k]]^{\mathrm{T}}$ is a $P \times 1$ vector containing the k^{th} samples of the users' signals, and $\mathbf{Z}[k]$ is a $P \times M$ linear combiner on subcarrier k. If we deploy the zero-forcing (ZF) combiner, $\mathbf{Z}[k]$ can be calculated as

$$\mathbf{Z}[k] = (\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k])^{-1}\boldsymbol{\mathcal{H}}^{\mathrm{H}}[k], \tag{4.36}$$

where $\boldsymbol{\mathcal{H}}[k] = [\bar{\mathbf{h}}_0[k], \bar{\mathbf{h}}_1[k], \cdots, \bar{\mathbf{h}}_{P-1}[k]]$ is an $M \times P$ matrix whose p^{th} column is $\bar{\mathbf{h}}_p[k] = [\bar{h}_{0,p}[k], \bar{h}_{1,p}[k], \cdots, \bar{h}_{M-1,p}[k]]^{\mathrm{T}}$, and $\bar{\mathbf{h}}_{m,p} = \mathbf{F}_N \mathbf{h}_{m,p}$.

Note that in the asymptotic regime, $P \times P$ normalization matrix $\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k]$ becomes a diagonal matrix,

$$\begin{aligned}\mathbf{D}_k &= \boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k] \\ &= \mathrm{diag}\left\{\left\|\bar{\mathbf{h}}_0[k]\right\|^2, \left\|\bar{\mathbf{h}}_1[k]\right\|^2, \cdots, \left\|\bar{\mathbf{h}}_{P-1}[k]\right\|^2\right\},\end{aligned} \tag{4.37}$$

and when M tends to infinity, according to the law of large numbers, $\mathbf{D}_k \to M\mathbf{I}_P$. Therefore, the combiner output for user q can be written as

$$\bar{y}_q[k] = \frac{1}{M}\bar{\mathbf{h}}_q^{\mathrm{H}}[k]\bar{\mathbf{r}}[k], \tag{4.38}$$

for $k = 0, 1, \cdots, N-1$. Substituting $\bar{\mathbf{r}}$ from (4.34) into (5.6), we can rewrite the q^{th} user's signal at the combiner output as

$$\bar{\mathbf{y}}_q = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\sum_{p=0}^{P-1}\mathbf{E}_p\widetilde{\mathbf{H}}_{m,p}\mathbf{d}_p + \tilde{\mathbf{n}}, \tag{4.39}$$

where $\tilde{\mathbf{n}} = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\bar{\mathbf{n}}_m$. Then, defining an $N \times N$ interference matrix as

$$\boldsymbol{\Omega}_{q,p} = \frac{1}{M}\sum_{m=0}^{M-1}\widetilde{\mathbf{H}}_{m,q}^{\mathrm{H}}\mathbf{E}_p\widetilde{\mathbf{H}}_{m,p}, \tag{4.40}$$

we have

$$\bar{\mathbf{y}}_q = \mathbf{\Omega}_{q,q}\mathbf{d}_q + \sum_{\substack{p=0 \\ p\neq q}}^{P-1} \mathbf{\Omega}_{q,p}\mathbf{d}_p + \tilde{\mathbf{n}}, \tag{4.41}$$

It is worth noting that since different users' channels are uncorrelated, the elements of $\mathbf{\Omega}_{q,p}$ for $p \neq q$ tend to zero as M increases to infinity, and multi-user interference as well as the noise term will be averaged out. Moreover, while in the absence of CFO, $\mathbf{\Omega}_{q,q}$ is the identity matrix, the presence of CFO makes it non-diagonal and banded, with the off-diagonal elements modeling the inter-carrier interference (ICI) effect. Furthermore, with some manipulations, the interference matrix in equation (4.40) can be represented as

$$\mathbf{\Omega}_{q,p} = \mathbf{E}_p \odot \mathbf{B}_{q,p}, \tag{4.42}$$

where

$$B_{q,p}[l,k] = \frac{1}{M}\sum_{m=0}^{M-1} \widetilde{H}^*_{m,q}[l,l]\widetilde{H}_{m,p}[k,k], \tag{4.43}$$

for $l,k = 0,\cdots,N-1$. Therefore, in the asymptotic regime $\mathbf{B}_{q,q}$ can be obtained as

$$\begin{aligned} B_{q,q}[l,k] &= \mathbb{E}\left\{\widetilde{H}^*_{m,q}[l,l]\widetilde{H}_{m,q}[k,k]\right\} \\ &= \mathbb{E}\left\{\bar{h}^*_{m,q}[l]\bar{h}_{m,q}[k]\right\} \\ &= \mathbb{E}\left\{\sum_{i=0}^{N-1}\sum_{i'=0}^{N-1} h^*_{m,q}[i]h_{m,q}[i']e^{-j\frac{2\pi}{N}(ki'-li)}\right\}. \end{aligned} \tag{4.44}$$

and $\mathbb{E}\left\{h^*_{m,q}[i]h_{m,q}[i']\right\} = 0$ for $i \neq i'$ since $\mathbf{h}_{m,q}$ is an i.i.d. random vector, and we have

$$B_{q,q}[l,k] = \sum_{i=0}^{N-1} \rho_q[i]e^{-j\frac{2\pi}{N}(k-l)i} = \bar{\rho}_q[k-l], \tag{4.45}$$

where $\bar{\boldsymbol{\rho}}_q = \mathbf{F}_N\boldsymbol{\rho}_q$ contains the N-point DFT samples of the channel PDP of the q^{th} user, $\boldsymbol{\rho}_q$, and since the channel PDPs are real-valued functions, $\bar{\rho}_q[i] = \bar{\rho}_q[i-N]$. Hence, we can conclude that $\mathbf{B}_{q,q}$ is a circulant matrix with the first column $\bar{\rho}_q[-l]$ for $l = 0,\cdots,N-1$. [1] $\mathbf{E}_q = \mathbf{F}_N\mathbf{\Phi}_q\mathbf{F}_N^{\mathrm{H}}$ is also a circulant matrix as $\mathbf{\Phi}_q$ is diagonal. Accordingly, $\mathbf{\Omega}_{q,q}$ derived in equation (4.42), is a circulant matrix and can be written as $\mathbf{\Omega}_{q,q} = \mathbf{F}_N\mathbf{Q}_{q,q}\mathbf{F}_N^{\mathrm{H}}$, where $\mathbf{Q}_{q,q}$ is diagonal. Therefore, in order to calculate the inverse of the interference matrix, $\mathbf{\Omega}_{q,q}$, we only need to calculate the inverse of the diagonal matrix $\mathbf{Q}_{q,q}$, and calculate, $\mathbf{\Omega}^{-1}_{q,q} = \mathbf{F}_N\mathbf{Q}^{-1}_{q,q}\mathbf{F}_N^{\mathrm{H}}$. Moreover, due to the circulant property of the matrix $\mathbf{\Omega}_{q,q}$, calculating its first column, which is the

[1] Note that if the distance between the BS antennas is less than half of the wavelength of the transmitted signal, $\lambda/2$, $\mathbf{h}_{m,q}$ will not be an i.i.d. random vector, and $\mathbb{E}\left\{h^*_{m,q}[i]h_{m,q}[i']\right\}$ will be equal to the covariance of these two variables. Consequently, $\mathbf{B}_{q,q}$ will not be a circulant matrix.

element-wise multiplication of the first columns of the matrices $\mathbf{E}_p$ and $\mathbf{B}_{q,p}$, is sufficient to form the rest of the matrix $\mathbf{\Omega}_{q,q}$. Since an element-wise multiplication in the frequency domain is equivalent to a circular convolution in the time domain, the diagonal elements of the matrix $\mathbf{Q}_{q,q}$ can be obtained as

$$\mathrm{diag}(\mathbf{Q}_{q,q}) = \mathrm{diag}(\mathbf{\Phi}_q) \otimes \tilde{\boldsymbol{\rho}}_q, \tag{4.46}$$

where $\tilde{\boldsymbol{\rho}}_q[i] = \boldsymbol{\rho}_q[-i] = \boldsymbol{\rho}_q[N-i]$. Hence, the q^{th} user's signal can be obtained as

$$\hat{\mathbf{d}}_q = \mathbf{F}_N \mathbf{Q}_{q,q}^{-1} \mathbf{F}_N^{\mathrm{H}} \bar{\mathbf{y}}_q. \tag{4.47}$$

Considering the symbol indices, one realizes that $\mathbf{Q}_{q,q}$ varies for different symbols as

$$\mathrm{diag}(\mathbf{Q}_{q,q}^{\kappa}) = \mathrm{diag}(\mathbf{\Phi}_q^{\kappa}) \otimes \tilde{\boldsymbol{\rho}}_q. \tag{4.48}$$

To deal with this issue, we suggest to calculate the matrix $\mathbf{Q}_{q,q}$ once for the first symbol, where $\Phi_q^0[l,l] = e^{j\frac{2\pi}{N}\epsilon_p(l+N_{\mathrm{CP}})}$ for $l = 0, 1, \cdots, N-1$, and for other symbols add a phase shift correction step. Thus, the κ^{th} data symbol can be estimated as

$$\hat{\mathbf{d}}_q^{\kappa} = e^{-j\frac{2\pi}{N}\epsilon_q(N+N_{\mathrm{CP}})\kappa} \mathbf{F}_N \mathbf{Q}_{q,q}^{-1} \mathbf{F}_N^{\mathrm{H}} \bar{\mathbf{y}}_q^{\kappa}, \tag{4.49}$$

It is worth to note that our proposed CFO compensation technique is performed only once for each user, and as a result, its computational complexity remains constant as the number of BS antennas increases.

4.3 Frequency synchronization enhancement

In Section 4.2.1, we have assumed that the CFO estimation is accurate. However in practice, it may not be a realistic assumption. In this section, we analyze the effect of CFO estimation errors and investigate the efficacy of the proposed technique. Then, we show that the possible error due to an inaccurate CFO estimation can be simply removed by an iterative phase correction algorithm.

4.3.1 CFO estimation error analysis

We consider the estimated CFO for user p as $\hat{\epsilon}_p = \epsilon_p + \tilde{\epsilon}_p$ where $\tilde{\epsilon}_p$ is the estimation error. Thus, the estimated CFO matrix can be written as

$$\hat{\mathbf{\Phi}}_p = \tilde{\mathbf{\Phi}}_p \mathbf{\Phi}_p, \tag{4.50}$$

where $\tilde{\mathbf{\Phi}}_p$ is a diagonal matrix with the elements $\tilde{\Phi}_p^{\kappa}[l,l] = e^{j\frac{2\pi}{N}\tilde{\epsilon}_p(l+(N+N_{\mathrm{CP}})\kappa+N_{\mathrm{CP}})}$ for $l = 0, \cdots, N-1$. After dropping the superscript κ for the sake of simplicity without loss of generality, the estimated channel at the m^{th} BS antenna can be obtained as

$$\begin{aligned}\hat{\mathbf{h}}_m =&(\hat{\mathbf{\Lambda}}^{\mathrm{H}}\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{\Lambda}}^{\mathrm{H}}\mathbf{r}_m \\ =&(\hat{\mathbf{\Lambda}}^{\mathrm{H}}\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{\Lambda}}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{h}_m + (\hat{\mathbf{\Lambda}}^{\mathrm{H}}\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{\Lambda}}^{\mathrm{H}}\mathbf{n}_m,\end{aligned} \tag{4.51}$$

where $\hat{\mathbf{\Lambda}} = [\hat{\mathbf{\Lambda}}_0, \hat{\mathbf{\Lambda}}_1, \cdots, \hat{\mathbf{\Lambda}}_{P-1}]$ and $\hat{\mathbf{\Lambda}}_p = \hat{\mathbf{\Phi}}_p\mathbf{X}_p$. Let us define $NP \times NP$ matrices $\mathbf{\Gamma}_1 \triangleq \hat{\mathbf{\Lambda}}^{\mathrm{H}}\hat{\mathbf{\Lambda}}$ and $\mathbf{\Gamma}_2 \triangleq \hat{\mathbf{\Lambda}}^{\mathrm{H}}\mathbf{\Lambda}$, with the elements

$$\begin{aligned}\Gamma_1[i,j] =& \sum_{l=0}^{N-1}\hat{\Lambda}_p^*[i,l]\hat{\Lambda}_q[l,j] \\ =& \sum_{l=0}^{N-1}X_p^*[l,i-pN]\Phi_p^*[l,l]\tilde{\Phi}_p^*[l,l]\tilde{\Phi}_q[l,l]\Phi_q[l,l] \\ &\times X_q[l,j-qN],\end{aligned} \tag{4.52}$$

and

$$\begin{aligned}\Gamma_2[i,j] =& \sum_{l=0}^{N-1}\hat{\Lambda}_p^*[i,l]\Lambda_q[l,j] \\ =& \sum_{l=0}^{N-1}X_p^*[l,i-pN]\Phi_p^*[l,l]\tilde{\Phi}_p^*[l,l]\Phi_q[l,l]X_q[l,j-qN].\end{aligned} \tag{4.53}$$

Here, p and q can be obtained as $p = \lfloor \frac{i}{N} \rfloor$ and $q = \lfloor \frac{j}{N} \rfloor$. Therefore, if $|i-j| < N$, $q = p$ and we have

$$\Gamma_1[i,j] = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases} \tag{4.54}$$

and

$$\Gamma_2[i,j] = \sum_{l=0}^{N-1}\tilde{\Phi}_p^*[l,l]X_p^*[l,i-pN]X_p[l,j-pN]. \tag{4.55}$$

It is worth noting that when the estimation error, $\tilde{\epsilon}_p$, has a small value, for $i \neq j$, $\Gamma_2[i,j] \approx 0$, and for $i = j$, we have

$$\Gamma_2[i,i] = \sum_{l=0}^{N-1}\tilde{\Phi}_p^*[l,l]\Big|X_p[l,i-pN]\Big|^2. \tag{4.56}$$

Then, according to the Parseval's theorem, we can rewrite equation (4.56) as

$$\Gamma_2[i,i] = \frac{1}{N}\sum_{k=0}^{N-1} \widetilde{E}_p^*[k,0]\Big|d_p[k]\Big|^2, \tag{4.57}$$

where $|d_p[k]| = 1$, and $\widetilde{\mathbf{E}}_p = \mathbf{F}_N\tilde{\mathbf{\Phi}}_p\mathbf{F}_N^{\mathrm{H}}$ is a circulant matrix, and its first column is the Fourier transform of $\mathrm{diag}(\tilde{\mathbf{\Phi}}_{\mathrm{p}})$. Moreover, with the assumption of small $\tilde{\epsilon}_p$, the first column of $\widetilde{\mathbf{E}}_p$ can be considered as an impulse, i.e., $\widetilde{E}_p^*[k,0] \approx \widetilde{E}_p^*[0,0]\delta[k]$. Hence, for $|i-j| < N$,

$$\Gamma_2[i,j] \approx \begin{cases} \varphi_p^*, & i = j \\ 0, & i \neq j, \end{cases} \tag{4.58}$$

where

$$\varphi_p = \frac{1}{N}\widetilde{E}_p[0,0] = \frac{1}{N}\sum_{l=0}^{N-1}\tilde{\Phi}_p[l,l]. \tag{4.59}$$

Since the estimation is performed on the first symbol, i.e. $\kappa = 0$, which is the pilot, we can expand equation (4.59) as

$$\begin{aligned}\varphi_p &= \frac{1}{N}\sum_{l=0}^{N-1} e^{j\frac{2\pi}{N}\tilde{\epsilon}_p(l+N_{\mathrm{CP}})} \\ &= \frac{1}{N}\sum_{l=0}^{N-1}\left\{\cos\left(\frac{2\pi}{N}\tilde{\epsilon}_p l\right) + j\sin\left(\frac{2\pi}{N}\tilde{\epsilon}_p l\right)\right\} e^{j\frac{2\pi}{N}\tilde{\epsilon}_p N_{\mathrm{CP}}}.\end{aligned} \tag{4.60}$$

Due to the small value of $\tilde{\epsilon}_p$, we can replace $\cos\left(\frac{2\pi}{N}\tilde{\epsilon}_p l\right) = 1$ and $\sin\left(\frac{2\pi}{N}\tilde{\epsilon}_p l\right) = \frac{2\pi}{N}\tilde{\epsilon}_p l$. Thus,

$$\begin{aligned}\varphi_p &= \frac{1}{N}\sum_{l=0}^{N-1}\left\{1 + j\frac{2\pi}{N}\tilde{\epsilon}_p l\right\} e^{j\frac{2\pi}{N}\tilde{\epsilon}_p N_{\mathrm{CP}}} \\ &= \left(1 + j\frac{2\pi}{N}\tilde{\epsilon}_p\frac{N-1}{2}\right)e^{j\frac{2\pi}{N}\tilde{\epsilon}_p N_{\mathrm{CP}}} \\ &= \left\{\cos\left(\frac{2\pi}{N}\tilde{\epsilon}_p\frac{N-1}{2}\right) + j\sin\left(\frac{2\pi}{N}\tilde{\epsilon}_p\frac{N-1}{2}\right)\right\} e^{j\frac{2\pi}{N}\tilde{\epsilon}_p N_{\mathrm{CP}}} \\ &= e^{j\frac{2\pi}{N}\tilde{\epsilon}_p(\frac{N-1}{2}+N_{\mathrm{CP}})} \\ &= \tilde{\Phi}_p[l,l]\big|_{l=\frac{N-1}{2}}.\end{aligned} \tag{4.61}$$

Therefore, with a similar argument, if we replace $\tilde{\Phi}_p[l,l]$ for $l = 0, 1, \cdots, N-1$ with $\tilde{\Phi}_p[\frac{N}{2}, \frac{N}{2}]$ in equation (4.53), we can also consider $\Gamma_2[i,j] \approx \varphi_p^*\Gamma_1[i,j]$ for $|i-j| \geq N$. Note that since N is set as an even number, and consequently, $\frac{N-1}{2}$ is not an integer, we have chosen $\frac{N}{2} = \left\lceil\frac{N-1}{2}\right\rceil$. Then, we can conclude $\mathbf{\Gamma}_1^{-1}\mathbf{\Gamma}_2 \approx [\varphi_0^* I_N, \varphi_1^* I_N, \cdots, \varphi_P^* I_N]$. Hence, the estimated channels that are used for the ZF

combiner can be written as

$$\hat{\mathbf{h}}_{m,p} = \varphi_p^* \mathbf{h}_{m,p} + \breve{\mathbf{n}}_m, \tag{4.62}$$

where $\breve{\mathbf{n}}_m = (\hat{\mathbf{\Lambda}}^{\mathrm{H}}\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{\Lambda}}^{\mathrm{H}}\mathbf{n}_m$. By ignoring the MUI and the noise term in equation (4.41) due to their small values, the combined signal of user q is given by

$$\bar{\mathbf{y}}_q = \widetilde{\mathbf{\Omega}}_{q,q}\mathbf{d}_p, \tag{4.63}$$

where $\widetilde{\mathbf{\Omega}}_{q,q} = \mathbf{E}_q \odot \widetilde{\mathbf{B}}_{q,q}$ and the elements of $\widetilde{\mathbf{B}}_{q,q}$ can be obtained as

$$\begin{aligned}
\widetilde{B}_{q,q}[l,k] =& \mathbb{E}\left\{\sum_{i=0}^{N-1} \hat{h}^*_{m,q}[i]h_{m,q}[i]e^{-j\frac{2\pi}{N}(k-l)i}\right\} \\
=& \varphi_q \sum_{i=0}^{N-1} \mathbb{E}\left\{h^*_{m,q}[i]h_{m,q}[i]\right\} e^{-j\frac{2\pi}{N}(k-l)i} \\
=& \varphi_q \bar{\rho}_q[k-l],
\end{aligned} \tag{4.64}$$

With a similar argument as in Section 4.2.1, one can prove that $\widetilde{\mathbf{\Omega}}_{q,q}$ is a circulant matrix and can be written as $\widetilde{\mathbf{\Omega}}_{q,q} = \mathbf{F}_N\widetilde{\mathbf{Q}}_{q,q}\mathbf{F}_N^{\mathrm{H}}$, where $\widetilde{\mathbf{Q}}_{q,q}$ is a diagonal matrix with the main diagonal elements

$$\begin{aligned}
\widetilde{Q}_{q,q}[i,i] &= \sum_{n=0}^{N-1} \varphi_q \tilde{\rho}_q[i-n]\Phi_q[n,n] \\
&= \varphi_q Q_{q,q}[i,i].
\end{aligned} \tag{4.65}$$

Consequently, in the proposed CFO compensation, we need to calculate the interference matrix using the estimated values of the CFOs and channel impulse responses, i.e. $\widehat{\mathbf{\Omega}}_{q,p} = \widehat{\mathbf{E}}_q \odot \widehat{\mathbf{B}}_{q,p}$. Similarly, the elements of the matrix $\widehat{\mathbf{B}}_{q,p}$ are calculated as

$$\begin{aligned}
\widehat{B}_{q,q}[l,k] =& \mathbb{E}\left\{\sum_{i=0}^{N-1} \hat{h}^*_{m,q}[i]\hat{h}_{m,q}[i]e^{-j\frac{2\pi}{N}(k-l)i}\right\} \\
=& \bar{\rho}_q[k-l],
\end{aligned} \tag{4.66}$$

which is equal to $B_{q,q}[l,k]$ in the case of accurate CFO estimation. Using the result in (4.61), it can be shown that

$$\widehat{\mathbf{E}}_q = \mathbf{F}_N\hat{\mathbf{\Phi}}_p\mathbf{F}_N^{\mathrm{H}} = \varphi_q\mathbf{E}_q, \tag{4.67}$$

and consequently, $\widehat{\mathbf{Q}}_{q,q} = \widetilde{\mathbf{Q}}_{q,q} = \varphi_q\mathbf{Q}_{q,q}$. As a result, the calculated interference matrix from the estimated values, $\widehat{\mathbf{\Omega}}_{q,q}$, is equal to the interference matrix in the

case of inaccurate CFO estimation, $\widetilde{\mathbf{\Omega}}_{q,q}$. Thus, by multiplying $\widehat{\mathbf{\Omega}}_{q,q}^{(-1)}$ to equation (4.63), the q^{th} user's signal, $\mathbf{d}_q$, can be extracted. As it is mentioned in Section 4.2.1, $\widehat{\mathbf{Q}}_{q,q}$ is calculated only for the first symbol, and the phase shifts of other symbols are corrected later. Thus, the data of the κ^{th} symbol can be obtained as

$$\begin{aligned}\hat{\mathbf{d}}_q^{\kappa} &= e^{-j\frac{2\pi}{N}(\hat{\epsilon}_q-\epsilon_q)(N+N_{\text{CP}})\kappa}\widehat{\mathbf{\Omega}}_{q,q}^{(-1)}\widetilde{\mathbf{\Omega}}_{q,q}\mathbf{d}_q^{\kappa} \\ &= e^{-j\frac{2\pi}{N}\tilde{\epsilon}_q(N+N_{\text{CP}})\kappa}\mathbf{d}_q^{\kappa},\end{aligned} \tag{4.68}$$

Therefore, by applying the proposed CFO compensation technique, the estimation error is absorbed into the interference matrix, and after compensation only a phase shift remains.

It is worth mentioning that though the residual phase shift in this proposed compensation technique is small, it progressively increases symbol by symbol, and it can result in a large rotation of the constellation. In the next section, a phase correction technique is proposed to calculate and effectively eliminate this error.

4.3.2 CFO estimation error correction

In this section, we suggest an approach to efficiently compensate the phase shift due to CFO estimation error. As it was discussed earlier, this error is zero for the first symbol and we cannot use the pilot to calculate this shift. Thus, we need to compare the users' signals with the constellation points and find the closest point as

$$\begin{aligned}\check{d}_q^{\kappa}[k] =& \underset{c}{\text{argmin}}|\hat{d}_q^{\kappa}[k]-c|. \\ &\text{s.t.} \quad c\in\mathbb{S}\end{aligned} \tag{4.69}$$

where $\mathbb{S}$ is the set of all the possible constellation points. Then, from equation (4.68), the estimation error can be calculated as

$$\tilde{\epsilon}_q = \frac{1}{2\pi(N+N_{\text{CP}})\kappa}\sum_{k=0}^{N-1}\angle\check{d}_q^{\kappa}[k]-\angle\hat{d}_q^{\kappa}[k]. \tag{4.70}$$

where $\kappa\neq 0$. After calculating the error, we can correct the phase shift error for the current symbol as

$$\hat{\mathbf{d}}_p^{\kappa} := e^{j\frac{2\pi}{N}\tilde{\epsilon}_q(N+N_{\text{CP}})\kappa}\hat{\mathbf{d}}_q^{\kappa}, \tag{4.71}$$

and update the estimated CFO for the phase correction step of the following symbols as $\hat{\epsilon}_q := \hat{\epsilon}_q-\tilde{\epsilon}_q$. In the asymptotic regime, the error can be calculated precisely. Thus, after updating the estimated CFO, $\hat{\epsilon}_q$, for one symbol, there will be no phase shift

error on the following symbols. However, in practice when M is not very large, MUI and noise are not completely averaged out, and the CFO estimation error in general cannot be accurately calculated and corrected. To deal with this issue, we can execute the correction loop for more than one symbol. The CFO compensation process with error correction for N_{cor} number of symbols from N_{sym} transmitted symbols is presented in Algorithm 1.

Algorithm 1 CFO compensation with error correction

Calculate $\mathbf{Q}_{q,q}^{0}$ from equation (4.48) **for** $\kappa = 0, 1, \cdots, N_{\text{sym}}$ **do**
  $\hat{\mathbf{d}}_q^{\kappa} = e^{-j\frac{2\pi}{N}\hat{\epsilon}_p(N+N_{\text{CP}})\kappa}\mathbf{F}_N(\mathbf{Q}_{q,q}^{0})^{-1}\mathbf{F}_N^{\text{H}}\bar{\mathbf{y}}_q^{\kappa}$ **if** $1 \leq \kappa \leq N_{\text{cor}}+1$ **then**
    Calculate $\check{d}_q^{\kappa}$ from equation (4.69)
    Calculate $\tilde{\epsilon}_q$ from equation (4.70)
    $\hat{\mathbf{d}}_p^{\kappa} := e^{j\frac{2\pi}{N}\tilde{\epsilon}_q(N+N_{\text{CP}})\kappa}\hat{\mathbf{d}}_q^{\kappa}$
    $\hat{\epsilon}_q := \hat{\epsilon}_q - \tilde{\epsilon}_q$
  end
end

It should be noted that since $\tilde{\epsilon}_q$ is small, most of the elements in the received symbol fall in the correct decision areas. However, for large constellation sizes, even a small phase shift might cause some errors. We suggest to repeat the process of error calculation and correction until we get the error equal to zero. In fact, in each iteration, since we turn the constellation, a larger number of points will fall in their respective decision areas, and an improved phase shift error calculation can be achieved. Algorithm 2 summarizes the CFO compensation process with iterative error correction, where ξ indicates the convergence tolerance.

Algorithm 2 CFO compensation with iterative error correction

Calculate $\mathbf{Q}_{q,q}^{0}$ from equation (4.48) **for** $\kappa = 0, 1, \cdots, N_{\text{sym}}$ **do**
  $\hat{\mathbf{d}}_q^{\kappa} = e^{-j\frac{2\pi}{N}\hat{\epsilon}_p(N+N_{\text{CP}})\kappa}\mathbf{F}_N(\mathbf{Q}_{q,q}^{0})^{-1}\mathbf{F}_N^{\text{H}}\bar{\mathbf{y}}_q^{\kappa}$ **if** $1 \leq \kappa \leq N_{\text{cor}}+1$ **then**
    while $\tilde{\epsilon}_q > \xi$ **do**
      Calculate $\check{d}_q^{\kappa}$ from equation (4.69)
      Calculate $\tilde{\epsilon}_q$ from equation (4.70)
      $\hat{\mathbf{d}}_p^{\kappa} := e^{j\frac{2\pi}{N}\tilde{\epsilon}_q(N+N_{\text{CP}})\kappa}\hat{\mathbf{d}}_q^{\kappa}$
      $\hat{\epsilon}_q := \hat{\epsilon}_q - \tilde{\epsilon}_q$
    end
  end
end

4.4 Computational Complexity analysis

In this section, we calculate the computational complexity [2] of our proposed technique and compare it to the other techniques in the literature. According to our proposed frequency synchronization technique, the CFO estimation is performed in the time domain, and the CFO compensation takes place after combining the received signals over all the BS antennas in the frequency domain. It means that unlike the other existing synchronization techniques, in our case, one set of DFT operations can be used for all the user signals and the separate receivers are not required. Therefore, ignoring the complexity of the channel estimation, the total number of complex multiplications (CMs) done at the receiver with the proposed synchronization technique can be denoted as

$$\mathcal{C}_{\mathrm{p}} = \mathcal{C}_{\mathrm{p}}^{\mathrm{Est}} + M\mathcal{C}^{\mathrm{DFT}} + P\mathcal{C}^{\mathrm{Comb}} + P\mathcal{C}_{\mathrm{p}}^{\mathrm{Comp}}, \tag{4.72}$$

where $\mathcal{C}_{\mathrm{p}}^{\mathrm{Est}}$, $\mathcal{C}^{\mathrm{DFT}}$, $\mathcal{C}^{\mathrm{Comb}}$, and $\mathcal{C}_{\mathrm{p}}^{\mathrm{Comp}}$ are the number of CMs required for our proposed CFO estimation technique, DFT operation, signal combiner, and our proposed CFO compensation technique, respectively.

For our proposed CFO estimation, the matrix $\mathbf{R}$ in equation (4.2) is calculated with MN^2 CMs. Then, calculating the cost function for each trial CFO requires N^2 CMs for the matrix in equation (4.7) and N CMs for the Frobenius norm of the resulting matrix. Therefore, the total computational complexity of our proposed estimation is given by

$$\mathcal{C}_{\mathrm{p}}^{\mathrm{Est}} = MN^2 + i_c P(N^2 + N), \tag{4.73}$$

[2]The computational complexity or simply complexity of an algorithm is the amount of resources required to run it, where the main resources are time and memory [122]. The complexity of an algorithm can be calculated differently based on the level of our scientific computing and the type of our problem (whether it is narrow or broad). For high-level scientific computing, a simple and machine-independent measure suffices. Thus, the amount of computational complexity of an algorithm in signal processing can be calculated in terms of floating-point operations (FLOPs) [123]. A FLOP serves as a basic unit of computation. It can be defined as an addition, subtraction, multiplication or division of two floating point numbers. However in practice, as multiplication is more expensive to compute, we can approximate one multiplication to one FLOP to simplify the overall calculations, and achieve a rough measure of how expensive an algorithm can be. It is worth noticing that many more aspects need to be taken into account to accurately estimate practical run-time including the implementation technique and the hardware used. But FLOPs counting is quite useful in comparing the complexity of different algorithms. Therefore, in this manuscript, we have calculated the computational complexity based on the number of complex multiplication operations required to run an algorithm. A complex multiplication consists of four real multiplications, and the run time of one real multiplication depends on the multiplication algorithm and the size of the multiplied values. For example, the time-complexity of multiplication of two n-digit numbers with Schoolbook long multiplication, Mixed-level Toom–Cook and Harvey-Hoeven algorithm are $O(n^2)$, $O(2^{\sqrt{2\log n}} n \log n)$ and $O(n \log n)$ [124–126], where $O(.)$ notation is used to indicate how an algorithm run time or space requirements grow as the input size grows.

where i_c is the number of trial CFO in the Golden search algorithm. Afterward, the complexity of DFT operation is $\mathcal{C}^{\text{DFT}} = \frac{N}{2}\log_2 N$ if FFT is exploited, and the complexity of the utilized ZF combiner is $\mathcal{C}^{\text{Comb}} = 3NMP^2$.

For our proposed CFO compensation, first, we need to obtain the matrix $\boldsymbol{Q}_{q,q}$, which requires NL number of CMs due to the circular convolution process introduced in equation (4.46). Note that $\mathbf{B}_{q,q}$ and $\mathbf{E}_q$ are both circulant matrices with the first column $\bar{\boldsymbol{\rho}}^*$ and the DFT of the main diagonal of $\boldsymbol{\Phi}_q$, i.e., $\mathbf{F}_N \text{diag}(\boldsymbol{\Phi}_q)$, respectively. Thus, we only need two DFT operations to calculate their first columns, and then we can form the entire matrices. Since the multiplication of two circulant matrices is also a circulant matrix, for the element-wise multiplication in (4.42), only N number of CMs are required to obtain the first column of $\boldsymbol{\Omega}_{q,q}$, which is the DFT of the main diagonal of $\mathbf{Q}_{q,q}$. Afterward, as it is indicated in equation (4.49), we take $\bar{\mathbf{y}}_q$ to the time domain, divide it element-wise by the elements on the main diagonal of $\mathbf{Q}_{q,q}$, and bring it back to the frequency domain. Therefore, two N-point IDFT operations, N number of CMs and one N-point DFT operation are needed to obtain the compensated signal. At the end, $i_p N N_{\text{cor}}$ number of CMs is required if the phase correction algorithm performs over N_{cor} OFDM symbols, where i_p is the number of iterations. Therefore, the number of required CMs for our proposed compensation technique is

$$\mathcal{C}_p^{\text{Comp}} = N(\log_2 N + L + 1 + i_p N_{\text{cor}}). \tag{4.74}$$

As a comparison, in all the other existing synchronization techniques, the CFO compensation is performed in the time domain by multiplying the compensation vector of the desired user, i.e., $\text{diag}(\boldsymbol{\Phi}_{\text{q}}^{\text{H}})$ to the received signals at the BS antennas. Thus, $\mathcal{C}_{\text{t}}^{\text{Comp}} = MN$ number of CMs are required for CFO compensation, which increase linearly with the number of BS antennas, M; the computational complexity of our CFO compensation technique is instead independent of the number of BS antennas and stays constant if M grows. In addition, each user needs a separate set of DFT operations. Hence, the total computational complexity of the receiver can be obtained by

$$\mathcal{C} = \mathcal{C}^{\text{Est}} + PMC^{\text{DFT}} + P\mathcal{C}^{\text{Comb}} + P\mathcal{C}_{\text{t}}^{\text{Comp}}, \tag{4.75}$$

where $\mathcal{C}^{\text{DFT}}$ and $\mathcal{C}^{\text{Comb}}$ are the same as in equation (4.72).

Regarding the CFO estimation, the number of required CMs of the lowest computationally complex technique in the literature, [33], is given by

$$\mathcal{C}_1^{\text{Est}} = 2M(N - PL) + P. \tag{4.76}$$

However, the performance of the technique in [33] relies on the long pilot sequences. As it is presented in Section 4.5, with the same pilot length, our proposed technique can provide a considerably better performance. Later, the work in [34], proposed a rather low-complexity CFO estimation technique in which the number of required CMs is obtained by

$$\mathcal{C}_2^{\text{Est}} = PMN^{(5/2)}. \tag{4.77}$$

Note that by increasing the number of BS antennas, the computational complexity of the technique in [34] grows with the slope of $PN^{(5/2)}$, while the slope in our case is only N^2 which is also independent of the number of users. More recently a scattered pilot-based frequency synchronization was proposed in [29] that uses N_p out of N subcarriers as pilots. It is also considered that the channels and CFOs of all users stay constant over L_b OFDM blocks. Then, the number of required CMs for the CFO estimation is obtained as

$$\begin{aligned}\mathcal{C}_3^{\text{Est}} &= i_c P(L_b NM \log_2 N + M(L_b - 1)N_p + 6P \\ &\quad \times (L_b NM \log_2 N + (L_b - 1)M^2 N_p/2 + M^3)) \\ &\quad + L_b NM^2/2 + M^3.\end{aligned} \tag{4.78}$$

In addition, a computationally efficient blind CFO estimation is proposed in [31] that considers N_0 out of N subcarriers as null subcarriers. Then, its total computational complexity is equal to

$$\begin{aligned}\mathcal{C}_4^{\text{Est}} &= P(i_c(L_b MN \log_2 N + M^3 + M^2 N_0 L_b/2) \\ &\quad + 3(L_b MN \log_2 N + M^2 N_0 L_b/2 + L_b M_e MN + M^3) \\ &\quad + 11(L_b M_e N \log_2 N + M_e L_b N + NM_e^2 L_b/2) \\ &\quad + 3\tbinom{M_e}{2}8N + 8NM_e^3),\end{aligned} \tag{4.79}$$

where $M_e = M - (P-1)L$.

As an illustration, a numerical comparison of all the aforementioned CFO estimation techniques is provided in table 4.1. We consider $P = 4$ users, $N = 256$ subcarriers, and CIR length of $L = 8$. According to the suggested values in [29] and [31], we consider $L_b = 2$, $N_p = N/2 = 128$, $N_0 = 8$. For the sake of fairness, we present the number of CMs as a function of the required number of iterations, i_c, for iterative algorithms. However, in [29] and [31], $i_c = 120$ is suggested, while in our case, $i_c = 11$ is sufficient. Furthermore, table 4.2 presents the computational complexity of the receiver with different CFO synchronization techniques. As it is justified in section 4.5, in the case of 16-QAM modulation, we can consider $N_{\text{cor}} = 4$ and $i_p = 3$ for our iterative error correction in Algorithm 2. As it is shown, our

Table 4.1 Computational complexity of the different CFO estimation techniques

Techniques	M=100	M=200	M=400
Proposed technique	$(0.2i_c + 6.5) \times 10^6$	$(0.2i_c + 13) \times 10^6$	$(0.2i_c + 26) \times 10^6$
Low-complexity [33], $\mathcal{C}_1^{\text{Est}}$	4.5×10^4	9×10^4	1.8×10^5
Constant-envelop pilot [34], $\mathcal{C}_2^{\text{Est}}$	4.2×10^8	8.3×10^8	1.7×10^9
Scattered pilot [29], $\mathcal{C}_3^{\text{Est}}$	$(20i_c + 0.4) \times 10^7$	$(10i_c + 0.2) \times 10^8$	$(7.3i_c + 0.1) \times 10^9$
Efficient Blind [31], $\mathcal{C}_4^{\text{Est}}$	$(0.6i_c + 4000) \times 10^7$	$(0.4i_c + 5000) \times 10^8$	$(0.3i_c + 5000) \times 10^9$

Table 4.2 Computational complexity of the receiver with different CFO synchronization techniques

Techniques	M=100	M=200	M=400
Proposed technique	$(0.2i_c + 26) \times 10^6$	$(0.2i_c + 53) \times 10^6$	$(0.2i_c + 100) \times 10^6$
Low-complexity [33]	2×10^7	4×10^7	8×10^7
Constant-envelop pilot [34]	4.4×10^8	9×10^8	1.8×10^9
Scattered pilot [29]	$(20i_c + 2.3) \times 10^7$	$(10i_c + 0.6) \times 10^8$	$(7.3i_c + 0.2) \times 10^9$
Efficient Blind [31]	$(0.6i_c + 4000) \times 10^7$	$(0.4i_c + 5000) \times 10^8$	$(0.3i_c + 5000) \times 10^9$

proposed CFO estimation technique has lower computational complexity than other techniques except for the one proposed in [33]. However, it is demonstrated in section 4.5 that our proposed technique provides a considerably higher performance as compared to the technique in [33].

4.5 Simulation results

In this section, we confirm our theoretical developments in the previous sections through numerical simulations. We assume that $P = 4$ users are communicating with a BS with $M = 200$ antennas. We also assume perfect power control for all the users, 16-QAM modulation, and $N_{\text{sym}} = 10$ OFDM symbols in each data packet. The extended typical urban (ETU) channel model, employed in Long Term Evolution (LTE) standard, is considered, [127], and the CP length is $N/8$. The normalized CFO is randomly generated from a uniform distribution within the range

$[-0.5, 0.5)$. In our simulations, the transmitted symbols are randomly chosen from 16-QAM constellation points. The PDP of the users' channels and the variance of the AWGN noise are assumed to be unchanged during one frame with 10 OFDM symbols.

First, we study the proposed CFO estimation with closed-form formula. Fig. 4.1 presents the mean square error (MSE) of this proposed CFO estimation technique. As a comparison, we also plot the MSE of the CFO estimation method proposed in [33], which has the lowest computational complexity among other techniques. All the MSE curves presented in this section are the average of the MSEs obtained from 10^5 number of iterations. As it is shown, having the same pilot length for a fair comparison, we can achieve more accurate estimation by using the proposed rectangular shape pilot sequences, i.e. around 6 dB and 4 dB improvements in MSE curves for $N = 128$ and $N = 256$ subcarriers, respectively. In fact, by using the impulse-like pilots proposed in [33], the covariance matrix of the received signals is a set of diagonal matrices. Instead by using the proposed rectangular shaped pilots, the covariance matrix is a set of banded matrices, and consequently, we have a larger number of samples from which we can estimate the users' CFOs. In Fig. 4.2, it is depicted that the performance of our proposed technique is not affected by increasing the number of users. Moreover, the accuracy of the CFO estimation can be further enhanced by increasing the number of BS antennas. The bit error rate (BER) curves are depicted in Fig. 4.3 for $M = 200$ BS antennas, $P = 4$ users and different number of subcarriers. In the simulation, 4-QAM constellation is chosen, and MRC is used as a channel equalizer. The LS channel estimator is employed to estimate the channel between users and each BS antenna [128]. We have compared the BER performance of our proposed technique with that of the perfect synchronization as a benchmark.

Then, we investigate the performance of the proposed CFO estimation technique with unimodal cost function which is more general compared to the previous one. It can also work with a wider range of CFO values as it is shown in Fig. 4.4, where we depict the MSE performance of this proposed estimation technique for different CFO ranges for SNR$= -1$ dB. In order to obtain the curves in Fig. 4.4, the normalized CFO is randomly generated from a uniform distribution within the range $[-\epsilon_{\max}, \epsilon_{\max}]$, where $\epsilon_{\max}$ is the maximum CFO. It is shown that the performance of our proposed technique without correction is the same for all CFO values. Our results in Fig. 4.4 show that a substantial improvement can be achieved through application of our proposed error correction loop. Here, we have set $N_{\text{cor}} = N_{\text{sym}}$ and $\xi = 10^{-10}$. The MSE deterioration for large CFO values around 100 is due to the residual interference from noise and MUI, which can be reduced by deploying a

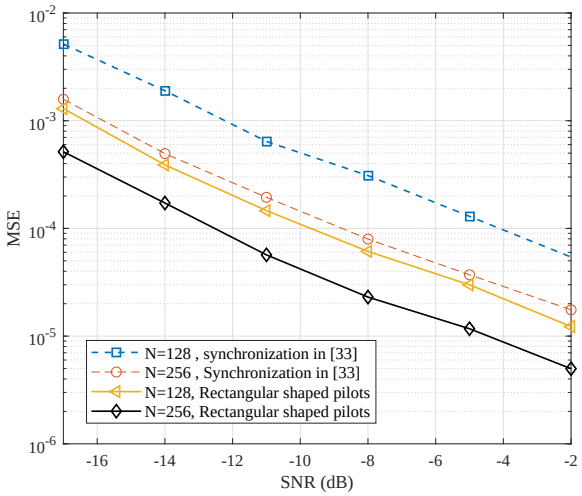

Figure 4.1 MSE of the proposed CFO estimation technique for $M = 200$ BS antennas, $P = 4$ users and different number of subcarriers, N.

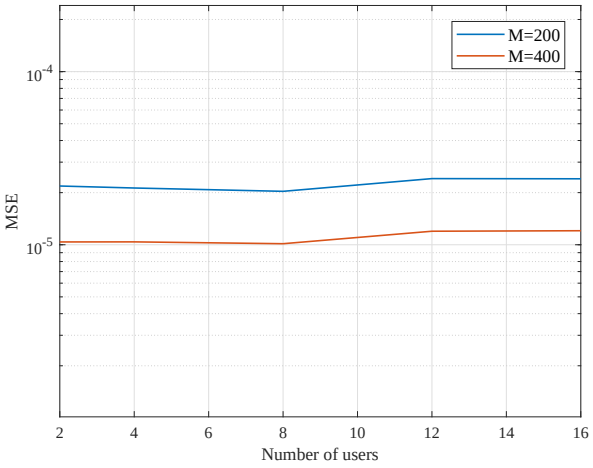

Figure 4.2 MSE of the proposed CFO estimation technique with respect to the number of users, P, for $N = 512$ subcarriers, and SNR $= -10$ dB.

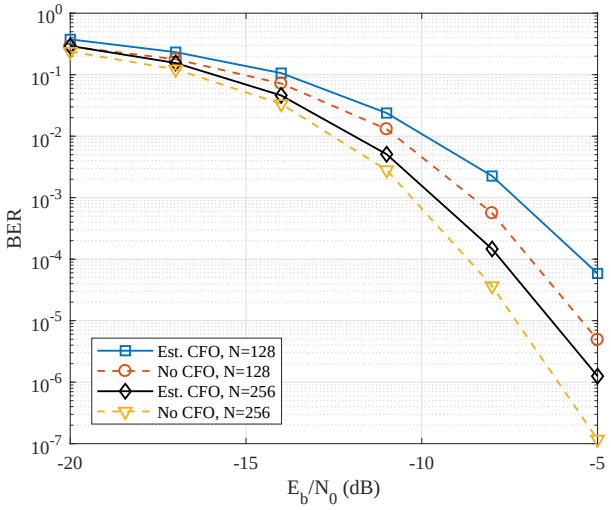

Figure 4.3 BER performance of the proposed CFO estimation for $M = 200$ BS antennas, $P = 4$ users.

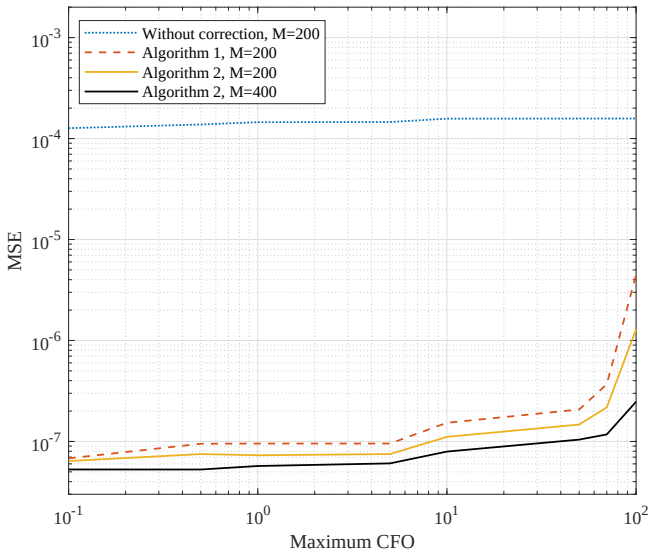

Figure 4.4 MSE of the proposed CFO estimation technique as a function of CFO for $N = 256$ subcarriers, 16-QAM modulation, and SNR$= -1$.

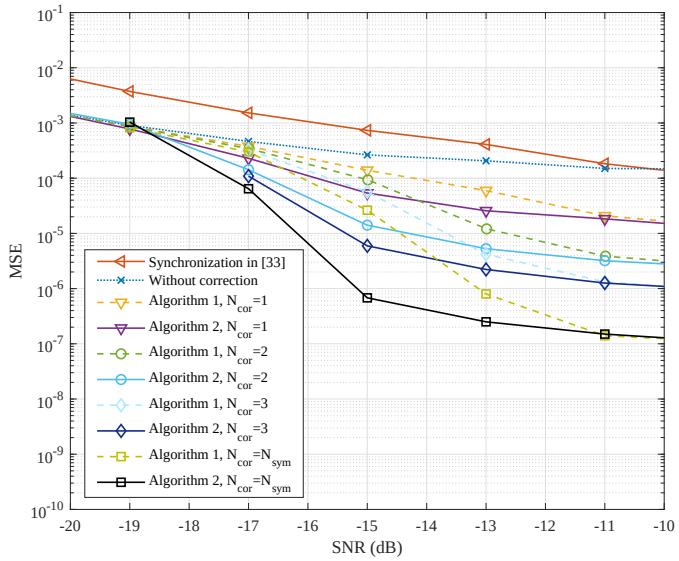

Figure 4.5 MSE of our proposed CFO estimation technique for $M = 200$ BS antennas, $P = 4$ users, $N = 256$ subcarriers, and 4-QAM modulation.

larger number of BS antennas. Since the integer part of CFO can be estimated correctly and the error mainly comes from the fractional part of the CFO, we consider $\epsilon_{\max} = 0.5$ for the rest of our simulations.

In Fig. 4.5, the MSE performance of the system with 4-QAM modulation is depicted. Again, we compare the MSE performance with the one in [33], since it has the lowest computational complexity among other techniques. It is shown that while the performance of our proposed CFO estimation technique without error correction is close to the technique in [33], applying the estimation error correction algorithms lead to around three orders of magnitude improvement in the CFO estimation accuracy. Note that the error correction algorithm only corrects the phase shift error due to the CFO estimation error and does not change the CFO compensation matrix. Therefore, this performance improvement proves our claim that our proposed CFO compensation technique can effectively eliminate the scattering error and the phase shift is the only source of error. Note that this phase correction cannot enhance the performance of a system with the conventional CFO compensation, because the constellation of the received signal is scattered due to the CFO estimation error. In Fig. 4.6, we demonstrate the BER performance of the system to evaluate our proposed CFO estimation and compensation techniques. Without loss of generality, our simulation results are obtained by transmission of raw data without application of any error correction coding technique. We compare our results with the perfect synchronization case as a benchmark. This figure shows that the BER performance is significantly improved through the proposed error correction algorithms, and for

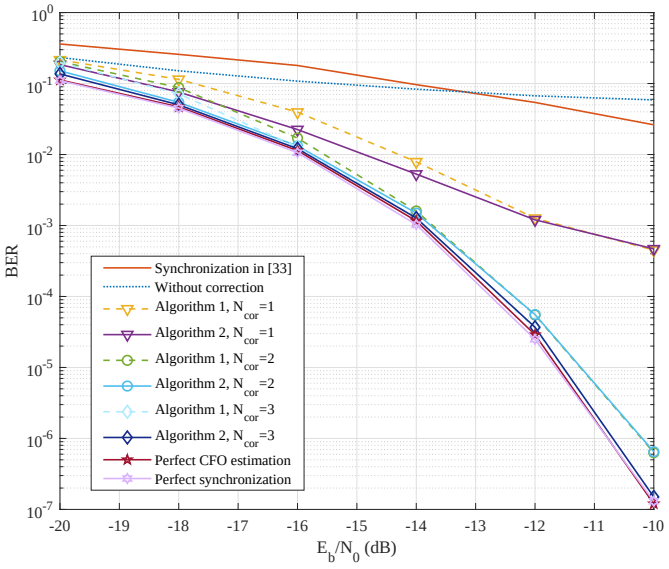

Figure 4.6 BER performance for $M = 200$ BS antennas, $P = 4$ users, $N = 256$ subcarriers, and 4-QAM modulation.

$N_{\text{cor}} = 3$ iterations, it almost matches the curve of the perfect CFO estimation, where $\tilde{\epsilon} = 0$.

The iterative correction in Algorithm 2 is more beneficial for larger constellation sizes, where the constellation points are closer to one another. Thus, in order to highlight the difference between the two proposed algorithms, Fig. 4.7 presents the MSE performance for the system with 16-QAM modulation. As it is shown, the iterative correction in Algorithm 2 seems necessary to achieve an MSE of 10^{-7}. Note that according to equation (4.41), although the noise is negligible at higher SNRs, the MUI remains. Therefore, by taking into account the residual MUI, equation (4.49) can be updated as

$$\hat{\mathbf{d}}_q^{\kappa} = e^{-j\frac{2\pi}{N}\epsilon_q(N+N_{\text{CP}})\kappa}\mathbf{\Omega}_{q,q}^{(-1)}(\mathbf{d}_q^{\kappa} + \sum_{\substack{p=0 \\ p\neq q}}^{P-1}\mathbf{\Omega}_{q,p}\mathbf{d}_p). \tag{4.80}$$

As the number of BS antennas grows large, the MUI term will be almost averaged out. It is worth noting that even a small amount of the residual MUI scatters the received signal constellation. This prevents us from achieving a perfect error calculation, and reflects as an error floor in the MSE curves. Furthermore, the BER performance of the system with 16-QAM modulation is studied in Fig. 4.8. As one can see, Algorithm 1 can improve the performance and outperform the technique in [33]. However, by increasing N_{cor} in the Algorithm 2, we can get the BER curve close to the perfect synchronization case. In addition, Fig 4.9 demonstrates the BER

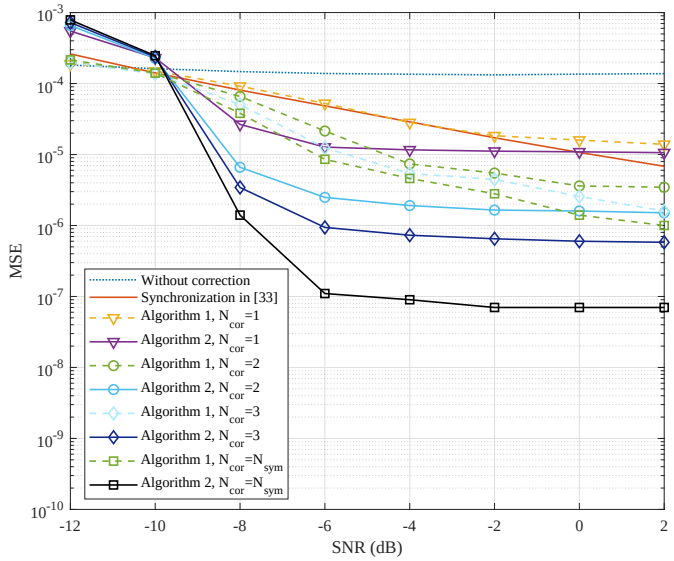

Figure 4.7 MSE of the proposed CFO estimation technique for $M = 200$ BS antennas, $P = 4$ users and $N = 256$ subcarriers with 16-QAM modulation.

performance of the proposed synchronization technique for a large constellation size, i.e. 64-QAM. From Fig 4.9, one can realize that, the proposed synchronization technique is effective enough to work for large constellation sizes. However, the scattering effect of the residual MUI is more problematic for 64-QAM than smaller constellations and preventing us from achieving a perfect synchronization.

Furthermore, Fig. 4.10 shows how increasing the number of BS antennas can affect the MSE performance of our proposed CFO estimation technique. Regarding the proposed CFO estimation technique in Section 4.1.2, a large number of BS antennas is required to have the noise terms efficiently averaged out and MUI diminished. As one can notice from Fig. 4.10, after a point where these requirements are fulfilled, no further enhancement can be achieved by increasing the number of BS antennas. On the other hand, the number of BS antennas should be large enough to provide an accurate equalization and CFO compensation, and separate the users' signals. Otherwise, the input signal to the error correction algorithms would be very scattered and lead to a poor phase shift error correction. The iterative error calculation in Algorithm 2 can deal with scattered data and provide the maximum MSE.

Finally, the MSE performance of our proposed CFO estimation technique with respect to the number of users is presented in Fig. 4.11. This figure demonstrates that our proposed synchronization technique can support a large number of users. In fact, since in massive MIMO systems, users share the whole available bandwidth, increasing the number of users affects the system performance as if the noise level is

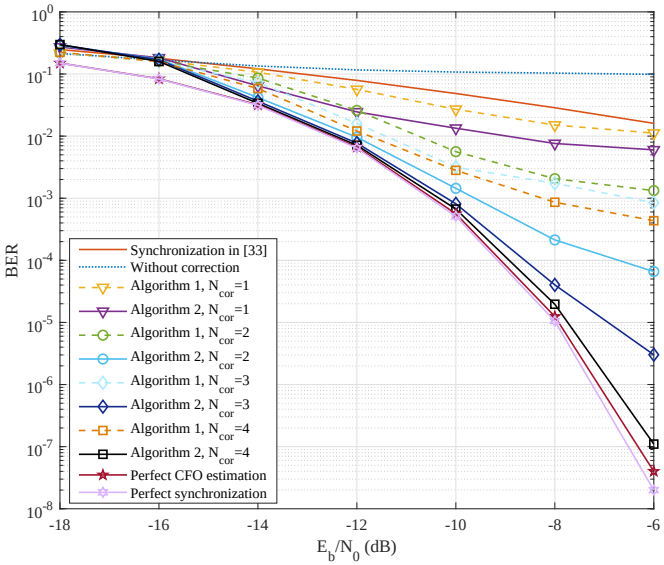

Figure 4.8 BER performance for $M = 200$ BS antennas, $P = 4$ users, $N = 256$ subcarriers and 16-QAM modulation.

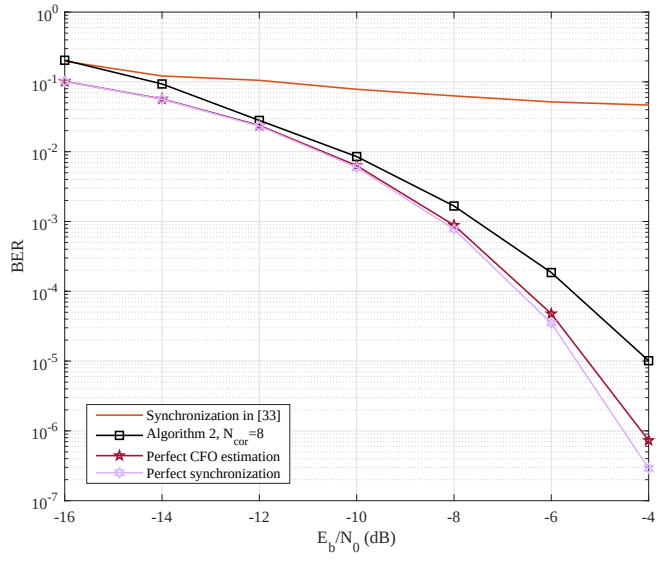

Figure 4.9 BER performance for $M = 200$ BS antennas, $P = 4$ users, $N = 256$ subcarriers and 64-QAM modulation.

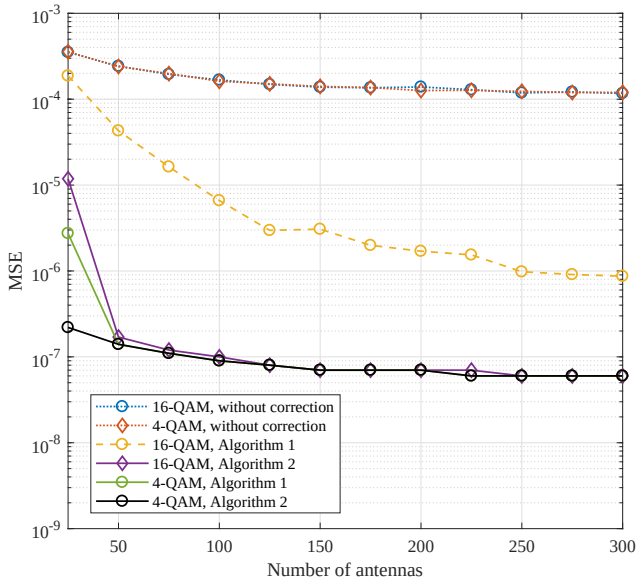

Figure 4.10 MSE of the proposed CFO estimation technique versus number of BS antennas for $SNR = -1$ dB, $P = 4$ users and $N = 256$ subcarriers.

increased. As it is shown, larger number of BS antennas can average out the MUI and improve the performance.

4.6 Conclusion

We proposed a new frequency synchronization technique for the uplink of multi-user OFDM-based massive MIMO systems. First, we proposed a CFO estimation technique with closed-form formula, which needs rectangular shaped pilots. This technique has a low computational complexity, and unlike the low-complexity technique proposed in [33], it does not require a long pilot. However, it suffers from high PAPR issue. Therefore, we proposed another CFO estimation technique whose complexity increases also linearly with respect to the number of BS antennas. Then, we proposed a CFO compensation technique that takes place after combining the received signals at all the BS antennas. Consequently, one receiver is sufficient for all the users. Moreover, its computational complexity is independent of the number of BS antennas.

We also proved that after performing our CFO compensation technique, the scattering effect of CFO is corrected and only a constant phase shift remains which is due to CFO estimation error. Then, we proposed an algorithm to efficiently calculate and remove this phase shift error. Numerical results were presented to verify the performance of our proposed synchronization technique. It was shown that the BER performance of our proposed technique is very close to that of a fully synchronous

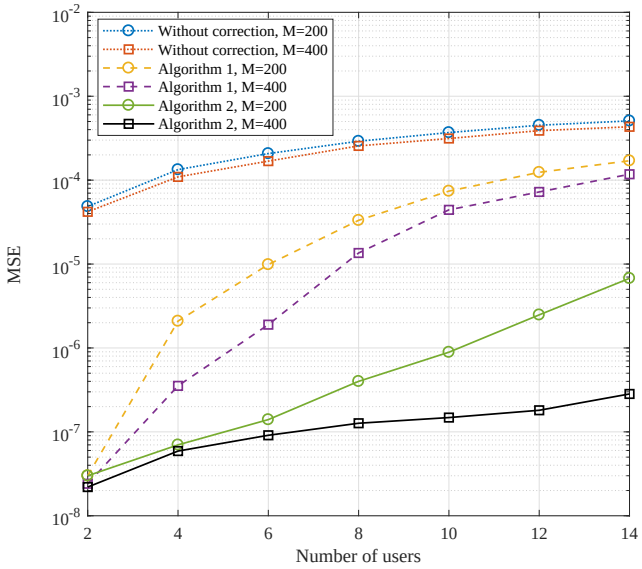

Figure 4.11 MSE of the proposed CFO estimation technique respect to number of users for $N = 256$ subcarriers, and $SNR = -1$ dB.

system. We also evaluated the CFO estimation accuracy and showed that MSE enhancement of three orders of magnitude can be achieved through application of our proposed error correction technique.

Chapter 5

Deep learning aided blind channel estimation for massive MIMO systems

CSI acquisition is one of the most important task in massive MIMO systems. This is because, as explained in section 2.1.1, the orthogonality between users is granted by spatial diversity, and therefore channel impulse responses are necessary to distinguish users' signals. Moreover, the knowledge of all the channels between users and a base station is required for precoding in downlink transmission, as discussed in section 2.1.2. However, as the scale of the network or the antennas array increases the estimation of CSI becomes one of the key obstacles for the utilization of massive MIMO systems. In addition, when the pilot sequences are reused in neighboring cells, the pilot contamination prevents the accurate channel estimation and degrades the system performance. The thorough review of the existing works on channel estimation for massive MIMO systems is presented in section 2.3.

In this chapter, we develop a DL-based blind channel estimation technique for massive MIMO systems that is performed in three steps; (i), the multi-user interference (MUI) is canceled by calculating the orthogonal complement space matrix of the MUI, (ii), based on the asymptotic orthogonality of the channel vectors in very large MIMO systems, the data of the first symbol of all the users are extracted separately using a DL algorithm, (iii), using the detected data as virtual pilots, we estimate the channel impulse responses between all the users and the BS antennas. At the end, we detect the transmitted symbols based on the estimated channels. Our proposed virtual pilot detection technique can be considered as a model-driven block-structured DL-based approach, since we use the prior knowledge about the asymptotic behavior of the received signals and propose some pre-processing and

mathematical manipulations alongside the DNN. More details are presented in the following sections.

5.1 System model

In this chapter, we also consider the uplink of an OFDM-based massive MIMO system as in section 2.1.1. Moreover, we assume each user employs N_{a} active subcarriers out of N number of subcarriers and reserves $N_{\mathrm{n}} = N - N_{\mathrm{a}}$ subcarriers as null subcarriers. I indicate the active and null subcarrier indexes of user p as $\mathcal{D}_p = \{\mathcal{D}_{p,0}, \mathcal{D}_{p,1}, \cdots, \mathcal{D}_{p,(N_{\mathrm{a}}-1)}\}$ and $\mathcal{V}_p = \{\mathcal{V}_{p,0}, \mathcal{V}_{p,1}, \cdots, \mathcal{V}_{p,(N_{\mathrm{n}}-1)}\}$, respectively, where $\mathcal{D}_p$ and $\mathcal{V}_p$ are subsets from the super-set $\{0, 1, \cdots, N-1\}$. Then, the data subcarrier assignment matrix $\mathbf{T}_p^{(\mathrm{d})}$ is an $N \times N_{\mathrm{a}}$ matrix whose i^{th} row is equal to the i^{th} row of the identity $I_{N_{\mathrm{a}}}$ if i is in $\mathcal{D}_p$ and otherwise is all zeros. Thus, the κ^{th} OFDM symbol for user p before CP insertion can be obtained as

$$\mathbf{x}_p^{\kappa} = \mathbf{F}_N^{\mathrm{H}} \mathbf{T}_p^{(\mathrm{d})} \mathbf{d}_p^{\kappa}, \tag{5.1}$$

where $\mathbf{d}_p^{\kappa}$ is an $N_{\mathrm{a}} \times 1$ vector containing the κ^{th} data symbol of user p, and its entries are taken from a zero mean i.i.d. data set with the variance of unity. To form the baseband transmit signal, a CP with the length of N_{CP} that is longer than the channel impulse response length, L, is added to the OFDM symbols. Considering perfect synchronization in both time and frequency, after CP removal, the OFDM symbol received at the m^{th} BS antenna can be written as

$$\mathbf{r}_m^{\kappa} = \sum_{p=0}^{P-1} \mathbf{X}_p^{\kappa} \mathbf{h}_{m,p} + \mathbf{n}_m^{\kappa}, \tag{5.2}$$

where $\mathbf{X}_p^{\kappa}$ is an $N \times L$ matrix including the first L columns of the circulant matrix $\mathrm{circ}(\mathbf{x}_p^{\kappa})$, $\mathbf{n}_m^{\kappa} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_N)$ is the complex additive white Gaussian noise (AWGN) with the variance of σ_n^2 at the m^{th} BS antenna. We consider the same noise variance for all the BS antennas. $\mathbf{h}_{m,p}$ is the $L \times 1$ CIR vector between user p and BS antenna m. I assume the channel taps to be a set of i.i.d. random variables that follow the complex normal distribution $\mathcal{CN}(0, \boldsymbol{\rho}_p)$, where $\boldsymbol{\rho}_p$ is an $L \times 1$ vector representing the p^{th} user's channel PDP.

At the receiver side, once all of the channel impulse responses are estimated, a linear combiner is used to detect transmitted data. To this end, I define the vector $\bar{\mathbf{r}}[k] = [\bar{r}_0[k], \bar{r}_1[k], \cdots, \bar{r}_{M-1}[k]]^{\mathrm{T}}$ where $\bar{r}_m[k]$ is the k^{th} sample of the received signals at the m^{th} BS antenna in the frequency domain for $k = 0, 1, \cdots, N-1$. Then, the

combiner output is given by

$$\hat{\mathbf{x}}^{\kappa}[k] = \mathbf{Z}[k]\bar{\mathbf{r}}^{\kappa}[k], \tag{5.3}$$

where $\bar{\mathbf{r}}^{\kappa}[k] = [\bar{r}_0^{\kappa}[k], \bar{r}_1^{\kappa}[k], \cdots, \bar{r}_{M-1}^{\kappa}[k]]^{\mathrm{T}}$ is an $M \times 1$ vector containing the k^{th} samples of the users' signals on the κ^{th} OFDM symbol, and $\mathbf{Z}[k]$ is a $P \times M$ linear combiner on subcarrier k. If we deploy the zero-forcing (ZF) combiner, $\mathbf{Z}[k]$ can be calculated as

$$\mathbf{Z}[k] = (\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k])^{-1}\boldsymbol{\mathcal{H}}^{\mathrm{H}}[k], \tag{5.4}$$

where $\boldsymbol{\mathcal{H}}[k] = [\bar{\mathbf{h}}_0[k], \bar{\mathbf{h}}_1[k], \cdots, \bar{\mathbf{h}}_{P-1}[k]]$ is an $M \times P$ matrix whose p^{th} column is $\bar{\mathbf{h}}_p[k] = [\bar{h}_{0,p}[k], \bar{h}_{1,p}[k], \cdots, \bar{h}_{M-1,p}[k]]^{\mathrm{T}}$, and $\bar{\mathbf{h}}_{m,p} = \mathbf{F}_N\mathbf{h}_{m,p}$.

Note that in the asymptotic regime, the $P \times P$ normalization matrix, $\boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k]$, becomes diagonal, i.e.,

$$\begin{aligned}\mathbf{D}_k &= \boldsymbol{\mathcal{H}}[k]^{\mathrm{H}}\boldsymbol{\mathcal{H}}[k] \\ &= \mathrm{diag}\left\{\left\|\bar{\mathbf{h}}_0[k]\right\|^2, \left\|\bar{\mathbf{h}}_1[k]\right\|^2, \cdots, \left\|\bar{\mathbf{h}}_{P-1}[k]\right\|^2\right\},\end{aligned} \tag{5.5}$$

and when M tends to infinity, according to the law of large numbers, $\mathbf{D}_k \to M\mathbf{I}_P$. Therefore, the combiner output for user q can be written as

$$\hat{x}_q^{\kappa}[k] = \frac{1}{M}\bar{\mathbf{h}}_q^{\mathrm{H}}[k]\bar{\mathbf{r}}^{\kappa}[k], \tag{5.6}$$

for $k = 0, 1, \cdots, N-1$. Note that the first N_{v} OFDM symbols of each user include N_{n} number of null subcarriers. Thus, by removing null subcarriers, the estimated data symbols are given as

$$\hat{\mathbf{d}}_q^{\kappa} = (\mathbf{T}_q^{(\mathrm{d})})^{\mathrm{H}}\hat{\mathbf{x}}_q^{\kappa}, \tag{5.7}$$

for $0 \le \kappa < N_{\mathrm{v}}$. For the remaining of symbols, i.e. $N_{\mathrm{v}} \le \kappa$, the transmitted data is equal to the combiner output, $\hat{\mathbf{d}}_q^{\kappa} = \hat{\mathbf{x}}_q^{\kappa}$. In the following, we present the three steps of our proposed blind channel estimation technique.

5.2 Multiuser interference cancellation

Stacking the received signals at the BS antennas in an $M \times N$ matrix as

$$R^{\kappa} = [\mathbf{r}_0^{\kappa}, \mathbf{r}_1^{\kappa}, \cdots, \mathbf{r}_{M-1}^{\kappa}]^{\mathrm{T}}, \tag{5.8}$$

we can rewrite equation (5.2) as

$$R^{\kappa} = \sum_{p=0}^{P-1} \mathbf{H}_p (\mathbf{X}_p^{\kappa})^{\mathrm{T}} + \mathbf{N}^{\kappa}, \tag{5.9}$$

where $\mathbf{H}_p = [\mathbf{h}_{0,p}, \mathbf{h}_{1,p}, \cdots, \mathbf{h}_{M-1,p}]^{\mathrm{T}}$, and $\mathbf{N}^{\kappa} = [\mathbf{n}_0^{\kappa}, \mathbf{n}_1^{\kappa}, \cdots, \mathbf{n}_{M-1}^{\kappa}]^{\mathrm{T}}$. Let us take the matrix R^{κ} to the frequency domain, and then, extract the null subcarriers of the user q as

$$\mathbf{\Upsilon}_q^{\kappa} = R^{\kappa} \mathbf{F}_N \mathbf{T}_q^{(\mathrm{n})}, \tag{5.10}$$

where $\mathbf{T}_q^{(\mathrm{n})}$ is an $N \times N_{\mathrm{n}}$ null subcarrier assignment matrix of user q whose i^{th} row is equal to the i^{th} row of the identity matrix $I_{N_{\mathrm{n}}}$ if i is in $\mathcal{V}_p$ and otherwise is all zeros. I assume that there is no overlap between the null subcarriers of different users, i.e. $\mathcal{V}_p \bigcap \mathcal{V}_q = \emptyset$ for $p \neq q$. In other words, at the position of null subcarriers of one user, other users are transmitting their data symbols. Hence, $\mathbf{\Upsilon}_q^{\kappa}$ does not contain null subcarriers of any user except user q.

Based on the subspace theory, the signal and noise subspaces can be obtained from the eigenvalue decomposition of the covariance matrix of $\mathbf{\Upsilon}_q^{\kappa}$ in equation (5.10), which can be obtained as

$$\mathbf{R}_q^{(\gamma)} = \mathbb{E}\left[\mathbf{\Upsilon}_q^{\kappa} (\mathbf{\Upsilon}_q^{\kappa})^{\mathrm{H}}\right], \tag{5.11}$$

The expectation is taken with respect to the data symbols and noise and can be approximated by

$$\mathbf{R}_q^{(\gamma)} \approx \sum_{\kappa=0}^{N_{\mathrm{v}}-1} \mathbf{\Upsilon}_q^{\kappa} (\mathbf{\Upsilon}_q^{\kappa})^{\mathrm{H}}, \tag{5.12}$$

where N_{v} is the number of OFDM symbols with N_{n} number of null subcarriers. $\mathbf{R}_q^{(\gamma)}$ can also be represented as

$$\mathbf{R}_q^{(\gamma)} = \sum_{p=0}^{P-1} \mathbf{H}_p \mathbf{R}_{p,q}^{(x)} \mathbf{H}_p^{\mathrm{H}} + \widetilde{\mathbf{N}}^{\kappa}, \tag{5.13}$$

where $\widetilde{\mathbf{N}}^{\kappa}$ is the noise effect, and $\mathbf{R}_{p,q}^{(x)}$ is

$$\begin{aligned} \mathbf{R}_{p,q}^{(x)} =& \mathbb{E}\left\{(\mathbf{X}_p^{\kappa})^{\mathrm{T}} \mathbf{F}_N \mathbf{T}_q^{(\mathrm{n})} (\mathbf{T}_q^{(\mathrm{n})})^{\mathrm{H}} \mathbf{F}_N^{\mathrm{H}} (\mathbf{X}_p^{\kappa})^{*}\right\} \\ =& \mathbb{E}\Big\{\mathbf{F}_L \mathrm{diag}\big(\mathbf{T}_p^{(\mathrm{d})} \mathbf{d}_p^{\kappa}\big) \mathbf{T}_q^{(\mathrm{n})} \\ & \times (\mathbf{T}_q^{(\mathrm{n})})^{\mathrm{H}} \mathrm{diag}\big((\mathbf{d}_p^{\kappa})^{*} \mathbf{T}_q^{(\mathrm{n})}\big) \mathbf{F}_L^{\mathrm{H}}\Big\}. \end{aligned} \tag{5.14}$$

where $\mathbf{F}_L$ is formed by the first L columns of $\mathbf{F}_N$. Since $\mathbf{R}_{p,q}^{(x)}$ is a $L \times L$ matrix, its rank is equal or smaller than L. In addition, if $N_{\mathrm{n}} \geq L$, and consequently,

$|\mathcal{D}_p \bigcap \mathcal{V}_q| \geq L$ for $p \neq q$, $\mathbf{R}_{p,q}^{(x)}$ has L non-zero column vectors and the rank of L. Note that for $p = q$, we have $\mathcal{D}_q \bigcap \mathcal{V}_q = \emptyset$, and $\mathbf{R}_{q,q}^{(x)} = 0$. Therefore, the rank of the matrix $\mathbf{R}_q^{(\gamma)}$ in (5.13) is $(P-1)L$.

Then, performing eigenvalue decomposition of $\mathbf{R}_q^{(\gamma)}$, we have

$$\mathbf{R}_q^{(\gamma)} = \left[\mathbf{U}_q^{(\mathrm{s})}, \mathbf{U}_q^{(\mathrm{n})}\right] \mathbf{S}_q \left[\mathbf{U}_q^{(\mathrm{s})}, \mathbf{U}_q^{(\mathrm{n})}\right]^{\mathrm{H}}, \tag{5.15}$$

where $\mathbf{S}_q$ is a diagonal matrix with the eigenvalues of matrix $\mathbf{R}_q^{(\gamma)}$ on its diagonal elements, and $M \times (P-1)L$ matrix $\mathbf{U}_q^{(\mathrm{s})}$ and $M \times M_{\mathrm{e}}$ matrix $\mathbf{U}_q^{(\mathrm{n})}$ correspond to signal and noise space, respectively, where $M_{\mathrm{e}} = M - (P-1)L$. Since the eigenvectors are orthogonal to each other, $\mathbf{U}_q^{(\mathrm{n})}$ is orthogonal to $\mathbf{U}_q^{(\mathrm{s})}$ and consequently, to the signal subspace of all the users other than user q. Hence, if we multiply $\mathbf{U}_q^{(\mathrm{n})}$ to the received signal, we have

$$\begin{aligned} \breve{R}_q^{\kappa} =& \mathbf{U}_q^{(\mathrm{n})} R^{\kappa} \\ =& \mathbf{U}_q^{(\mathrm{n})} \mathbf{H}_q (\mathbf{X}_q^{\kappa})^{\mathrm{T}} + \sum_{\substack{p=0 \\ p \neq q}}^{P-1} \mathbf{U}_q^{(\mathrm{n})} \mathbf{H}_p (\mathbf{X}_p^{\kappa})^{\mathrm{T}} + \mathbf{U}_q^{(\mathrm{n})} \mathbf{N}^{\kappa}. \end{aligned} \tag{5.16}$$

Note that $\mathbf{U}_q^{(\mathrm{n})}$ is orthogonal to the signal subspace of all the users $p \neq q$ and the second term on the right hand side of equation (5.16), which corresponds to MUI, is zero. Thus,

$$\breve{R}_q^{\kappa} = \breve{\mathbf{H}}_q (\mathbf{X}_q^{\kappa})^{\mathrm{T}} + \breve{\mathbf{N}}^{\kappa}, \tag{5.17}$$

where $\breve{\mathbf{H}}_q = (\mathbf{U}_q^{(\mathrm{n})})^{\mathrm{H}} \mathbf{H}_q$, is the $M_{\mathrm{e}} \times L$ equivalent channel matrix of user q, and $\breve{\mathbf{N}}^{\kappa} = \mathbf{U}_q^{(\mathrm{n})} \mathbf{N}^{\kappa}$ is the $M_{\mathrm{e}} \times N$ equivalent noise matrix. Once the multiuser interference is canceled, the data detection can be performed individually for each user.

5.3 Virtual pilot detection

I consider the first transmitted data symbol of each user as a virtual pilot and blindly detect it to be used in the channel estimation stage. To keep the formulation simple and without loss of generality, let us drop the symbol index κ in our derivations. Then, the m^{th} row of the received signal for user q in equation (5.17) can be written as

$$\breve{\mathbf{r}}_{m,q} = \mathbf{X}_q \breve{\mathbf{h}}_{m,q} + \breve{\mathbf{n}}_m, \tag{5.18}$$

where $\breve{\mathbf{h}}_{m,q}$ is an $L \times 1$ vector equal to the m^{th} row of the equivalent channel matrix $\breve{\mathbf{H}}_q$, and $N \times 1$ vector $\breve{\mathbf{n}}_m$ consists of the m^{th} row of equivalent noise matrix $\breve{\mathbf{N}}$.

After DFT operation, the covariance matrix of the received signal in the frequency domain can be calculated as

$$\begin{aligned}\mathbf{R}_q^{(\mathrm{r})} &= \frac{1}{M_\mathrm{e}} \sum_{m=0}^{M_\mathrm{e}-1} \mathbf{F}_N \breve{\mathbf{r}}_{m,q} (\mathbf{F}_N \breve{\mathbf{r}}_{m,q})^\mathrm{H} \\ &= \mathbf{F}_N \mathbf{X}_q \mathbf{C}_{\breve{h},\breve{h}} \mathbf{X}_q^\mathrm{H} \mathbf{F}_N^\mathrm{H} + \mathbf{F}_N \mathbf{V}_q \mathbf{F}_N^\mathrm{H} + \mathbf{F}_N \mathbf{C}_{\breve{n},\breve{n}} \mathbf{F}_N^\mathrm{H}, \end{aligned} \tag{5.19}$$

with

$$\mathbf{C}_{\breve{h},\breve{h}} = \frac{1}{M_\mathrm{e}} \sum_{m=0}^{M_\mathrm{e}-1} \breve{\mathbf{h}}_{m,q} \breve{\mathbf{h}}_{m,q}^\mathrm{H}, \tag{5.20}$$

and

$$\mathbf{V}_q = \mathbf{X}_q \mathbf{C}_{\breve{h},\breve{n}} + (\mathbf{X}_q \mathbf{C}_{\breve{h},\breve{n}})^\mathrm{H}, \tag{5.21}$$

is the noise effect where

$$\mathbf{C}_{\breve{h},\breve{n}} = \frac{1}{M} \sum_{m=0}^{M-1} \breve{\mathbf{h}}_{m,p} \breve{\mathbf{n}}_m^\mathrm{H}, \tag{5.22}$$

and

$$\mathbf{C}_{\breve{n},\breve{n}} = \frac{1}{M} \sum_{m=0}^{M-1} \breve{\mathbf{n}}_m \breve{\mathbf{n}}_m^\mathrm{H}. \tag{5.23}$$

We can express $\breve{\mathbf{h}}_{m,q}$ as

$$\begin{aligned}\breve{\mathbf{h}}_{m,q} &= (\mathbf{u}_{m,q}^{(\mathrm{n})\,\mathrm{H}} \mathbf{H}_q)^\mathrm{T} \\ &= \mathbf{H}_q^\mathrm{T} (\mathbf{u}_{m,q}^{(\mathrm{n})})^*, \end{aligned} \tag{5.24}$$

where $\mathbf{u}_{m,q}^{(\mathrm{n})}$ is the m^{th} column of matrix $\mathbf{U}_q^{(\mathrm{n})}$. Then, if we substitute the value of $\breve{\mathbf{h}}_{m,q}$ in equation (5.20), we can expand $\mathbf{C}_{\breve{h},\breve{h}}$ as

$$\begin{aligned}
\mathbf{C}_{\breve{h},\breve{h}} &= \frac{1}{M_e}\sum_{m=0}^{M_e-1}(\mathbf{H}_q^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)(\mathbf{H}_q^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)^{\mathrm{H}} \\
&= \frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}\mathbf{H}_q^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*(\mathbf{u}_{m,q}^{(\mathrm{n})})^{\mathrm{T}}\mathbf{H}_q^* \\
&= \mathbf{H}_q^{\mathrm{T}}\left\{\frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}(\mathbf{u}_{m,q}^{(\mathrm{n})}\mathbf{u}_{m,q}^{(\mathrm{n})\,\mathrm{H}})^*\right\}\mathbf{H}_q^* \\
&= (\mathbf{H}_q^{\mathrm{H}}\mathbf{I}_M\mathbf{H}_q)^* \\
&= \mathbf{C}_{h,h}^*. \qquad (5.25)
\end{aligned}$$

According to the law of large numbers, as M grows large, $C_{h,h}[i,j] \to \mathbb{E}\left\{h_{m,q}[i]h_{m,q}^*[j]\right\}$. Since $h_{m,q}[i]$ is a set of i.i.d. random variables, $\mathbb{E}\left\{h_{m,q}[i]h_{m,q}^*[j]\right\} = 0$ for $i \neq j$, and $\mathbb{E}\left\{h_{m,q}[i]h_{m,q}^*[i]\right\} = \rho_q[i]$. Thus, $\mathbf{C}_{h,h}$ is a real-valued diagonal matrix with the main diagonal equal to $\boldsymbol{\rho}_q$. Following a similar line of derivations as above, we have

$$\begin{aligned}
\mathbf{C}_{\breve{h},\breve{n}} &= \frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}(\mathbf{H}_q^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)(\mathbf{N}^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)^{\mathrm{H}} \\
&= \frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}\mathbf{H}_q^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*(\mathbf{u}_{m,q}^{(\mathrm{n})})^{\mathrm{T}}\mathbf{N}^* \\
&= \mathbf{H}_q^{\mathrm{T}}\left\{\frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}(\mathbf{u}_{m,q}^{(\mathrm{n})}\mathbf{u}_{m,q}^{(\mathrm{n})\,\mathrm{H}})^*\right\}\mathbf{N}^* \\
&= (\mathbf{H}_q^{\mathrm{H}}\mathbf{I}_M\mathbf{N})^* \\
&= \mathbf{C}_{h,n}^*. \qquad (5.26)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{C}_{\breve{n},\breve{n}} &= \frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}(\mathbf{N}^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)(\mathbf{N}^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*)^{\mathrm{H}} \\
&= \frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}\mathbf{N}^{\mathrm{T}}(\mathbf{u}_{m,q}^{(\mathrm{n})})^*(\mathbf{u}_{m,q}^{(\mathrm{n})})^{\mathrm{T}}\mathbf{N}^* \\
&= \mathbf{N}^{\mathrm{T}}\left\{\frac{1}{M_{\mathrm{e}}}\sum_{m=0}^{M_e-1}(\mathbf{u}_{m,q}^{(n)}\mathbf{u}_{m,q}^{(\mathrm{n})\,\mathrm{H}})^*\right\}\mathbf{N}^* \\
&= (\mathbf{N}^{\mathrm{H}}\mathbf{I}_M\mathbf{N})^* \\
&= \mathbf{C}_{n,n}^*. \qquad (5.27)
\end{aligned}$$

Note that $n_m[i]$ and $h_{m,q}[i]$ are also independent and both are identically distributed. Thus, as M grows large, $C_{h,n}[i,j] \to \mathbb{E}\{h_{m,q}[i]n_m^*[j]\} = 0$, and $C_{n,n}[i,j] \to \mathbb{E}\{n_m[i]n_m^*[j]\}$ which is zero for $i \neq j$ and σ_n^2 for $i = j$. Note that the last term on the right hand side of equation (5.19) can be removed by subtracting $\sigma_n^2\mathbf{I}_N$.

Moreover, considering $\mathbf{F}_N^{\mathrm{H}}\mathbf{F}_N = \mathbf{I}_N$, equation (5.19) can be rewritten as

$$\mathbf{R}_q^{(\mathrm{r})} = (\mathbf{F}_N\mathbf{X}_q\mathbf{F}_N^{\mathrm{H}})(\mathbf{F}_N\mathbf{C}_{h,h}\mathbf{F}_N^{\mathrm{H}})(\mathbf{F}_N\mathbf{X}_q^{\mathrm{H}}\mathbf{F}_N^{\mathrm{H}}) + \mathbf{F}_N\mathbf{V}_q\mathbf{F}_N^{\mathrm{H}} \tag{5.28}$$

It is worth noting that since $\mathbf{X}_q$ is a circulant matrix, $\mathbf{F}_N\mathbf{X}_q\mathbf{F}_N^{\mathrm{H}}$ is a diagonal matrix with the main diagonal equal to $\mathbf{T}_q^{(\mathrm{d})}\mathbf{d}_q$. Hence, we have

$$\begin{aligned}\mathbf{R}_q^{(\mathrm{r})} &= \mathrm{diag}(\mathbf{T}_{\mathrm{q}}^{(\mathrm{d})}\mathbf{d}_{\mathrm{q}})\widetilde{\mathbf{C}}_{\mathrm{h,h}}(\mathrm{diag}(\mathbf{d}_{\mathrm{q}}^{\mathrm{H}}(\mathbf{T}_{\mathrm{q}}^{(\mathrm{d})})^{\mathrm{H}}) + \mathbf{F}_{\mathrm{N}}\mathbf{V}_{\mathrm{q}}\mathbf{F}_{\mathrm{N}}^{\mathrm{H}} \\ &= \widetilde{\mathbf{C}}_{h,h} \odot \left\{\mathbf{T}_q^{(\mathrm{d})}\mathbf{d}_q\mathbf{d}_q^{\mathrm{H}}(\mathbf{T}_q^{(\mathrm{d})})^{\mathrm{H}}\right\} + \mathbf{F}_N\mathbf{V}_q\mathbf{F}_N^{\mathrm{H}},\end{aligned} \tag{5.29}$$

where $\widetilde{\mathbf{C}}_{h,h} = \mathbf{F}_N\mathbf{C}_{h,h}\mathbf{F}_N^{\mathrm{H}} = \mathrm{circ}(\bar{\boldsymbol{\rho}}_{\mathrm{q}})$ is a circulant matrix as $\mathbf{C}_{h,h}$ is diagonal, and $\bar{\boldsymbol{\rho}}_q$ is the Fourier transform of $\boldsymbol{\rho}_q$. Furthermore, if we element-wise divide the matrix $\mathbf{R}_q^{(\mathrm{r})}$ by the elements of the matrix $\widetilde{\mathbf{C}}_{h,h}$, and then, multiply to $\mathbf{T}_q^{(\mathrm{d})}$, we have

$$\begin{aligned}\mathbf{G}_q &= (\mathbf{T}_q^{(\mathrm{d})})^{\mathrm{H}}\left\{\widetilde{\mathbf{C}}_{h,h} \oslash \mathbf{R}_q^{(\mathrm{r})}\right\}\mathbf{T}_q^{(\mathrm{d})} \\ &= \mathbf{d}_q\mathbf{d}_q^{\mathrm{H}} + \widetilde{\mathbf{V}}_q,\end{aligned} \tag{5.30}$$

where

$$\widetilde{\mathbf{V}}_q = (\mathbf{T}_q^{(\mathrm{d})})^{\mathrm{H}}\left\{\widetilde{\mathbf{C}}_{h,h} \oslash (\mathbf{F}_N\mathbf{V}_q\mathbf{F}_N^{\mathrm{H}})\right\}\mathbf{T}_q^{(\mathrm{d})}. \tag{5.31}$$

Let us assume that the data on the first subcarrier of the first OFDM symbol of the users, i.e. $d_p[0]$ for $p = 0, 1, \cdots, P-1$, is known as a reference data. Then, we can form a matrix $\mathbf{Y}_q$ as

$$\begin{aligned}Y_q[i,j] &= d_q[0] \times \frac{G_q[i,j]}{G_q[0,j]} \\ &= d_q[0] \times \frac{d_q[i]d_q[j] + \widetilde{V}_q[i,j]}{d_q[0]d_q[j] + \widetilde{V}_q[0,j]} \\ &= W[j]d_q[i] + \widehat{V}_q[i,j],\end{aligned} \tag{5.32}$$

where

$$W_q[j] = \frac{d_q[0]d_q[j]}{d_q[0]d_q[j] + \widetilde{V}_q[0,j]}, \tag{5.33}$$

and

$$\widehat{V}_q[i,j] = \frac{d_q[0]\widetilde{V}_q[i,j]}{d_q[0]d_q[j] + \widetilde{V}_q[0,j]}. \tag{5.34}$$

One can see from equation (5.32) that all the N_a elements in the i^{th} row of the matrix $\mathbf{Y}_q$ contain the information of $d_q[i]$. Moreover, in case the number of BS antennas, M, is large and noise effect is averaged out, $V_q[i,j]$ in equation (5.21) is zero, and we have $W_q[j] \approx 1$ and $\widehat{V}_q[i,j] \approx 0$. Therefore, the data symbols $d_q[i]$ can be estimated as

$$\hat{d}_q[i] = \frac{1}{N_\mathrm{a}} \sum_{j=0}^{N_\mathrm{a}-1} Y_q[i,j], \tag{5.35}$$

for $i = 0, 1, \cdots, N_\mathrm{a} - 1$ without any phase ambiguity. However, in practice, due to the practical limitations on the number of BS antennas, the noise effects will not completely fade away and will therefore distort the received symbols.

5.4 DL-aided virtual pilot detection

To address this issue, we propose applying CNN-based DL algorithms to mitigate the noise effect and extract the information in virtual pilots with higher accuracy. The DL algorithm can learn the pattern of both the additive and the multiplicative noises and eliminate them. Moreover, with joint denoising and demodulation, the DNN can be trained to learn the pattern of the noises based on the symbol that is transmitted on the desired subcarrier and also other subcarriers of the OFDM symbol. In addition, unlike other conventional denoising techniques which are iterative and time consuming, a DNN is trained offline, and during online data transmission, it only consists of some simple operations like addition and multiplication. Therefore, after training when the parameters/weights of the network are tuned and saved in the memory, its implementation on FPGA is quite straightforward.

In the following, first, we propose applying DnCNN algorithm to our virtual pilot detection technique. Second, we propose another DL-based algorithm named U-net which can provide a better denoising performance compared to DnCNN. Third, we propose applying ResNet architecture followed by a fully connected feed forward neural network to jointly denoise and detect the transmitted symbols.

5.4.1 Denoising with DnCNN

DnCNN is based on a CNN which contains three types of layers. The first convolutional layer uses 64 different $3 \times 3 \times 1$ filters and is followed by a rectified linear unit (ReLU). Second, there are 18 convolutional layers with 64 filters of size $3 \times 3 \times 64$, batch-normalization, [58], and a ReLU. Final layer utilizes one $3 \times 3 \times 64$ filter to reconstruct the output. DnCNN is capable to learn either the original mapping or the residual mapping. In original mapping, DnCNN directly detects the data from a noisy observation. While in residual mapping, first, it detects the additive noise, and then, subtracts it from the input to recover the original data. This method is known as residual learning, [60].

Here, we deploy DnCNN for the residual mapping. I separate the real and the imaginary parts of the matrix $\mathbf{Y}_q$ into two $N_\mathrm{a} \times N_\mathrm{a}$ matrices and give them to DnCNN as two different inputs. Then, DnCNN learns the additive noise for each part, and we can obtain the complex noise from the outputs of DnCNN corresponding to the real and imaginary parts as

$$\widehat{\mathbf{V}}_q^{(\mathrm{DnCNN})} = \widehat{\mathbf{V}}_{q,\mathrm{real}}^{(\mathrm{DnCNN})} + \widehat{\mathbf{V}}_{q,\mathrm{imag}}^{(\mathrm{DnCNN})}, \tag{5.36}$$

that should be subtracted from $\mathbf{Y}_q$. Therefore, we have the denoised version of the matrix $\mathbf{Y}_q$ as

$$Y_q^{(\mathrm{DnCNN})}[i,j] = Y_q[i,j] - \widehat{V}_q^{(\mathrm{DnCNN})}[0,j], \tag{5.37}$$

and the data symbols $d_q[i]$ can be estimated as

$$\hat{d}_q^{(\mathrm{DnCNN})}[i] = \frac{1}{N_\mathrm{a}} \sum_{j=0}^{N_\mathrm{a}-1} Y_q^{(\mathrm{DnCNN})}[i,j], \tag{5.38}$$

for $i = 0, 1, \cdots, N_\mathrm{a} - 1$. It is worth noticing that since the massive MIMO systems are supposed to be able to work at low SNRs, the level of noise in our case is higher than the one in image processing problems. Therefore, we propose another DL-based algorithm as a denoiser that can outperform DnCNN at low SNRs.

5.4.2 Denoising with U-Net

As it is explained in section 2.4.1.3, U-Net consists of several convolutional layers as DnCNN, but it has several unitary shortcuts which take into account the features from multiple layers, and consequently, help remove the noise more effectively. Moreover, the size of tensors does not remain constant as in DnCNN. In contracting path of the U-Net architecture, the weight and height of the tensors are halved and

the size of the feature channels is doubled every other convolutional layer. Then, in the expanding path, this large number of feature channels allows the network to better propagate the context information to higher resolution layers. Therefore, even in the high noise level scenarios, this ability of U-Net architecture helps us better extract the features of the received matrix $\mathbf{Y}_q$.

Similar to the case of DnCNN, we separate the real and the imaginary parts of the matrix $\mathbf{Y}_q$ into two $N_\mathrm{a} \times N_\mathrm{a}$ matrices and give them to U-Net as two different inputs. The transmitted data symbols can be estimated from the output of the U-Net denoiser, $\mathbf{Y}_q^{(\mathrm{U-Net})}$, as

$$\hat{d}_q^{(\mathrm{U-Net})}[i] = \frac{1}{N_\mathrm{a}} \sum_{j=0}^{N_\mathrm{a}-1} Y_q^{(\mathrm{U-Net})}[i,j], \tag{5.39}$$

for $i = 0, 1, \cdots, N_\mathrm{a} - 1$. Both DnCNN and U-Net denoisers can reconstruct the patterns in the matrix $\mathbf{Y}_q$. However, they may not converge to the exact expected value on each row, which leads to an inaccurate data detection. To address this problem, we propose an alternative technique which is presented in the following.

5.4.3 ResNet detector

In order to force the network to converge to the right point while reconstructing the original pattern in the matrix, we add a fully connected feedforward neural network at the end of the denoiser. In fact, the DL algorithm jointly denoises and demodulates the received signals. To this end, the received data on each subcarrier should be detected separately. Because if we consider all the subcarriers, which means the whole matrix $\mathbf{Y}_q$, the network will be too big to be trained. For instance, considering 16-QAM modulation, the number of the number of classes, and consequently, the number of neurons on the output layer will be N^4. Therefore, we take each row of the matrix $\mathbf{Y}_q$ as the input of the neural network.

I utilize the ResNet architecture introduced in section 2.4.1.2 instead of a simple convolutional neural network to take advantage of the unitary shortcuts for denoising. In addition, we consider a contracting path such as in the U-Net architecture to benefit from the large number of feature channels. Note that there is no need for the successive expanding path to increase the resolution since the classification network can extract the transmitted data symbol from the resulting tensors. In addition, we propose that instead of separating the real and the imaginary parts, to consider the imaginary part as the information of the second channel of the input image to the ResNet network. It means that considering the three channels of red, green and blue of each image (RGB), the real part of the input matrix is on the red channel,

the imaginary part is allocated on the green channel, and the blue channel remains zero. Therefore, the number of classes is 16 for a system with 16-QAM modulation. In this technique, the output of the network indicates the estimated data symbols $\hat{\mathbf{d}}_q^{(\text{ResNet})}$, and can be used as virtual pilots to estimate the channels.

5.5 Channel Estimation

After detecting the users' first symbols, we can treat them as virtual pilots and employ a simple maximum likelihood (ML) channel estimation to estimate the channel impulse responses. To this end, we rearrange equation (5.2) as

$$\mathbf{r}_m = \mathbf{X}\mathbf{h}_m + \mathbf{n}_m, \tag{5.40}$$

where $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_{P-1}]$, and $\mathbf{h}_m = [\mathbf{h}_{m,0}^{\mathrm{T}}, \mathbf{h}_{m,1}^{\mathrm{T}}, \cdots, \mathbf{h}_{m,P-1}^{\mathrm{T}}]^{\mathrm{T}}$. Then, the logarithm of the conditional probability density function (PDF) of $\mathbf{r}_m$ given $\mathbf{h}_{m,p}$ can be written as

$$\begin{aligned} \ln p(\mathbf{r}_m \,|\mathbf{h}_{m,p}) = \Theta_0 - \Theta_1 \left[\mathbf{r}_m - \mathbf{X}\mathbf{h}_{m,p}\right]^{\mathrm{H}} \\ \times \left[\mathbf{r}_m - \mathbf{X}\mathbf{h}_{m,p}\right], \end{aligned} \tag{5.41}$$

where $\Theta_0 = -N\ln(2\pi\sigma_n^2)/2$, and $\Theta_1 = 1/2\sigma_n^2$. By taking the derivative of equation (5.41) with respect to $\mathbf{h}_{m,p}$ and setting it equal to zero, the estimated channel can be obtained as

$$\hat{\mathbf{h}}_m = (\mathbf{X}^{\mathrm{H}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{H}}\mathbf{r}_m. \tag{5.42}$$

5.6 Simulation results

In this section, we confirm our theoretical developments in the previous sections through numerical simulations. We consider the uplink transmission of a multi-user OFDM-based massive MIMO system with perfect power control for all the users. I assume that all the users simultaneously communicate to a BS with $M = 100$ antennas over $N = 256$ subcarriers using 16 QAM. In our simulations, we use the 3GPP Long Term Evolution (LTE) extended typical urban (ETU) channel model, [127], and a CP length of $\lfloor N/14 \rfloor$. We set the coherence time to 1ms, which is equal to the length of two sub-frames including 14 OFDM symbols. For MUI cancellation, we consider the distributed null subcarrier assignment scheme, where the null subcarrier index set of user q is $\mathcal{V}_q = \left\{q, q + \frac{N}{N_{\mathrm{n}}}, \cdots, q + \frac{N(N_{\mathrm{n}}-1)}{N_{\mathrm{n}}}\right\}$. Since it is desirable to keep the number of null subcarriers as low as possible to save

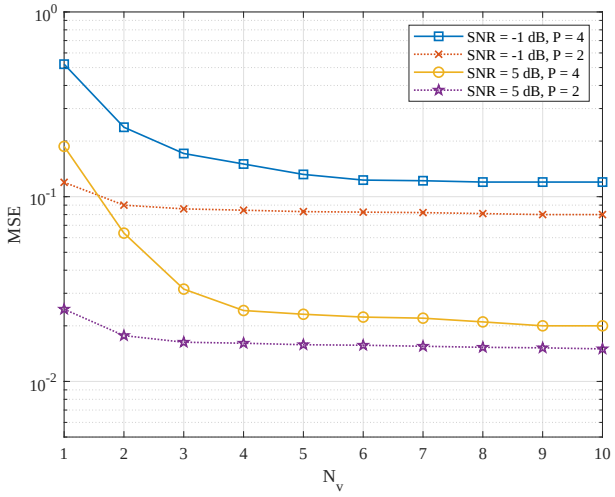

Figure 5.1 MSE with respect to the number of symbols with null subcarriers, N_{v}, for $M = 100$ BS antennas, $N = 256$ subcarriers, and 16-QAM modulation.

more subcarriers for data transmission, we set $N_{\mathrm{n}} = L = 8$. We have conducted a thorough analysis to select the value of the number of symbols with null subcarriers, N_{v}, by studying its impact on MSE when varying the SNR and the number of users, as shown in Fig. 5.1. The same trend on MSE can be observed regardless of the SNR and number of users, i.e. the MSE decreases with N_{v} until a fixed level is reached. The minimum value of N_{v} for which a minimum MSE is reached is 4, and this is the value used in all the simulations presented in this section.

For training the proposed deep neural networks, the training and validation sets contain 80000 and 16000 samples for DnCNN and U-net denoiser, and 179200 and 89600 samples for ResNet detector. The input data to these networks, at both offline and online setups, is scaled to the range of $[0, 1]$. In offline training, the network parameters are tuned according to the training set to minimize the cost function. The validation set is a data set that the network does not see during the tuning. Thus, in the offline training, the validation set can evaluate the performance of the network for an unseen data. If a network does not have overfitting or underfitting problem, its performance for training and validation sets is almost the same. Adam optimizer, [50], is utilized to train the networks. The training rate is set to be 10^{-3} initially, and then, it is dropped 10 times after any 10 epochs. As it is observed, the validation error stops improving after 40 epochs. After offline training, the trained networks is used in an online setup, where new data set is generated and transmitted through new channels and processed in the receiver before reaching the neural network.

As a benchmark, we consider the data-aided channel estimation technique in [11], which is introduced in section 2.3.1. As in LTE standard, the subcarrier spacing is

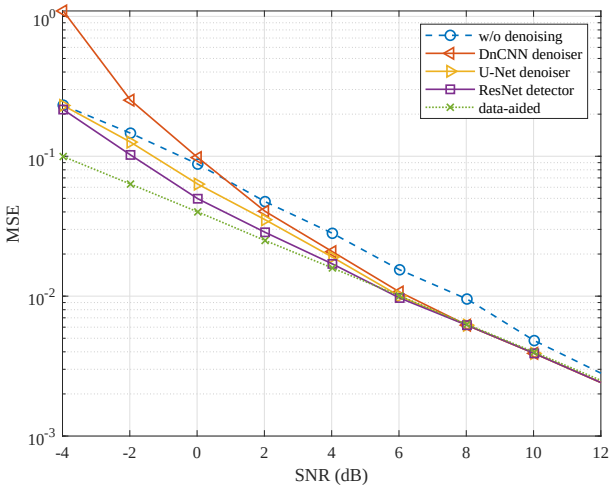

Figure 5.2 MSE with respect to SNR for $M = 100$ BS antennas, $N = 256$ subcarriers, $P = 4$ users, and 16-QAM modulation.

$\Delta f = 15$ kHz. Since 14 OFDM symbols are transmitted in each 1ms, one OFDM symbol interval is $T_{\mathrm{s}} = 1000/14 = 71.4$ microseconds, the useful symbol duration is $T_{\mathrm{u}} = 1/\Delta f = 66.7$ microseconds, and consequently, the guard interval is $T_{\mathrm{g}} = T_{\mathrm{s}} - T_{\mathrm{u}} = 4.76$ microseconds. Therefore, the frequency smoothness interval in equation (2.18) is $N_{\mathrm{smooth}} = 14$ subcarriers. As mentioned before, each coherent block must host the number of pilots greater or equal to the number of users, thus, we consider $\tau_{\mathrm{p}} = 4$.

Fig. 5.2 displays the MSE performance of our proposed blind channel estimation techniques with respect to SNR for $P = 4$ users. As the SNR increases, the MSE of the proposed technique without denoiser coincides with the benchmark. By deploying the DnCNN algorithm, the MSE curve approaches the benchmark at lower SNRs. However, its MSE can be higher than the MSE of the one without denoising at very low SNRs. Instead applying U-net denoiser can improve the MSE curve at all the SNRs. Moreover, As it is depicted, the utilization of ResNet detector gives an MSE very close to the benchmark. In Fig. 5.3, we study the performance of the proposed techniques for a system with $P = 2$ users. As it is shown, the MSE curve of the ResNet detector almost matches the benchmark. This means if there is less amount of MUI, the proposed technique can provide a performance similar to that of the system with data-aided channel estimation.

I also analyze the bit error rate (BER) performance of the system with the proposed techniques in Fig. 5.4. As it can also be seen in this figure, DnCNN can improve the performance of the system, and U-net denoiser can outperform the DnCNN, especially at lower values of $\mathrm{E_b/N_0}$. Moreover, it is shown that the blind

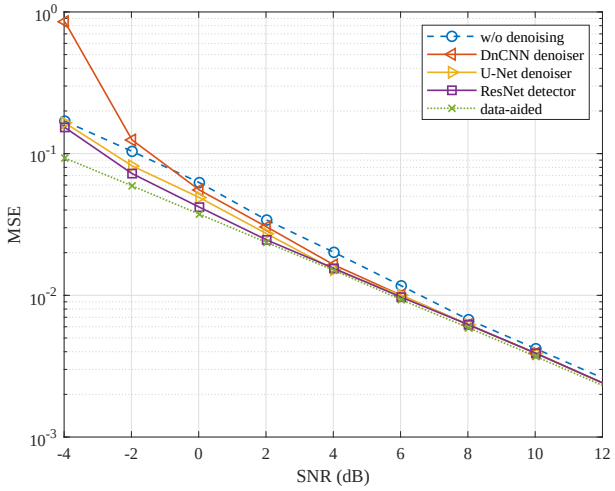

Figure 5.3 MSE with respect to SNR for $M = 100$ BS antennas, $N = 256$ subcarriers, $P = 2$ users, and 16-QAM modulation.

estimation technique with ResNet detector approaches that of the system with the data-aided channel estimation.

Fig. 5.5 presents the MSE of our proposed techniques with respect to the number of BS antennas for different values of SNR. It demonstrates that by increasing the number of BS antennas, the MSE of our proposed technique without denoiser is improved and approaches the benchmark. Moreover, by deploying the proposed denoiser techniques, an MSE similar to the data-aided channel estimation technique can be achieved with fewer BS antennas. As one can see, $M = 100$ BS antennas are sufficient for the system with ResNet detector to achieve an MSE close to that of the system with data-aided channel estimation.

Finally, Fig. 5.6 presents the throughput of the system with our proposed channel estimation techniques. In our simulations, we calculated the number of successfully detected bits of all the users' signals, and divide it by the duration of the transmission, i.e. 1ms, and the communication bandwidth of 2.5 MHz. It is demonstrated that our proposed technique, offers a higher throughput than the data-aided channel estimation. In addition, when it is assisted by the proposed denoiser algorithms, can reach to the maximum throughput at lower values of $\mathrm{E_b/N_0}$.

5.7 Conclusion

In this chapter, we proposed a deep learning-aided blind channel estimation technique for multi-user OFDM-based massive MIMO systems. Our proposed technique consists of three steps: MUI cancellation, virtual pilot detection, and channel estimation. For MUI cancellation, some null subcarriers are inserted in each user

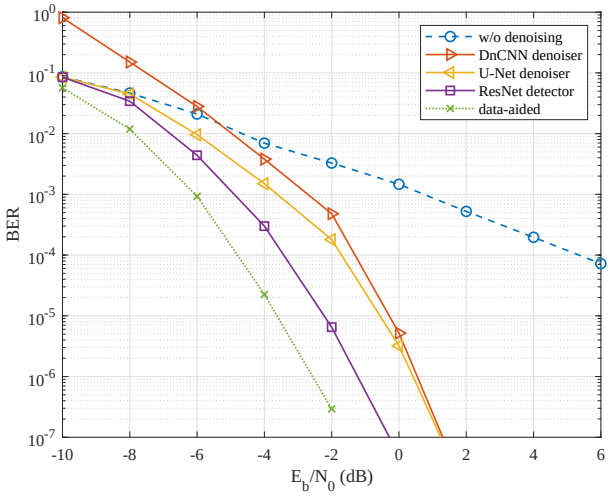

Figure 5.4 BER for $M = 100$ BS antennas, $N = 256$ subcarriers, and 16-QAM modulation, and different number of users.

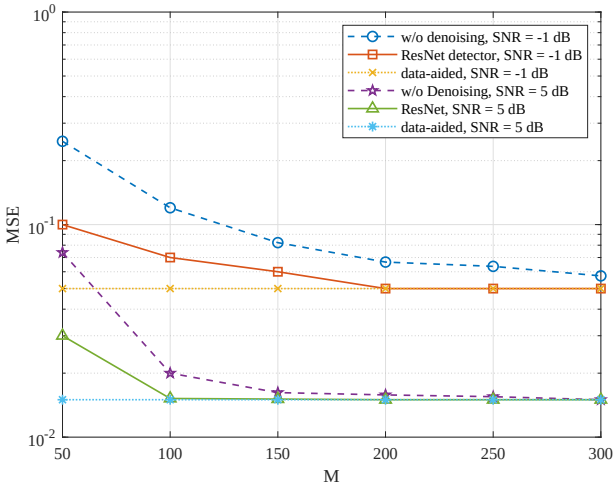

Figure 5.5 MSE with respect to the number of BS antennas, M, for $P = 4$ users, $N = 256$ subcarriers, and 16-QAM modulation at SNR $= -1$dB.

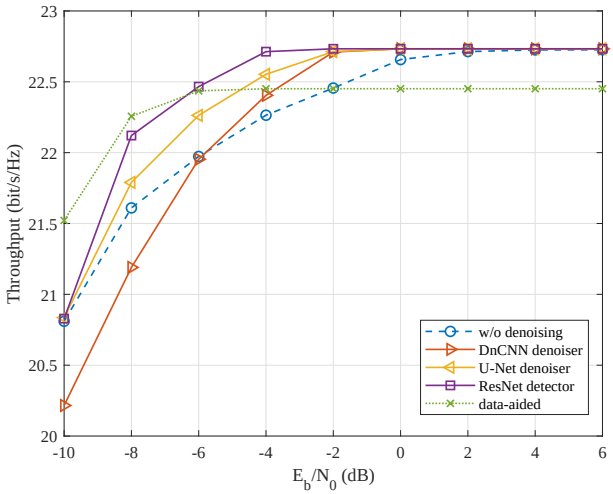

Figure 5.6 Spectral efficiency for $P = 4$, $M = 100$ BS antennas, $N = 256$ subcarriers, and 16-QAM modulation.

signal, where other users are transmitting their data. Then, using the subspace theory, a matrix orthogonal to the signal space of all the users other than the desired one can be calculated. Thus, we canceled the MUI by multiplying this matrix to the received signal at the BS antennas. Afterward, we proposed a virtual pilot detection technique that can detect the transmitted signal without the knowledge of channel impulse responses. Since in the proposed technique, several noise affected observations of the transmitted symbol are obtained, a DL-based algorithm can be used as a denoiser to improve the system performance. Then, first, we applied DnCNN algorithm as a denoiser that enhanced the accuracy of virtual pilot detection. Second, we proposed U-Net denoiser which outperforms DnCNN and leads to further improvement. Third, we proposed a joint denoising and detection technique based on the ResNet architecture which can effectively remove both the additive and the multiplicative noise. In our simulations, we demonstrated that the MSE performance of our proposed blind channel estimation technique meets that of the data-aided one at high SNRs. Then, we showed that by deploying the mentioned denoiser techniques, the MSE of our proposed technique can meet the MSE curve of the data-aided technique at lower SNRs.

Chapter 6

Conclusions and future work

This thesis addresses some of the main challenges of massive MIMO systems to prepare them for the next generation of wireless networks. In general, the goal of this thesis has been to explore the specific features of massive MIMO channels to enable these systems to work with low computational complexity, high spectral and energy efficiency, while preserving the optimal performance.

6.1 Summary of contributions and conclusions

In this section, the contributions of the thesis that were previously discussed in detail in chapters 3, 4 and 5 are briefly presented.

6.1.1 Performance analysis of FBMC-PAM for massive MIMO systems

Due to the drawbacks of OFDM modulation commonly used in multicarrier massive MIMO systems, FBMC-PAM is proposed as an alternative. FBMC-PAM has the same data rate as OFDM, while it does not need CP which makes it more spectrally efficient than OFDM. Moreover, FBMC-PAM has lower out of band emission, and consequently, it is more robust against frequency synchronization error. In addition, it has the lowest computational complexity among the traditional FBMC-based waveforms because of its minimum overlapping factor.

As discussed in chapter 3, while the low overlapping factor of FBMC-PAM causes some shortcomings in terms of equalization, self-equalization property of massive MIMO compensates this weakness and completely removes the error floor in the BER. Therefore, due to the absence of CP, FBMC-PAM offers a better BER performance compared to OFDM.

However, as the number of BS antennas grows, performance of massive MIMO system with FBMC-PAM saturates. It is explained that although the noise and MUI are averaged out in the asymptotic regime, some residual interference due to the multipath channel remains and causes the SIR saturation. Then, the level of this saturation, which is the upper bound of the system performance, is mathematically calculated. It is also shown that by increasing the number of subcarriers, a higher level of saturation can be achieved.

It is worth noticing that the effect of multipath channel can be completely removed in OFDM modulation due to the CP assignment. Thus, such a saturation level is not observed in OFDM-based massive MIMO systems, and their performance linearly grows as the number of BS antennas increases. Therefore, for the rest of this thesis, we focused on the OFDM-based massive MIMO systems.

6.1.2 Low-complexity frequency synchronization for massive MIMO systems

Since OFDM modulation is very sensitive to the frequency offset, an effective frequency synchronization technique is crucial for multi-user OFDM-based massive MIMO systems. Therefore, in chapter 4, two low-complexity CFO estimation techniques are proposed that are based on the covariance matrix of the received signals at the base station. In the first one, CFO can be estimated using a closed-form equation and there is no need to solve an optimization problem. However, it is limited to real valued pilots, while the second one is more general and works with any arbitrary pilot sequence. In the second technique, an optimization problem is proposed for CFO estimation that has a unimodal objective function, and can be simply solved by Golden search algorithm.

Then, since the time domain CFO compensation increases the complexity of the system, a frequency domain CFO compensation technique is proposed that takes place after combining the received signals. Performance of the proposed CFO estimation and compensation techniques is investigated in terms of SIR, and it is proved that even in the presence of CFO estimation error, the scattering effect of CFO is eliminated and only a phase shift remains. Hence, two iterative phase correction algorithms are proposed to mitigate the remaining error and achieve a high accuracy synchronization.

6.1.3 DL-aided blind channel estimation for massive MIMO systems

Another important concern in massive MIMO systems is CSI acquisition. As the scale of antenna array increases or the number of users in the network grows, channel estimation can limit massive MIMO system performance. Data-aided channel estimation techniques which employ the pilot information waste a large amount of resources and energy on transmitting pilot signals. Moreover, if the pilots are reused in the neighboring cells, pilot contamination causes inter-cell interference that strongly affects the performance of channel estimation. There are some blind channel estimation techniques in the literature that are presented in section 2.3.2. Those works either are limited with some constraints, such as the sparsity level of the channels, or impose high computational complexity to the system.

In chapter 5, a DL-aided blind channel estimation technique is proposed which consists of three steps. First, MUI cancellation is performed by calculating the orthogonal complement space matrix of the MUI. After that the system can be treated as a single user system. Second, the covariance matrix of the received signals in the resulting equivalent single user system is calculated. Then, with some mathematical manipulations and relying on the favorable propagation property of massive MIMO systems, the first transmitted symbols of all the users can be estimated individually. These detected signals can be considered as virtual pilots, and channels can be estimated based on the virtual pilots as the third step.

Ideally, noise and MUI are averaged out in massive MIMO system, and the channels can be accurately estimated with the aforementioned steps. However, in practice, noise is not completely eliminated and causes considerable amount of distortion. To deal with this issue, first the DnCNN is applied to suppress the noise effect before virtual pilot detection. Although, DnCNN can bring some improvement to the estimation technique, it does not work in the low SNR scenarios. Thus, U-Net, an alternative DL-based algorithm, is proposed that can offer a better performance than DnCNN at low SNRs. Then, we noticed that these denoising algorithms can learn the pattern in the input matrix and recover it, but they do not give the exact desired value at the output. In fact, they can only cope with additive noise, while there is some multiplicative noise as well. Therefore, we proposed the ResNet architecture followed by a feedforward fully connected neural network that can be trained to converge to the desired value and provide much more accurate virtual pilot detection. As a result, an MSE very close to that of a data-aided channel estimation technique can be achieved.

6.2 Future works

The work that is presented in this thesis can be extended, by pursuing a number of directions such as:

- The frequency synchronization technique proposed in this thesis is designed for fully digital massive MIMO systems. Similar approach can be applied to the hybrid massive MIMO, where digital domain combining/precoding is replaced by a combination of analog and digital processing to reduce the number of required analog to digital convertors (ADC). Hybrid massive MIMO systems are mostly applied to millimeter wave systems.
- Study the possibility of applying proposed denoising techniques for other use cases such as pilot decontamination can be useful.
- The main barrier of the deep learning-based end-to-end communication system is dealing with channel layer. As it is proved in this thesis, the transmitted symbols in massive MIMO systems can be detected without the knowledge of CSIs. Therefore, the possibility of replacing the whole system with an deep learning algorithm by exploiting unique properties of massive MIMO can be investigated.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[3] M. Series, "IMT vision-framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU*, pp. 2083–0, 2015.

[4] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE communications magazine*, vol. 52, no. 2, pp. 122–130, 2014.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[7] E. Bjornson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, 2018.

[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[9] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, 2017.

[10] G. Strang, G. Strang, G. Strang, and G. Strang, *Introduction to linear algebra*. Wellesley-Cambridge Press Wellesley, MA, 1993, vol. 3.

[11] T. L. Marzetta, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[12] M. Ganji and H. Jafarkhani, "On the performance of MRC receiver with unknown timing mismatch-a large scale analysis," *arXiv preprint arXiv:1703.10422*, 2017.

[13] V. Kotzsch and G. Fettweis, "Interference analysis in time and frequency asynchronous network MIMO OFDM systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010, pp. 1–6.

[14] C. Shan, Y. Zhang, L. Chen, X. Chen, and W. Wang, "Performance analysis of large scale antenna system with carrier frequency offset, quasi-static mismatch and channel estimation error," *IEEE Access*, vol. 5, pp. 26 135–26 145, 2017.

[15] K. Raghunath and A. Chockalingam, "SIR analysis and interference cancellation in uplink OFDMA with large carrier frequency/timing offsets," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, 2009.

[16] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE transactions on communications*, vol. 45, no. 12, pp. 1613–1621, 1997.

[17] A. Farhang, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Low complexity CFO compensation in uplink OFDMA systems with receiver windowing." *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2546–2558, 2015.

[18] H. Minn, V. K. Bhargava, and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 822–839, 2003.

[19] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, "The world's first real-time testbed for massive MIMO: Design, implementation, and validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.

[20] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE transactions on signal processing*, vol. 56, no. 8, pp. 4008–4019, 2008.

[21] Y. Wu, J. W. Bergmans, and S. Attallah, "Carrier frequency offset estimation for multiuser MIMO OFDM uplink using CAZAC sequences: performance and sequence optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 570680, 2011.

[22] Y.-R. Tsai, H.-Y. Huang, Y.-C. Chen, and K.-J. Yang, "Simultaneous multiple carrier frequency offsets estimation for coordinated multi-point transmission in OFDM systems," *IEEE transactions on wireless communications*, vol. 12, no. 9, pp. 4558–4568, 2013.

[23] V. Kotzsch, J. Holfeld, and G. Fettweis, "Joint detection and CFO compensation in asynchronous multi-user MIMO OFDM systems," in *IEEE Vehicular Technology Conference (VTC)*, 2009, pp. 1–5.

[24] H. Huang, W. G. Wang, and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system," in *IEEE International Conference on Communications (ICC)*, 2015, pp. 1280–1285.

[25] H. V. Cheng and E. G. Larsson, "Some fundamental limits on frequency synchronization in massive MIMO," in *IEEE Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 1213–1217.

[26] H. Hojatian, M. J. Omidi, H. Saeedi-Sourck, and A. Farhang, "Joint CFO and channel estimation in OFDM-based massive MIMO systems," in *IEEE International Symposium on Telecommunications (IST)*, 2016, pp. 343–348.

[27] W. Zhang, F. Gao, and H.-M. Wang, "Frequency synchronization for massive MIMO multi-user uplink," in *IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[28] W. Zhang, F. Gao, S. Jin, and H. Lin, "Frequency synchronization for uplink massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 235–249, 2018.

[29] W. Zhang, F. Gao, H. Minn, and H.-M. Wang, "Scattered pilots-based frequency synchronization for multiuser OFDM systems with large number of receive antennas," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1733–1745, 2017.

[30] W. Zhang and F. Gao, "Blind frequency synchronization for multiuser OFDM uplink with large number of receive antennas," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2255–2268.

[31] Y. Feng, W. Zhang, F. Gao, and Q. Sun, "Computationally efficient blind CFO estimation for massive MIMO uplink," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7795–7799, 2018.

[32] W. Zhang, Q. Yin, and F. Gao, "Computationally efficient blind estimation of carrier frequency offset for MIMO-OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7644–7656, 2016.

[33] S. Mukherjee and S. K. Mohammed, "Low-complexity CFO estimation for multi-user massive MIMO systems," in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.

[34] S. Mukherjee, S. K. Mohammed, and I. Bhushan, "Impact of CFO estimation on the performance of ZF receiver in massive MU-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9430–9436, 2016.

[35] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2006, pp. 359–363.

[36] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2015.

[37] X. Yuan, C. Fan, and Y. J. Zhang, "Fundamental limits of training-based uplink multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7544–7558, 2018.

[38] E. Bjornson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2016.

[39] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," *Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep*, vol. 5, pp. 10–15, 2007.

[40] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[41] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 104–118, 2014.

[42] W. Yan and X. Yuan, "Semi-blind channel-and-signal estimation for uplink massive MIMO with channel sparsity," *arXiv preprint arXiv:1903.04163*, 2019.

[43] J. Zhang, X. Yuan, and Y.-J. A. Zhang, "Blind signal detection in massive MIMO: Exploiting the channel sparsity," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 700–712, 2017.

[44] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE transactions on information theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[45] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *arXiv preprint arXiv:1203.5924*, 2012.

[46] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1461–1476, 2015.

[47] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, 2014.

[48] A. Mezghani and L. Swindlehurst, "Blind estimation of sparse multi-user massive MIMO channels," in *WSA 2017; 21th International ITG Workshop on Smart Antennas*. VDE, 2017, pp. 1–5.

[49] J. A. Franklin, T. M. Mitchell, and S. Thrun, *Recent advances in robot learning*. Springer, 1996.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[51] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv:1512.01274*, 2015.

[52] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[53] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[54] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky *et al.*, "Theano: A python framework for fast computation of mathematical expressions," *arXiv preprint arXiv:1605.02688*, 2016.

[55] F. Chollet *et al.*, "Keras," 2015.

[56] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS workshop*, no. CONF, 2011.

[57] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[59] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[61] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[62] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 873–880.

[63] T. Ohea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[64] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 298–303.

[65] H. Ye, G. Y. Li, B.-H. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2018, pp. 1–5.

[66] H. Ye, L. Liang, G. Y. Li, and B.-H. F. Juang, "Deep learning based end-to-end wireless communication systems with conditional GAN as unknown channel," *arXiv preprint arXiv:1903.02551*, 2019.

[67] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2278–2281, 2018.

[68] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[69] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based MIMO communications," *arXiv preprint arXiv:1707.07980*, 2017.

[70] T. Erpek, T. J. O'Shea, and T. C. Clancy, "Learning a physical layer scheme for the mimo interference channel," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–5.

[71] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2326–2330.

[72] V. Papyan, Y. Romano, J. Sulam, and M. Elad, "Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 72–89, 2018.

[73] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

[74] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[75] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, 2019.

[76] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2905–2917, 2018.

[77] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8549–8560, 2018.

[78] J. Mairal, F. R. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration." in *ICCV*, vol. 29. Citeseer, 2009, pp. 54–62.

[79] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[80] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[81] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black, "Efficient belief propagation with learned higher-order markov random fields," in *European conference on computer vision*. Springer, 2006, pp. 269–282.

[82] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[83] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

[84] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 80–93, 2014.

[85] M. Schellmann, Z. Zhao, H. Lin, P. Siohan, N. Rajatheva, V. Luecken, and A. Ishaque, "FBMC-based air interface for 5g mobile: Challenges and proposed solutions," in *IEEE International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 2014, pp. 102–107.

[86] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding digital subscriber line technology*. Prentice Hall PTR, 1999.

[87] R. v. Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House, Inc., 2000.

[88] G. Fettweis, M. Krondorf, and S. Bittner, "GFDM-generalized frequency division multiplexing," in *IEEE Vehicular Technology Conference (VTC Spring)*, 2009, pp. 1–4.

[89] G. P. Fettweis, M. Krondorf, and S. Bittner, "GFDM-generalized frequency division multiplexing." in *VTC Spring*. Citeseer, 2009, pp. 1–4.

[90] M. J. Abdoli, M. Jia, and J. Ma, "Weighted circularly convolved filtering in OFDM/OQAM," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2013, pp. 657–661.

[91] A. M. Tonello, "A novel multi-carrier scheme: Cyclic block filtered multitone modulation," in *2013 IEEE International Conference on Communications (ICC)*. IEEE, 2013, pp. 5263–5267.

[92] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 66–70.

[93] M. Huemer, C. Hofbauer, and J. B. Huber, *The potential of unique words in OFDM*. na, 2010.

[94] Z. Wang, X. Ma, and G. B. Giannakis, "Optimality of single-carrier zero-padded block transmissions," in *2002 IEEE Wireless Communications and Networking Conference Record. WCNC 2002 (Cat. No. 02TH8609)*, vol. 2. IEEE, 2002, pp. 660–664.

[95] D. Van Welden and H. Steendam, "Near optimal iterative channel estimation for KSP-OFDM," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4948–4954, 2010.

[96] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, and J.-F. Frigon, "Universal-filtered multi-carrier technique for wireless systems beyond LTE," in *2013 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2013, pp. 223–228.

[97] M. Bellanger, "FS-FBMC: An alternative scheme for filter bank based multi-carrier transmission," in *IEEE International Symposium on Communications Control and Signal Processing (ISCCSP)*, 2012, pp. 1–4.

[98] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *Signal Processing Magazine, IEEE*, vol. 28, no. 3, pp. 92–112, 2011.

[99] R. Chang, "High-speed multichannel data transmission with bandlimited orthogonal signals," *Bell Sys. Tech. J*, vol. 45, no. 10, pp. 1775–1796, 1966.

[100] A. Farhang, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Filter bank multicarrier for massive MIMO," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2014, pp. 1–7.

[101] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1170–1183, 2002.

[102] D. Mattera, M. Tanda, and M. Bellanger, "Filter bank multicarrier with PAM modulation for future wireless systems," *Signal Processing*, vol. 120, pp. 594–606, 2016.

[103] C. Kim, K. Kim, Y. H. Yun, Z. Ho, B. Lee, and J.-Y. Seol, "QAM-FBMC: A new multi-carrier system for post-OFDM wireless communications," in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.

[104] Y. H. Yun, C. Kim, K. Kim, Z. Ho, B. Lee, and J.-Y. Seol, "A new waveform enabling enhanced QAM-FBMC systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 116–120.

[105] M. Bellanger, D. Mattera, and M. Tanda, "A filter bank multicarrier scheme running at symbol rate for future wireless systems," in *Wireless Telecommunications Symposium (WTS)*. IEEE, 2015, pp. 1–5.

[106] H. S. Malvar, *Signal processing with lapped transforms*. Artech House, Inc., 1992.

[107] P. P. Vaidyanathan, *Multirate systems and filter banks*. Pearson Education India, 1993.

[108] H. Malvar, "Optimal pre-and postilters in noisy sampled-data systems," Ph.D. dissertation, PhD. dissertation, Dep. Elec. Eng, Massachusetts Inst. Technol., Cambridge, 1986.

[109] M. Bellanger, D. Mattera, and M. Tanda, "Lapped-OFDM as an alternative to CP-OFDM for 5G asynchronous access and cognitive radio," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, 2015, pp. 1–5.

[110] D. Mattera, M. Tanda, and M. Bellanger, "Performance analysis of some timing offset equalizers for FBMC/OQAM systems," *Signal Processing*, vol. 108, pp. 167–182, 2015.

[111] R. W. Young and N. G. Kingsbury, "Frequency-domain motion estimation using a complex lapped transform," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 2–17, 1993.

[112] J. P. Princen, A. Johnson, and A. B. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *IEEE International Conference on ICASSP'87 Acoustics, Speech, and Signal Processing*, vol. 12, 1987, pp. 2161–2164.

[113] S. Jošilo, M. Pejović, B. Đorđević, M. Narandžić, and S. Nedić, "Multicarrier waveforms with I/Q staggering: uniform and nonuniform FBMC formats," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–15, 2014.

[114] B. Farhang-Boroujeny and C. H. Yuen, "Cosine modulated and offset QAM filter bank multicarrier techniques: a continuous-time prospect," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 6, 2010.

[115] K. Hari, D. Baum, T. Kaitz, and B. D. Trinkwon, "Project ieee 802.16 broadband wireless access working group, http://ieee802. org/16, channel models for fixed wireless applications," 2001.

[116] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[117] B. Emil, *A look at an LTE-TDD Massive MIMO product*, Accessed 2019-08-07. [Online]. Available: http://ma-mimo.ellintech.se/2018/08/27/a-look-at-an-lte-tdd-massive-mimo-product/

[118] P. von Butovitsch, D. Astely, C. Friberg, A. Furuskär, B. Göransson, B. Hogan, J. Karlsson, and E. Larsson, "Advanced antenna systems for 5G networks," *Ericsson White Paper*, 2018.

[119] S. U. Six, "5g-ready cities; significant milestone toward launching first 5G mobile network in the US," 2019.

[120] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Golden section search in one dimension," *Numerical Recipies in C: the Art of Scientific Computing*, vol. 2, 1992.

[121] K. Scharnhorst, "Angles in complex vector spaces," *Acta Applicandae Mathematica*, vol. 69, no. 1, pp. 95–103, 2001.

[122] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[123] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Siam, 1997, vol. 50.

[124] D. E. Knuth, *The art of computer programming*. Pearson Education, 1997, vol. 3.

[125] D. Harvey and J. Van Der Hoeven, "Integer multiplication in time o (n log n)," 2019.

[126] E. Klarreich, "Multiplication hits the speed limit," 2019.

[127] W.-B. Yang, W.-B. Yang, and M. Souryal, *LTE physical layer performance analysis*. US Department of Commerce, National Institute of Standards and Technology, 2014.

[128] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.