UNIVERSITY OF CALIFORNIA,
IRVINE

Asynchronous Transmission in Multiuser Networks

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical Engineering

by

Mehdi Ganji

Dissertation Committee:
Chancellor's Professor Hamid Jafarkhani, Chair
Professor Ender Ayanoglu
Professor Lee Swindlehurst

2020

DEDICATION

I dedicate this thesis to my beloved parents. I cannot imagine what my life would have been without their sacrifice.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Prof. Jafarkhani, for his guidance and support throughout my Ph.D. studies. I am truly honored to have the opportunity of being his student. He is a true role model. He constantly encouraged me to pursue my research goals. I truly appreciate his help in building my confidence and teaching me to become a researcher. I could not have imagined having a better advisor and mentor for my Ph.D. study. I would like to thank the rest of my thesis and advancement committee members, Prof. Ayanoglu, Prof. Swindlehurst, Prof. Eltawil, and Prof. Jafar for their insightful comments and feedback which motivated me to expand and deepen my research from various perspectives.

I am also grateful to my friends, who were always there for me throughout this chapter of my life. I want to thank Pouya Pezeshkpour, Arash Gholami, Ali Tazarv, Hessam Pirzadeh, Behnam Pourghassemi, Navid Rezazadeh, Korosh Vatanparvar, Forough Arabshahi, Laleh Aghababaie Beni, and Mahdi Abbaspour Tehrani for standing by my side and supporting me. I want to thank my fantastic lab mates in our research group: Xun Zou, Lisi Jiang, Weiqi Li, Saeed Karimi Bidhendi, Carles Diaz, Jun Guo, Biao He, Xiaoyi Liu, Sina Poorkasmaei, Mahdi HajiAghayi, Michael R. Avendi, Erdem Koyuncu, Murat Demirtas for their instructive comments and helpful discussions. Notably, I want to thank Xun Zou, from whom I acquired a vast amount of knowledge and pleasure during our fruitful collaboration.

Another gratitude goes out to the National Science Foundation (NSF) Award Computing and Communication Foundations (CCF)-1526780 for providing funding for my research. I'm also grateful for the Department of Electrical Engineering and Computer Science (EECS) for offering me the opportunity to pursue a doctorate degree.

Finally, I would like to thank my parents, my brother, Behnam, and my sister, Behnaz, who have always pushed me to be curious, open-minded, and eager to improve. My father taught me to be disciplined and hard-working, my mother taught me to be kind and passionate, Behnam taught me to be ambitious, and Behnaz taught me to be reliable and supportive of the loved ones. I am always grateful for their supports and sacrifices. I cannot express how fortunate and blessed I am to have my other half, Mehran, as my best friend my entire life. Mehran was my security blanket since childhood because I knew I was never alone. Without his constant guidance and support, I would not have been able to attain the achievements that I have today.

Last but not least, I want to express my gratitude and love for my dear wife, Shabnam, who has supported me throughout this journey. I sincerely appreciate her for pushing me to be a stronger and better person every day. I cannot thank her enough for never giving up on me, supporting me unwaveringly through my darkest times, and showing me what unconditional love looks like.

VITA

Mehdi Ganji

EDUCATION

Ph.D. in Electrical Engineering — **2020**
University of California, Irvine — *Irvine, California*

M.Sc. in Electrical Eng. & Computer Science — **2015**
University of California, Irvine — *Irvine, California*

B.Sc. in Electrical Engineering — **2013**
Sharif University of Technology — *Tehran, Iran*

RESEARCH EXPERIENCE

Graduate Research Assistant — **2014–2020**
University of California, Irvine — *Irvine, California*

Intern-ship — **Summer 2018**
SoC Lab, Samsung Semiconductor Inc. — *San Diego, California*

Reviewer-ship
IEEE JSTSP, IEEE TWC, IEEE TVT, IEEE WCL

TEACHING EXPERIENCE

Discrete Mathematics for Computer Science — **Spring 2020**
ICS Department, University of California, Irvine — *Irvine, California*

Engineering Probability — **Winter 2016, 2017, 2020**
EECS Department, University of California, Irvine — *Irvine, California*

Computer Netwroks — **Fall 2019**
ICS Department, University of California, Irvine — *Irvine, California*

Random Process — **Fall 2015**
EECS Department, University of California, Irvine — *Irvine, California*

Digital Signal Processing — **Spring 2013**
EE Department, Sharif University of Technology — *Tehran, Iran*

REFEREED JOURNAL PUBLICATIONS

On the Capacity of Faster than Nyquist Signaling **2020**
IEEE Communications Letters

Exploiting Time Asynchrony in Multiuser Transmit Beamforming **2020**
IEEE Transactions on Wireless Communications

Cooperative Asynchronous Non-Orthogonal Multiple Access with Power Minimization Under QoS Constraints **2019**
IEEE Transactions on Wireless Communications

Trellis-Coded Non-Orthogonal Multiple Access **2019**
IEEE Wireless Communications Letters

Improving NOMA Multi-Carrier Systems with Intentional Frequency Offsets **2019**
IEEE Wireless Communications Letters

REFEREED CONFERENCE PUBLICATIONS

Downlink Asynchronous Non-Orthogonal Multiple Access Systems with Imperfect Channel Information **2019**
IEEE Global Communications Conference

A Block-Based Non-Orthogonal Multi-carrier Scheme **2019**
IEEE Global Communications Conference

Time Asynchronous NOMA for Downlink Transmission **2019**
IEEE Wireless Communications and Networking Conference

On the Performance of MRC Receiver with Unknown Timing Mismatch-A Large Scale Analysis **2018**
IEEE International Conference on Communications

Exploiting Asynchronous Signaling for Multiuser Cooperative Networks with Analog Network Coding **2017**
IEEE Wireless Communications and Networking Conference

Interference Mitigation Using Asynchronous Transmission and Sampling Diversity **2016**
IEEE Global Communications Conference

ABSTRACT OF THE DISSERTATION

Asynchronous Transmission in Multiuser Networks

By

Mehdi Ganji

Doctor of Philosophy in Electrical Engineering

University of California, Irvine, 2020

Chancellor's Professor Hamid Jafarkhani, Chair

Time asynchrony inherently exists in many wireless communication systems, especially in multiuser scenarios, where the users are located in various locations. Different locations and paths impose different delays on the received signals, resulting in asynchronous reception at the receiver. In most of the works in the literature, perfect synchronization among received signals is a common presumption. However, it might be impossible to synchronize signals at all the nodes in the network even if an ideal infrastructure using control signals is considered. For example, assume that the receiver encompasses multiple receive antennas or multiple distributed base stations. Then, although the synchronization can be realized at one of them, realizing synchronization at all of the base stations/antennas might be impossible. Thus, it is essential to investigate the effect of the time asynchrony in the wireless systems, particularly multiuser systems.

Asynchrony naturally imposes some performance degradation in a system designed optimally based on having synchronized incoming signals. One immediate solution is to eliminate the time asynchrony and achieve almost perfect synchronization among the received signals. This question is analyzed under the notion of time synchronization. There are many methods proposed and analyzed in literature aiming for that goal, and one of the

common ones is to use control signals to achieve synchronization among different users. However, apart from this approach's feasibility, the questions which are atypical but of importance to answer are: Can a system be designed such that it provides performance improvement under an asynchronous condition? Can a system, which is designed based on asynchronous assumption, outperform its counterpart system, which is designed based on the assumption of perfect synchronization?

We thoroughly investigate these questions in this thesis. First, we theoretically analyze the performance bounds of multi-antenna, multiuser systems under the asynchrony assumptions. We show that by exploiting inherent time delays between different users in a multiuser/multi-antenna scenario, we can improve the performance. Besides, we propose to intentionally add timing offsets in the systems that are not inherently impaired by time asynchrony. We introduce the optimal transceiver designs under asynchronous assumptions and analyze the performance improvement provided. We consider various multiuser networks, including broadcasting networks, multiple access networks, and cooperative networks, and examine the advantages and disadvantages of having asynchrony in such multiuser networks.

Chapter 1

Introduction

1.1 Time Synchronization

Time synchronization is an inseparable ingredient of wireless communications. Timing synchronization is the process by which the proper sampling instants are acquired by the receiver. If the receiver cannot grasp the right times to sample the incoming signal, the resulting performance will be degraded [1]. Generally, in wireless systems, the transmitted signals are distorted by changes in amplitude, phase, and timing, which are a priori unknown to the receiver node. Usually, the channel parameters are estimated in the channel estimation process using pilot based or blind methods. For example, in SISO systems, the received signal at the receiver antenna is affected by a single timing offset (TO). The receiver needs to estimate this parameter and compensate for its effects to be able to decode the transmitted symbols.

On the other hand, in a multiuser system, data is transmitted from distributed users. The received signal at a receive antenna is given by a linear combination of the distorted

and delayed version of the data symbols transmitted from different users resulting in multiple timing offsets (MTOs) [2]. In practice, the users perform timing correction using the timing advance (TA) estimate received in the physical downlink control channel (PDCCH). For example, in LTE standards, synchronization is achieved through periodically transmitted primary and secondary synchronization signals from the base station. Any user who has not yet acquired the uplink synchronization can use the primary and secondary synchronization signals to achieve synchronization in the downlink. Therefore, a significant amount of overhead signaling is used to eliminate MTOs. Moreover, having multiple distributed antennas make the synchronization process challenging and, in some cases, even impossible [3].

Besides the additional overhead signaling, waste of resources, and other challenges to remove MTOs at all receive nodes; the main question is that whether it is necessary to enforce perfect time synchronization with such a high price? Some preliminary works in the literature motivate the use of time asynchrony to improve the system performance [4, 5, 6, 7, 8, 9]. In this thesis, we thoroughly investigate the possibilities and challenges that time asynchrony provides in multiuser networks.

1.2 Multiuser Systems: Inter-user Interference

Another issue in multiuser systems is inter-user interference (IUI), which is the result of sharing a common medium to transmit data. Numerous examples of multi-access communication include uplink transmission of a single cell in a cellular system, a group of twisted-pair copper subscriber lines transmitting data to the same switching office, multiple ground stations communicating with a satellite and interactive cable television networks. The key challenge in these types of networks is interference from unwanted sig-

nals. Over several decades, many methods have been introduced to address this problem [10, 11, 12, 13]. The currently prevailing approach for mitigating IUI lies in the category of orthogonal multiple access (OMA). For example, time division multiple access (TDMA) protocols allocate different time slots to different users to mitigate interference. The same concept can be applied by partitioning the frequency spectrum among different users, called frequency division multiple access (FDMA). Code division multiple access (CDMA) is another scheme used to surpass IUI in which users are multiplexed by distinct codes rather than by orthogonal frequency bands, or by orthogonal time slots [14]. Multiple antennas are used to take advantage of the spatial domain to cancel interference [15], [16]. More recently, non-orthogonal multiple access (NOMA) [17, 18] is proposed as a candidate for future radio access to partially fulfill the requirements of future networks.

The significant advantage of OMA methods is that their complexity is merely the complexity of single-user encoders/decoders. On the other hand, assigning resource blocks exclusively to each user can be very inefficient (in terms of achievable rate-regions) and may pose a severe fairness problem among users. In contrast to OMA, NOMA allows users to utilize the same resource blocks for transmission simultaneously and therefore is potentially more efficient. When evaluated under the LTE system characteristics, NOMA demonstrates significant gains over OMA systems [19, 20]. The NOMA methods are generally based on superposition coding and successive interference cancellation (SIC) [21, 22]. Thus, the mitigation of IUI is critical in modern multiuser networks.

In this thesis, we investigate the timing asynchrony between received signals as an additional resource to address the problem of IUI. We show that by exploiting time delays between users and proper sampling, the IUI can be effectively reduced. Besides, by applying the oversampling method, a system model with increased independent dimensions can be obtained. The increased dimensionality combined with proper receiver design can mitigate IUI and enhance system performance. It is shown that by exploiting time delays

between the received signal and employing appropriate sampling methods and detection methods, not only can the IUI be appropriately managed, but the resulting system outperforms the synchronous counterpart. The content of this thesis is summarized next.

1.3 Content

In Chapter 2, we analyze the performance degradation caused by time asynchrony in a large scale multiuser MIMO (MU-MIMO) systems. With a large number of receive antennas, one major challenge is to receive perfectly aligned signals at all the receive antennas, particularly in a distributed scenario where the receive antennas are not collocated. Inevitable timing mismatch between received signals results in imperfect sampling and hence creates inter-symbol interference (ISI). Considering practical challenges for delay acquisition in a large scale MU-MIMO system, we assume that the timing mismatch values are random, and the receiver only knows their distribution. We develop a mathematical model that explicitly accounts for the random timing mismatch among the received signals. We quantify the uplink achievable rates obtained by the MRC receiver with perfect channel state information (CSI) and imperfect CSI while random time delays impair the system performance. After quantifying the detrimental impact of asynchrony on the MRC receiver, we design new algorithms to alleviate the effects of timing mismatch. Finally, to verify our analytical results, we present simulation results that thoroughly investigate the performance of the traditional MRC receiver and the introduced designs under asynchrony assumption.

The goals in Chapter 2 are two-fold. The first goal is to show the harmful effects of time asynchrony on the system performance while the time delays are neglected. The second goal is to shed light on the importance of sampling methods and receiver design

in dealing with time asynchrony. If a proper sampling method is not used, some parts of information are lost, and the performance is degraded by time asynchrony, no matter what process we apply afterward. Therefore, in Chapter 3, we express more details about time synchronization and the possible opportunities that it provides. More importantly, we introduce the proper structure and optimal sampling methods in order to exploit time asynchrony. We present various system models accounting for time asynchrony and provide their corresponding features. We also explain the reasons behind the possible advantages of time asynchrony in providing performance improvement.

In Chapter 4, we show that by investigating inherent time delays between different users in a multiuser scenario, we can cancel interference more efficiently. Time asynchrony provides another tool to cancel interference, which results in preserving other resources like frequency, time, and code. By taking advantage of the sampling methods explained in Chapter 3, we derive the system model, which is similar to an ISI channel. Then, to equalize the ISI introduced by timing mismatch, sequence detection methods such as the Viterbi algorithm are applied. We also present successive interference cancellation with hard and soft decision passing. We show that by exploiting the time asynchrony, proper sampling, and suitable detection algorithms, a superior Bit Error Probability (BEP) can be achieved compared with the synchronous model. Also, the diversity of zero-forcing detection methods is analyzed. It is shown that with asynchronous transmission, full diversity can be achieved while the synchronous counterpart suffers from loss of diversity due to IUI cancellation. Simulation results are presented to verify our analysis.

In Chapter 4, the inherent time delays in an uplink multiuser are exploited. However, the same concept can be applied to a downlink scenario where the timing offset is intentionally added among the transmitted signals. In Chapter 5, we investigate the effect of time asynchrony in NOMA schemes for downlink transmissions. First, we analyze the benefit of adding intentional timing offsets to the conventional power domain-NOMA (P-NOMA).

This method, called Asynchronous-Power Domain-NOMA (AP-NOMA), introduces artificial symbol-offsets between packets destined for different users. It reduces the mutual interference, which results in enlarging the achievable rate-region of the conventional P-NOMA. Then, we propose a precoding scheme that fully exploits the degrees of freedom provided by the time asynchrony. We call this multiple access scheme T-NOMA, which provides higher degrees of freedom for users than the conventional P-NOMA or even the modified AP-NOMA. T-NOMA adopts precoding at the base station and a linear preprocessing scheme at the receiving user, which decomposes the broadcast channel into parallel channels circumventing the need for Successive Interference Cancellation (SIC). The numerical results show that T-NOMA outperforms AP-NOMA and both outperform the conventional P-NOMA.

In chapter 5, it is shown that asynchronous transmission can improve the capacity region of the broadcast channel. Mainly, with the use of the proposed precoding, significant capacity enlargement can be achieved. However, this is not obtained free. In fact, the improved capacity region is obtained in exchange for out-of-band (OOB) emission. In Chapter 6, we analyze the OOB emission and propose a simple solution that provides a trade-off between increased capacity and OOB emission. We analyze the OOB emission in the context of faster than Nyquist (FTN) signaling, which resembles the asynchronous downlink transmission. FTN refers to transmitting the desired symbols faster than the Nyquist rate with shorter time intervals, which was first introduced in [23]. Asynchronous transmission in which multiple streams of information are superimposed and transmitted with some timing offsets resemble FTN. In Chapter 6, we consider the capacity computations of faster-than-Nyquist (FTN) signaling. We calculate the theoretical capacity of FTN signaling, which is obtained by a correlated input. The capacity-achieving power spectral density (PSD) is derived, and its superiority over the independent input is shown. The practical issue imposed by the capacity-achieving PSD, i.e., OOB emission, is shown.

To solve this issue, we introduce an upper-bound for the input PSD to limit the OOB emission. The new optimization problem is solved, and the constrained PSD is obtained. The introduced PSD captures the trade-off between the obtained capacity and the OOB emission.

In Chapter 7, we analyze the benefits of intentionally adding timing mismatch in the downlink transmit beamforming for wireless transmission. Transmit beamforming enables the so-called space-division multiple access (SDMA), where multiple spatially separated users are served simultaneously. The optimal beamforming vectors can be found to minimize the average transmit power under each user's Quality-of-Service (QoS) constraint. We show that intentionally adding timing offsets between the transmitted signals can significantly reduce the average transmission power compared with the conventional optimal beamforming method while providing the same QoSs for users. The frequency-selectivity in communication channels provides the opportunity to exploit intelligent design for performance improvement. The frequency-selectivity is limited in environments with line-of-sight links or little scattering. In such environments, we propose adding intentional time delays to induce frequency-selectivity that can be exploited. We provide three different methods exploiting the artificially induced frequency-selectivity, which improves the performance with a computational complexity similar to that of the optimal synchronous beamforming. We derive the expressions for the achievable rates using the proposed methods and then provide efficient algorithms to solve the minimum power optimization. We show analytically and numerically that our proposed methods outperform the conventional optimal transmit beamforming.

In Chapter 8, we thoroughly analyze the rate-region provided by the asynchronous transmission in multiple access channels (MACs). We derive the theoretical capacity regions which apply to a wide range of pulse shaping methods. We analytically prove that asynchronous transmission enlarges the capacity region of MACs. We show that although

successive interference cancellation (SIC) is sum-rate achieving for the conventional uplink NOMA methods, it is unable to achieve the boundary of the capacity region for the asynchronous transmission. We also demonstrate that for the asynchronous transmission, the optimal SIC decoding order to achieve the maximum sum-rate is based on the channel strengths of the users. This optimal ordering is in contrast to the conventional uplink NOMA, where the order of decoding does not affect the sum-rate. In Chapter 9, final remarks and summary of main contributions are provided. Besides, possible future works are mentioned.

Chapter 2

Performance Degradation Caused by Time Asynchrony: A Large-Scale Analysis

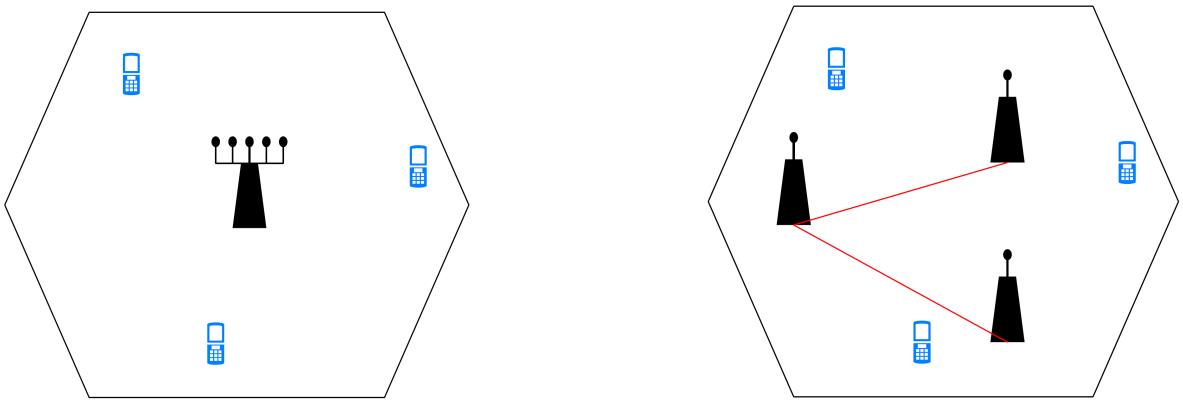
This chapter shows how time asynchrony can degrade the performance of a distributed multiuser system. One of the main issues, particularly in distributed massive MIMO setting, is timing asynchrony among different users. We quantify the uplink achievable rates obtained by the MRC receiver with perfect channel state information (CSI) and imperfect CSI while random time delays impair the system performance. Furthermore, some solutions are provided to alleviate the degradation caused by time asynchrony.

2.1 Introduction

Introducing multiple-input multiple-output (MIMO) systems was a breakthrough in communication systems. MIMO communications were studied extensively during the past two decades [24, 25]. One of the MIMO systems' applications is in multiuser scenarios where K users communicate with a common multiple-antenna receiver. Besides traditional problems in point to point communication, due to the distributed nature of multiuser-MIMO (MU-MIMO) systems, new challenges exist like timing mismatch between received signals from different users [11]. When the number of users and the number of receive antennas are moderate, this issue is often handled by synchronization methods [26, 27, 28]. However, increasing the number of receive antennas and users makes the time delay estimation or synchronization challenging, especially in the context of massive MIMO systems[29].

In large scale MU-MIMO systems, the base station is equipped with a very large number of receive antennas and communicates with tens of users at the same time and frequency. The benefits of the massive MIMO settings including, near-optimal performance using simple processing like maximum ratio combining (MRC), increased spectral efficiency, and energy efficiency, have been studied in the literature[30, 31, 32]. Two models can be adopted to implement massive MIMO in wireless networks: centralized, where antennas are co-located at the base station (BS), as illustrated in Fig. 2.1a, and distributed, where BS antennas are deployed at different geographical locations while connected through high-capacity back-haul links such as fiber-optic cables, as shown in Fig. 2.1b.

From a practical point of view, compared with the distributed topology, the centralized configuration is easier to deploy [33]. The distributed setting, however, exhibits several advantages over the centralized setting, such as lower transmit power, higher multiplexing gain, higher spectral efficiency, enhanced coverage area, and ease of network planning [34, 35]. Nevertheless, many challenges need to be addressed before the gains can be



(a) Centralized (b) Distributed

Figure 2.1: Different implementations of massive MIMO systems

realized in practice [36, 37]. For hundreds of receive antennas, one major challenge is the fact that it is impossible to receive perfectly aligned signals at all the receive antennas, especially in a distributed scenario where the receive antennas are not collocated. In [38], the authors showed that when multiple base stations (BSs) communicate with their corresponding users, the interference is inherently asynchronous, meaning that BSs cannot align all the interfering signals at each user because of the different propagation times between the BSs and users. This phenomenon is what happens in uplink, where multiple users communicate with a base station with multiple distributed receive antennas. Even if the users perform timing correction using the timing advance (TA) estimate received in the physical downlink control channel (PDCCH), the synchronization can be realized at only one receive antenna. However, due to different propagation delays, the other receive antennas will experience asynchrony. Therefore, it is of great importance to investigate timing mismatch in large scale distributed MU-MIMO systems. Inevitable timing mismatch between received signals results in imperfect sampling and hence creates inter-symbol interference (ISI), as illustrated in Fig. 2.2. Considering practical challenges for delay acquisition in a large scale MU-MIMO system, we assume that the timing mismatch values are random, and the receiver only knows their distribution.

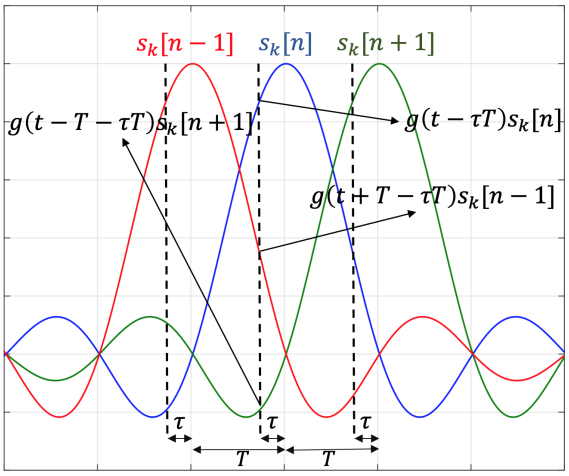


Figure 2.2: Demonstration of ISI caused by imperfect sampling

It is shown in the literature that in large scale MU-MIMO systems, a low complexity MRC receiver can approach near-optimal performance and even outperforms its complex counterparts, i.e., ZF and MMSE receivers, at low SNR[31]. The MRC receiver also follows the power scaling law. The power scaling law indicates that, for all SNRs, to maintain the same quality-of-service of a single-user SISO scenario with no interference, the transmitting power of single-antenna users to a 100-antenna BS would be almost 1 % of that of the single-user SISO system [39]. As we shall see, ignoring the asynchrony can significantly degrade the performance of the MRC receiver. We develop a mathematical model that explicitly accounts for the random timing mismatch among the received signals. We then quantify the detrimental impact of asynchrony on the MRC receiver and suggest how to mitigate it by making some modifications to the MRC receiver.

The rest of this chapter is organized as follows: In Section 2.2, first, we introduce the system model and preliminary results for the perfect synchronized scenario, then, modify the system model to include the existence of random time delays. We analyze the achievable rates obtained by the MRC receiver with random time delays in Section 2.3. In Section 2.4, the MRC-ZF receiver structure is introduced, and the corresponding achievable rates are presented. Next, simulation results are presented in Section 2.5 to verify our analysis.

Finally, we summarize our contributions in Section 2.6. The materials in this chapter are partially published in [40, 41].

2.2 System Model

2.2.1 Transmit Signal Model

We consider a single-cell multiuser distributed massive MIMO system. In this system, there is one BS equipped with M antennas which are spatially distributed. The BS serves K users, each equipped with single antenna. The number of antennas M is assumed to be large, e.g., a few hundreds. We assume that the distributed BS antennas are connected with high capacity back-haul (e.g., optical fiber channels) and have ideal cooperation with each other [42, 43]. The single carrier uplink communication is considered, where the K users transmit their data in the same time-frequency resource to the BS. The signal transmitted from User k is described by:

$$s_k(t) = \sqrt{\rho_d} \sum_{n=1}^{N} s_k[n] p(t - nT) \tag{2.1}$$

where T, ρ_d and $p(.)$ represent the symbol period, the transmit power, and the pulse-shaping filter with non-zero duration of T_p, respectively. For the rectangular pulse shape, $T_p = T$, and for Nyquist pulse shapes truncated with I significant adjacent side lobes, $T_p = 2(I + 1)T$. The number of significant adjacent side lobes is specified based on the desired stop band attenuation. More number of side lobes results in higher stop band attenuation. For example, we assume 3 side lobes in our simulation which results in about 25 dBm stop band attenuation for the square-root raised cosine pulse shape. Also, N is the frame length and $s_k[i]$ is the transmitted symbol by User k in the ith time slot.

2.2.2 Received Signal Model: Perfect Synchronization

Denote the channel coefficient between the mth antenna of the BS and the kth user as c_{mk}. We consider both the path-loss and small-scale fading as follows [43]:

$$c_{mk} = \sqrt{\beta_{mk}} h_{mk}, \quad k = 1, \cdots, K, \quad m = 1, \cdots, M \tag{2.2}$$

where h_{mk} is the small-scale fading coefficient, which is modeled as a random variable with zero-mean and unit-variance. Without loss of generality, Rayleigh fading is adopted in the simulation, where h_{mk}s are independent, identically distributed (i.i.d), following circularly symmetric complex Gaussian distribution, i.e., $h_{mk} \sim CN(0, 1)$. β_{mk} models the path-loss, given as:

$$\beta_{mk} = \left(\frac{r_h}{r_{mk}} \right)^v \tag{2.3}$$

where v is the path-loss exponent, and r_{mk} is the distance between the kth user and the mth BS antenna. r_h is the minimum distance between a user and a BS antenna, which exists due to physical implementation. In centralized configuration, $r_{1k} = \cdots = r_{Mk}$, thus $\beta_{1k} = \cdots = \beta_{Mk}$, which indicates that the large-scale fading coefficients from the kth user to all BS antennas are the same. Assuming fixed channel coefficients for a block, the continuous received signal at the mth receive antenna can be represented by:

$$y_m(t) = \sum_{k=1}^{K} c_{mk} s_k(t) + n_m(t) \tag{2.4}$$

where $n_m(t)$ is the additive white Gaussian noise (AWGN) with zero mean and variance of σ_n^2. To obtain the discrete samples of the received signal, first, the continuous received

signal should be passed through a matched filter and its output can be written as follows:

$$\hat{y}_m(t) = \sqrt{\rho_d} \sum_{k=1}^{K} c_{mk} \sum_{n=1}^{N} s_k[n] g(t - nT) + n_m(t) * p(t) \tag{2.5}$$

where $g(t) = p(t) * p(t)$. Then, the output of the matched filter is sampled at the instants of, $t_i^s = iT, i = 1, \cdots, N$. The obtained samples at the sampler of the mth receive antenna, denoted by $y_m[i] = \hat{y}_m(t)|_{t=t_i^s}$, can be written as:

$$y_m[i] = \sqrt{\rho_d} \sum_{k=1}^{K} \sqrt{\beta_{mk}} h_{mk} \sum_{n=1}^{N} s_k[n] g((i-n)T) + \hat{n}_m(t)\big|_{t=iT} \tag{2.6}$$

where $\hat{n}_m(t) = n_m(t) * p(t)$. Due to the ideal assumption of perfect synchronization, there is no ISI caused by adjacent symbols, i.e., $y_m[i] = \sqrt{\rho_d} \sum_{k=1}^{K} c_{mk} s_k[i] + n_m[i]$. Thus, the time index can be discarded and for each time instant, the $M \times 1$ collective vector of the received signals at time instant i, at the BS can be denoted as:

$$\boldsymbol{y} = \sqrt{\rho_d} \boldsymbol{C} \boldsymbol{s} + \boldsymbol{n} \tag{2.7}$$

where $\boldsymbol{s} = (s_1, s_2, \cdots, s_K)^T$ is the transmitted symbols by all users at time instant i and $\boldsymbol{n} = (n_1, n_2, \cdots, n_M)^T$ is the noise vector with covariance matrix of $\sigma_n^2 \boldsymbol{I_M}$, where $\boldsymbol{I_M}$ is an $M \times M$ identity matrix. Also, $\boldsymbol{C}$ is the channel matrix whose elements are defined as $\boldsymbol{C}(m,k) = c_{mk}$. Note that the index i can be discarded for the perfect synchronized case and the block fading scenario.

2.2.3 Review of Performance Results

We first consider the case that the BS has perfect CSI. Considering the MRC receiver, the BS processes its received signal vector by multiplying it with $\frac{1}{M}\boldsymbol{C}^H$ [29]. Then, the

corresponding sample for the lth user can be denoted as:

$$y_{l,p}^{mrc-ideal} = \frac{1}{M}\sum_{m=1}^{M} c_{ml}^{*} y_m \tag{2.8}$$

The achievable rate of the lth user in the single-cell multiuser distributed MIMO uplink can be approximated as [33, 39]:[1]

$$\tilde{R}_{l,p}^{mrc-ideal} \approx \log_2\left(1 + \frac{\rho_d(M\overline{\beta_l}^2 + \overline{\beta_{ll}})}{\rho_d \sum\limits_{\substack{k=1 \\ k\neq l}}^{K} \overline{\beta_{lk}} + \overline{\beta_l}}\right) \tag{2.9}$$

where $\overline{\beta_k}$ and $\overline{\beta_{lk}}$ are defined as follows:

$$\overline{\beta_k} = \frac{1}{M}\sum_{m=1}^{M} \beta_{mk}, \quad \overline{\beta_{lk}} = \frac{1}{M}\sum_{m=1}^{M} \beta_{ml}\beta_{mk} \tag{2.10}$$

As M increases and approaches infinity, the achievable rate will be infinity. This is the underlying reason for defining the power scaling law. The power scaling law for the massive MIMO setting with perfect CSI can be stated as follows.

Corollary 2.1. *Using MRC receivers with M receive antennas and perfect CSI, cutting the transmit power of each user by a factor of $\frac{1}{M}$, the system still achieves the performance of the SISO system with no interference. In other words, the achievable uplink rate for each user approaches:*

$$R_{l,p}^{mrc-ideal} \to \log_2\left(1 + E_d\overline{\beta_l}^2\right) \tag{2.11}$$

as $M \to \infty$ and $\rho_d = \frac{E_d}{M}$ [31].

[1] Throughout this chapter, the subscripts p and ip are used for perfect CSI and imperfect CSI, respectively.

In real situations, the channel matrix is estimated at the BS. We assume the channel coefficients are estimated by sending known sequences of symbols, called pilot sequences. Each user assigns its first N_p symbols of each frame to send pilot symbols. We denote the assigned pilot sequence to the kth user as $\boldsymbol{p_k} = (p_k[1], \cdots, p_k[N_p])$. It is common in the literature to assume that the assigned pilot sequences for different users are mutually orthogonal, i.e., $\langle \boldsymbol{p_i}.\boldsymbol{p_j} \rangle = \delta[i-j]$, where $\langle \ . \ \rangle$ shows the inner product. In addition, N_p should be equal to or greater than the number of users and its optimal value is shown to be $N_p = K$ [44]. The mutual orthogonality enables all the users to send the pilot symbols simultaneously without interfering with each other. The $K \times N_p$ matrix that contains all the pilot sequences is represented by:

$$\boldsymbol{\Phi} = \begin{pmatrix} p_1[1] & \cdots & p_1[N_p] \\ p_2[1] & \cdots & p_2[N_p] \\ \vdots & \ddots & \vdots \\ p_K[1] & \cdots & p_K[N_p] \end{pmatrix}_{K \times N_p} \tag{2.12}$$

Due to the orthogonality of the rows, the pilot matrix is unitary, i.e., $\boldsymbol{\Phi}\boldsymbol{\Phi}^H = \boldsymbol{I_K}$. In the ideal case of perfect synchronization, the received signal can be written as:

$$\boldsymbol{Y_p} = \sqrt{\rho_p}\boldsymbol{C}\boldsymbol{\Phi} + \boldsymbol{N} \tag{2.13}$$

where $\boldsymbol{Y_p}$ and $\boldsymbol{N}$ are $M \times N_p$ matrices of received samples and noise samples, respectively. Also, ρ_p is the power assigned to the transmission of pilot sequences and is equal to $\rho_p = N_p\rho_d$. The least squares estimate of the channel matrix $\boldsymbol{C}$ can be calculated as [45]:

$$\tilde{\boldsymbol{C}} = \frac{1}{\sqrt{\rho_p}}\boldsymbol{Y_p}\boldsymbol{\Phi}^H \tag{2.14}$$

The corresponding sample for the detection of the lth user's symbol can be denoted as

$y_{l,ip}^{mrc-ideal} = \frac{1}{M}\sum_{m=1}^{M} \tilde{c}_{ml}^{*} y_m$. The achievable rate of the lth user in the single-cell multiuser distributed MIMO uplink with estimated channel coefficients can be approximated as [33, 39]:

$$\tilde{R}_{l,ip}^{mrc-ideal} \approx \kappa \log_2 \left(1 + \frac{N_p \rho_d^2 (M \overline{\beta_l}^2 + \overline{\beta_{ll}})}{N_p \rho_d^2 \sum\limits_{\substack{k=1 \\ k \neq l}}^{K} \overline{\beta_{lk}} + \rho_d \sum\limits_{k=1 k \neq l}^{K} \overline{\beta_k} + \rho_d (N_p + 1) \overline{\beta_l} + 1} \right) \tag{2.15}$$

where $\kappa = \frac{N - N_p}{N}$ accounts for the spectral efficiency loss due to channel estimation. The power scaling law for the massive MIMO setting with imperfect CSI can be stated as follows.

Corollary 2.2. *Using MRC receivers with M receive antennas and imperfect CSI, cutting the transmit power of each user by a factor of $\frac{1}{\sqrt{M}}$, the system still achieves the performance of the SISO system with no interference. In other words, the achievable uplink rate for each user approaches*

$$R_{l,ip}^{mrc-ideal} \to \log_2 \left(1 + N_p E_d^2 \overline{\beta_l}^2\right) \tag{2.16}$$

as $M \to \infty$ and $\rho_d = \frac{E_d}{\sqrt{M}}$ [31].

These results are obtained based on the assumption of having perfect synchronization. However, due to the distributed locations of the received antennas, synchronization at all receive antennas is impossible. Therefore, we modify the system model to the asynchronous scenario and show inability of the conventional MRC receivers to hold the power scaling law. Then, we introduce a simple solution which restores the power scaling law.

2.2.4 Received Signal Model: Timing Mismatch

Due to the distributed locations of users and receive antennas, the transmitted signal from the kth user to the mth receive antenna is received with a relative delay of $d_{mk}T + \tau_{mk}$. Then, the continuous received signal at the mth receive antenna can be represented by:

$$y_m(t) = \sqrt{\rho_d} \sum_{k=1}^{K} c_{mk} s_k(t - d_{mk}T - \tau_{mk}T) + n_m(t) \tag{2.17}$$

We assume that the values of the frame asynchrony, i.e., d_{mk} are known at the receiver [28]. Therefore, without loss of generality, we assume $d_{mk} = 0$ in Eq. (2.17). However, the symbol-level asynchrony, i.e., τ_{mk} is unknown and is treated as a random variable between $(0, 1)$. The reason behind this assumption is that the precision needed for the symbol-level synchronization is much higher than the frame synchronization, which becomes important particularly in the massive MIMO context, where using estimation algorithms with high precision and thus long acquisition time is infeasible. Also, the value of d_{mk} is discrete and easily represent-able by a finite number of bits, while τ_{mk} is continuous and its exact representation needs infinite number of feedback bits which is impossible. Therefore, always, there will be some residual error that can be modeled as a random variable τ_{mk}. Then, the output of the matched filter can be written as follows:

$$\hat{y}_m(t) = \sqrt{\rho_d} \sum_{k=1}^{K} \sqrt{\beta_{mk}} h_{mk} \sum_{n=1}^{N} s_k[n] g(t - nT - \tau_{mk}T) + n_m(t) * p(t) \tag{2.18}$$

We sample the output of the matched filter at the instants of $t_i^s = iT, i = 1, \cdots, N$. The obtained samples at the sampler of the mth receive antenna, denoted by $y_m[i] = \hat{y}_m(t)|_{t=t_i^s}$, can be written as:

$$y_m[i] = \sqrt{\rho_d} \sum_{k=1}^{K} \sqrt{\beta_{mk}} h_{mk} \sum_{n=1}^{N} s_k[i] g((i-n)T - \tau_{mk}T) + \hat{n}_m(t)\big|_{t=iT} \tag{2.19}$$

where $\hat{n}_m(t) = n_m(t) * p(t)$. We can put the obtained samples together and form the system model equation as follows:

$$\boldsymbol{y_m} = \sqrt{\rho_d} \sum_{k=1}^{K} \sqrt{\beta_{mk}} h_{mk} \boldsymbol{G_{mk}} \boldsymbol{s_k} + \boldsymbol{n_m} \tag{2.20}$$

where $\boldsymbol{s_k} = (s_k[1], s_k[2], \cdots, s_k[N])^T$ is the transmitted frame by the kth user and $\boldsymbol{n_m} = (n_m[1], n_m[2], \cdots, n_m[N])^T$ is the noise vector containing samples of $\hat{n}_m(t)$, i.e., $n_m[i] = \hat{n}_m(t)|_{t=iT}, 1 \leq i \leq N$. Also, $\boldsymbol{G_{mk}}$ is an $N \times N$ matrix defined as:

$$\boldsymbol{G_{mk}} = \begin{pmatrix} g(-\tau_{mk}T) & g(-T-\tau_{mk}T) & \cdots & g((1-N)T_s-\tau_{mk}T) \\ g(T-\tau_{mk}T) & g(-\tau_{mk}T) & \cdots & g((2-N)T-\tau_{mk}T) \\ \vdots & \ddots & \ddots & \vdots \\ g((N-1)T_s-\tau_{mk}T) & \cdots & g(T-\tau_{mk}T) & g(-\tau_{mk}T) \end{pmatrix} \tag{2.21}$$

Note that for the synchronous case, matrix $\boldsymbol{G_{mk}}$ turns into an identity matrix and the system model simplifies to Eq. (2.7), because there is no ISI in the systems. However, for the asynchronous scenario, the off-diagonal elements are nonzero due to the timing offsets. Defining $\boldsymbol{T_{mk}} = \sqrt{\beta_{mk}} h_{mk} \boldsymbol{G_{mk}}$, Eq. (2.20) can be written in the following short form:

$$\boldsymbol{y_m} = \sqrt{\rho_d} \sum_{k=1}^{K} \boldsymbol{T_{mk}} \boldsymbol{s_k} + \boldsymbol{n_m} \tag{2.22}$$

Denoting $\tilde{c}_{ml}$ as the estimate of the channel coefficient between the lth user and the mth receive antenna, the MRC output vector for the detection of the lth user's symbols, i.e., $\boldsymbol{y_l^{mrc}} = \frac{1}{M} \sum_{m=1}^{M} \tilde{c}_{ml}^* \boldsymbol{y_m}$, can be expressed as:

$$\boldsymbol{y_l^{mrc}} = \sqrt{\rho_d} \sum_{k=1}^{K} \boldsymbol{T_{lk}^{mrc}} \boldsymbol{b_k} + \boldsymbol{n_l^{mrc}} \tag{2.23}$$

where the effective channel matrices and the resulting noise vector are denoted by $\boldsymbol{T_{lk}^{mrc}}$ and $\boldsymbol{n_l^{mrc}}$, respectively, and will be defined later based on the available CSI and detection

methods. In the next section, we analyze the performance of the MRC detection with timing mismatch considering perfect or estimated CSI at the receiver.

2.3 The Achievable Rate of MRC Receiver with Random Timing Mismatch

2.3.1 Perfect CSI

By assuming perfect CSI, i.e., $\tilde{c}_{mk} = c_{mk}$, the effective channel matrix, i.e., $\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc}}$ and effective noise vector, i.e., $\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}}$ can be represented as follows

$$\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc}} = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\beta_{ml}\beta_{mk}}h_{ml}^{*}h_{mk}\boldsymbol{G}_{\boldsymbol{mk}}, \tag{2.24}$$

$$\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}} = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\beta_{ml}}h_{ml}^{*}\boldsymbol{n}_{\boldsymbol{m}} \tag{2.25}$$

The achievable rate for the corresponding system model is found in the next theorem.

Theorem 2.1. *The achievable rate of the MRC receiver for User l, when there is random time delays between received signals, can be approximated as:*

$$\tilde{R}_{l,p}^{mrc} \approx \log_2\left(1+\frac{\rho_d M\overline{\beta_l}^2 g'[0]}{\rho_d\left(g''\sum_{k=1}^{K}\overline{\beta_{lk}}+(g''-g')\overline{\beta_{ll}}\right)+\rho_d M\overline{\beta_l}^2(g'-g'[0])+\overline{\beta_l}}\right) \tag{2.26}$$

where $g'[i]$ and $g''[i]$ are defined based on the pulse shape and delay distributions as follows:

$$g'[i] = E_{f(\tau)}^{2}\{g(iT-\tau_{mk})\}$$

$$g''[i] = E_{f(\tau)}\{g^2(iT - \tau_{mk})\} \tag{2.27}$$

where $f(\tau)$ is the distribution of time delays. Assuming the same distribution for all time delays, the receive antenna and user indexes are discarded after taking expectations. g' and g'' are defined as $g' = \sum_{i=-I}^{I} g'[i]$ and $g'' = \sum_{i=-I}^{I} g''[i]$, where I is the number of significant side lobes of the pulse shape.

Proof. The proof is presented in Appendix A.1. □

By ignoring the inevitable timing mismatch, the promised benefit of power scaling law in a massive MIMO setting vanishes as described in the next corollary.

Corollary 2.3. *In the presence of random timing mismatches, the power scaling law of the large scale MRC receiver is degraded by ISI. In more details, if we put $\rho_d = \frac{E_d}{M}$ in Eq. (2.26) and let M go to infinity, we will have:*

$$R_{l,p}^{mrc} \to \log_2\left(1 + \frac{E_d\overline{\beta_l}g'[0]}{E_d\overline{\beta_l}(g' - g'[0]) + 1}\right) \tag{2.28}$$

The achievable rate in Eq. (2.28) is limited by ISI, and by increasing the transmit power it will be saturated to a constant value, i.e.:

$$R_{l,p}^{mrc} \to \log_2\left(1 + \frac{g'[0]}{g' - g'[0]}\right) \tag{2.29}$$

Therefore, at high SNR regime, no matter how much transmit power is used, the achievable rate converges to a fixed value independent of the transmit power. This fixed value depends on the delay distribution and the pulse shape. However, by optimizing the sampling instants, the asymptotic values can be maximized.

2.3.2 Imperfect CSI

In what follows, we provide a similar analysis for the channel estimation when the misalignment exists between received signals. Time delays modify Eq. (2.13) to:

$$\boldsymbol{Y_p} = \sqrt{\rho_p} \sum_{i=-I}^{I} \boldsymbol{C^i \Phi^i} + \boldsymbol{N} \tag{2.30}$$

where $\boldsymbol{C^i}_{M\times K}$ and $\boldsymbol{\Phi^i}_{K\times N_p}$ are defined as follows:

$$\boldsymbol{C^i} = \begin{pmatrix} g(iT-\tau_{11}T)\sqrt{\beta_{11}}h_{11} & \cdots & g(iT-\tau_{1K}T)\sqrt{\beta_{1K}}h_{1K} \\ g(iT-\tau_{21}T)\sqrt{\beta_{21}}h_{21} & \cdots & g(iT-\tau_{2K}T)\sqrt{\beta_{2K}}h_{2K} \\ \vdots & \ddots & \vdots \\ g(iT-\tau_{M1}T)\sqrt{\beta_{M1}}h_{M1} & \cdots & g(iT-\tau_{MK}T)\sqrt{\beta_{MK}}h_{MK} \end{pmatrix} \tag{2.31}$$

$$\boldsymbol{\Phi^{i\leq 0}} = \begin{pmatrix} p_1[1-i] & \cdots & p_1[N_p] & 0 & \cdots & 0 \\ p_2[1-i] & \cdots & p_2[N_p] & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & & \\ p_K[1-i] & \cdots & p_K[N_p] & 0 & \cdots & 0 \end{pmatrix}, \tag{2.32}$$

$$\boldsymbol{\Phi^{i\geq 0}} = \begin{pmatrix} 0 & \cdots & 0 & p_1[1] & \cdots & p_1[N_p-i] \\ 0 & \cdots & 0 & p_2[1] & \cdots & p_2[N_p-i] \\ \vdots & \ddots & \vdots & & & \\ 0 & \cdots & 0 & p_K[1] & \cdots & p_K[N_p-i] \end{pmatrix} \tag{2.33}$$

The process of de-spreading, which is multiplying the received pilot signal by $\frac{1}{\sqrt{\rho_p}}\boldsymbol{\Phi}^H$, yields the following channel matrix estimator:

$$\tilde{\boldsymbol{C}} = \sum_{i=-I}^{I} \boldsymbol{C^i}\boldsymbol{\Phi^i}\boldsymbol{\Phi}^H + \tilde{\boldsymbol{N}} \tag{2.34}$$

where $\tilde{\boldsymbol{N}}$ is the estimation noise. We denote $\boldsymbol{\Phi^i}\boldsymbol{\Phi}^H$ by $\boldsymbol{\Upsilon^i}$ which is equal to $\boldsymbol{I_K}$ for $i = 0$, and for the other values of i can be calculated as:

$$(\boldsymbol{\Upsilon^{i<0}})^T = \boldsymbol{\Upsilon^{i>0}} = \begin{pmatrix} \langle\ \boldsymbol{p_1}[1{:}N_p{-}i].\boldsymbol{p_1}[1{+}i{:}N_p]\rangle & \cdots & \langle\ \boldsymbol{p_1}[1{:}N_p{-}i].\boldsymbol{p_K}[1{+}i{:}N_p]\rangle \\ \vdots & \ddots & \vdots \\ \langle\ \boldsymbol{p_K}[1{:}N_p{-}i].\boldsymbol{p_1}[1{+}i{:}N_p]\rangle & \cdots & \langle\ \boldsymbol{p_K}[1{:}N_p{-}i].\boldsymbol{p_K}[1{+}i{:}N_p]\rangle \end{pmatrix} \tag{2.35}$$

where $\boldsymbol{p}[i:j]$ represents the vector $(p[i], p[i+1], \cdots, p[j])$. After some calculations, the estimate of the User l's channel coefficient to the mth receive antenna can be represented as:

$$\tilde{c}_{ml} = \sum_{j=1}^{K} \lambda_{lmj} c_{mj} + \tilde{n}_{ml} \tag{2.36}$$

where λ_{lmj} is the leakage from User j to the estimation of the User l's channel coefficient to receive antenna m and is equal to:

$$\lambda_{lmj} = \sum_{i=-I}^{I} g(iT - \tau_{mj}T)\boldsymbol{\Upsilon^i}(j,l) \tag{2.37}$$

This phenomenon is similar to the "pilot contamination" effect, i.e., the channel estimation of each user to the mth receive antenna is contaminated by the channel coefficients of other users. In the "pilot contamination" scenario, the reason behind contamination is reusing the same pilot sequences in different cells, however, here, the reason is random timing mismatches between the received signals. Due to the time asynchrony between the received signals, the orthogonality between pilot sequences is not preserved anymore

and the de-spreading matrix is not able to eliminate the effect of interfering users. The effective channel matrix and noise vector after MRC can be expressed, respectively, as:

$$\boldsymbol{T}_{\boldsymbol{lk,ip}}^{\boldsymbol{mrc}} = \frac{1}{M}\sum_{m=1}^{M}\sum_{j=1}^{K}\lambda_{lmj}\sqrt{\beta_{mj}\beta_{mk}}h_{mj}^{*}h_{mk}\boldsymbol{G}_{\boldsymbol{mk}} \tag{2.38}$$

$$\boldsymbol{n}_{\boldsymbol{l,ip}}^{\boldsymbol{mrc}} = \frac{\sqrt{\rho_d}}{M}\sum_{m=1}^{M}\tilde{n}_{ml}\sum_{k=1}^{K}\sqrt{\beta_{mk}}h_{mk}\boldsymbol{G}_{\boldsymbol{mk}}\boldsymbol{b}_{\boldsymbol{k}} + \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{j=1}^{K}\lambda_{lmj}\sqrt{\beta_{mj}}h_{mj}^{*} + \tilde{n}_{ml}\right)\boldsymbol{n}_{\boldsymbol{m}} \tag{2.39}$$

The corresponding achievable rate results are presented in the next theorem.

Theorem 2.2. *The achievable rate by the MRC receiver using orthogonal channel estimation, when there is random time delays between received signals can be approximated as follows:*

$$R_{l,ip}^{mrc} \approx \kappa\log_2\left(1 + \frac{N_p\rho_d^2\overline{\beta_l}^2 M\gamma'_{lll}[0]}{\sum_{k=1}^{K}\Delta_k - N_p\rho_d^2\overline{\beta_l}^2 M\gamma'_{lll}[0] + 1}\right) \tag{2.40}$$

where $\kappa = \frac{N-N_p}{N}$ *accounts for the spectral efficiency loss due to channel estimation, and* Δ_k *is defined as follows:*

$$\Delta_k = N_p\rho_d^2\left(\sum_{j=1}^{K}\overline{\beta_{jk}}\gamma''_{ljk} + \overline{\beta_{kk}}(\gamma''_{lkk} - \gamma'_{lkk})\right) + N_p\rho_d^2 M\overline{\beta_k}^2\gamma'_{lkk} + \rho_d(g'' + N_p\lambda''_{lk})\overline{\beta_k}$$

where $\gamma'_{ljk}[i]$,$\gamma''_{ljk}[i]$ *and* λ''_{lk} *are defined based on the pulse shape, delay distributions and pilot sequences by*

$$\gamma'_{ljk}[i] = E^2_{f(\tau_j,\tau_k)}[\lambda_{lmj}g(iT - \tau_{mk})] \tag{2.41}$$

$$\gamma''_{ljk}[i] = E_{f(\tau_j,\tau_k)}[\lambda^2_{lmj}g^2(iT - \tau_{mk})] \tag{2.42}$$

$$\lambda''_{lk} = E_{f(\tau)}[\lambda^2_{lmk}] \tag{2.43}$$

Assuming the same distribution for all time delays, the receive antenna index is discarded after taking expectations. γ'_{ljk} *and* γ''_{ljk} *are denoted as* $\gamma'_{ljk} = \sum_{i=-I}^{I} \gamma'_{ljk}[i]$ *and* $\gamma''_{ljk} = \sum_{i=-I}^{I} \gamma''_{ljk}[i]$*, where* I *is the number of significant side lobes of the pulse shape.*

Proof. The proof is presented in Appendix A.2. □

These results are general for any pilot matrices and delay distributions. The values of $\gamma'_{ljk}[i], \gamma''_{ljk}[i]$ and λ''_{lk} only depend on the pulse shape, pilot sequences and the delay distribution which can be calculated analytically or numerically. Thus, we can insert any delay distribution, any pilot matrix and any value of K in the above formulas and calculate the achievable rates. Due to the existence of timing mismatch, the promised power scaling law is lost which is presented in the next corollary.

Corollary 2.4. *If we reduce the transmit power by an order of* $\frac{1}{\sqrt{M}}$ *and let* M *go to infinity, then the achievable rate for each user becomes:*

$$R_{l,ip}^{mrc} \to \kappa \log_2 \left(1 + \frac{N_p E_d^2 \overline{\beta_l}^2 \gamma'_{lll}[0]}{N_p E_d^2 \left(\sum_{k=1}^{K} \overline{\beta_k}^2 \gamma'_{lkk} - \overline{\beta_l}^2 \gamma'_{lll}[0] \right) + 1} \right) \tag{2.44}$$

By increasing the transmit power, the achievable rate saturates at the following fix value:

$$R_{l,ip}^{mrc} \to \kappa \log_2 \left(1 + \frac{\overline{\beta_l}^2 \gamma'_{lll}[0]}{\sum_{k=1}^{K} \overline{\beta_k}^2 \gamma'_{lkk} - \overline{\beta_l}^2 \gamma'_{lll}[0]} \right) \tag{2.45}$$

For any given pulse shape and pilot sequence, the performance in Eq. (2.45) can be maximized by optimizing the sampling instants.

While such an optimization can increase the achievable rates in Theorems 1 and 2, the

main benefit of massive MIMO setting which is having unlimited achievable rate by using asymptotically large number of receive antennas will still be out of reach in the presence of random time delays. Therefore, we design two receiver structures, for perfect CSI and imperfect CSI scenarios, to remove the unwanted effects of ISI and IUI imposed by random time delays. The details are presented next.

2.4 The Achievable Rate of MRC-ZF Receiver

In an ideal distributed massive MIMO system, the achievable uplink rate grows unbounded when M grows large. Therefore, we can scale down the power of each user by the ratios of M and $\sqrt{M}$ for the perfect and imperfect CSI, respectively, to achieve the single-user performance with no interference. However, in a realistic distributed massive MIMO system where timing mismatch among received signals is inevitable, the uplink achievable rate using the MRC receiver approaches a constant value when M grows large, as shown in Theorems 2.1 and 2.2. Therefore, the power scaling law, provided by the MRC receiver in a perfectly synchronized massive MIMO system, is not achievable with random time delays. To understand the underlying reason, we express the asymptotic values of $y_{l,p/ip}^{mrc-ideal}$ and $\boldsymbol{y_{l,p/ip}^{mrc}}$, as $M \to \infty$ in the next lemma.

Lemma 2.1. *By using the law of large numbers, the asymptotic values of* $y_{l,p/ip}^{mrc-ideal}$ *and* $\boldsymbol{y_{l,p/ip}^{mrc}}$ *approaches:*

$$y_{l,p/ip}^{mrc-ideal} \to \sqrt{\rho_d}\overline{\beta_l}s_l, \quad \boldsymbol{y_{l,p}^{mrc}} \to \sqrt{\rho_d}\overline{\beta_l}\boldsymbol{Z}\boldsymbol{s_l}, \quad \boldsymbol{y_{l,ip}^{mrc}} \to \sqrt{\rho_d}\sum_{k=1}^{K}\overline{\beta_k}\boldsymbol{\Gamma_{lk}}\boldsymbol{s_k} \tag{2.46}$$

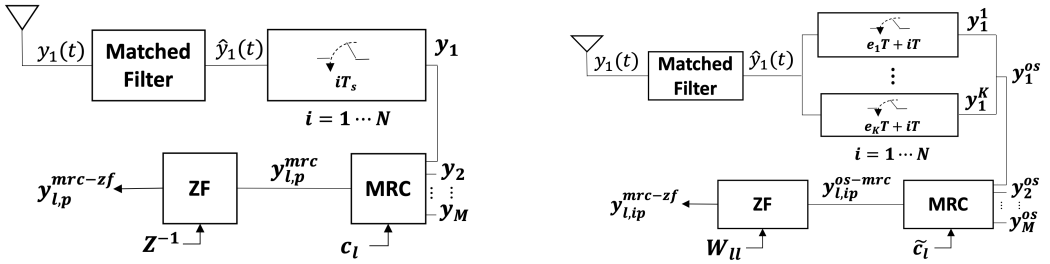
, as $M \to \infty$. $\boldsymbol{Z}$ and $\boldsymbol{\Gamma_{lk}}$ are $N \times N$ matrices defined as follows:

$$\boldsymbol{Z} = \begin{pmatrix} g'[0] & \cdots & g'[-I] & 0 & \cdots & 0 \\ \vdots & g'[0] & \ddots & \ddots & \ddots & \vdots \\ g'[I] & \ddots & g'[0] & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & g'[-I] \\ \vdots & \ddots & \ddots & \ddots & g'[0] & \vdots \\ 0 & \cdots & 0 & g'[I] & \cdots & g'[0] \end{pmatrix}^{\odot \frac{1}{2}} \tag{2.47}$$

$$\boldsymbol{\Gamma_{lk}} = \begin{pmatrix} \gamma'_{lkk}[0] & \cdots & \gamma'_{lkk}[-I] & 0 & \cdots & 0 \\ \vdots & \gamma'_{lkk}[0] & \ddots & \ddots & \ddots & \vdots \\ \gamma'_{lkk}[I] & \ddots & \gamma'_{lkk}[0] & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \gamma'_{lkk}[-I] \\ \vdots & \ddots & \ddots & \ddots & \gamma'_{lkk}[0] & \vdots \\ 0 & \cdots & 0 & \gamma'_{lkk}[I] & \cdots & \gamma'_{lkk}[0] \end{pmatrix}^{\odot \frac{1}{2}} \tag{2.48}$$

where $\boldsymbol{M}^{\odot r}$ is the element-wise rth-exponent of matrix $\boldsymbol{M}$. The results in Lemma 2.1 are immediately derived by applying the law of large numbers. As you can see, in the perfect synchronized scenario, because of the law of large numbers and also orthogonality between channel vectors, the inter-user interference is vanished and infinite rate can be achieved as M goes to infinity. However, when unknown time delays exist between received signals, particularly in the distributed scenario, ISI and IUI will degrade the performance. In the perfect CSI scenario, the main impairment is ISI, indicated by the nonzero off-diagonal elements of matrix $\boldsymbol{Z}$. In the imperfect CSI scenario, due to loss of orthogonality in the estimation process and the "contamination" effect, not only ISI but also IUI, indicated by nonzero matrices of $\boldsymbol{\Gamma_{lk}}$, will degrade the performance.

To cancel the effect of these impairments, we propose to implement some post-processings. In this work, we use linear post-processing and, in particular, zero forcing (ZF) processing, however, other methods like minimum mean squared error (MMSE) and successive interference cancellation (SIC) can also be used. The summary of the proposed structures are explained next.



(a) MRC-ZF Receiver with perfect CSI (b) MRC-ZF Receiver with imperfect CSI

Figure 2.3: MRC-ZF Receiver

2.4.1 MRC-ZF structures

- Perfect CSI: For the perfect CSI scenario, where ISI is the the main source of degradation, its effect can be canceled by multiplying the output sample of the MRC receiver by the inverse of matrix $\boldsymbol{Z}_{N\times N}$.

- Imperfect CSI: For the imperfect CSI scenario, where both ISI and IUI degrades the performance, we need extra sets of equations to cancel all the interference terms. Therefore, we utilize the concept of oversampling as explained in [46, 47, 48]. After performing oversampling and MRC, the vector of samples is multiplied by the proper processing matrix, denoted as $\boldsymbol{W}_{ll\,N\times NK}$.

The receiver structure for the proposed methods are shown in Figs. 2.3a and 2.3b. The detailed description and calculation of matrix $\boldsymbol{W}_{ll}$ is provided in Appendix A.3. Next, the achievable rates by MRC-ZF structures are analyzed.

2.4.2 Perfect CSI

As shown in Fig. 2.3a, the output samples of the MRC receiver are multiplied by $\boldsymbol{Z}^{-1}$ to cancel the effect of the averaged ISI. Note that matrix $\boldsymbol{Z}$ is pre-calculated once based on

the pulse shape, sampling origin and delay distributions and then, it can be used during the entire transmission. We call this receiver MRC-ZF whose output samples are:

$$\boldsymbol{y}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}} = \boldsymbol{Z}^{-1}\boldsymbol{y}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}} = \sqrt{\rho_d}\sum_{k=1}^{K}\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc-zf}}\boldsymbol{s_k} + \boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}} \tag{2.49}$$

where $\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc-zf}} = \boldsymbol{Z}^{-1}\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc}}$ and $\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}} = \boldsymbol{Z}^{-1}\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}}$. For the special case of symbol-level synchronization, i.e., $f(\tau) = \delta(\tau)$, $\boldsymbol{Z}$ will be the identity matrix; meaning that no additional processing is required. To show the effectiveness of the proposed processing, it can be easily shown that when $M \to \infty$ the asymptotic value of $\boldsymbol{y}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}}$ approaches $\sqrt{\rho_d}\overline{\beta_l}\boldsymbol{I_N}\boldsymbol{s_l}$. Thus, the effect of ISI is eliminated as M goes to infinity. The approximation of the achievable rate by the MRC-ZF receiver is presented in the next theorem.

Theorem 2.3. *The achievable rate by each user using the MRC-ZF receiver can be approximated by:*

$$R_{l,p}^{mrc-zf} \approx \log_2\left(1 + \frac{M\rho_d\overline{\beta_l}^2}{\rho_d\hat{g}''\sum_{k=1}^{K}\overline{\beta_{lk}} + \rho_d\left(\hat{g}''-1\right)\overline{\beta_{ll}} + \epsilon_0\overline{\beta_l}}\right) \tag{2.50}$$

where $\hat{g}''$ depends on the distribution of time delays and is defined in Appendix A.4. The value of ϵ_0 is the noise enhancement factor due to ZF processing and is equal to the diagonal element of matrix $\left(\boldsymbol{Z}^H\boldsymbol{Z}\right)^{-1}$. These are only functions of the distribution of delays and the pulse shape.

Proof. The proof is presented in Appendix A.4. □

By using the MRC-ZF receiver which exploits the statistics of the time delays, the effect of the averaged ISI is vanished. If the number of receive antennas goes to infinity, the achievable rate goes to infinity. Therefore, the power scaling law is restored.

Corollary 2.5. *In the presence of random time delays, using MRC-ZF receivers with M receive antennas and perfect CSI, cutting the transmit power of each user by a factor of $\frac{1}{M}$, the system still achieves the performance of the SISO system with no interference. In other words, the achievable uplink rate becomes*

$$R_{l,p}^{mrc-zf} \to \log_2\left(1+\frac{E_d\overline{\beta_l}}{\epsilon_0}\right) \tag{2.51}$$

as $M \to \infty$ and $\rho_d = \frac{E_d}{M}$.

Hence, even in the presence of random time delays, the power scaling law is held for the MRC-ZF receiver.

2.4.3 Imperfect CSI

As shown in Fig. 2.3b, the output samples of the MRC receiver are multiplied by $\boldsymbol{W_{ll}}$ to cancel the effect of the averaged ISI and IUI. Note that matrix $\boldsymbol{W_{ll}}$ is pre-calculated once based on the pulse shape, sampling origins and delay distributions and then, it can be used during the entire transmission. We call this receiver MRC-ZF whose output samples are:

$$\boldsymbol{y_{l,ip}^{mrc-zf}} = \boldsymbol{W_{ll}}\boldsymbol{y_{l,ip}^{os-mrc}} = \sqrt{\rho_d}\sum_{k=1}^{K}\boldsymbol{T_{lk,ip}^{mrc-zf}}\boldsymbol{s_k} + \boldsymbol{n_{l,ip}^{mrc-zf}} \tag{2.52}$$

where $\boldsymbol{T_{lk,ip}^{mrc-zf}} = \boldsymbol{W_{ll}}\hat{\boldsymbol{T}}_{\boldsymbol{lk}}$, $\boldsymbol{n_{l,ip}^{mrc-zf}} = \boldsymbol{W_{ll}}\boldsymbol{n_{l,ip}^{os-mrc}}$ and $\hat{\boldsymbol{T}}_{\boldsymbol{lk}} = \left((\boldsymbol{T_{lk}^{1}})^T,\ldots,(\boldsymbol{T_{lk}^{K}})^T\right)^T$. It can be easily shown that when $M \to \infty$ the asymptotic value of $\boldsymbol{y_{l,ip}^{mrc-zf}}$ approaches $\sqrt{\rho_d}\overline{\beta_l}\boldsymbol{I_N}\boldsymbol{s_l}$. Thus, the effects of ISI and IUI are eliminated as M goes to infinity. The achievable rates for the aforementioned system is presented in the next theorem.

Theorem 2.4. *The achievable rate by the MRC-ZF receiver using estimated channel coef-*

ficients, when there is random time delays between the received signals can be approximated by:

$$R_{l,ip}^{mrc-zf} \approx \kappa \log_2 \left(1 + \frac{N_p \rho_d^2 \overline{\beta_l}^2 M}{\sum_{k=1}^{K} \Delta_k - N_p \rho_d^2 \overline{\beta_{ll}} + v_{l0}} \right) \tag{2.53}$$

where Δ_k *is defined as:*

$$\Delta_k = N_p \rho_d^2 \left(\sum_{j=1}^{K} \overline{\beta_{jk}} \hat{\gamma}''_{ljk} + \overline{\beta_{kk}} \hat{\gamma}''_{lkk} \right) + \rho_d (u_{l0} + v_{l0} N_p \lambda''_{lk}) \overline{\beta_k}$$

$\hat{\gamma}''_{ljk}$ *depends on the distribution of time delays and is defined in Appendix A.5. The values of* u_{l0} *and* v_{l0} *are the noise enhancement factors due to ZF processing and depends on the matrix* $\boldsymbol{W_u}$.

Proof. The proof is presented in Appendix A.5. □

If the number of receive antennas goes to infinity, the achievable rate goes to infinity. Thus, by using the MRC-ZF receiver which exploits the statistics of time delays, the effect of averaged ISI and IUI is vanished and the power scaling law can be obtained.

Corollary 2.6. *In the presence of random time delays, using MRC-ZF receivers with* M *receive antennas and imperfect CSI, cutting the transmit power of each user by a factor of* $\frac{1}{\sqrt{M}}$*, the system still achieves the performance of the SISO system with no interference. In other words, the achievable uplink rate becomes*

$$R_{l,ip}^{mrc-zf} \to \kappa \log_2 \left(1 + \frac{N_p E_d^2 \overline{\beta_l}}{v_{l0}} \right) \tag{2.54}$$

as $M \to \infty$ *and* $\rho_d = \frac{E_d}{\sqrt{M}}$.

Hence, even in the presence of random time delays, the power scaling law is held by using

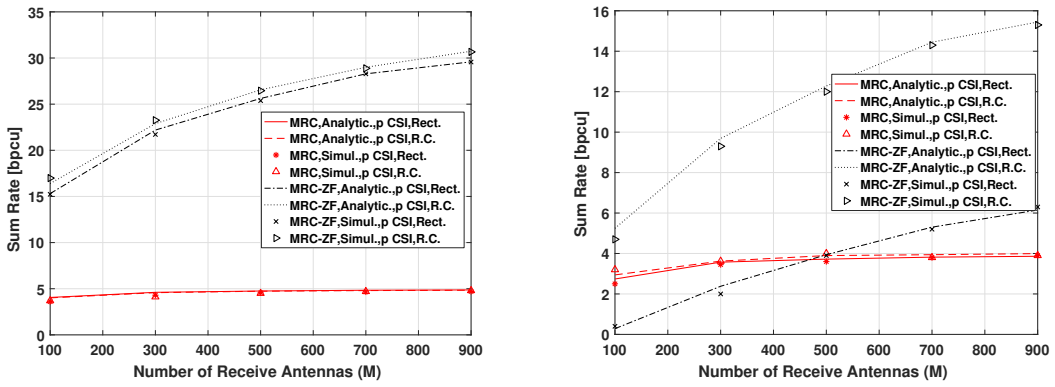
the MRC-ZF receiver. In all schemes presented in this section, the noise enhancement by ZF can be mitigated by using other cancellation methods like MMSE and SIC.

2.5 Numerical Results and Discussions

In this section, we present and discuss simulations results, compared with numerical ones pertaining to the analysis developed previously. The simulation experiments, of Monte Carlo type, are performed in the platform Matlab. The system considered consists of a single cell with a distributed BS having M antennas. The number of users served by the BS is $K = 5$. The small-scale fading coefficients follow an i.i.d. Rayleigh block-fading channel model drawn from an i.i.d. complex Gaussian distribution, i.e., $h_{mk} \sim CN(0, 1)$. To generate path loss coefficients, a cell with a radius of $R = 1000$ is considered and r_{mk}s are distributed randomly and uniformly over the cell, with the exclusion of a central disk of radius $r_h = 100$, and the path loss exponent is assumed to be equal to 2. Time delays follow the uniform distribution and noise samples are also distributed as $CN(0, 1)$. The MC simulations are conducted for $10,000$ independent channel, time delay and noise realizations.

In Fig. 2.4, the performance of the MRC and MRC-ZF receivers with perfect CSI and imperfect CSI is presented by theoretical approximation in Theorems 2.1, 2.2, 2.3 and 2.4 and via simulation. The sum rate for 5 users are plotted with respect to the number of receive antennas. The results include Rectangular (Rect.) pulse shape and Raised Cosine (R.C.) pulse shape with roll-off factor of $\beta = 0.5$ truncated at 3 side lobes. Our analytical approximation and simulation results match. It also shows that, unknown time delays limit the performance of MRC receivers, and by increasing M, the sum rate is saturated for both perfect and imperfect scenarios. However, the performance of the

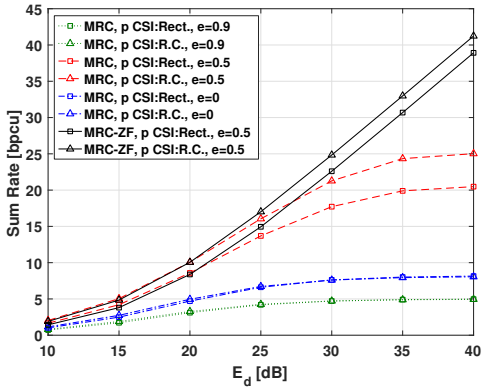
MRC-ZF receiver is not saturated and by increasing M, the sum rate increases. For MRC receivers, the results are the same for both Rect. and R.C. pulse shapes, however, for MRC-ZF structures, the pulse shape has a significant effect, particularly, in imperfect CSI scenario. In the perfect scenario case, the effect of noise-interference enhancement induced by matrix $\boldsymbol{Z}$ is roughly the same for both Rect. and R.C. pulse shapes. On the other hand, in the imperfect scenario case, the choice of pulse shape has a great effect on the condition number of matrix $\boldsymbol{\Gamma}$. Using Rect. pulse shape results in huge noise-interference enhancement by $\boldsymbol{W}_{ll}$, while for more practical pulse shape of R.C., the noise-interference enhancement is much less. In Fig. 2.5, the power scaling law is depicted



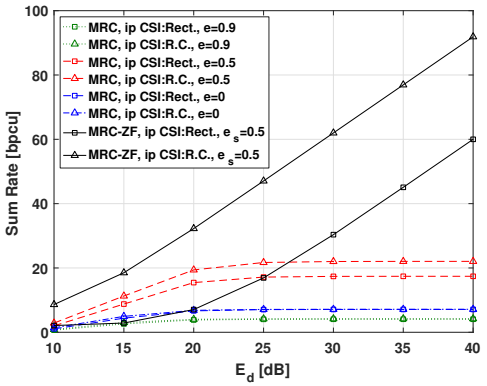
(a) perfect CSI (b) imperfect CSI

Figure 2.4: Performance of the MRC and MRC-ZF receivers with respect to the number of receive antennas, using perfect/imperfect CSI and $\rho_d = 10$ dB

for perfect and imperfect scenarios. In Fig. 2.5a, the performance of the MRC and the MRC-ZF receivers are presented using perfect CSI while $\rho_d = \frac{E_d}{M}$ and in Fig. 2.5b, a similar analysis is presented for imperfect CSI with $\rho_d = \frac{E_d}{\sqrt{M}}$. For MRC receivers, by increasing E_d, the sum rate is saturating as also shown by the analytical results. On the other hand, by using the MRC-ZF receiver, the achievable sum rate increases by E_d which verifies that the power-scaling law can be achieved by MRC-ZF receivers. Fig. 2.5 also shows the effect of e, which is the sampling adjustment variable, on the performance of MRC receivers.



(a) perfect CSI, $\rho_d = \frac{E_d}{M}$ and $M \to \infty$



(b) imperfect CSI, $\rho_d = \frac{E_d}{\sqrt{M}}$ and $M \to \infty$

Figure 2.5: Analysis of Power Scaling Law for the MRC and the MRC-ZF receiver

2.6 Conclusion

In this chapter, we obtained the general formula for the achievable rate of the MRC receiver when random timing mismatch exists. We showed that unknown time delays degrade the performance substantially. In other words, in the presence of random timing mismatch, the achievable rate by each user is interference limited, and the power scaling law is hindered. Therefore, intelligent system design is required to manage the inherent time delays in a multiuser network, which is discussed in the next chapters.

Chapter 3

Oversampling Explained: The General System Model for an Asynchronous Transceiver and Its Properties

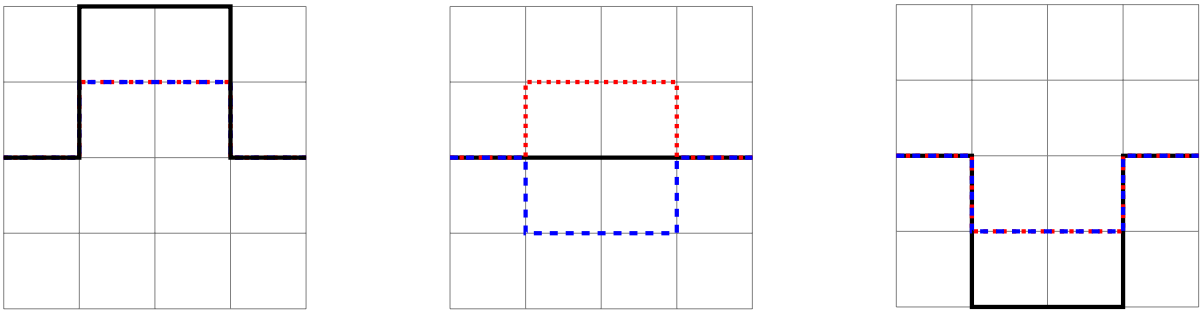
In this chapter, we try to explore the possible advantages of asynchronous transmission and its underlying reasons. In Section 3.1, the notion of time asynchrony and concepts behind its possible advantages are explained. In Section 3.2, we introduce the concept of oversampling and present sufficient statistics to detect the transmitted symbols in an asynchronous scenario. Finally, in section 3.3, the critical features of the resulting system model are provided.

3.1 The Concept Behind Advantages of Asynchrony

In this section, we try to provide some insights into possible benefits of asynchronous transmission. First, we provide a toy example to demonstrate the advantages of asynchronous transmission. Assume that we have two single-antenna users, namely, U_1 and U_2, trying to transmit their corresponding symbols from alphabet $\{-1, 1\}$ to a common single-antenna receiver at a same time and same frequency. Assuming AWGN channel, the received signal at the receiver can be denoted as:

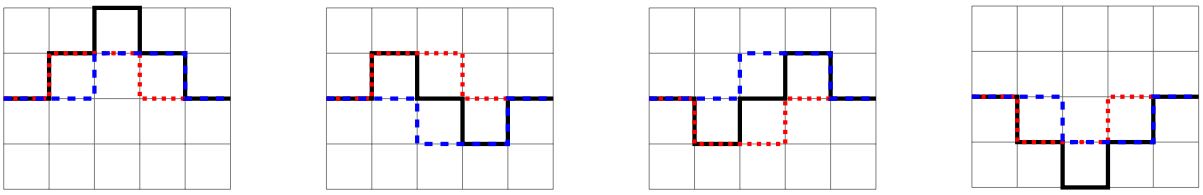
$$y = s_1 + s_2 + n. \tag{3.1}$$

Considering Rectangular pulse shape and perfectly synchronized scenario, the possible outcomes for different permutations of input symbols are shown in Fig. 3.1. As it can be seen, no matter how small the additive noise is, it is impossible to distinguish between transmission of $\{1, -1\}$ and $\{-1, 1\}$. As a consequence of IUI, we are unable to decode the transmitted symbols in synchronous fashion. However, assume that either we add intentional time delay to transmission of one of the users or channel itself impose different delays to the transmitted signals. Then, the resulting combination of transmitted symbols can be one of the options in Fig. 3.2 .



(a) $\{1, 1\}$ (b) $\{-1, 1\}, \{1, -1\}$ (c) $\{-1, -1\}$

Figure 3.1: Possible outcomes for synchronous transmission



(a) $\{1, 1\}$ (b) $\{1, -1\}$ (c) $\{-1, 1\}$ (d) $\{-1, -1\}$

Figure 3.2: Possible outcomes for asynchronous transmission

Thanks to the timing offset between the two received signals, all possible combinations of the transmitted symbols are distinguishable and, hence, decodable. However, if we follow conventional sampling methods like sampling the middle point or calculate the integral over the range of transmission, then the benefits of asynchrony will be lost and the performance might even be worse. Thus, the proper sampling method is of great importance in order to take advantage of asynchronous transmission. Although in this simple example, there is 150% increase in the time interval which results in the reduced rate, by using sufficiently long frames, the loss due to adding time delays will be negligible.

The metric to numerically express "distinguishability" is the minimum distance between the resulting signals. The distance between two continuous and integrable signals, e.g., $f(t)$ and $g(t)$, is defined as:

$$D(f, g) = \int_{-\infty}^{\infty} |f(t) - g(t)|^2 dt \tag{3.2}$$

Based on the aforementioned example, the minimum distance between the resulting signals in the synchronous scenario is zero while in the asynchronous scenario, it depends on the introduced delay and is equal to $\min\{\frac{2T}{T+\tau}, \frac{4\tau}{T+\tau}\}$ where T and τ are the symbol length and the time delay, respectively.(the result is normalized by the transmission interval). Assume that the time delay is a portion of the symbol length, i.e., $\tau = \kappa T$, then, the minimum distance can be shown with respect to κ as follows:

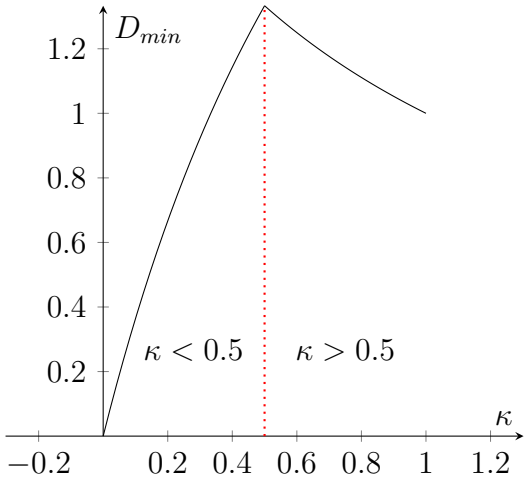


Figure 3.3: D_{min} with respect to $\kappa = \frac{\tau}{T}$

Maximizing the the minimum distance between the received signals results in minimizing the BER. Therefore, roughly speaking, we can interpret from Fig. 3.3 that the timing offset equal to half of the symbol length results in the best BER performance. These concepts are explained in more details in next chapters.

Next, we provide another example using the truncated root raised cosine (r.r.c.) pulse shape. Assume that two symbols modulated by truncated r.r.c. are transmitted through two different channels, namely, h_1, h_2. Then, the received signal after matched filtering is equal to:

$$y(t) = h_1 s_1(t) + h_2 s_2(t) + n(t). \tag{3.3}$$

where $s_i(t) = s_i g(t)$, $g(t)$ being the truncated r.c. pulse shape. If the two signals are perfectly aligned, then the received signal is sampled at the peak point as shown in Fig. 3.4. Assuming a normalized pulse shape, the discrete sample can be written as $y_1 = h_1 s_1 + h_2 s_2 + n_1$. Note that any extra sample at any other instant, e.g., t' will result in $y_2 = \alpha h_1 s_1 + \alpha h_2 s_2 + n_2$ where α is the amplitude of the pulse shape at the new sampling point t'. Now, assume that there is a time delay between the received signals.

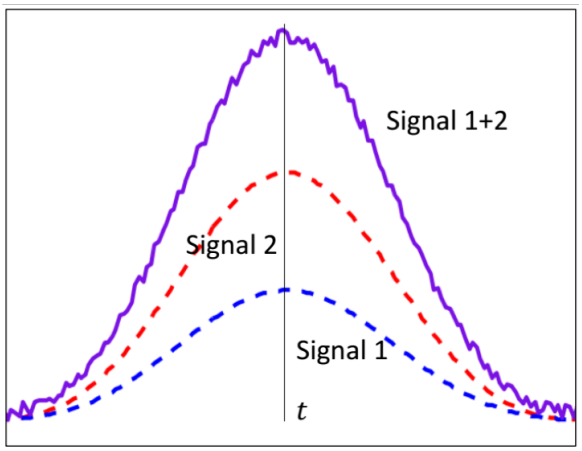


Figure 3.4: Illustration of synchronous reception

Then, based on the sampling method shown in Fig. 3.5, the two samples are obtained as $y_1 = h_1 s_1 + \beta h_2 s_2 + n_1$ and $y_2 = \gamma h_1 s_1 + h_2 s_2 + n_2$ where β and γ are the amplitude of the normalized pulse shape at the times t_1 and t_2, respectively. If we put the samples in

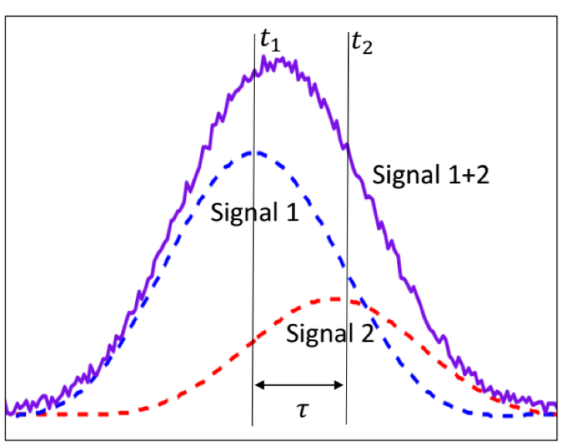


Figure 3.5: Illustration of asynchronous reception

the matrix form, for both synchronous and asynchronous scenarios, we will have:

$$\boldsymbol{y_{synch}} = \begin{pmatrix} h_1 & h_2 \\ \alpha h_1 & \alpha h_2 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \boldsymbol{n_{synch}}, \quad \boldsymbol{y_{asynch}} = \begin{pmatrix} h_1 & \beta h_2 \\ \gamma h_1 & h_2 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \boldsymbol{n_{asynch}}. \tag{3.4}$$

$$= \boldsymbol{H_{synch}} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \boldsymbol{n_{synch}}, \qquad = \boldsymbol{H_{asynch}} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \boldsymbol{n_{asynch}}. \tag{3.5}$$

The main difference between the synchronous and asynchronous transmissions is in the structure of the resulting offset matrices denoted as $\boldsymbol{H_{synch}}$ and $\boldsymbol{H_{asynch}}$. One critical characteristic of the offset matrix is its rank which is usually called the degree of freedom or multiplexing gain of the channel. In other words, by using proper processing, any channel matrix $\boldsymbol{H}$ can be divided into $rank(\boldsymbol{H})$ number of parallel sub-channels. It can be easily shown that the rank of matrices $\boldsymbol{H_{synch}}$ and $\boldsymbol{H_{asynch}}$ are 1 and 2, respectively, i.e., $rank(\boldsymbol{H_{synch}}) = 1$, $rank(\boldsymbol{H_{asynch}}) = 2$ (channel coefficients are assumed to be independent). Therefore, synchronous transmission results in one degree of freedom while the asynchronous transmission provides two degrees of freedom.

3.2 The Sampling Method: Oversampling

In the conventional point to point Nyquist-rate transmission, after performing encoding and modulation, the modulated symbols are shaped with appropriate waveforms suited to the communication channel. Let the block length be equal to N, then the transmitted signal will be:

$$s(t) = \sum_{n=1}^{N} s[n]p(t - nT) \tag{3.6}$$

where $p(t)$ is the pulse shape, e.g., r.r.c., which is truncated and its length is denoted by T_p. T is the symbol interval and is usually equal to $\frac{1}{2W}$ where W is the occupied bandwidth in which most of the transmitted power is concentrated. Let us assume a K user system with a common receiver where each user's signal experiences a time delay

of $\tau_k,\ 0 \leq \tau_1 < \tau_2 < \cdots < \tau_K < 1$ to be received at the receiver. By assuming AWGN channel, the received signal is described as:

$$y(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} s_k[n] p(t - nT - \tau_k T) + n(t) \tag{3.7}$$

To detect the transmitted symbols $x_k[n]$, instead of working with the continuous random process $y(t)$, we use a set of statistics, $Z_j = r_j(y(t)), j = 1, \cdots, J$ that are sufficient for the detection of the transmitted symbols. Intuitively, $Z_1, Z_2, \cdots, Z_J$ are jointly sufficient statistics if the statistician who knows the values of $Z_1, Z_2, \cdots, Z_J$ can do just as good a job of estimating the transmitted symbols as the statistician who knows the entire random process $y(t)$. We use the well-known factorization theorem to find the sufficient statistics.

Theorem 3.1. *Let $Y_1, \cdots, Y_n$ be random variables with joint density $f(y_1, y_2, \cdots, y_n|\theta)$. The statistics*

$$Z_j = r_j(Y_1, Y_2, \cdots, Y_n), j = 1, \cdots, J \tag{3.8}$$

are jointly sufficient to estimate θ if and only if the joint density can be factored as follows:

$$f(y_1, y_2, \cdots, y_n|\theta) = u(y_1, y_2, \cdots, y_n).v(z_1, \cdots, z_J, \theta)$$

where u and v are non-negative functions [49].

The density of $y(t)$ given the transmitted symbols is calculated as follows:

$$f(y(t)|\{s_k[n]\}) = c \exp\left[\int_{-\infty}^{\infty} w^2(t) dt\right]$$

where $w(t) = y(t) - \sum_{k=1}^{K} \sum_{n=1}^{N} s_k[n] p(t - nT - \tau_k T)$ and c is a constant value. By expanding the likelihood function, it can be observed that the transmitted symbol $s_k[n]$

is related to $y(t)$ through the value of $\int_{-\infty}^{\infty} y(t)p(t-nT-\tau_k T)dt$. Therefore, using the factorization theorem stated above, we can conclude that sufficient statistics for detecting the transmitted symbols are:

$$y_l[m] = \int_{-\infty}^{\infty} y(t)p(t-mT-\tau_l T)dt \tag{3.9}$$

which is known as the matched filter in the literature and can also be implemented using convolution followed by a sampler, i.e.,:

$$y_l[m] = y(t) * p(t)|_{t=mT+\tau_l T} \tag{3.10}$$

Denoting $p(t) * p(t)$ as $g(t)$, the sufficient statistics can be represented as:

$$y_l[m] = \sum_{k=1}^{K}\sum_{n=1}^{N} s_k[n]g((m-n)T + (\tau_l - \tau_k)T) + n_l[m] \tag{3.11}$$

where $n_l[m] = n(t) * p(t)|_{t=mT+\tau_l T}$. Therefore, the knowledge of timing offsets are critical in obtaining the sufficient statistics and are assume to be known at the receiver. The NK obtained samples can be put into a matrix form in two different ways. If we define $\boldsymbol{y_l} = (y_l[1], \cdots, y_l[N])^T$ and $\boldsymbol{s_k} = (s_k[1], \cdots, s_k[N])^T$, the input-output relation of the system can be presented in matrix form as follows:

$$\begin{pmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_K} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R_{11}} & \boldsymbol{R_{12}} & \cdots & \boldsymbol{R_{1K}} \\ \boldsymbol{R_{21}} & \boldsymbol{R_{22}} & \cdots & \boldsymbol{R_{2K}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R_{K1}} & \boldsymbol{R_{K2}} & \cdots & \boldsymbol{R_{KK}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s_1} \\ \boldsymbol{s_2} \\ \vdots \\ \boldsymbol{s_K} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_1} \\ \boldsymbol{n_2} \\ \vdots \\ \boldsymbol{n_K} \end{pmatrix} \tag{3.12}$$

where $\boldsymbol{R_{lk}}$ is the $N \times N$ constructing sub-block whose elements are defined as:

$$\boldsymbol{R_{lk}}(m,n) = g((m-n)T + (\tau_l - \tau_k)T) \tag{3.13}$$

Matrix $\boldsymbol{R}$ is a Hermitian matrix whose sub-blocks, i.e., $\boldsymbol{R_{lk}}$ are Toeplitz. Two examples of matrix $\boldsymbol{R}$ are provided for Rectangular and r.r.c. pulse shapes, $K = 3, \boldsymbol{\tau} = [0, 0.2, 0.7]$, and $N = 3$.

$$\boldsymbol{R}_{Rect.} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0.8 & 0 & 0 & 0.3 & 0 & 0 \\ 0 & 1 & 0 & 0.2 & 0.8 & 0 & 0.7 & 0.3 & 0 \\ 0 & 0 & 1 & 0 & 0.2 & 0.8 & 0 & 0.7 & 0.3 \\ \hline 0.8 & 0.2 & 0 & 1 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0.8 & 0.2 & 0 & 1 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.8 & 0 & 0 & 1 & 0 & 0.5 & 0.5 \\ \hline 0.3 & 0.7 & 0 & 0.5 & 0.5 & 0 & 1 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 & 0.5 & 0.5 & 0 & 1 & 0 \\ 0 & 0 & 0.3 & 0 & 0 & 0.5 & 0 & 0 & 1 \end{array}\right] \tag{3.14}$$

$$\boldsymbol{R}_{r.r.c.} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0.93 & -0.11 & 0.02 & 0.33 & -0.07 & 0.01 \\ 0 & 1 & 0 & 0.2 & 0.93 & -0.11 & 0.84 & 0.33 & -0.07 \\ 0 & 0 & 1 & -0.04 & 0.2 & 0.93 & -0.13 & 0.84 & 0.33 \\ \hline 0.93 & 0.2 & -0.04 & 1 & 0 & 0 & 0.6 & -0.12 & 0.02 \\ -0.11 & 0.93 & 0.2 & 0 & 1 & 0 & 0.6 & 0.6 & -0.12 \\ 0.02 & -0.11 & 0.93 & 0 & 0 & 1 & -0.12 & 0.6 & 0.6 \\ \hline 0.33 & 0.84 & -0.13 & 0.6 & 0.6 & -0.12 & 1 & 0 & 0 \\ -0.07 & 0.33 & 0.84 & -0.12 & 0.6 & 0.6 & 0 & 1 & 0 \\ 0.01 & -0.07 & 0.33 & 0.02 & -0.12 & 0.6 & 0 & 0 & 1 \end{array}\right] \tag{3.15}$$

If we define $\boldsymbol{y[m]} = (y_1[m], \cdots, y_K[m])^T$ and $\boldsymbol{s[m]} = (s_1[m], \cdots, s_K[m])^T$, the input-

output relation of the system can be presented in matrix form as follows:

$$\begin{pmatrix} \boldsymbol{y}[\mathbf{1}] \\ \boldsymbol{y}[\mathbf{2}] \\ \vdots \\ \boldsymbol{y}[\boldsymbol{N}] \end{pmatrix} = \begin{pmatrix} \boldsymbol{R'_{11}} & \boldsymbol{R'_{12}} & \cdots & \boldsymbol{R'_{1N}} \\ \boldsymbol{R'_{21}} & \boldsymbol{R'_{22}} & \cdots & \boldsymbol{R'_{2K}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R'_{N1}} & \boldsymbol{R'_{N2}} & \cdots & \boldsymbol{R'_{NN}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}[\mathbf{1}] \\ \boldsymbol{s}[\mathbf{2}] \\ \vdots \\ \boldsymbol{s}[\boldsymbol{N}] \end{pmatrix} + \begin{pmatrix} \boldsymbol{n}[\mathbf{1}] \\ \boldsymbol{n}[\mathbf{2}] \\ \vdots \\ \boldsymbol{n}[\boldsymbol{N}] \end{pmatrix} \tag{3.16}$$

where $\boldsymbol{R'_{mn}}$ is the $K \times K$ constructing sub-block whose elements are defined as:

$$\boldsymbol{R'_{mn}}(l,k) = g((m-n)T + (\tau_l - \tau_k)T) \tag{3.17}$$

Matrix $\boldsymbol{R'}$ is a Hermitian block-Toeplitz matrix, i.e., $\boldsymbol{R'_{ij}} = \boldsymbol{R'_{i'j'}}$ if $i-j = i'-j'$, whose sub-blocks are not necessarily Toeplitz. For simpler representation, and due to the block Toeplitz structure of the matrix $\boldsymbol{R'}$,we re-name the constructive block of $\boldsymbol{R'_{ij}} = \boldsymbol{R'_{i-j}}$. Thus, we can have an alternative representation of:

$$\begin{pmatrix} \boldsymbol{y}[\mathbf{1}] \\ \boldsymbol{y}[\mathbf{2}] \\ \vdots \\ \boldsymbol{y}[\boldsymbol{N}] \end{pmatrix} = \begin{pmatrix} \boldsymbol{R'_{0}} & \boldsymbol{R'_{-1}} & \cdots & \boldsymbol{R'_{1-N}} \\ \boldsymbol{R'_{1}} & \boldsymbol{R'_{0}} & \cdots & \boldsymbol{R'_{2-N}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R'_{N-1}} & \boldsymbol{R'_{N-2}} & \cdots & \boldsymbol{R'_{0}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}[\mathbf{1}] \\ \boldsymbol{s}[\mathbf{2}] \\ \vdots \\ \boldsymbol{s}[\boldsymbol{N}] \end{pmatrix} + \begin{pmatrix} \boldsymbol{n}[\mathbf{1}] \\ \boldsymbol{n}[\mathbf{2}] \\ \vdots \\ \boldsymbol{n}[\boldsymbol{N}] \end{pmatrix} \tag{3.18}$$

Two examples of matrix $\boldsymbol{R'}$ are provided for Rectangular and r.r.c. pulse shapes, $K = 3, \boldsymbol{\tau} = [0, 0.2, 0.7]$ and $N = 3$.

$$\boldsymbol{R'}_{Rect.} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0.8 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 1 & 0.3 & 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.3 & 1 & 0.7 & 0.5 & 0 & 0 & 0 & 0 \\ \hline 0 & 0.2 & 0.7 & 1 & 0.8 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.8 & 1 & 0.3 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.3 & 1 & 0.7 & 0.5 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.2 & 0.7 & 1 & 0.8 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.8 & 1 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.3 & 1 \end{array}\right] \tag{3.19}$$

$$\boldsymbol{R'}_{r.r.c.} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0.93 & 0.33 & 0 & -0.11 & -0.07 & 0 & 0.02 & 0.01 \\ 0.93 & 1 & 0.6 & 0.2 & 0 & -0.12 & -0.04 & 0 & 0.02 \\ 0.33 & 0.6 & 1 & 0.84 & 0.6 & 0 & -0.13 & -0.12 & 0 \\ \hline 0 & 0.2 & 0.84 & 1 & 0.93 & 0.33 & 0 & -0.11 & -0.07 \\ -0.11 & 0 & 0.6 & 0.93 & 1 & 0.6 & 0.2 & 0 & -0.12 \\ -0.07 & -0.12 & 0 & 0.33 & 0.6 & 1 & 0.84 & 0.6 & 0 \\ \hline -0 & -0.04 & -0.13 & 0 & 0.2 & 0.84 & 1 & 0.93 & 0.33 \\ 0.02 & 0 & -0.12 & -0.11 & 0 & 0.6 & 0.93 & 1 & 0.6 \\ 0.01 & 0.017 & 0 & -0.07 & -0.12 & 0 & 0.33 & 0.6 & 1 \end{array}\right] \tag{3.20}$$

If the time delays are equi-spaced, i.e., $\tau_k = (k-1)\frac{T}{K},\ k = 1, \cdots, K$, then the matrix $\boldsymbol{R}$ turns into a block-Toeplitz matrix which each blocks is also Toeplitz. On the other hand, the matrix $\boldsymbol{R'}$ will turn into a Toeplitz matrix. These two structures are used interchangeably throughout the thesis based on the application.

In the conventional transmission scheme, the set of modulated signals, hence, the set of matched filters are orthogonal, which results in independent noise samples. However, due

to timing offsets and the overlapping among the pulse shapes, the set of matched filters are not orthogonal anymore. Therefore, the noise samples are not independent anymore and their covariance is not identity matrix. Remember that:

$$n_l[m] = \int_{-\infty}^{\infty} n(t)p(t - mT - \tau_l T)dt \tag{3.21}$$

Thus, the covariance between $n_l[m]$ and $n_k[n]$, denoted as δ_{lk}^{mn}, will be equal to:

$$\begin{aligned}
\delta_{lk}^{mn} &= E\{n_l[m]n_k^*[n]\} \\
&= E\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(t)p(t - mT - \tau_l T)n^*(s)p(s - nT - \tau_k T)dtds\right\} \\
&= \sigma_n^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(t-s)p(t - mT - \tau_l T)p(s - nT - \tau_k T)dtds \\
&= \sigma_n^2 \int_{-\infty}^{\infty} p(t - mT - \tau_l T)p(t - nT - \tau_k T)dt \\
&= \sigma_n^2 p(t) * p(t)|_{t=(m-n)T+(\tau_l-\tau_k)T} \\
&= \sigma_n^2 g((m-n)T + (\tau_l - \tau_k))
\end{aligned} \tag{3.22}$$

Therefore, the covariance matrix of noise vectors $\boldsymbol{n}$ and $\boldsymbol{n'}$ are equal to $\boldsymbol{R}\sigma_n^2$ and $\boldsymbol{R'}\sigma_n^2$, respectively. However, we can use a different set of matched filters to produce independent noise samples which circumvent the noise whitening procedure involving Cholesky decomposition. For example, one way is to break down the integrals corresponding to the sampling in Fig. 3.6 to define a new sampling method as shown in Fig. 3.7 [46, 47]. The corresponding output samples are written as follows where τ_{k+1} is an auxiliary variable equal to T_p.

$$\begin{aligned}
y_l[m] = &\int_{\tau_l+mT}^{\tau_{(l+1)}+mT} \sum_{k=1}^{K}\sum_{n=1}^{N} s_k[n]p(t - mT - \tau_l T)p(t - nT - \tau_k T)dt \\
&+ \int_{\tau_l+mT}^{\tau_{(l+1)}+mT} n(t)p(t - mT - \tau_l T)dt \qquad 1 \le l \le K,\ 1 \le m \le N+1
\end{aligned}$$

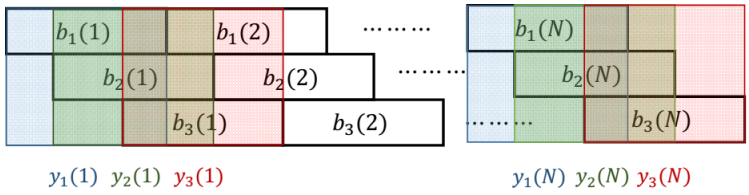


Figure 3.6: Sampling method with correlated pulse shapes

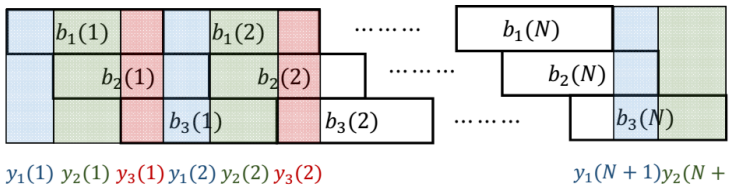


Figure 3.7: Sampling method with independent pulse shapes

By defining intermediate variables $u_{mn}(l,k)$ and noise samples $v_l[m]$ as follows:

$$u_{mn}(l,k) = \int_{\tau_l+mT}^{\tau_{(l+1)}+mT} p(t-mT-\tau_l T)p(t-nT-\tau_k T)dt \tag{3.23}$$

$$\nu_l[m] = \int_{\tau_l+mT}^{\tau_{(l+1)}+mT} n(t)p(t-mT-\tau_l T)dt \tag{3.24}$$

, respectively, we can write the output samples in a more compact way:

$$y_l[m] = \sum_{k=1}^{K}\sum_{n=1}^{N} s_k[n]u_{mn}(l,k) + \nu_l[m] \tag{3.25}$$

$$m = 1,\ldots,N+1 \quad \text{sampling time index}$$

$$l = 1,\ldots,K \quad \text{index of the matched user}$$

Defining $\boldsymbol{y[m]} = (y_1[m], y_2[m], \ldots, y_K[m])^T$ and $\boldsymbol{s[n]} = (s_1[n], s_2[n], \ldots, s_K[n])^T$, then, $\boldsymbol{y[m]}$ for different values of m can be written as:

$$\boldsymbol{y[m]} = \sum_{n=1}^{N} \boldsymbol{U_{mn}s[n]} + \boldsymbol{\nu[m]} \quad 1 \le m \le N+1 \tag{3.26}$$

where $\boldsymbol{\nu[m]} = [\nu_1(j), \nu_2(j), \ldots, \nu_K(j)]^T$ and $\boldsymbol{U_{mn}}$ is a $K \times K$ matrix whose elements are

defined as $\boldsymbol{U_{mn}}(l,k) = u_{mn}(l,k)$. The next step is to put all vectors of $\boldsymbol{y[m]}$ together and define $\boldsymbol{y}$ as $(\boldsymbol{y[1]}^T, \boldsymbol{y[2]}^T, \ldots, \boldsymbol{y[N+1]}^T)^T$. Then, $\boldsymbol{y}$ can be written as:

$$\begin{aligned}\boldsymbol{y} &= \begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{U_{12}} & \cdots & \boldsymbol{U_{1N}} \\ \boldsymbol{U_{21}} & \boldsymbol{U_{22}} & \cdots & \boldsymbol{U_{2N}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{U_{N1}} & \cdots & \boldsymbol{U_{N(N-1)}} & \boldsymbol{U_{NN}} \\ \boldsymbol{U_{(N+1)1}} & \cdots & \boldsymbol{U_{(N+1)(N-1)}} & \boldsymbol{U_{(N+1)N}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s[1]} \\ \boldsymbol{s[2]} \\ \vdots \\ \boldsymbol{s[N]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu[1]} \\ \boldsymbol{\nu[2]} \\ \vdots \\ \boldsymbol{\nu[N]} \end{pmatrix} \\ &= \boldsymbol{U s} + \boldsymbol{\nu} \end{aligned} \tag{3.27}$$

Block Toeplitz structure of $\boldsymbol{U}$ originates from the fact that $u_{(j+m)(i+m)}(l,k) = u_{ji}(l,k)$. This can be verified by a change of variable in Eq. (3.23). Defining $\boldsymbol{U_{i-j}} = \boldsymbol{U_{ij}}$, we can have an alternative representation of:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{U_0} & \boldsymbol{U_{-1}} & \cdots & \boldsymbol{U_{1-N}} \\ \boldsymbol{U_1} & \boldsymbol{U_0} & \cdots & \boldsymbol{U_{2-N}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{U_{N-1}} & \cdots & \boldsymbol{U_1} & \boldsymbol{U_0} \\ \boldsymbol{U_N} & \cdots & \boldsymbol{U_2} & \boldsymbol{U_1} \end{pmatrix} \begin{pmatrix} \boldsymbol{s[1]} \\ \boldsymbol{s[2]} \\ \vdots \\ \boldsymbol{s[N]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu[1]} \\ \boldsymbol{\nu[2]} \\ \vdots \\ \boldsymbol{\nu[N]} \end{pmatrix}$$

The important fact about this sampling method is that the covariance matrix of noise samples is diagonal. With a small abuse of notation, we denote $Diag(\boldsymbol{U_0})$ as a diagonal matrix including diagonal elements of $\boldsymbol{U_0}$. Then, it can be shown that $E[\boldsymbol{\nu\nu}^H]$ is equal to $\sigma^2(\boldsymbol{I_{N+1}} \otimes Diag(\mathbf{U_0}))$, where $\boldsymbol{I_n}$ is an $n \times n$ identity matrix and $(\otimes)$ is Kronecker product.

Since the statistically sufficient samples in Fig. 3.6 can be created from samples in Fig. 3.7, the samples in Fig. 3.7, are sufficient statistics too. Both of these sampling methods

introduce intentional ISI and impose memory on the system; however, they have some differences:

1. Since sampling intervals are disjoint in Fig. 3.7, noise samples are independent. However, due to sampling overlap, the noise samples in Fig. 3.6 are correlated.

2. The sampling in Fig. 3.7 results in an overdetermined system, while the number of output samples in Fig. 3.6 is equal to the number of input symbols.

Note that, the notion of increasing signaling dimension by performing oversampling, introduces ISI in the system. Hence, the benefit of having more degree of freedom in the system model is obtained in the expense of dealing with ISI which necessitates the need for designing low complexity receiver architectures. The question of whether the offset matrix, i.e., $\boldsymbol{R}/\boldsymbol{R'}$, is invertible or not and how it behaves asymptotically as the block length N tends to infinity will have important consequences on the performance of system and the design of encoders and decoders [50]. Hence, we will investigate this question in the next section.

3.3 Features of the Offset Matrix

To understand the asymptotic behavior of matrix $\boldsymbol{R}/\boldsymbol{R'}$, we will utilize the Szego theorem which states that [51, 52]:

Theorem 3.2. *Let* $\boldsymbol{T_N} = [t_{k-j}; k, j = 0, 1, 2, \cdots, N-1]$ *be a sequence of Hermitian Toeplitz matrices whose generating functions is defined as* $f(w) = \sum_{k=-\infty}^{\infty} t_k e^{ikw}, w \in [0, 2\pi]$. *Also,* $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1}$ *are the sorted eigenvalues of matrix* $\boldsymbol{T_N}$. *Then, for*

any function F that is continuous on the range of f:

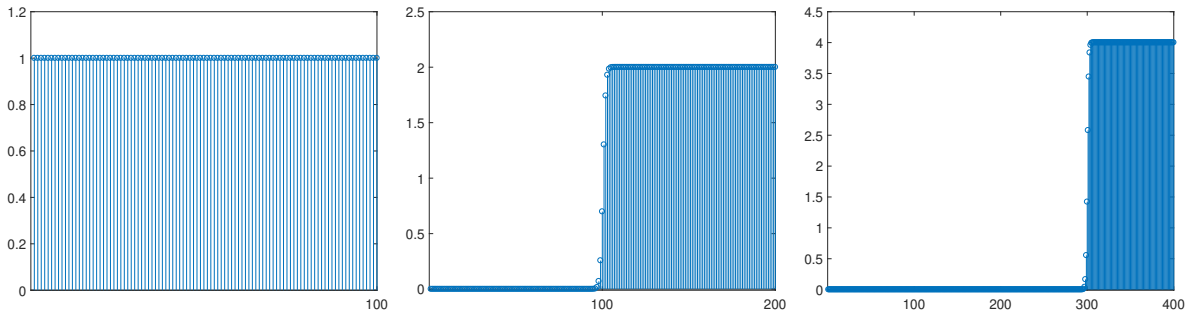
$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(\lambda_k) = \frac{1}{2\pi} \int_0^{2\pi} F(f(w))dw \tag{3.28}$$

In addition, the largest and smallest eigenvalues asymptotically converge to:

$$\lim_{N\to\infty} \lambda_0 = \max_w f(w)$$

$$\lim_{N\to\infty} \lambda_{N-1} = \min_w f(w)$$

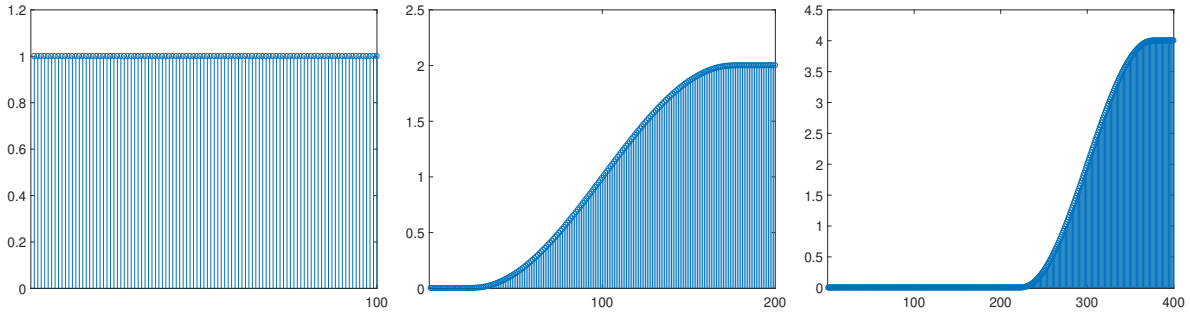
The proof comes from asymptotic equivalence of sequences of Hermitian Toeplitz matrices and their corresponding circulant versions which result in asymptotic convergence of the their eigenvalues [53]. Therefore, this theorem can be utilized to specify the eigenvalues of matrix $\boldsymbol{R'}$ when the time delays are uniformly distributed. As N tends to infinity, the asymptotic eigenvalues of matrix $\boldsymbol{R'}$ with uniformly distributed time delays approach the equispaced samples of the $\frac{K}{T}$-folded spectrum of the pulse shape. $\frac{K}{T}$-folded spectrum is defined the same as the conventional folded spectrum except that the frequency shifts are $\frac{K}{T}$ [54].



(a) $N = 100$ and $K = 1$ (b) $N = 100$ and $K = 2$ (c) $N = 100$ and $K = 4$

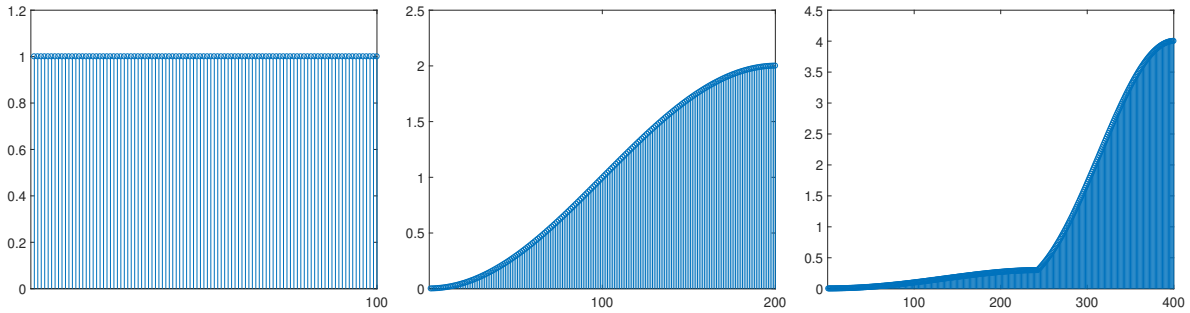
Figure 3.8: Eigenvalues of matrix $\boldsymbol{R'}$ with Sinc pulse shape

Eigenvalues of matrix $\boldsymbol{R'}$ with uniform time delays and different pulse shapes including Sinc, r.r.c. and Rectangular are shown in Figs. 3.8, 3.9 and 3.10, respectively. Due to



(a) $N = 100$ and $K = 1$ (b) $N = 100$ and $K = 2$ (c) $N = 100$ and $K = 4$

Figure 3.9: Eigenvalues of matrix $\boldsymbol{R'}$ with r.r.c. pulse shape ($\beta = 0.75$)



(a) $N = 100$ and $K = 1$ (b) $N = 100$ and $K = 2$ (c) $N = 100$ and $K = 4$

Figure 3.10: Eigenvalues of matrix $\boldsymbol{R'}$ with Rectangular pulse shape

the time limited transmission (infinite-time pulse shapes can not be practically realized), the spectrum is theoretically non-zero for every frequency, thus the extra eigenvalues provided by the asynchronous transmission are always positive. However, the used pulse shape has a huge impact on the magnitude of the extra eigenvalues. For example, for Sinc function, adding asynchrony is not very beneficial, because the extra eigenvalues have very small values. However, for other pulse shapes which have some nonzero spectrum outside the $\frac{1}{T}$ bandwidth, like, Rectangular and r.r.c. pulse shapes, adding asynchrony provides additional nonzero eigenvalues. For example, for r.r.c. pulse shape, β percent additional eigenvalues are available to exploit. For Rectangular pulse shape, due to having unlimited spectrum, there are unlimited number of eigenvalues to exploit, however, by increasing K, the additional eigenvalues get closer to zero. In summary, the benefits of the asynchronous

transmission depends on the utilized pulse shape and its $\frac{K}{T}$-folded spectrum.

Fortunately, the Szego Theorem can also be extended to Toeplitz block matrices like $\boldsymbol{R}$ [54]. The generalized Szego Theorem relates the collective behavior of the eigenvalues to the generalized generating function, $\boldsymbol{R(w)}$, which is defined as:

$$\boldsymbol{R(w)} = \begin{pmatrix} f_{11}(w) & f_{12}(w) & \cdots & f_{1K}(w) \\ f_{21}(w) & f_{22}(w) & \cdots & f_{2K}(w) \\ \vdots & \ddots & \ddots & \vdots \\ f_{K1}(w) & f_{K2}(w) & \cdots & f_{KK}(w) \end{pmatrix} \tag{3.29}$$

where $f_{lk}(w)$ is the generating function for the corresponding Toeplitz block $\boldsymbol{R_{lk}}$. The generalized Szego Theorem states that for any continuous function F [54]:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{NK-1} F[\lambda_k(\boldsymbol{R})] = \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=0}^{K-1} F[\lambda_j(\boldsymbol{R(w)})]dw \tag{3.30}$$

In particular, for $F(x) = x$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{NK-1} \lambda_k(\boldsymbol{R}) = \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=0}^{K-1} \lambda_j(\boldsymbol{R(w)})dw \tag{3.31}$$

Moreover, same convergence results can be obtained for the largest and smallest eigenvalues, i.e,:

$$\lim_{N\to\infty} \lambda_0(\boldsymbol{R}) = \max_w \lambda_0(\boldsymbol{R(w)})$$

$$\lim_{N\to\infty} \lambda_{NK-1}(\boldsymbol{R}) = \min_w \lambda_{K-1}(\boldsymbol{R(w)})$$

Therefore, some of the properties of matrix $\boldsymbol{R}$ can be deducted by properties of matrix $\boldsymbol{R(w)}$ when the block length is sufficiently large. For example, it is proved in the liter-

ature that for time limited transmission, matrix $\boldsymbol{R}(\boldsymbol{w})$ is positive definite with bounded eigenvalues, thus, eigenvalues of matrix $\boldsymbol{R}$ are nonzero and bounded. On the other hand, when the pulse shapes are strictly band-limited, matrix $\boldsymbol{R}(\boldsymbol{w})$ is singular which results in singularity of $\boldsymbol{R}$ [7].

Chapter 4

Multiuser Systems and Time Asynchrony: Receiver Design

This chapter shows that by investigating inherent time delays between different users in a multiuser scenario, we can cancel interference more efficiently. Time asynchrony provides another tool to cancel interference, which results in preserving other resources like frequency, time, and code. By taking advantage of the sampling methods explained in the previous chapter, a system model with memory is derived, which resembles an ISI channel. Thus, the asynchronous transmission with oversampling enables the use of sequence detection methods such as the Viterbi algorithm and forward-backward belief propagation, which provides superior BEP performance. Next, the diversity of zero-forcing (ZF) detection is analyzed. It is shown that with asynchronous transmission, full diversity can be achieved by ZF detection. Simulation results are also presented to verify our analysis.

4.1 Introduction

There are many applications where multiple users share a common channel to transmit data to a receiver. Numerous examples of multi-access communication include uplink transmission of a single cell in a cellular system, a group of twisted-pair copper subscriber lines transmitting data to the same switching office, and multiple ground stations communicating with a satellite and interactive cable television networks. The critical challenge in multiuser transmissions or multiple access channels is Interuser Interference. Over several decades, many methods have been introduced to address this problem [10], [11]. Most of these methods are based on assigning orthogonal dimensions to different users to separate them and prevent interference. For example, time division multiple access (TDMA) protocols allocate different time slots to different users to mitigate interference. The same concept can be applied by partitioning the frequency spectrum among different users, called frequency division multiple access (FDMA). Code division multiple access is another scheme used to surpass interuser interference in which users are multiplexed by distinct codes rather than by orthogonal frequency bands, or by orthogonal time slots [14]. More recently, multiple receive antennas are utilized at the receive side to take advantage of the spatial domain in order to cancel interference [15], [16].

In this chapter, we investigate the timing mismatch between users as an additional resource to address the problem of interuser interference. By exploiting time delays between users and employing an appropriate sampling method, we design detection methods which not only cancel the interference but also outperform the synchronous ones. There are also other examples in the literature that motivates the asynchronous transmission. For example, in [9, 4], the ZF receiver structure with the asynchronous transmission is considered. However, the proposed receiver is not BEP wise optimal. In this chapter, we consider the BEP wise optimal ZF receiver and prove its full diversity gain. The design of asyn-

chronous differential decoding methods that outperform their synchronous counterparts is discussed in [46], [47]. This chapter presents sampling diversity and provides several detection methods to gain advantages from the asynchronous transmission. We analytically prove that our ZF method provides full diversity.

In Section 4.2, the system model is presented. In Section 4.3, the receiver design is explained including MLSD, ZF and SIC methods. In Section 4.4, the performance analysis is provided, and the diversity gain of the ZF method is analyzed. Finally, simulation results are provided in Section 4.5. The materials in this chapter are partially published in [48].

4.2 System Model

4.2.1 General Settings

We consider a system with K users, transmitting data to a common receiver simultaneously, which can have one receive antenna or multiple ones. Due to different physical locations of users, their signal is received with various time delays. It is assumed that each data stream is received with an arbitrary delay smaller than the symbol interval and only the receiver knows the time delays. The signal transmitted from User k is described by:

$$s_k(t) = \sum_{n=1}^{N} s_k[n]p(t - nT) \tag{4.1}$$

where T is the symbol length and $p(.)$ is the pulse-shaping filter with non-zero duration of T_p. Also, N is the frame length and $s_k[n]$ is the transmitted symbol by User k in the nth

time slot. The transmitted signals are received with a relative delay of τ_k and a channel path gain of h_k. Then, the received signal can be represented by:

$$y(t) = \sum_{k=1}^{K} h_k s_k(t - \tau_k T) + n(t) \tag{4.2}$$

where K is the number of users and $n(t)$ is the white noise with variance of σ_n^2. Without loss of generality, we assume that $0 = \tau_1 < \tau_2 < \cdots < \tau_K < 1$.

4.2.2 Output Samples

The sampling methods explained in the previous chapter can be used to provide sufficient statistics for detecting the transmitted symbols. By using the sampling method in Fig. 3.6, the corresponding output samples can be written as:

$$\begin{aligned} \boldsymbol{y} &= \begin{pmatrix} \boldsymbol{U_0} & \boldsymbol{U_{-1}} & \cdots & \boldsymbol{U_{1-N}} \\ \boldsymbol{U_1} & \boldsymbol{U_0} & \cdots & \boldsymbol{U_{2-N}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{U_{N-1}} & \cdots & \boldsymbol{U_1} & \boldsymbol{U_0} \\ \boldsymbol{U_N} & \cdots & \boldsymbol{U_2} & \boldsymbol{U_1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{H}} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{H}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\mathcal{H}} & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\mathcal{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s[1]} \\ \boldsymbol{s[2]} \\ \vdots \\ \boldsymbol{s[N]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu[1]} \\ \boldsymbol{\nu[2]} \\ \vdots \\ \boldsymbol{\nu[N]} \end{pmatrix} \\ &= \boldsymbol{UHs} + \boldsymbol{\nu} \end{aligned} \tag{4.3}$$

where $\boldsymbol{\mathcal{H}}$ is a $K \times K$ matrix equivalent to $\text{diag}(h_1, h_2, \ldots, h_K)$. As explained in the previous chapter, the advantage of this sampling is to generate independent noise samples. Based on the relation between T and T_p, different numbers of adjacent symbols interfere with each other. For example, for rectangular pulse shapes, i.e., $T = T_p$, at each instant only current and previous symbols cause interference. In other words, only $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$ are nonzero. Without loss of generality, we assume that $T = 1$, therefore $\boldsymbol{U_0}$ and $\boldsymbol{U_1}$ are

defined as follows:

$$\boldsymbol{U_0} = \begin{pmatrix} \tau_{2,1} & 0 & \dots & 0 \\ \tau_{3,2} & \tau_{3,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{K,K-1} & \dots & \tau_{K,K-1} & 0 \\ 1-\tau_K & \dots & 1-\tau_K & 1-\tau_K \end{pmatrix}, \quad \boldsymbol{U_1} = \begin{pmatrix} 0 & \tau_{2,1} & \dots & \tau_{2,1} \\ 0 & 0 & \dots & \tau_{3,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \tau_{K,K-1} \\ 0 & \dots & 0 & 0 \end{pmatrix} \tag{4.4}$$

where $\tau_{i,j} = \tau_i - \tau_j$. Equivalently, the sampling method in Fig. 3.7 can be utilized to generate the samples of:

$$\begin{aligned} \boldsymbol{y} &= \begin{pmatrix} \boldsymbol{R'_0} & \boldsymbol{R'_{-1}} & \dots & \boldsymbol{R'_{1-N}} \\ \boldsymbol{R'_1} & \boldsymbol{R'_0} & \dots & \boldsymbol{R'_{2-N}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R'_{N-1}} & \boldsymbol{R'_{N-2}} & \dots & \boldsymbol{R'_0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{H}} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{H}} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{\mathcal{H}} & \boldsymbol{0} \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\mathcal{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}[1] \\ \boldsymbol{s}[2] \\ \vdots \\ \boldsymbol{s}[N] \end{pmatrix} + \boldsymbol{n} \\ &= \boldsymbol{R'Hs} + \boldsymbol{n} \end{aligned} \tag{4.5}$$

where $\boldsymbol{R'_0}$, and $\boldsymbol{R'_{-1}} = \boldsymbol{R'^T_1}$ are defined as:

$$\boldsymbol{R'_0} = \begin{pmatrix} 1 & 1-\tau_{2,1} & \dots & 1-\tau_{K,1} \\ 1-\tau_{2,1} & 1 & \dots & 1-\tau_{K,2} \\ \vdots & \ddots & \ddots & \vdots \\ 1-\tau_{K-1,1} & \dots & 1 & 1-\tau_{K,K-1} \\ 1-\tau_{K,1} & \dots & 1-\tau_{K,K-1} & 1 \end{pmatrix} \boldsymbol{R'_{-1}} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \tau_{2,1} & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \tau_{K-1,1} & \tau_{K-1,2} & \dots & 0 & 0 \\ \tau_{K,1} & \tau_{K,2} & \dots & \tau_{K,K-1} & 0 \end{pmatrix}$$

Because of intersection between sampling intervals, noise samples are correlated and noise whitening procedure needs to be performed before symbol detection. We use the first system model to design the receiver architectures and use the second system model to analyze the system performance.

4.3 Receiver Design

In this section, we introduce different detection methods which take advantage of distinct features of the proposed sampling methods. One of these features is converting a memoryless system into a system with memory and independent noise samples. This enables us to implement the Viterbi algorithm based on samples in Eq. (4.3). The other feature

is that this sampling method provides a over-determined system model which can be used to improve detection methods. For example, the extra samples make it possible to carry out successive interference cancellation (SIC) backward and forward. Also, by means of the additional available degrees of freedom, ZF detection can be performed even with one receive antenna, which is impossible in synchronous multiuser transmission. In what follows, we will show how asynchronous multiuser transmission can outperform synchronous multiuser transmission.

The problem is to recover the vector $\boldsymbol{s} \in \mathbb{R}^{NK}$ from an observation vector in the form of:

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{\nu} \tag{4.6}$$

where the noise vector is distributed as $\boldsymbol{\nu} \sim N(0, \sigma_n^2 \boldsymbol{\Sigma_\nu})$. Thanks to the new sampling method introduced in the previous section, the covariance matrix of noise vector is a diagonal matrix. For the sake of notational simplicity, we can multiply the obtained vector with ${\boldsymbol{\Sigma_\nu}}^{-1/2}$ to equalize the variance of noise for different sub-channels. Thus, we will have $\bar{\boldsymbol{y}} = \bar{\boldsymbol{U}}\boldsymbol{s} + \bar{\boldsymbol{\nu}}$ where $\bar{\boldsymbol{U}} = {\boldsymbol{\Sigma_\nu}}^{-1/2}\boldsymbol{U}\boldsymbol{H}$ and $\bar{\boldsymbol{\nu}} \sim N(0, \sigma_n^2 \boldsymbol{I_{(N+1)K}})$. The elements of $\boldsymbol{s}$ belong to a finite alphabet Ω of size $|\Omega|$. Therefore, there are $|\Omega|^{NK}$ possible vectors of $\boldsymbol{s}$. Detecting $\boldsymbol{s}$ in the maximum-likelihood (ML) sense is equivalent to:

$$\min_{\boldsymbol{s} \in \Omega^{NK}} \left\| \bar{\boldsymbol{y}} - \bar{\boldsymbol{U}}\boldsymbol{s} \right\| \tag{4.7}$$

Eq. (4.7) is a finite-alphabet-constrained least-squares (LS) problem, which is known to be nondeterministic polynomial-time (NP)-hard. The complicating factor is of course the constraint $\boldsymbol{x} \in \Omega^{NK}$, otherwise it would be just the classical LS regression [55]. Due to the distinct feature of the sampling method shown in Fig. 5.3 , i.e., $\bar{\boldsymbol{U}}(i,j) = 0, j > i$,

the metric in Eq. (4.7) can be rewritten as:

$$\min_{\boldsymbol{s}\in\Omega^{NK}}\{f_1(s_1)+f_2(s_1,s_2)+\cdots+f_{NK}(s_1,\cdots,s_{NK})\} \tag{4.8}$$

where

$$f_n(s_1,\cdots,s_n)=\left(\bar{y}_n-\sum_{m=1}^{n}\bar{\boldsymbol{U}}(n,m)s_m\right)^2 \tag{4.9}$$

The new representation can be visualized as a decision tree with NK+1 layers, $|\Omega|$ branches emanating from each node, and $|\Omega|^{NK}$ leaf nodes. To any branch, we associate a hypothetical decision on s_n , and the branch metric $f_n(s_1,\cdots,s_n)$. Also, to any node, we associate the cumulative metric which is just the sum of all branch metrics accumulated when traveling to that node from the root. Finally, to each node, we associate the symbols $\{s_1,s_2\cdots,s_n\}$ it takes to reach there from the root.

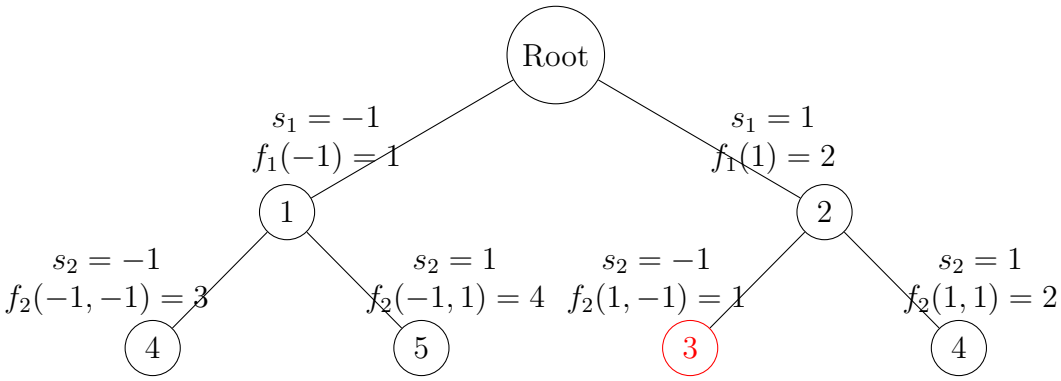


Figure 4.1: Tree representation of the minimization problem

Clearly, a naive but valid way of solving the minimization in Eq. (4.8) would be to traverse the entire tree to find the leaf node with the smallest cumulative metric. However, such a brute-force search is extremely inefficient, since there are $|\Omega|^n$ leaf nodes to examine. We will now review some efficient, popular, but approximate solutions for this problem setting [56, 57].

4.3.1 Maximum-Likelihood Sequence Detection

Simply stated, Maximum-Likelihood Sequence Detection (MLSD) finds the sequence through the trellis that looks most like the received output sequence, or, in other words, minimizes the relation in Eq. (4.8) [58]. As a simple example, the highlighted path in Fig. 4.2 is the MLSD decision for the sequence that best matches the received outputs. The receiver need only wait until the entire sequence is received and then compare it against all possible sequences.

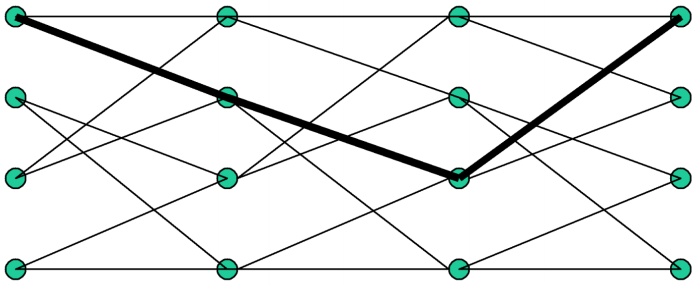


Figure 4.2: Illustration of the concept of MLSD in a trellis

The concept is simple, but the complexity grows exponentially with length of the sequence. The Viterbi Algorithm reduces this complexity through a recursive elimination procedure where trellis paths that are worse than other paths are eliminated early from further consideration. At each stage of the trellis, the equilizer keep only the best "surviving" path into each state, eliminating all the rest into that state at that time. Thanks to the memory introduced in the system by the time delays, we can implement MLSD by using Viterbi algorithm. As a result, the complexity order of $\mathcal{O}(|\Omega|^{NK})$ offered by exhaustive search reduces to complexity order of $\mathcal{O}(|\Omega|^{K})$[59].

Based on the recursive relation between the input and output which is described as:

$$\bar{\boldsymbol{y}}[\boldsymbol{j}] = \bar{\boldsymbol{U}}_{\boldsymbol{0}}\boldsymbol{s}[\boldsymbol{j}] + \bar{\boldsymbol{U}}_{\boldsymbol{1}}\boldsymbol{s}[\boldsymbol{j}-\boldsymbol{1}] + \bar{\boldsymbol{\nu}}[\boldsymbol{j}] \qquad 2 \leq j \leq N$$

where $\bar{\boldsymbol{U}}_{\boldsymbol{i}} = \boldsymbol{U}_{\boldsymbol{i}}\boldsymbol{\mathcal{H}}$. The trellis diagram of the system includes $|\Omega|^K$ states with $|\Omega|^K$ outgoing paths to the next states, and $|\Omega|^K$ incoming paths from previous states. To calculate the metric for each path, we need to calculate the likelihood function as follows:

$$\begin{aligned} Pr(\bar{\boldsymbol{y}}[\boldsymbol{j}]|\boldsymbol{s}[\boldsymbol{j}], \boldsymbol{s}[\boldsymbol{j-1}]) =& Pr(\bar{\boldsymbol{\nu}}[\boldsymbol{j}] = \bar{\boldsymbol{y}}[\boldsymbol{j}] - \bar{\boldsymbol{U}}_{\boldsymbol{0}}\boldsymbol{s}[\boldsymbol{j}] - \bar{\boldsymbol{U}}_{\boldsymbol{1}}\boldsymbol{s}[\boldsymbol{j-1}]) \\ =& \frac{1}{\sqrt{(2\pi)^K \sigma_n^2}} \exp\left(-\frac{1}{2\sigma_n^2} \boldsymbol{x_j}^H \boldsymbol{x_j}\right) \end{aligned}$$

where $\boldsymbol{x_j} = \bar{\boldsymbol{y}}[\boldsymbol{j}] - \bar{\boldsymbol{U}}_{\boldsymbol{0}}\boldsymbol{s}[\boldsymbol{j}] - \bar{\boldsymbol{U}}_{\boldsymbol{1}}\boldsymbol{s}[\boldsymbol{j-1}]$. By discarding common terms and simple calculations, the metric for each path can be defined as $\sum_{i=1}^{K} |\boldsymbol{x_j}(i)|^2$. After calculating the path metrics, the final goal is to find the surviving path and trace it back to detect the transmitted symbols.

4.3.2 Zero-Forcing (ZF)

The ZF detector first solves the problem mentioned in Eq. (4.7) neglecting the finite alphabet constraint.

$$\tilde{\boldsymbol{s}} = \underset{\boldsymbol{s}}{\operatorname{argmin}} \left\| \bar{\boldsymbol{y}} - \bar{\boldsymbol{U}}\boldsymbol{s} \right\| \tag{4.10}$$

$$= \bar{\boldsymbol{U}}^{-1} \bar{\boldsymbol{y}} \tag{4.11}$$

Note that $\bar{\boldsymbol{U}}^{-1}$ does not need to be explicitly calculated. For example, one way is to use Gaussian elimination, i.e.,

$$\bar{s}_1 = \frac{\bar{y}_1}{\bar{\boldsymbol{U}}(1,1)} \tag{4.12}$$

$$\bar{s}_2 = \frac{\bar{y}_2 - \bar{\boldsymbol{U}}(2,1)\bar{s}_1}{\bar{\boldsymbol{U}}(2,2)} \tag{4.13}$$

$$\vdots \tag{4.14}$$

$$\bar{s}_n = \frac{\bar{y}_n - \sum_{m=1}^{n-1} \bar{\boldsymbol{U}}(n,m)\bar{s}_m}{\bar{\boldsymbol{U}}(n,n)} \tag{4.15}$$

ZF detector, then approximate the answers by projecting each $\bar{s}_n$ onto the closet point in the constellation point, i.e.,

$$\tilde{s}_n = \underset{s_k \in \Omega}{\operatorname{argmin}} \|s_k - \bar{s}_k\| \tag{4.16}$$

It can be seen that $\bar{\boldsymbol{s}} = \boldsymbol{s} + \bar{\boldsymbol{U}}^{-1}\bar{\boldsymbol{\nu}}$ which means that all the ISI has been completely removed. However, unfortunately ZF works poorly unless $\bar{\boldsymbol{U}}$ is well conditioned. This can be improved by using MMSE estimate but it does not overcome the fundamental problem of the approach [55].

4.3.3 Successive Interference Cancellation with Hard Decision Passing

Successive interference cancellation (SIC) detection that takes a serial approach to cancel interference can be used to reduce complexity. SIC is sometimes called ZF detector with decision feedback (ZF-DF). ZF-DF detector performs Gaussian elimination method mentioned before to detect the transmitted symbol. However, it performs the Gaussian elimination with the modification that it projects the symbols onto the constellation Ω in each step of the Gaussian elimination, rather than afterwards. In more details:

$$\tilde{s}_1 = \underset{s_1 \in \Omega}{\operatorname{argmin}} f_1(s_1) \tag{4.17}$$

$$= \left\lfloor \frac{\bar{y}_1}{\bar{\boldsymbol{U}}(1,1)} \right\rceil \tag{4.18}$$

where [.] means the projection onto the nearest point in the constellation. For the next symbols, the effect of previously detected symbols are removed by subtracing them from the corresponding samples, i.e.,

$$\tilde{s}_n = \underset{s_n \in \Omega}{\operatorname{argmin}} \{ f_n(\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_{n-1}, s_n) \} \tag{4.19}$$

$$= \left[\frac{\bar{y}_n - \sum_{m=1}^{n-1} \bar{\boldsymbol{U}}(n,m) \tilde{s}_m}{\bar{\boldsymbol{U}}(n,n)} \right] \tag{4.20}$$

In the decision-tree perspective, ZF-DF can be considered as just examining one single path down from the root. When deciding on s_n, it considers $s_1, s_2, \cdots, s_{n-1}$ known and takes the s_n that corresponds to the smallest branch metric. Clearly, after NK steps we end up at one of the leaf nodes, but not necessarily in the one with the smallest cumulative metric. For example, in Fig. 4.1, considering the branch metrics for s_1 will result in detecting $s_1 = -1$. Next, deciding on s_2 given that $s_1 = -1$ will yield the detected vector as $(-1, -1)$ which is not the optimal answer, i.e., $(1, -1)$.

Thus, the problem with ZF-DF is error propagation. If, due to noise, an incorrect decision is taken in any of the steps, then this error will propagate and many of the subsequent symbols are likely to be detected wrong as well. The detection order can be optimized to minimize the effects of error propagation. It is best to start with the symbol for which ZF produces the most reliable result, i.e., the sub-channel with smallest noise variance, and then proceed to less and less reliable symbols. However, even with the optimal ordering, error propagation severely limits the performance [60].

Using the sampling method in Fig. 5.3, this serial approach can be either a forward SIC initiated from the first transmitted symbol, i.e., $s_1[1]$, or a backward SIC started from the last transmitted symbol, i.e., $s_K[N]$. For example, for forward processing, $s_1[1]$ can be detected by using $y_1[1]$ without interference, then $s_2[1]$ can be detected by canceling

the interference of $s_1[1]$ from $y_2[1]$, and so on. The same procedure can be performed backwards. One can also combine forward and backward operations. However, when hard decisions are used, such a combination will not result in a noticeable gain. On the other hand, by using soft decisions, combining the forward and backward operations will improve the results as explained in the following section.

4.3.4 Forward-Backward Belief Propagation Detection

In the previous section we introduced an SIC method which was performed by passing hard decisions of previously detected symbols to cancel the interference. In this section, we introduce a similar detection method which passes likelihood values.

In practice, each symbol s_k typically is composed of information-carrying bits. It is then of interest to take decisions on the individual bits, and often, also to quantify how reliable these decisions are. Such reliability information about a bit is called a "soft decision," and is typically expressed via the probability ratio. Hard decisions does not distinguish between two events of $p(s = 0|y) = 0.49, p(s = 0|y) = 0.51$ and $p(s = 0|y) = 0.01, p(s = 0|y) = 0.99$, while, the soft decisions take this difference into account. Therefore, soft decisions carry more information with respect to the hard decisions. By using likelihood values, instead of hard decisions, performance can be improved. Additionally, this method provides the opportunity to exploit benefits of backward processing as well. We explain the strategy of detecting BPSK modulation and $K = 2$, but it can be also generalized to other modulations and other values of K. We also assume that transmitted symbols have the same prior probabilities and calculate the conditional probabilities as follows:

$$\begin{cases} a = P(y_1[1]|s_1[1]=0) = \frac{1}{\sqrt{2\pi\rho_1\sigma_n^2}} \exp\left(-\frac{|y_1[1]+h_1\rho_1|^2}{2\rho_1\sigma_n^2}\right) \\ b = P(y_1[1]|s_1[1]=1) = \frac{1}{\sqrt{2\pi\rho_1\sigma_n^2}} \exp\left(-\frac{|y_1[1]-h_1\rho_1|^2}{2\rho_1\sigma_n^2}\right) \\ P_0^{fw}(s_1[1]) = P(s_1[1]=0|y_1[1]) = \frac{a}{a+b} \\ P_1^{fw}(s_1[1]) = P(s_1[1]=1|y_1[1]) = \frac{b}{a+b} \end{cases}$$

$$\begin{cases} c = P(y_2[1]|s_2[1]=0, y_1[1]) = \\ P_0^{fw}(s_1[1]) \frac{1}{\sqrt{2\pi\rho_2\sigma_n^2}} \exp\left(-\frac{|y_2[1]+h_1\rho_1+h_2\rho_2|^2}{2\rho_2\sigma_n^2}\right) + P_1^{fw}(s_1[1]) \frac{1}{\sqrt{2\pi\rho_2\sigma_n^2}} \exp\left(-\frac{|y_2[1]-h_1\rho_1+h_2\rho_2|^2}{2\rho_2\sigma_n^2}\right) \\ d = P(y_2[1]|s_2[1]=1, y_1[1]) = \\ P_0^{fw}(s_1[1]) \frac{1}{\sqrt{2\pi\rho_2\sigma_n^2}} \exp\left(-\frac{|y_2[1]+h_1\rho_1-h_2\rho_2|^2}{2\rho_2\sigma_n^2}\right) + P_1^{fw}(s_1[1]) \frac{1}{\sqrt{2\pi\rho_2\sigma_n^2}} \exp\left(-\frac{|y_2[1]-h_1\rho_1-h_2\rho_2|^2}{2\rho_2\sigma_n^2}\right) \\ P_0^{fw}(s_2[1]) = P(s_2[1]=0|y_1[1], y_2[1]) = \frac{c}{c+d} \\ P_1^{fw}(s_2[1]) = P(s_2[1]=1|y_1[1], y_2[1]) = \frac{d}{c+d} \end{cases}$$

where $\rho_i = \boldsymbol{U_0}(i,i)$. Using these successive calculations, $P_0^{fw}(s_k[n])$ and $P_1^{fw}(s_k[n])$ can be found for all values of $1 \leq n \leq N$ and $1 \leq k \leq K$. As explained before, due to the structure of the sampling method in Fig. 5.3, the last transmitted symbol can also be detected without interference and the same procedure can be applied backward to find $P_0^{bw}(s_k[n])$ and $P_1^{bw}(s_k[n])$. Using either of these likelihood sets as a detection metric will result in an improvement over the hard-decision SIC method that was presented in the previous section. Moreover, the performance can be furthered improved if we use forward and backward operations together and define the detection metric as:

$$P_0(s_k[n]) = P_0^{fw}(s_k[n]) P_0^{bw}(s_k[n])$$
$$P_1(s_k[n]) = P_1^{fw}(s_k[n]) P_1^{bw}(s_k[n])$$

4.4 Performance Analysis

In this section, the achievable performance by the asynchronous transmission is analyzed. We use Bit Error Probability (BEP) as the performance criterion. In the analysis of BEP, one important factor is diversity gain which specifies the reliability of the transmission.

Unlike the AWGN channel, the fading channel suffers from sudden declines in the power. This is due to the destructive addition of multi-path signals in the propagation media. Therefore, the effective signal-to-noise ratio (SNR) at the receiver can go through deep fades and be dropped dramatically. Usually we can assume a threshold for the received SNR in which the receiver can reliably detect and detect the transmitted signal. If the received SNR is lower than such a threshold, a reliable recovery of the transmitted signal is impossible. and this event is called an "outage." The probability of outage, which is the probability of having a received power lower than the given threshold, can be calculated based on the statistical model of the channel or based on the actual measurements of the channel.

One way to combat the outage events is to provide different replicas of the transmitted signal to the receiver, which is the main idea behind "diversity". If these different replicas fade independently, it is less probable to have all copies of the transmitted signal in deep fade simultaneously. Therefore, the receiver can reliably detect the transmitted signal using these received signals. To define diversity quantitatively, we use the relationship between the received SNR, denoted by δ , and the probability of error, denoted by P_e. A tractable definition of the diversity, or diversity gain, is

$$D = -\lim_{\delta \to \infty} \frac{\log P_e}{\log \delta} \tag{4.21}$$

where P_e is the error probability at an SNR equal to δ . In other words, diversity is the

slope of the error probability curve in terms of the received SNR in a log-log scale [25].

4.4.1 Diversity Analysis for a Receiver with Multiple Antennas and ZF detection

There is a fundamental difference between synchronous and asynchronous ZF. In the synchronous ZF, we need more number of receive antennas than transmit antennas/users to be able to cancel interference completely. However, in the asynchronous ZF, even with one receive antenna, we are able to remove all the interference completely. The required dimension at the receiver signal space is provided by asynchronous transmission and oversampling. Denoting M as the number of receiver antenna, it is well-known in the literature that, synchronous ZF provides $M-K+1$ diversity orders when K independent symbols are transmitted. $K-1$ diversity orders are wasted to remove the interference [61]. In more details, transmitting S ($S \leq K \leq M$) independent symbols will result in $(M-S+1)(K-S+1)$ diversity orders [62]. On the other hand, asynchronous ZF provides additional degrees of freedom by means of oversampling which enable us to enjoy full diversity of M while transmitting K independent symbols.

Here, we present the system model when multiple receive antennas are used at the receiver. To represent the system model we use the sampling method described in Eq. (4.5). By stacking output samples of all receive antennas together we can represent the system model as follows:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R'} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R'} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{R'} \end{bmatrix} \begin{bmatrix} \boldsymbol{H_1} \\ \boldsymbol{H_2} \\ \vdots \\ \boldsymbol{H_M} \end{bmatrix} \boldsymbol{s} + \begin{bmatrix} \boldsymbol{n_1} \\ \boldsymbol{n_2} \\ \vdots \\ \boldsymbol{n_M} \end{bmatrix}$$

$$\boldsymbol{y_T} = \boldsymbol{U_T H_T s} + \boldsymbol{n_T}$$

where M is number of receive antennas, $\boldsymbol{H_m}$, $\boldsymbol{y_m}$ and $\boldsymbol{n_m}$ are the effective channel matrix, the received sample vector and the noise vector, respectively, at mth receive antenna. Then, by performing the ZF detection, we have:

$$\tilde{\boldsymbol{y}} = \boldsymbol{s} + \tilde{\boldsymbol{n}} \tag{4.22}$$

where $\boldsymbol{COV_{\tilde{n}}}$ can be calculated as:

$$\boldsymbol{COV_{\tilde{n}}} = E\{\tilde{\boldsymbol{n}}\tilde{\boldsymbol{n}}^H\} = \sigma_n^2(\sum_{m=1}^{M} \boldsymbol{H_m^* R' H_m})^{-1} \tag{4.23}$$

Unfortunately, due to the complex structure of $\left(\sum_{m=1}^{M} \boldsymbol{H_m^* R' H_m}\right)^{-1}$ for $M \geq 1$, finding the exact expression of BEP for $M \geq 1$ is not easy. We derive an upper bound on BEP by finding an upper bound on the diagonal elements of $\boldsymbol{COV_{\tilde{n}}}$ and show that full diversity is achieved. Because $\boldsymbol{R'}$ is positive definite, for every $1 \leq m \leq M$, $\boldsymbol{H_m^* R' H_m}$ is also positive definite. Therefore, we can apply the following lemma.

Lemma 4.1. *For n positive definite matrices $\boldsymbol{A_i}, 1 \leq i \leq n$, we have:*

$$(\sum_{i=1}^{n} \boldsymbol{A_i})^{-1} \leq \sum_{i=1}^{n} \boldsymbol{A_i}^{-1} \tag{4.24}$$

where $\boldsymbol{B} \leq \boldsymbol{C}$ means that $\boldsymbol{C} - \boldsymbol{B}$ is positive semidefinite.

Proof. This lemma is a straightforward result of the following inequality, which can be found in [63].

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} \leq \boldsymbol{A}^{-1} \quad \boldsymbol{A}, \boldsymbol{B}: \text{ positive definite matrices}$$

□

As a result, we can conclude that $\boldsymbol{COV}_{\tilde{\boldsymbol{n}}} \leq \sigma_n^2 \sum_{i=1}^{M} (\boldsymbol{H_i}^* \boldsymbol{R'} \boldsymbol{H_i})^{-1}$. This inequality implies that the diagonal elements of the covariance matrix of noise are upper bounded as follows:

$$\boldsymbol{COV}_{\tilde{\boldsymbol{n}}}(i,i) \leq \frac{\sigma_n^2 \boldsymbol{R'}^{-1}(i,i)}{\sum_{m=1}^{M} |h_{\pi(i),m}|^2} \quad 1 \leq i \leq NK \tag{4.25}$$

where $\pi(.)$ is the assigning function that assigns each sub-channel to a specific user, i.e., $\pi : \mathcal{I} \rightarrow \mathcal{K}$, where $\mathcal{I} = \{1, \cdots, NK\}$ and $\mathcal{K} = \{1, \cdots, K\}$ are the set of sub-channel and user indices, respectively. Also, $h_{k,m}$ represents the channel coefficient between User k and Receive Antenna m.

The upper-bound for the BEP expression for each sub-channel is equal to:

$$BEP_{u,i} = \frac{\sqrt{\frac{2\delta_0}{\pi \boldsymbol{R'}^{-1}(i,i)}}}{2\left(1 + \frac{2\delta_0}{\boldsymbol{R'}^{-1}(i,i)}\right)^{M+\frac{1}{2}}} \frac{\Gamma(M+\frac{1}{2})}{\Gamma(M+1)} \times {}_2F_1(1, M+\frac{1}{2}; M+1; \frac{1}{1+\frac{2\delta_0}{\boldsymbol{R'}^{-1}(i,i)}}) \tag{4.26}$$

where $\delta_0 = \frac{E[|s_k[i]|^2]}{\sigma_n^2}$. The details of derivation can be found in Appendix B.1. The actual BEP for each subchannel is upper-bounded by $BEP_{u,i}$, i.e., $BEP_i \leq BEP_{u,i}$. If we define $D_i = -\lim_{\delta_0 \rightarrow \infty} \frac{\log BEP_i}{\log \delta_0}$ and $d_i = -\lim_{\delta_0 \rightarrow \infty} \frac{\log BEP_{u,i}}{\log \delta_0}$, it is clear that $D_i \geq d_i$. By using the fact that the hypergeometric function of form ${}_2F_1(1, m+\frac{1}{2}; m+1; \frac{1}{1+c})$ converges to one as c grows large [64], we can calculate that $d_i = M$. Therefore, the diversity of the ith subchannel is greater than or equal to M. On the other hand, M is the maximum available diversity for this system, which completes the proof of achieving full diversity, i.e. $D_i = M$.

4.4.2 Effect of Time Delays on Performance

In this section, we calculate the optimal values of delays for the ZF detection in order to achieve the lowest average BEP with one receive antenna at high SNR. Because for $M = 1$ the inequality in Eq. (4.25) turns into equality, the exact BEP expression for each sub-channel can be obtained as:

$$BEP_i = \frac{\sqrt{\frac{2\delta_0}{\pi \boldsymbol{R'}^{-1}(i,i)}}}{2\left(1+\frac{2\delta_0}{\boldsymbol{R'}^{-1}(i,i)}\right)^{3/2}} \frac{\Gamma(3/2)}{\Gamma(2)}\ {}_2F_1(1,3/2;2;\frac{1}{1+\frac{2\delta_0}{\boldsymbol{R'}^{-1}(i,i)}})$$

Approximating BEP_{avg} at high SNR for one receive antenna results in: (see Appendix B.2 for more details)

$$\widetilde{BEP}_{avg} = \frac{1}{4\sqrt{\pi}NK} \frac{\Gamma(3/2)}{\Gamma(2)} \times \frac{\sum_i \boldsymbol{R'}^{-1}(i,i)}{\delta_0},$$

For a fixed number of users and frame length, in order to minimize $\widetilde{BEP}_{avg}$, we need to minimize the trace($\boldsymbol{R'}^{-1}$) which is related to time delays between different users. In what follows, we derive the relationship between the trace($\boldsymbol{R'}^{-1}$) and time delays, and consequently find the optimum time delays.

Lemma 4.2. *the sum of the diagonal elements of the inverse of matrix $\boldsymbol{R'}$ is equal to:*

$$trace(\boldsymbol{R'}^{-1}) = \frac{(N-1)(N+1)}{3(1+\tau_{1,K})} + \frac{2N+1}{3(N+1+\tau_{1,K})} + \frac{N(N+2)}{3}\sum_{i=1}^{K-1}\frac{1}{\tau_{i+1,i}} \quad (4.27)$$

where $\tau_{i+1,i} = \tau_{i+1} - \tau_i$. The proof is presented in Appendix B.3.

Theorem 4.1. *The optimum time delays which result in the lowest average BEP for ZF*

detection at high SNR are: (τ_1 is assumed to be zero)

$$\tau_{i-1} = \frac{i-2}{i-1} \times \tau_i \quad 3 \leq i \leq K \tag{4.28}$$

Also τ_K is found by solving the following equation:

$$A\tau_K^4 + B\tau_K^3 + C\tau_K^2 + D\tau_K + E = 0 \tag{4.29}$$

where

$$A = (1-(K-1)^2)\frac{(N+2)}{3}.$$
$$B = \frac{-2}{3}(1-(K-1)^2)N^2 + 2(4(K-1)^2-1)\frac{(N+1)}{3}.$$
$$C = \frac{1}{3}(1-(K-1)^2)N^3 + \frac{2}{3}(1-4(K-1)^2)N^2 - 2(K-1)^2(3N+2).$$
$$D = \frac{2}{3}(K-1)^2(N^3+5N^2+8N+4).$$
$$E = -\frac{1}{3}(K-1)^2(N^3+4N^2+5N+2).$$

The proof is easily obtained by taking the derivation of Eq. (4.27) with respect to time delays.

For $K = 2$, A will be zero and Eq. (4.29) is a polynomial of degree 3 which has a closed-form solution as follows:

$$\tau_{opt} = \frac{N+2-\sqrt[3]{N^3+1.5N^2-1.5N-1}}{3} \tag{4.30}$$

where N is the block length. However, for other values of K, Eq. (4.29) should be solved numerically. After finding τ_K, the remaining time delays are calculated recursively using Eq. (4.28). The optimum delay values for different K and N values are reported in Tables 4.1 and 4.2. Optimum time delays approach uniform time delays, i.e, $\tau_k = \frac{k-1}{K}$, $2 \leq k \leq K$, as N increases. The effects of time delay values on the performance are

Table 4.1: Optimum Time Delays when $K = 2$

Case	N=10	N=32	N=64	N=128	N $\to \infty$
K=2	0.5240	0.5077	0.5039	0.5019	0.5

Table 4.2: Optimum delays when $N = 128$

Case	N=128
K=4	[0.2505,0.5010,0.7514]
K=6	[0.1669,0.3338,0.5006,0.6675,0.8344]
K=8	[0.1251,0.2502,0.3754,0.5004,0.6256,0.7507,0.8758]

studied numerically in the following section.

In this section, we showed that unlike synchronous ZF, asynchronous ZF can provide full diversity of M. If we assume $M > K$, then transmitting independent symbols from different users and using ML detection can provide diversity order of M. However, conventional ZF detection wastes $K - 1$ diversity orders to null the interference resulting in diversity order of $M - K + 1$. On the other hand, if $M < K$, we are only able to transmit maximum number of M symbols which is apprehend as the multiplexing gain. At the receiver side, M diversity gain is available which can be exploited by ML detection, however, the ZF detection is unable to detect the transmitted symbols because the null space of the channel matrix is empty. These diversity results are summarize in Table 4.3 In summary,

Table 4.3: Receive Diversity Gain results

Diversity gain	$M > K$	$M < K$
synchronous ML	M	M
synchronous ZF	$M - K + 1$	unable to detect
Asynchronous	M	M

a MIMO system can provide two types of gains: diversity gain and spatial multiplexing gain. Given a MIMO channel, both gains can be simultaneously obtained, but there is a fundamental trade-off between them: higher multiplexing gain comes at the price of sac-

rificing diversity [65]. However, by means of asynchronous transmission, another domain for exploiting multiplexing gain is provided which let us enjoy full diversity provided by spatial domain without sacrificing the multiplexing gain.

4.5 Simulation Results and Discussions

In this section, we provide simulation results in order to validate our theoretical results and compare different methods. We present the BEP results in the Rayliegh fading channel. In all simulations, channel coefficients are independent Rayliegh fadings with variance one, fixed during the block and changing independently for each block. All users have the same average power of one and variance of noise (σ_n^2) is equal to $10^{\frac{-\text{SNR}}{10}}$ where SNR is in dB. To avoid inter-block interference, the last symbol of each block should be idle for asynchronous methods. This will reduce spectral efficiency, but it is negligible for large block lengths. In all simulations, the block length is 128 and the time delays are uniform except in the case where we report the time delays to study their effects on the performance. The number of users and the number of receive antennas is denoted by K and M, respectively. When M is not specified, the assumption is that only one receive antenna is used. Transmitted symbols are chosen from BPSK modulation and the comparing criterion is the average bit error rate among all the users.

In Fig. 4.3, we compare the performance of the asynchronous MLSD method with that of the synchronous ML. Asynchronous MLSD outperforms synchronous ML detection with similar complexity. Fig. 4.3 also includes the single-user bound for a better comparison. As can be seen in the figure, asynchronous MLSD for $K = 2$ achieves performance of the single user system at high SNR. Fig. 4.4 shows the performance of different SIC methods presented in Section 4.3.3. Our new forward-backward belief propagation method using

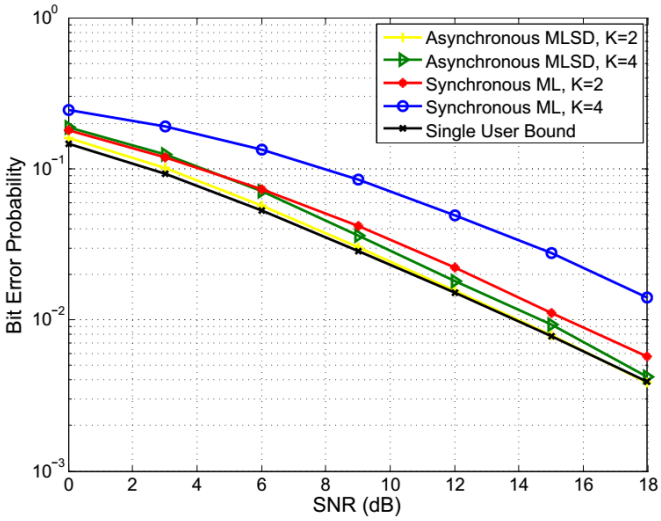


Figure 4.3: Comparing asynchronous MLSD and synchronous ML

the sampling method in Fig. 5.3 improves the performance of traditional SIC method by about 3 dB. Fig. 4.5 compares the performance of the synchronous and asynchronous

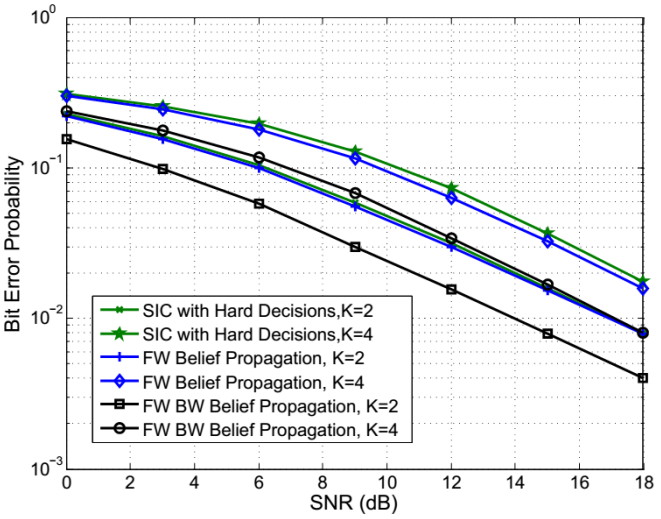


Figure 4.4: Performance of SIC method with hard decisions and soft decisions

ZF detectors. Although asynchronous ZF is even possible with one receive antenna, for fair comparison, we consider the cases where the number of receive antennas and users are the same. Since all users are assumed to have the same transmit power, synchronous ZF for $(K = 2, M = 2)$ and $(K = 4, M = 4)$ provides the same performance and both of

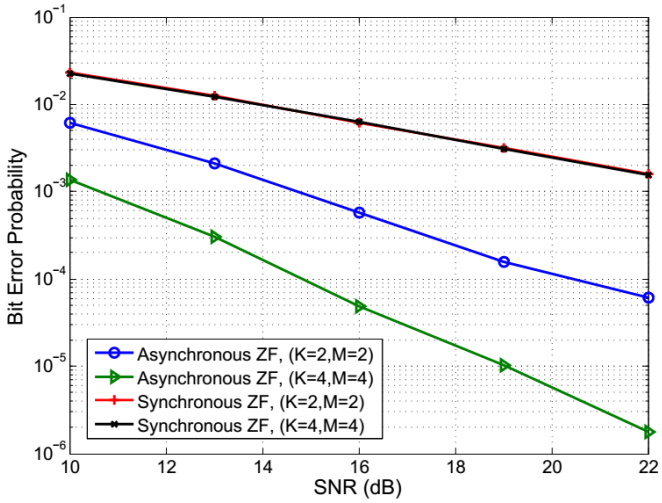


Figure 4.5: Comparing synchronous and asynchronous ZF

them have diversity of one. However, for asynchronous ZF detection, diversity of 2 and 4 is achieved for $(K = 2, M = 2)$ and $(K = 4, M = 4)$, respectively.

We study the effects of time delay values on the performance of a ZF system with $K = 4$ users and one receive antenna in Fig. 4.6. Note that a synchronous ZF solution does not exist in this case as we need at least $M = 4$ receive antennas. We show the results for six different sets of time delays. For optimum time delays we use the result of Section 4.4.2 as reported in Table 4.2. The curve associated with random time delays represents the average performance over uniformly distributed random time delays. The remaining sets of time delays are specified in the figure. The optimum time delays and time delays of $[0.01, 0.1, 0.9]$ have the best and worst performances, respectively. They also have the lowest and the highest trace$(\boldsymbol{R'}^{-1})$, respectively, which are presented along with other sets of time delays in Table 4.4. As can be seen, a lower trace$(\boldsymbol{R'}^{-1})$ results in a better performance. This observation is in line with the analysis in Section 4.4.2 where trace$(\boldsymbol{R'}^{-1})$ was introduced as a criterion to compare the performance of different time delays.

Finally, to compare different methods with each other, we include the performance of all detection methods for $K = 2$ in Fig. 4.7. Both MLSD and forward-backward BP

Table 4.4: Comparing trace($\boldsymbol{R'}^{-1}$) for different time delays in Fig. 4.6

Time delays	trace($\boldsymbol{R'}^{-1}$)
$[0.2505, 0.5010, 0.7514]$	8.8404×10^4
$[0.4, 0.6, 0.8]$	9.6639×10^4
$[0.1, 0.4, 0.7]$	1.1065×10^5
$[0.1, 0.2, 0.9]$	1.7347×10^5
$[0.01, 0.1, 0.9]$	6.7784×10^5

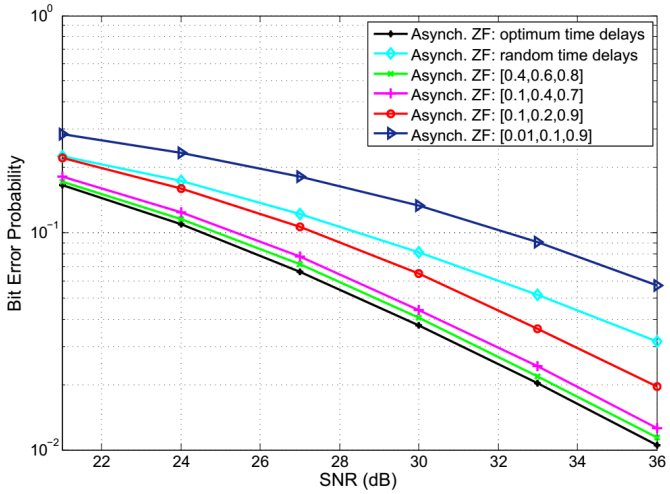


Figure 4.6: Effect of time delays in asynchronous ZF detection for $K = 4$

detection methods not only outperform the synchronous ML detection, but also achieve the performance of the single user system. In addition, the low complexity method of SIC with hard decisions also provides good performance.

4.6 Conclusion

In this chapter, we studied the benefits of asynchrony when multiple users are sending data simultaneously to a common receiver. Instead of treating asynchrony as a disruptive factor, we exploited it as an additional resource to cancel interference. We have shown that asynchrony between data streams adds a favorable ISI, which can be exploited to

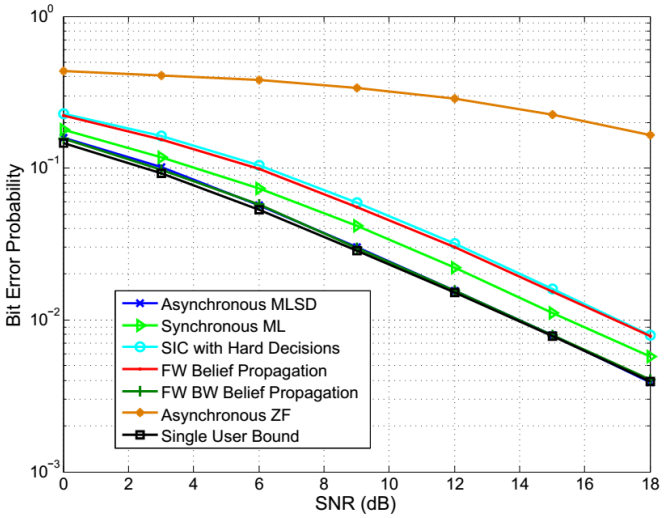


Figure 4.7: Comparison of all detection methods for $K = 2$

cancel the interference. It also introduces memory to the system, which can be exploited by methods such as maximum-likelihood sequence detection. Exact BEP expression for ZF detection was derived, and it was verified that a diversity equal to the number of receive antennas is achievable by asynchronous transmission.

Chapter 5

Effect of Asynchrony on the Achievable Rates of Broadcast Channels

In this chapter, we investigate the effect of time asynchrony in non-orthogonal multiple access (NOMA) schemes for downlink transmissions. First, we analyze the benefit of adding intentional timing offsets to the conventional power domain-NOMA (P-NOMA). This method, called Asynchronous-Power Domain-NOMA (AP-NOMA), introduces artificial symbol-offsets between packets destined for different users. It reduces the mutual interference, which results in enlarging the achievable rate-region of the conventional P-NOMA. Then, we propose a precoding scheme that fully exploits the degrees of freedom provided by the time asynchrony. We call this multiple access scheme T-NOMA, which provides higher degrees of freedom for users compared to the conventional P-NOMA or even the modified AP-NOMA. T-NOMA adopts precoding at the base station and a linear preprocessing scheme at the receiving user, which decomposes the broadcast channel into

parallel channels circumventing the need for Successive Interference Cancellation (SIC). The numerical results show that T-NOMA outperforms AP-NOMA and both outperform the conventional P-NOMA. We also compare the maximum sum-rate and fairness provided by these methods. Moreover, the impact of pulse shape and symbol offset on AP-NOMA and T-NOMA schemes' performance is investigated.

5.1 Introduction

For future radio access, significant gains in the system capacity/efficiency and quality of user experience are required. In particular, the multiple access approach is a crucial part of radio access technology [66]. In [17, 18], and the references therein, NOMA is proposed as a candidate for the future radio access to partially fulfill the requirements of next-generation networks. The currently prevailing approach for multiple access lies in the category of orthogonal multiple access (OMA). In 2G systems, time division multiple access (TDMA) is adopted. In the 3G mobile communication systems such as W-CDMA and CDMA2000, direct sequence-code division multiple access (DS-CDMA) is used, and the receiver is based on simple single-user detection using the Rake receiver. OMA based on orthogonal frequency division multiple access (OFDMA) or single carrier-frequency division multiple access (SC-FDMA) is used in the 4th generation mobile communication systems such as LTE and LTE-Advanced. These approaches first partition resources into orthogonal resource blocks and then assign each resource block exclusively to one user. After this, the problem reduces to a point-to-point (P2P) communication problem, and then well developed single-user encoders/decoders can be applied. The significant advantage of OMA methods is that their complexity is merely the complexity of single-user encoders/decoders. On the other hand, assigning resource blocks exclusively to one user can be very inefficient (in terms of achievable rate-regions) and may pose a severe

fairness problem among users. In contrast to OMA, NOMA allows users to utilize the same resource blocks for transmission simultaneously and therefore is potentially more efficient. When evaluated under the LTE system characteristics, NOMA demonstrates significant gains over OMA systems [19, 20].

The problem of communicating with many receivers arises in "downlink" scenarios such as communication from an access point to stations in WiFi or from a base station in cellular systems. Although OMA approaches eliminate interference between transmissions, it does not, in general, achieve the highest possible transmission rates for a given packet error rate [22]. Superposition Coding (SC) is a well-known non-orthogonal scheme that achieves the capacity on a scalar Gaussian broadcast channel [21, 67]. Superposition coding is a technique of simultaneously communicating information to several receivers by a single source. In other words, it allows the transmitter to send the information of multiple users at the same time and frequency. At the receiver's side, successive interference cancellation (SIC) is applied, which exploits the differences in signal strength among the signals of interest [25]. The basic idea behind SIC is that user signals are successively decoded. Superposition coding and SIC are the optimal encoder/decoder methods for degraded broadcast channels where users can be ordered in terms of the quality of the received signals [68, 22]. This concept is of particular importance in cellular systems where the channel conditions vary significantly among users due to the near-far effect[18]. From an information-theoretic perspective, NOMA with a SIC is an optimal multiple access scheme from the viewpoint of the achievable multiuser capacity region, in the downlink [69, 70, 71, 72, 73] and in the uplink [74].

Applications of NOMA in the downlink scenario have been widely studied [75]. In [76] and [77], various power allocation and user scheduling algorithms were proposed to improve the sum-rate of the NOMA-based multiuser system. In [78], cooperative NOMA scheme was investigated to improve the spectral efficiency and transmission reliability.

In [79], the trellis-coded NOMA method is introduced, and in [80], a block-based non-orthogonal multicarrier Scheme is proposed. More recently, the study of the combination of multiple-input and multiple-output (MIMO) and NOMA has received considerable attention [81], [82]. However, most of the previous research on NOMA only considered symbol-synchronous transmission. Often in the literature, timing mismatch is considered as an impairment, and different synchronization methods are applied to eliminate it [1]. However, in this work, we show that time asynchrony can indeed be beneficial. Time asynchrony can decrease interference and provide additional degrees of freedom using proper transmission and receiver design.

The usefulness of timing offset, or time asynchrony, has been studied in the literature. For example, the results in [83] show that time asynchrony can increase the capacity region in multiple-access channels. Also, the effect of time asynchrony is other contexts are discussed in [84, 46, 47, 48, 4, 40, 85, 86]. In this chapter, we introduce two schemes called AP-NOMA and T-NOMA which use time asynchrony to enlarge the downlink rate-region and thus improve the user experience [87].

The rest of the chapter is organized as follows. In Section 5.2, we provide an overview of the concepts of NOMA, including superposition coding and SIC. Next, we provide some insights regarding the benefits of time asynchrony. The concept behind AP-NOMA is the reduction in IUI by using intentional time delays, and the concept behind T-NOMA is exploiting extra degrees of freedom provided by time asynchrony, explained in Section 5.3. Then, we present the system model and its characteristics in Section 5.4. We present the achievable rate-region results for the conventional P-NOMA, AP-NOMA, and T-NOMA in Section 5.5. In the end, we provide the numerical results and the final remarks in Sections 5.6 and 5.7, respectively. The materials in this chapter are partially published in [88, 89].

5.2 Concepts of P-NOMA

Simultaneous transmission of information from one source to several receivers has been studied under the title of broadcast channel [90]. Superposition coding at the transmitter and SIC at the receivers provide the capacity achieving performance and thus play important roles in the P-NOMA. In this section, we briefly summarize the results in the literature and explain the concepts of superposition coding and SIC method. Consider the Gaussian broadcast channel

$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2$$

where Z_1 and Z_2 follow Gaussian distributions, i.e., $Z_1 \sim N(0, N_1)$ and $Z_2 \sim N(0, N_2)$, assuming that $N_2 > N_1$.

Theorem 5.1. *The capacity region for the Gaussian broadcast channel, with signal power constraint P, is given by:*

$$R_1 \leq \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_1}\right)$$
$$R_2 \leq \frac{1}{2}\log\left(1 + \frac{(1-\alpha)P}{\alpha P + N_2}\right)$$

where $\alpha \in [0, 1]$ represents the power allocation coefficient. This region is achieved by the superposition coding and SIC schemes described in [21].

To encode the messages, the transmitter generates two codebooks, one with power αP at rate R_1, and another codebook with power $(1-\alpha)P$ at rate R_2, where R_1 and R_2 lie in the capacity region above. Then, to send an index $w_1 \in \{1, 2, \cdots, 2^{nR_1}\}$ and

$w_2 \in \{1, 2, \cdots, 2^{nR_2}\}$ to Y_1 and Y_2, respectively, the transmitter takes the codeword $X(w_1)$ from the first codebook and codeword $X(w_2)$ from the second codebook and computes the sum. The transmitter sends the sum over the channel [68]. An example of superposition Coding using 4-PSK and 8-QAM constellations for two users are shown in Fig. 5.1. The input bits of the user with weaker channel are modulated with 4-PSK modulation as shown by the coarse points in Fig. 5.1 and the input bits for the user with stronger channel are modulated with 8-QAM modulation and sum of the modulated symbols results in a 32-QAM constellation shown by the fine points in Fig. 5.1.

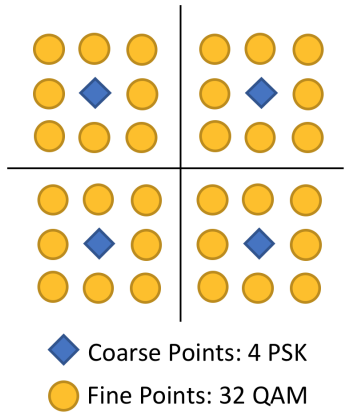
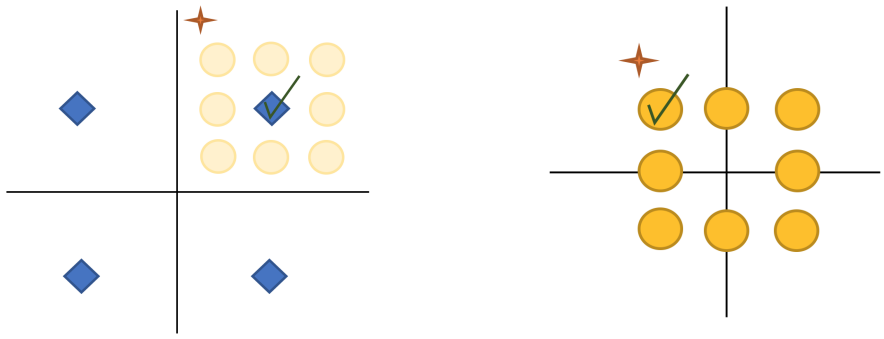


Figure 5.1: An Example of Superposition Coding

The receivers must now decode the messages. The weak receiver, Y_2, merely looks through the second codebook to find the closest codeword to the received vector Y_2. His effective signal-to-noise ratio is $\frac{(1-\alpha)P}{\alpha P+N_2}$, since Y_1's message acts as noise to Y_2. The strong receiver, Y_1, first decodes Y_2's codeword, which he can accomplish because of his lower noise N_1. After subtracting this codeword $\hat{X}_2$ from Y_1, the receiver looks for the codeword in the first codebook closest to $Y_1 - \hat{X}_2$. The resulting probability of error can be made as low as desired. A nice dividend of optimal encoding for degraded broadcast channels is that the strong receiver Y_1 always knows the message intended for receiver Y_2 in addition to its own message [68]. Fig. 5.2 presents the technique for decoding the superposed signal

(Fig. 5.1) at the receiving side. As shown in Fig. 5.2, the weak receiver only decodes the coarse points by mapping the received signal to the nearest point in the corresponding constellation (4-PSK). The stronger user is also able to decode the coarse points and after subtracting the decoded symbol from the received signal, the resulting signal is decoded using the corresponding constellation (8-QAM) as shown in Fig. 5.2b.



(a) Decoding at the weak user/first step at the strong user

(b) Second step of decoding at the strong user

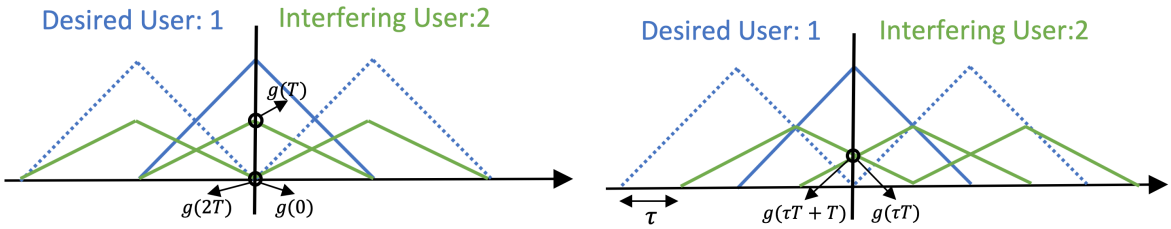
Figure 5.2: An example of SIC decoding

5.3 Motivations Behind Asynchronous Transmission

It is shown in the literature that time asynchrony which is the intrinsic feature of most of the wireless networks, not only is not disruptive, but also can be beneficial if the proper processing is employed [84, 46, 47, 48, 4, 40, 85]. We introduce two NOMA schemes enjoying the benefits of time asynchrony, i.e., AP-NOMA and T-NOMA. The details of these two methods will be elaborated later, but before that, let us briefly express the intuition and motivations behind each of the mentioned methods.

5.3.1 Motivation Behind AP-NOMA

In conventional P-NOMA, the underlying assumption is the reception with perfect synchronization. In the perfect synchronous scenario, the peak points of all users are aligned, however, by adding intentional time delays to each user, the peak points drift apart. The comparison between synchronous and asynchronous reception for two users is shown in Fig. 5.3.



(a) Synchronous Reception (b) Asynchronous Reception

Figure 5.3: Illustration of IUI for two cases of synchronous and asynchronous reception

Denoting the effective pulse shape as $g(t)$ and assuming normalized symbols, the interuser interference (IUI) power from interfering user can be calculated as:

$$IUI(\tau) = \sum_{i=-\infty}^{\infty} |g(\tau + iT)|^2 \tag{5.1}$$

where τ is the time delay and T is the symbol interval. In the next lemma, we show that adding intentional time delay will decrease the IUI power.

Lemma 5.1. *For any pulse shape with real spectrum in frequency domain (i.e., real and even in the time domain), we will have*

$$IUI(\tau) \leq IUI(0) \tag{5.2}$$

Proof. The proof is presented in Appendix C.1. □

Note that the conditions mentioned above encompasses almost all the the pulse shapes in the literature like rectangular and sinc pulse shapes and even practical pulse shapes like the raised cosine pulse shape. Thus, adding time asynchrony can decrease the IUI which is the main degradation in NOMA schemes. However, the benefits of time asynchrony is not limited to decreasing IUI power. In fact, as we will explain later, adding time asynchrony provides additional degrees of freedom which can be exploited to serve more users over the same time and frequency resources.

5.3.2 Motivation Behind T-NOMA

Here, we use Hilbert space formulation to show the potential of asynchronous transmission in providing additional degrees of freedom. Hilbert space generalizes the Euclidean space of real numbers to finite energy signals. Each finite energy signal can be represented by a vector in the Hilbert space with each coordinate given by an inner product with the corresponding orthonormal basis functions. In more details, any finite energy signal like $s(t)$ can be written as a linear combination of the orthonormal basis functions as:

$$s(t) = \sum_{n \in \Gamma} s[n] p_n(t) \tag{5.3}$$

where $p_n(t)$ is an orthonormal basis function, i.e, $\langle p_n(t).p_m(t) \rangle = \delta[n-m]$ and $s[n]$ is the corresponding coefficient in the direction of the basis function $p_n(t)$ which is obtained by the following inner product.

$$s[n] = \langle s(t).p_n(t) \rangle \tag{5.4}$$

If we further constrain our finite energy signals to strictly band-limited ones, then the Hilbert Space is called the Paley-Wiener space. The Nyquist sampling theorem states that

any signal in Paley-Wiener space whose Fourier transform is supported on $f \in (-W\ W)$ can be written as the linear combination of some sinc pulses, i.e,

$$s(t) = \sum_{n=-\infty}^{\infty} s[n] \left(\sqrt{2W} sinc(2W(t - nT)) \right) \tag{5.5}$$

where T is the Nyquist interval, i.e., $T = \frac{1}{2W}$ [91]. Due to completeness of the of sinc pulses, all band-limited signals, even sinc pulses that do not lie at integer multiples of T, e.g., their fractionally-shifted version, i.e., $sinc(2W(t - nT - \tau T))$, still lie completely in the Paley-Wiener space.

Assume that, in Eq. (5.5), $s[n]$ is the transmitted symbol modulated on the sinc pulse $p_n(t) = \sqrt{2W} sinc(2W(t-nT))$. In practice, pulses spanning an unlimited time domain are not feasible, hence, they are usually truncated within a desired interval. Assume that the transmission interval is truncated into NT seconds, then we are capable of transmitting approximately $2WNT$ symbols. In other words, $2WNT$ dimensions is used in the case of finite-time transmissions [92]. However, due to the truncation, the finite set of sinc pulses, i.e.:

$$S \equiv \left\{ \begin{array}{c} p_n(t) = \sqrt{2W} sinc(2W(t - nT)) \\ n = 0, \cdots, 2WNT - 1 \end{array} \right\} \tag{5.6}$$

is not complete anymore and does not span the whole signal space. Therefore, we can insert additional pulses to exploit more signaling dimensions which leads to higher data throughput [93].

For example, defining $b_{2WNT} = \sqrt{2W} sinc(2W(t - \tau T))$ and applying the well-known Gram-Schmidt orthogonalization process, provide us an orthonormal basis function with size $2WNT + 1$. The newly formed basis function exploits an additional signaling dimen-

sion. We provide a numerical example next.

Example 5.1. *Assume that* $W = 0.5, N = 5, T = 1$, *then* $p_n(t) = sinc(t-n)$ *for* $n = 0, \cdots, 4$ *with truncation length of* $5T$. *The Gram matrix of the aforementioned set is equal to:*

$$\boldsymbol{G_S} = \begin{pmatrix} 0.959 & 0.052 & -0.084 & 0.057 & -0.013 \\ 0.052 & 0.959 & 0.052 & -0.084 & 0.057 \\ -0.084 & 0.052 & 0.959 & 0.052 & -0.084 \\ 0.057 & -0.084 & 0.052 & 0.959 & 0.052 \\ -0.013 & 0.057 & -0.084 & 0.052 & 0.959 \end{pmatrix} \tag{5.7}$$

If the pulse shapes were strictly band-limited, i.e., unlimited time support, matrix $\boldsymbol{G_S}$ *would be the identity matrix. In addition, with unlimited time support, any other function like* $p_6(t) = sinc(t-0.5)$ *can be written as* $\sum_{n=-\infty}^{\infty} a_n sinc(t-n)$ *where* $a_n = sinc(n-0.5)$. *On the other hand, in a time-limited scenario,* $p_5(t)$ *cannot be written as the weighted sum of truncated sinc functions. Thus, performing the Gram-Schmidt process, we can get the following orthonormal set:*

$$\left\{ \begin{array}{c} p_0'(t) = sinc(t), e_0(t) = \frac{p_0'(t)}{|p_0'(t)|} \\ p_1'(t) = p_1(t) - 0.053e_0(t), e_1(t) = \frac{p_1'(t)}{|p_1'(t)|} \\ p_2'(t) = p_2(t) + 0.086e_0(t) - 0.058e_1(t), e_2(t) = \frac{p_2'(t)}{|p_2'(t)|} \\ p_3'(t) = p_3(t) - 0.059e_0(t) + 0.089e_1(t) - 0.064e_2(t), e_3(t) = \frac{p_3'(t)}{|p_3'(t)|} \\ p_4'(t) = p_4(t) + 0.014e_0(t) - 0.06e_1(t) + 0.091e_2(t) - 0.066e_3(t), e_4(t) = \frac{p_4'(t)}{|p_4'(t)|} \\ p_5'(t) = p_5(t) - 0.647e_0(t) - 0.612e_1(t) + 0.191e_2(t) - 0.099e_3(t) + 0.06e_4(t), e_5(t) = \frac{p_5'(t)}{|p_5'(t)|} \end{array} \right\}$$

Now we have a new set with six elements, i.e., $S' \equiv \{e_n(t), n = 0, \cdots, 5\}$, *for which the Gram matrix is almost the identity matrix.*

We can continue this procedure and take advantage of the rest of the available signaling dimensions. In fact, it is shown in the literature that the available degree of freedom in

a time-limited channel is unbounded [8]. We will use asynchronous transmission to fully exploit the available degrees of freedom in a NOMA framework.

5.4 System Model

After performing coding and modulation, the modulated symbols intended for each user, namely, User k, are shaped with appropriate waveforms suited to the communication channel. We denote the block length by N, then the intended signal for User k will be:

$$s_k(t) = \sum_{n=1}^{N} s_k[n]p(t - nT) \tag{5.8}$$

where $p(t)$ is the pulse shape, e.g., r.r.c., which is truncated and its length is denoted by T_p and T is the symbol interval. The transmit power assigned to $s_k(t)$ is denoted by P_k which is calculated as:

$$P_k = E\left[\int_{-\infty}^{\infty} s_k(t)s_k^*(t)dt\right] \tag{5.9}$$

The transmitted signal from BS will be the super-position of signals from all users, i.e.:

$$s_{synch}(t) = \sum_{k=1}^{K} s_k(t) \tag{5.10}$$

where the transmit power, i.e., $P_{synch} = E\left[\int_{-\infty}^{\infty} s_{synch}(t)s_{synch}^*(t)dt\right]$, satisfies the total power constraint for N time slots, i.e., $P_{synch} \leq NP$. To take advantage of asynchrony, assume that each sub-stream is shifted with a specific time delay $\tau_k T$ where $0 \leq \tau_k < 1$.

Then, the transmitted signal will be:

$$s_{asynch}(t) = \sum_{k=1}^{K} s_k(t - \tau_k T) \tag{5.11}$$

such that $P_{asynch} \leq NP$. By assuming flat fading and additive white Gaussian noise, the received signal at the desired user is described as:

$$y(t) = h_r \sum_{k=1}^{K} s_k(t - \tau_k T) + n(t) \tag{5.12}$$

where h_r is the channel coefficient for the corresponding user. We use the sampling method explained in Chapter 3 to obtain a set of samples which are sufficient statistics:

$$y_k[m] = \int_{-\infty}^{\infty} y(t) p(t - mT - \tau_k T) dt \tag{5.13}$$
$$k = 1, \cdots, K \quad m = 1, \cdots, N$$

Denoting $p(t) * p(t)$ as $g(t)$ and $g((m-n)T + (\tau_k - \tau_l)T)$ as $g_{kl}[m-n]$, the sufficient statistics can be represented as:

$$y_k[m] = h_r \sum_{l=1}^{K} \sum_{n=1}^{N} g_{kl}[m-n] s_l[n] + n_k[m] \tag{5.14}$$

where $n_k[m] = n(t) * p(t)|_{mT+\tau_k T}$. Defining the vectors $\boldsymbol{y_k} = (y_k[1], \cdots, y_k[N])^T$ and $\boldsymbol{s_k} = (s_k[1], \cdots, s_k[N])^T$, the input-output relation in matrix form is represented as follows:

$$\boldsymbol{y} = h_r \boldsymbol{R} \boldsymbol{s} + \boldsymbol{n} \tag{5.15}$$

where $\boldsymbol{y} = \left(\boldsymbol{y_1}^T, \cdots, \boldsymbol{y_K}^T\right)^T$, $\boldsymbol{s} = \left(\boldsymbol{s_1}^T, \cdots, \boldsymbol{s_K}^T\right)^T$, $\boldsymbol{n} = \left(\boldsymbol{n_1}^T, \cdots, \boldsymbol{n_K}^T\right)^T$ and the matrix $\boldsymbol{R}$ is defined as:

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R_{11}} & \boldsymbol{R_{12}} & \cdots & \boldsymbol{R_{1K}} \\ \boldsymbol{R_{21}} & \boldsymbol{R_{22}} & \cdots & \boldsymbol{R_{2K}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R_{K1}} & \boldsymbol{R_{K2}} & \cdots & \boldsymbol{R_{KK}} \end{pmatrix} \tag{5.16}$$

where the elements of each sub-matrix $\boldsymbol{R_{lk}}$ is defined as:

$$\boldsymbol{R_{kl}}(m,n) = g_{kl}[m-n] \tag{5.17}$$

Matrix $\boldsymbol{R}$ is a Hermitian matrix whose sub-blocks, i.e., $\boldsymbol{R_{lk}}$ are banded Toeplitz blocks of order u, where $u = \frac{T_p}{T}$. Due to oversampling, the noise samples are correlated and their covariance is equal to $\boldsymbol{R}\sigma_n^2$. Note that, in the perfect synchronous case, sub-blocks $\boldsymbol{R_{lk}}$ turn into identity matrices, i.e., $\boldsymbol{I_N}$. However, for the asynchronous case, only diagonal blocks are identity matrices, i.e., $\boldsymbol{R_{ll}} = \boldsymbol{I_N}$, and the other sub-blocks have non-zero off diagonal elements. It is proved in the literature that as N grows large, i.e., $N \to \infty$, matrix $\boldsymbol{R}$ becomes a positive definite matrix with bounded eigenvalues for time-limited pulse shapes. On the other hand, when the pulse shapes are strictly band-limited, matrix $\boldsymbol{R}$ is asymptotically singular [7].

5.4.1 Transmit Power Examination

In this section, we analyze the transmit power for synchronous and asynchronous scenarios. We calculate P_{asynch} in the next lemma and P_{synch} will be immediately found by assuming that the time delays are equal to zero.

Lemma 5.2. *The transmit power of the asynchronous transmission of K superimposed sub-streams defined in Eq. (5.11) can be calculated as:*

$$\begin{aligned} P_{asynch} &= \sum_{i=1}^{K}\sum_{j=1}^{K} trace(\boldsymbol{R_{ij}Q_{ij}}) \\ &= trace(\boldsymbol{RQ_s}) \end{aligned} \tag{5.18}$$

where $\boldsymbol{Q_{ij}} = COV\{\boldsymbol{s_i}, \boldsymbol{s_j}\}$ *and* $\boldsymbol{Q_s} = COV\{\boldsymbol{s}\}$.

Proof. The proof is presented in Appendix C.2 □

Remarks 5.1. *With no precoding, i.e.,* $\boldsymbol{Q_{ij}} = \boldsymbol{0_N}$ *and* $\boldsymbol{Q_{ii}} = P_i\boldsymbol{I_N}$*, the transmit power of both methods, the synchronous and asynchronous methods equal* $P_{asycnh} = P_{synch} = N\sum_{k=1}^{K} P_k$ *where* P_k *is the power assigned to each symbol of User k. Consequently, the power constraint simplifies to* $\sum_{k=1}^{K} P_k \leq P$.

Remarks 5.2. *With precoding, the power constraint depends on the transmission mode, or equivalently, on the choice of time delays. The power constraint becomes* $trace(\boldsymbol{RQ}) \leq NP$. *Therefore, with precoding, the choice of matrix* $\boldsymbol{R}$ *affects the power allocation, whereas, with no precoding, it has no effect on the transmit power.*

5.5 Achievable Rate-Region Analysis

This section analyzes the achievable rate-region of the proposed NOMA schemes, AP-NOMA and T-NOMA. For comparison purposes, we also present the achievable rate-regions of the synchronous P-NOMA.

5.5.1 Achievable Rates for Conventional P-NOMA

We begin by applying the SIC detection at each user in the synchronous NOMA scheme. The optimal detection sequence is $s_K, \cdots, s_2, s_1$ assuming $\frac{|h_1|^2}{\sigma_n^2} > \cdots > \frac{|h_K|^2}{\sigma_n^2}$. In particular, User 1 first decodes $s_2, \cdots, s_K$ and subtracts their components from the received signal y_1. Then, User 1 decodes s_1 without interference from other users. On the other hand, User K can directly decode s_K while considering other users as noise. Assuming successful decoding and no error propagation, the achievable rate-region can be represented as:

$$R_{P-NOMA} \triangleq \begin{cases} 0 \leq R_1 \leq \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\sigma_n^2}\right) \\ 0 \leq R_2 \leq \frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{P_1|h_2|^2+\sigma_n^2}\right) \\ \vdots \\ 0 \leq R_K \leq \frac{1}{2}\log_2\left(1+\frac{P_K|h_K|^2}{|h_K|^2\sum_{k=1}^{K-1}P_k+\sigma_n^2}\right) \end{cases} \tag{5.19}$$

5.5.2 Achievable Rates for AP-NOMA

In AP-NOMA, the SIC detection is the same as P-NOMA, however, the set of samples that is used to decode symbols of each user is different. In the synchronous case, there is only one set of samples at each user. However, based on the sufficient statistics derived in Eq. (5.13), in the asynchronous case, there are K sets of samples at each user, each of them matched to the timing offset of one of the users represented as:

$$\boldsymbol{y_j} = h_r \sum_{k=1}^{K} \boldsymbol{R_{jk}} \boldsymbol{s_k} + \boldsymbol{n_j}, \quad j = 1, \cdots, K \tag{5.20}$$

Based on the assumption of $\frac{|h_1|^2}{\sigma_n^2} > \cdots > \frac{|h_K|^2}{\sigma_n^2}$, User r can decode signals of Users $\{r+1, \cdots, K\}$ using the sample sets of $\{\boldsymbol{y_{r+1}}, \cdots, \boldsymbol{y_K}\}$ and subtract them from the

corresponding sample set of $\boldsymbol{y_r}$. Note that based on the ordering of channel strengths, User r only needs $K-r$ sets of samples. In particular, the strongest user needs all K sets of samples and the weakest user only needs its own corresponding set of samples. The resulting set of samples after subtraction, at User r, is calculated as:

$$\hat{\boldsymbol{y}}_{\boldsymbol{r}} = h_r \boldsymbol{I_N} \boldsymbol{s_r} + h_r \sum_{k=1}^{r-1} \boldsymbol{R_{rk}} \boldsymbol{s_k} + \boldsymbol{n_r} \tag{5.21}$$

Due to the Toeplitz structure of the sub-blocks $\boldsymbol{R_{kl}}$, the resulting sample at each time instant i can be written as:

$$\hat{y}_r[i] = h_r s_r[i] + h_r \sum_{k=1}^{r-1} \sum_{j=-u}^{u} g_{rk}[j] s_k[i-j] + n_r[i], \quad i = u+1, \cdots, N-u$$

Treating the remaining interferers as noise will result in the following achievable rate for User r:

$$0 \leq R_r \leq \frac{1}{2} \log_2 \left(1 + \frac{P_r |h_r|^2}{|h_r|^2 \sum_{k=1}^{r-1} \eta_{rk} P_k + \sigma_n^2} \right) \tag{5.22}$$

where $\eta_{rk} = \sum_{i=-u}^{u} g_{rk}^2[i]$. In general, the overall rate-region depends on how we assign the timing offsets to users with different channel strengths. For a specific assignment, namely ψ, the achievable rate-region, $R_{AP-NOMA}^{\psi}$ can be defined as:

$$R_{AP-NOMA}^{\psi} \triangleq \begin{cases} 0 \leq R_1 \leq \frac{1}{2} \log_2 \left(1 + \frac{P_1 |h_1|^2}{\sigma_n^2} \right) \\ 0 \leq R_2 \leq \frac{1}{2} \log_2 \left(1 + \frac{P_2 |h_2|^2}{\eta_{21} P_1 |h_2|^2 + \sigma_n^2} \right) \\ \vdots \\ 0 \leq R_K \leq \frac{1}{2} \log_2 \left(1 + \frac{P_K |h_K|^2}{|h_K|^2 \sum_{k=1}^{K-1} \eta_{Kk} P_k + \sigma_n^2} \right) \end{cases} \tag{5.23}$$

Note that there are $K!$ different assignments of time delays to users. Hence, the total

rate-region is the convex hull of all possible assignments, i.e.,

$$R_{AP-NOMA} \triangleq \bigcup_{\psi=1}^{K!} R_{AP-NOMA}^{\psi} \tag{5.24}$$

Using Lemma 1, it can be shown that $\eta_{rk} = \sum_{-u}^{u} g_{rk}^2[i] \leq 1$ where equality is obtained by putting $\tau = 0$. As a result, the rate-region for each assignment and thus the total rate-region for AP-NOMA is larger than that of the conventional P-NOMA. Note that for the 2-user scenario, the rate-region for both possible assignments is the same because $\eta_{21} = \eta_{12}$. However, this is not valid for more number of users unless the difference between time delays is equal which results in matrix $\boldsymbol{R}$ to be Hermitian Toeplitz.

5.5.3 Achievable Rates for T-NOMA

In this section, we derive the achievable rate-region for the T-NOMA method. As mentioned before, asynchrony provides additional degrees of freedom which will be exploited in T-NOMA by using precoding at the BS. T-NOMA method applies a simple precoding at the BS. Assuming that $\boldsymbol{Q_s} = COV\{\boldsymbol{s}\}$, the transmitted symbols can be written as $\boldsymbol{s} = \boldsymbol{U_T}\sqrt{\boldsymbol{P}}\boldsymbol{x}$ where $\boldsymbol{Q_x} = \boldsymbol{I_N}$ and $\boldsymbol{Q_s} = \boldsymbol{U_T}\boldsymbol{P}\boldsymbol{U_T}^H$. Matrix $\boldsymbol{U_T}$ is an orthogonal matrix representing the precoding matrix and $\sqrt{\boldsymbol{P}}$ denotes the element wise square-root of matrix $\boldsymbol{P}$ which is a diagonal matrix representing the power allocation matrix defined as:

$$\boldsymbol{P} = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{NK} \end{pmatrix}$$

Then, the received samples at User r is denoted as:

$$\boldsymbol{y} = h_r \boldsymbol{R} \boldsymbol{U_T} \sqrt{\boldsymbol{P}} \boldsymbol{x} + \boldsymbol{n} \tag{5.25}$$

where the power constraint is stated as $trace(\boldsymbol{R}\boldsymbol{U_T}\boldsymbol{P}\boldsymbol{U_T}^H) \leq NP$. The covariance matrix of the noise vector is $\boldsymbol{Q_n} = \boldsymbol{R}\sigma_n^2$. To find the proper precoding matrix and power allocation, let us consider the eigen-decomposition of matrix $\boldsymbol{R}$. Matrix $\boldsymbol{R}$ is a Hermitian matrix, thus its eigen-decomposition can be written as:

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{U_R} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{NK} \end{pmatrix} \boldsymbol{U_R}^H \\ &= \boldsymbol{U_R} \boldsymbol{\Lambda_R} \boldsymbol{U_R}^H \end{aligned} \tag{5.26}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{NK}$ are the eigenvalues of matrix $\boldsymbol{R}$ and $\boldsymbol{U_R}$ is a Unitary matrix. Therefore, the received signal can be rewritten as:

$$\boldsymbol{y} = h_r \boldsymbol{U_R} \boldsymbol{\Lambda_R} \boldsymbol{U_R^H} \boldsymbol{U_T} \sqrt{\boldsymbol{P}} \boldsymbol{x} + \boldsymbol{n} \tag{5.27}$$

Because the time delays are known at the transmitter, the matrix $\boldsymbol{R}$ is known at the transmitter. Hence, the transmitted symbols can be precoded in the direction of the eigen-vectors of matrix $\boldsymbol{R}$. In addition, the sub-channels can be decomposed at each user by a post-processing matrix $\boldsymbol{U_U}$. By choosing $\boldsymbol{U_T} = \boldsymbol{U_R}$ and $\boldsymbol{U_U} = \boldsymbol{U_R}^H$, we will have

$$\bar{\boldsymbol{y}} = h_r \boldsymbol{\Lambda_R} \sqrt{\boldsymbol{P}} \boldsymbol{x} + \bar{\boldsymbol{n}} \tag{5.28}$$

where $\bar{\boldsymbol{y}} = \boldsymbol{U}^H \boldsymbol{y}$ and $\bar{\boldsymbol{n}} = \boldsymbol{U}^H \boldsymbol{n}$. The covariance matrix of the noise vector is given by $\boldsymbol{Q}_{\bar{\boldsymbol{n}}} = \boldsymbol{\Lambda}_{\boldsymbol{R}} \sigma_n^2$. By using proper precoding and post processing at the destination, the channel has turned into NK independent sub-channels. In other words, we have used the available degrees of freedoms offered by asynchrony to decompose the transmitted symbols and eliminate interference. The precoding at the BS also affects the power constraint. By using the proposed precoding, the product of $\boldsymbol{RQ}$ simplifies to:

$$\boldsymbol{RQ} = \boldsymbol{U}_R \begin{pmatrix} p_1\lambda_1 & 0 & \cdots & 0 \\ 0 & p_2\lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{KN}\lambda_{KN} \end{pmatrix} \boldsymbol{U}_R^H \tag{5.29}$$

As a result, the power constraint will be:

$$\sum_{i=1}^{NK} p_i \lambda_i \leq NP \tag{5.30}$$

Based on the sample in Eq. 5.28, the achievable rate for User r can be denoted as

$$R_r = \frac{1}{2N} \sum_{i \in \mathcal{I}_r} log_2 \left(1 + \frac{p_i \lambda_i |h_r|^2}{\sigma_n^2}\right)$$

where $\mathcal{I}_r$ represents the set of sub-channels indices that are assigned to User r.

Lemma 5.3. *The achievable rate-region of T-NOMA method can be described as follows:*

$$R_k = \frac{1}{2} log_2 \left(1 + \frac{P_k |h_k|^2}{\sigma_n^2}\right), \quad k = 1, \cdots, K$$
$$s.t. \sum_{k=1}^{K} P_k \leq P \tag{5.31}$$

Proof. The proof is presented in Appendix C.3. □

Note that the achievable rate-region of T-NOMA method is independent of the pulse shape and the timing offsets. The only requirement to achieve this rate-region is the matrix $\boldsymbol{R}$ to be full rank which is satisfied as long as the pulse shapes are time limited and the timed delays are distinct. Also note that, using relatively band-limited signals like truncated sinc pulse shapes and truncated r.c. pulse shapes will limit the spectrum broadening caused by precoding [93]. Numerical results are provided in the next section.

5.6 Numerical Results

In this section, we present numerical results to show the effectiveness of using asynchrony in providing a larger rate-region. In particular, we show that AP-NOMA outperforms P-NOMA with a slight change of adding timing offset among transmitted symbols. However, T-NOMA which uses the degrees of freedom provided by time asynchrony results in the best performance. We assume that a transmit power of $P = 10$ is available at the base station which is serving two users. We consider the typical pulse shaping function in the literature, i.e., Rectangular pulse shape (Rect.), and a more practical pulse shaping function, i.e., r.r.c.. Theoretically, r.r.c. pulse shaping is unlimited in time, however, it is truncated in practice and we have adopted the truncated version with 4 side lobes. The symbol duration T is normalized to be 1, $\tau_1 = 0$ and $\tau_2 \in [0, 0.5]$ due to the symmetry.

We first consider the Gaussian channel where the channel coefficients are determined by $|h_1|^2/\sigma_n^2 = 10$ and $|h_2|^2/\sigma_n^2 = 1$. In Fig. 5.4, we show the achievable rate-regions of AP-NOMA and T-NOMA with different symbol offsets and r.c. pulse shape. Increasing the timing offset will improve the performance and $\tau_2 = 0.5$ results in the best performance. On the other hand, the performance of the T-NOMA method, which exploits the degrees of freedom available in the system, is independent of the pulse shape and time delays as long

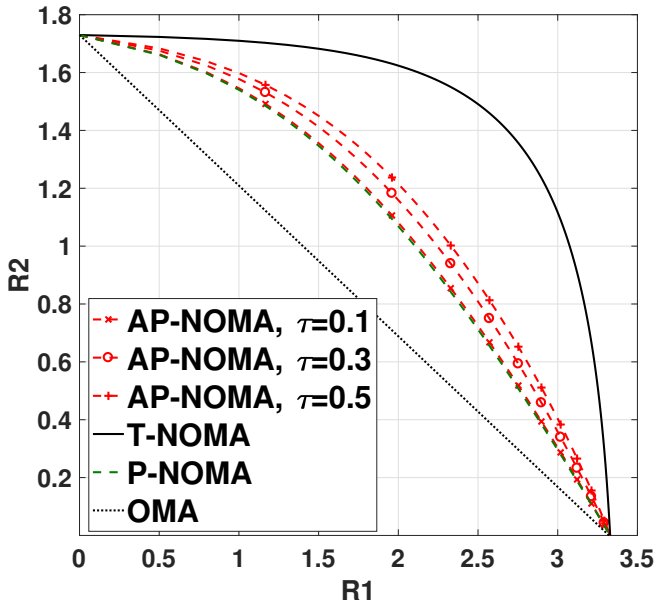


Figure 5.4: Achievable rate-regions of AP-NOMA and T-NOMA for different symbol offsets and a Gaussian channel.

as matrix $\boldsymbol{R}$ is full rank. Fig. 5.4 shows that both T-NOMA and AP-NOMA outperform the conventional P-NOMA. In Fig. 5.4, there is sufficient discrepancy between channel qualities of the two users to be exploited by P-NOMA and AP-NOMA methods. However, in Fig. 5.5, the channel coefficients are assumed to be the same, i.e., $|h_1|^2/\sigma_n^2 = 1$ and $|h_2|^2/\sigma_n^2 = 1$, thus the P-NOMA performance coincides with that of the OMA systems. In such a case, the AP-NOMA method provides slightly better performance; however, T-NOMA significantly improves the performance showing the capability of this method even without channel quality discrepancy.

The rate-region for 3 users with rectangular pulse shape and $\boldsymbol{\tau} = [0, 0.3, 0.7]$ is shown in Fig. 5.6. The rate-regions for P-NOMA and T-NOMA are calculated similar to those of the 2-user scenario. However, there are two main differences in calculating the rate-region of AP-NOMA. First, there are 3!=6 different assignments of time delays to users, and the total region is found by taking the convex hull of all possible assignments. The other difference is that when one of the power assignments is equal to zero, the remaining time

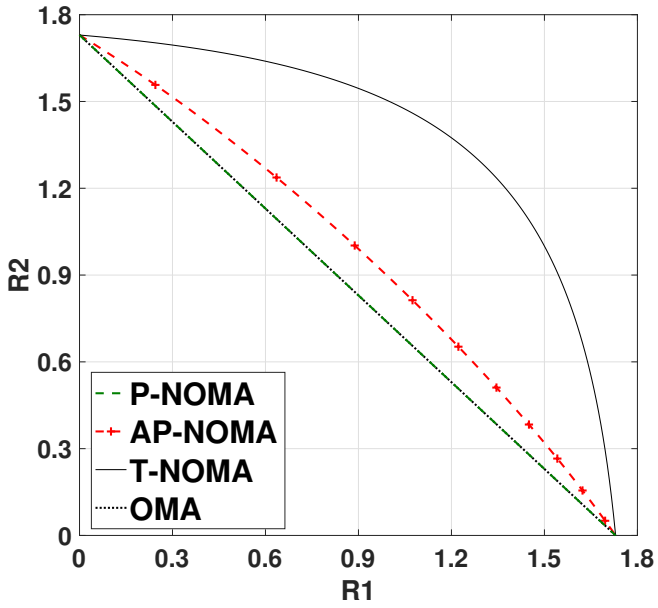


Figure 5.5: The maximum achievable rate-regions of three schemes: P-NOMA, AP-NOMA and T-NOMA in Gaussian channels with $|h_1|^2/\sigma_n^2 = 1$ and $|h_2|^2/\sigma_n^2 = 1$.

delays for the other two users need to be updated. In more details:

$$R_{AP-NOMA} = \begin{cases} R_{AP-NOMA}[0, 0.3, 0.7] & P_1 \neq 0, P_2 \neq 0, P_3 \neq 0 \\ R_{AP-NOMA}[0, 0.4] & P_1 = 0, P_2 \neq 0, P_3 \neq 0 \\ R_{AP-NOMA}[0, 0.7] & P_1 \neq 0, P_2 = 0, P_3 \neq 0 \\ R_{AP-NOMA}[0, 0.3] & P_1 \neq 0, P_2 \neq 0, P_3 = 0 \end{cases} \tag{5.32}$$

To have a better understanding of the 3 dimensional rate-region provided in Fig. (5.6), we show different two-dimensional cuts when $R_1 = 1$, $R_2 = 1$ and $R_3 = 1$ in Figs. (5.7a), (5.7b) and (5.7c), respectively.

Next, we consider the Rayleigh block-fading channel where the channel coefficients are independent Rayleigh distribution with unit variance, and noise variance is set to 0.1. The ergodic rate is averaged over 10^5 different realizations of the channel. In Fig. 5.8, we show the achievable rate-regions of P-NOMA, AP-NOMA and T-NOMA in Rayleigh block-

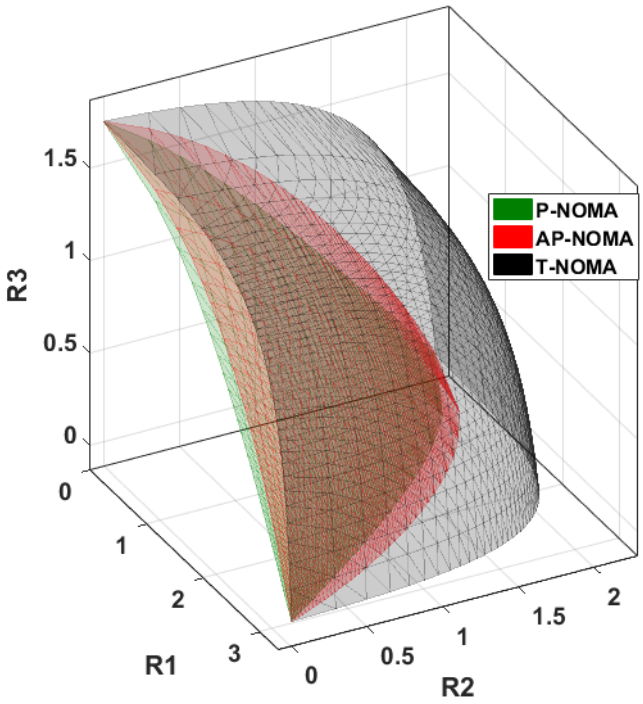
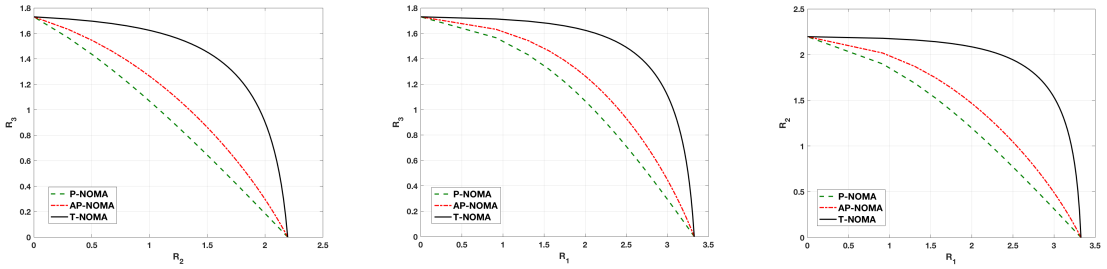


Figure 5.6: The maximum achievable rate-regions of three schemes: P-NOMA, AP-NOMA and T-NOMA in Gaussian channels with $|h_1|^2/\sigma_n^2 = 10$, $|h_2|^2/\sigma_n^2 = 2$ and $|h_3|^2/\sigma_n^2 = 1$.



(a) R_1=1 (b) R_2=1 (c) R_3=1

Figure 5.7: Different 2D cuts of the 3 dimensional rate-region in Fig. 5.6

fading channels. Similar to Gaussian channels, with fading channels, the achievable rate-region of P-NOMA is improved by adding asynchrony. In addition, T-NOMA provides a large improvement compared with other schemes.

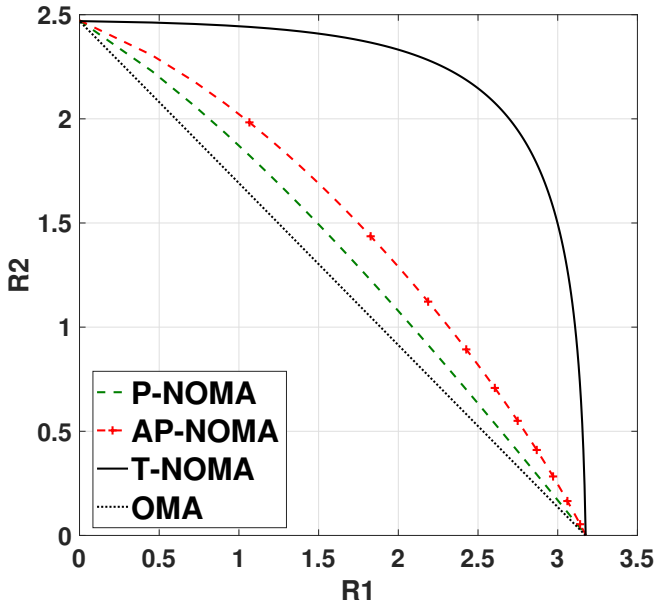


Figure 5.8: The maximum achievable rate-regions of three schemes: P-NOMA, AP-NOMA and T-NOMA in Rayleigh fading channels.

5.7 Conclusion

In this chapter, we proposed a novel symbol-asynchronous downlink NOMA schemes. In contrast to the conventional P-NOMA, we proposed to add timing offsets among superimposed symbols. The receiver architecture in AP-NOMA includes oversampling and a SIC scheme similar to the P-NOMA, however, asynchrony reduces IUI and improves the overall performance. T-NOMA exploits the degrees of freedom introduced by time asynchrony, using novel precoding and simple post-processing at users. In other words, T-NOMA decomposes the channel into independent sub-channels and eliminates the interference. Our analysis shows that both AP-NOMA and T-NOMA methods can improve the achievable rate-regions.

Chapter 6

Link to the Faster than Nyquist Signaling: Challenges and Solutions

Faster than Nyquist (FTN) signaling refers to transmitting the desired symbols faster than the Nyquist rate with shorter time intervals, which was first introduced in [23] where its minimum Euclidean distance property was analyzed. Asynchronous transmission in which multiple streams of information are superimposed and transmitted with some timing offsets resemble FTN. Notably, an asynchronous transmission method with $\tau = 0.5$ is similar to an FTN signaling method with a compression factor of $\delta = 0.5$. Although asynchronous transmission and FTN signaling can provide an increased achievable rate, the optimal input design imposes some practical issues such as out-of-band (OOB) emission, which is analyzed and addressed in this chapter.

In more detail, this chapter considers the capacity computations of faster-than-Nyquist (FTN) signaling. We calculate the theoretical capacity of FTN signaling, which is obtained by a correlated input. The capacity-achieving power spectral density (PSD) is derived, and its superiority over the independent input is shown. The practical issue im-

posed by the capacity-achieving PSD, i.e., OOB emission, is shown. To solve this issue, we introduce an upper-bound on the input PSD to limit the OOB emission. The new optimization problem is solved, and the constrained PSD is obtained. The introduced PSD captures the trade-off between the obtained capacity and the OOB emission.

6.1 Introduction

We consider the problem of transmitting data over bandlimited additive white Gaussian noise (AWGN) and fading channels employing a set of signals that are generated by linear modulation. Linear modulation signals have the form $x(t) = \sum a[n]p(t - n\delta T)$ where $\{a[n]\}$ are the symbols to be transmitted and can be independent or correlated. The pulse shape $p(t)$ usually has useful detection properties, e.g., orthogonality, commonly mentioned as No-ISI Nyquist condition. An example of such pulses is called r.c. pulse and is extensively used in communication standards, including the DVB-S2X in satellite communication [94].

Shannon's classical result states that the highest transmission rate over the AWGN channel is $W \log_2(1 + P/WN_0)$ where W is one-sided bandwidth, P is the average power, and $N_0/2$ is the power spectral density (PSD) of the white noise [22]. Such a capacity is achieved by the *sinc* pulse and $T = \frac{1}{2W}$. However, the Sinc pulse has serious realization problems and disadvantages in practice. Therefore, smoother orthogonal pulses are utilized, which introduce the excess bandwidth. This excess bandwidth remains unutilized by orthogonal signaling. However, the excess bandwidth can be utilized by Faster than Nyquist (FTN) Signaling [95].

FTN was first introduced in [23], where its minimum Euclidean distance property was analyzed. FTN has been extensively studied over the past few years [96]. For example,

asymptotic optimality of binary FTN signaling is shown in [97], non-binary FTN and the bit error rate (BER) performance are studied in [98], the detection and the receiver design are examined in [99], and the application of FTN in the multiuser broadcast channel is considered in [100]. The effect of time localization of signals on the performance of FTN signaling is considered in [101], and the multicarrier FTN is analyzed in [102]. The information rates of FTN signaling and the provided gain by faster signaling are thoroughly investigated in [95, 103]. In [95], an independent input with a constant PSD that satisfies the power constraint is considered. In [103], a more comprehensive set of inputs, including independent and correlated inputs, is considered. Denoting the input PSD as $S_a(f), f \in [-1/2, 1/2]$, the power constraint is expressed as $\int_{-1/2}^{1/2} S_a(f)df \leq P$, and the capacity-achieving input is found by using the well-known water-filling algorithm. Although the mentioned power constraint is correct for Nyquist signaling, the correlated input in conjunction with FTN signaling will alter the power constraint and change the capacity expressions.

Our goal in this chapter is to study the power constraint and the resulting capacity expression comprehensively. We derive the capacity-achieving PSD and show its enhancement over the independent input. To alleviate the practical issue of the capacity-achieving input, we modify the input PSD constraint to strike a trade-off between the independent and capacity-achieving inputs. In Section 6.2, the general system model is presented. In Section 6.3, some preliminary results are provided and in Section 6.4, the new capacity expression of FTN signaling is presented. In Section 6.5, the constrained capacity is expressed which provides a trade-of between the achieved capacity and OOB emission. Finally, in Section 6.7, the numerical results are presented. The materials in this chapter are originally published in [104].

6.2 System Model

We consider the fundamental single-carrier communication model. The transmitted linearly modulated baseband signal has the form

$$x(t) = \sum_{n=1}^{N} a[n]p(t - n\delta T), \tag{6.1}$$

where $\{a[n]\}$ is a stationary Gaussian process with power spectral density (PSD) of $\{S_a(f), f \in [-1/2, 1/2]\}$ which is periodic in f with period of 1. The symbol interval is denoted by δT where $\delta \in (0, 1]$ is the compression factor in FTN signaling and $\delta = 1$ yields the traditional Nyquist signaling with signaling rate of $1/T$. Also, $p(t)$ is a root-Nyquist pulse shaping filter with unit energy, i.e., $\int_{-\infty}^{\infty} |p(t)|^2 dt = 1$. Denoting $g(t) = p(t) \star p^*(-t)$, the Nyquist No-ISI condition is expressed as: $g(nT) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$. The Nyquist No-ISI condition can be equivalently expressed as $G_T(f) = \sum_{k=-\infty}^{\infty} G\left(f - \frac{k}{T}\right) = T$. The PSD of the transmitted signal $x(t)$ can be calculated as $S_x(f) = S_a(\delta T f)G(f)$. The transmission power is calculated as $P_x = \int_{-\infty}^{\infty} S_x(f)df$. Because $S_a(\delta T f)$ is periodic in f with period of $1/\delta T$, the transmission power can be written as:

$$P_x = \int_{\frac{-1}{2\delta T}}^{\frac{1}{2\delta T}} S_a(\delta T f) G_{\delta T}(f) df = \frac{1}{\delta T} \int_{\frac{-1}{2}}^{\frac{1}{2}} S_a(f) G_{\delta T}\left(\frac{f}{\delta T}\right) df, \tag{6.2}$$

where $G_{\delta T}(f) = \sum_{k=-\infty}^{\infty} G\left(f - \frac{k}{\delta T}\right)$. Note that for the Nyquist signaling, i.e., $\delta = 1$, $G_T(f) = T$, and thus, the transmit power reduces to $P_x = \int_{-1/2}^{1/2} S_a(f)df$ which is commonly used in the literature. However, for $\delta < 1$, the above equality does not hold. Also note that for independent symbols, i.e., $S_a(f) = \sigma_a^2$, the transmission power reduces to $P_x = \frac{\sigma_a^2}{\delta T} \int_{-1/2}^{1/2} G_{\delta T}(\frac{f}{\delta T})df = \sigma_a^2$. Assuming the AWGN channel, the received signal at the decoder is denoted as: $y(t) = \sum_{n=1}^{N} a[n]p(t - n\delta T) + n(t)$ where $n(t)$ is the white noise

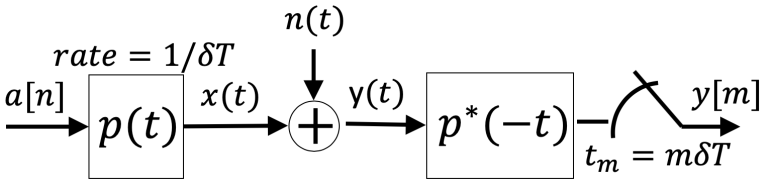


Figure 6.1: FTN system model.

with the PSD of $\sigma_n^2 = N_0/2$. At the decoder, the set of sufficient statistics is obtained by a matched filter followed by a sampler which samples at time instants $t_m = m\delta T$ as shown in Fig. 6.1. Thus, the samples can be written as:

$$y[m] = \sum_{n=1}^{N} a[n]g((m-n)\delta T) + n[m], \tag{6.3}$$

where $n[m] = \int_{-\infty}^{\infty} n(t)p^*(t - m\delta T)dt$ and the correlation of the noise sample can be denoted as $E\{n[m]n^*[i]\} = \sigma_n^2 g((m-i)\delta T)$. The input-output relation can be written in matrix form as:

$$\boldsymbol{y}^N = \boldsymbol{G}^N \boldsymbol{a}^N + \boldsymbol{n}^N, \tag{6.4}$$

where $\boldsymbol{y}^N = \{y[1], \cdots, y[N]\}$, $\boldsymbol{a}^N = \{a[1], \cdots, a[N]\}$, $\boldsymbol{n}^N = \{n[1], \cdots, n[N]\}$ and $\boldsymbol{G}^N$ is an $N \times N$ Toeplitz matrix whose elements are defined as:

$$[\boldsymbol{G}^N]_{m,n} = g((m-n)\delta T). \tag{6.5}$$

Note that with Nyquist signaling, the input-output relation is independent for each transmitted symbol denoted as $y[m] = a[m] + n[m]$.

6.3 Existing Results

In [95], the capacity of FTN signaling for independent symbols with power constraint of P is calculated as:

$$C(P,\delta T)=\int_0^{1/2\delta T}\log_2\left(1+\frac{P}{\sigma_n^2}G_{\delta T}(f)\right)df. \tag{6.6}$$

In [103], the capacity with correlated symbols is calculated as:

$$\begin{aligned} C(P,\delta T)=&\sup_{S_a(f)}\int_0^{1/2\delta T}\log_2\left(1+\frac{1}{\sigma_n^2}S_a(\delta T f)G_{\delta T}(f)\right)df.\\ &s.t.\quad \int_{-1/2}^{1/2}S_a(f)df=P, \end{aligned} \tag{6.7}$$

where the optimization problem is solved by the well-known water-filling strategy. However, in these results, the effect of FTN signaling on the power transmission was ignored which led to erroneous results. Hence, we revisit the capacity expression with the correct transmission power and reveal the main advantages of FTN signaling.

6.4 FTN Capacity Expressions Revisited

For completeness, we derive the capacity expressions for FTN signaling, following the steps taken in [105] and [83]. Based on [105], the capacity of the discrete-time system in (6.4), which resembles an ISI channel, is equal to $C_{DT}=\lim_{N\to\infty}\frac{1}{N}I(\boldsymbol{y}^N;\boldsymbol{a}^N)$ given the input symbols satisfy the power constraint in (6.2). The differential entropy of a stationary Gaussian n-vector, $\boldsymbol{x}^N$, with covariance matrix of $\boldsymbol{\Sigma}_x^N$ is equal to [68]:

$$h(\boldsymbol{x}^N)=\frac{N}{2}\log_2\left(2\pi e\left[det(\boldsymbol{\Sigma}_x^N)\right]^{1/N}\right). \tag{6.8}$$

In addition, according to the Toeplitz distribution theorem [51], as $N \to \infty$, the asymptotic differential entropy approaches:

$$\lim_{N\to\infty} \frac{1}{N} h(\boldsymbol{x}^N) = \frac{1}{2} \int_{-1/2}^{1/2} \log_2 \left(2\pi e \sigma_x(f)\right) df, \tag{6.9}$$

where $\sigma_x(f)$ is the generating function of the Toeplitz matrix $\boldsymbol{\Sigma}_x$ defined as $\sigma_x(f) = \sum_{n=-\infty}^{\infty} [\boldsymbol{\Sigma}_x]_n \, e^{-j2\pi f n}$ and $[\boldsymbol{\Sigma}_x]_n$ represents the nth diagonal element of $\boldsymbol{\Sigma}_x$.

Using (6.8) and (6.9), the asymptotic information rate of $C_{DT} = \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{y}^N; \boldsymbol{a}^N)$ can be calculated in the following steps:

$$\begin{aligned} C_{DT} &= \lim_{N\to\infty} \frac{1}{N} h(\boldsymbol{y}^N) - \lim_{N\to\infty} \frac{1}{N} h(\boldsymbol{n}^N) \\ &= \frac{1}{2} \int_{-1/2}^{1/2} \log_2 \left(\sigma_n^2 \hat{g}(f) + \hat{g}^2(f) S_a(f)\right) - \log_2 \left(\sigma_n^2 \hat{g}(f)\right) df \\ &= \frac{1}{2} \int_{-1/2}^{1/2} \log_2 \left(1 + \frac{1}{\sigma_n^2} \hat{g}(f) S_a(f)\right) df, \end{aligned} \tag{6.10}$$

where $\hat{g}(f) = \sum_{n=-\infty}^{\infty} g(n\delta T) e^{-j2\pi f n}$ and after some calculations, it can be shown that $\hat{g}(f) = G_{\delta T}(f/\delta T)$. Then, after normalizing by the signaling rate, the capacity of FTN signaling can be formulated as:

$$\begin{aligned} C(P, \delta T) = \sup_{S_a(f)} & \frac{1}{2\delta T} \int_{-1/2}^{1/2} \log_2 \left(1 + \frac{1}{\sigma_n^2} G_{\delta T}\left(\frac{f}{\delta T}\right) S_a(f)\right) df. \\ & s.t. \;\; \frac{1}{\delta T} \int_{-1/2}^{1/2} S_a(f) G_{\delta T}\left(\frac{f}{\delta T}\right) df \le P. \end{aligned} \tag{6.11}$$

Note that the power constraint in the new capacity formulation differs from the power constraint mentioned in (6.7). Considering Jensen's inequality, the capacity is achieved

by:

$$S_a(f) = \begin{cases} \frac{\delta TP}{\mu G_{\delta T}(f/\delta T)} & G_{\delta T}(f/\delta T) \neq 0 \\ 0 & o.w. \end{cases}, \tag{6.12}$$

where $\mu = \mathcal{L}\left(f \in [-1/2, 1/2] | G_{\delta T}(f/\delta T) \neq 0\right)$ and $\mathcal{L}(I)$ is the Lebesgue measure of the set I. By simple calculations, the capacity can be written as:

$$C(P, \delta T) = \frac{\mu}{2\delta T} \log_2 \left(1 + \frac{\delta TP}{\mu \sigma_n^2}\right). \tag{6.13}$$

Remarks 6.1. *For sinc function, the achieved capacity is* $C(P, \delta T) = \frac{1}{2T} \log_2 \left(1 + \frac{TP}{\sigma_n^2}\right)$.

Remarks 6.2. *For Nyquist signaling, i.e.,* $\delta = 1$ *with any No-ISI satisfying pulse shape, i.e.,* $G_T(f/T) = T$*, the capacity-achieving input and the achieved capacity are* $S_a(f) = P$ *and* $C(P, \delta T) = \frac{1}{2T} \log_2 \left(1 + \frac{TP}{\sigma_n^2}\right)$*, respectively. This verifies dissipation of the excess bandwidth with Nyquist signaling.*

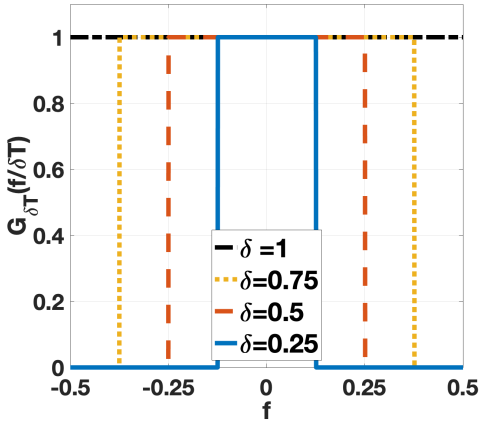
Remarks 6.3. *For any non-orthogonal signaling, i.e.,* $\delta < 1$*, the capacity-achieving input and the capacity expression are calculated by inserting* $\mu = \min\{1, f_{BW}\delta T\}$ *in (6.12) and (6.13), respectively, where* f_{BW} *is the frequency support of* $g(t)$*.*

Example 6.1. *Consider the r.c. pulse shape with a roll-of-factor of* β *whose Fourier transform is denoted as:*

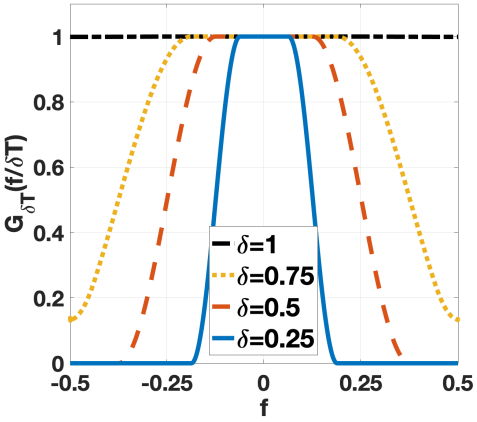
$$G(f) = \begin{cases} T & |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2}\left[1 + \cos\left(\frac{\pi T}{\beta}\left(|f| - \frac{1-\beta}{2T}\right)\right)\right] & \frac{1-\beta}{2T} < |f| \leq \frac{1+\beta}{2T} \\ 0 & o.w. \end{cases}.$$

The folded-scaled spectrum $G_{\delta T}(\frac{f}{\delta T}) = \sum_{k=-\infty}^{\infty} G\left(\frac{f}{\delta T} - \frac{k}{\delta T}\right)$ *is depicted in Fig. 6.2 for* $T = 1$ *and different values of* δ *and* β*. Observe that for* $\delta = 1$*, all the pulse shapes with*

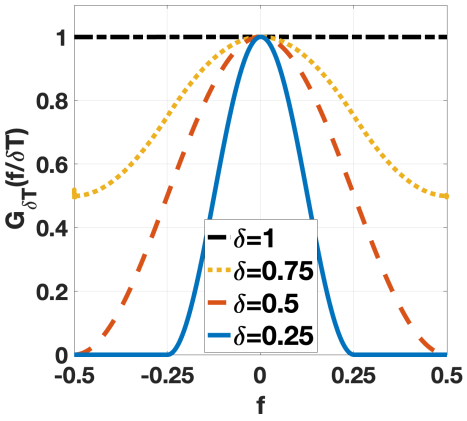
different roll-of-factors result in a constant folded-scaled spectrum. Decreasing δ alters the shape of the folded-scaled spectrum.



(a) $\beta = 0$



(b) $\beta = 0.5$



(c) $\beta = 1$

Figure 6.2: Illustration of the folded-scaled spectrum for different values of β and δ.

It can be calculated that $\mu = \min\{1, (1+\beta)\delta)\}$ *and as a result:*

$$C(P, \delta T) = \begin{cases} \frac{1}{2\delta T} \log_2\left(1 + \frac{\delta T P}{\sigma_n^2}\right) & \frac{1}{1+\beta} < \delta \leq 1 \\ \frac{1+\beta}{2T} \log_2\left(1 + \frac{TP}{(1+\beta)\sigma_n^2}\right) & \delta \leq \frac{1}{1+\beta} \end{cases},$$

which is obtained by the correlated input expressed in (6.12). With independent input, i.e., $S_a(f) = P$*, the achievable rate is equal to* $C_{ind}(P, \delta T) = \frac{1}{2\delta T} \int_{-1/2}^{1/2} \log_2\left(1 + \frac{P}{\sigma_n^2} G_{\delta T}\left(\frac{f}{\delta T}\right)\right) df$

also derived in [95] .

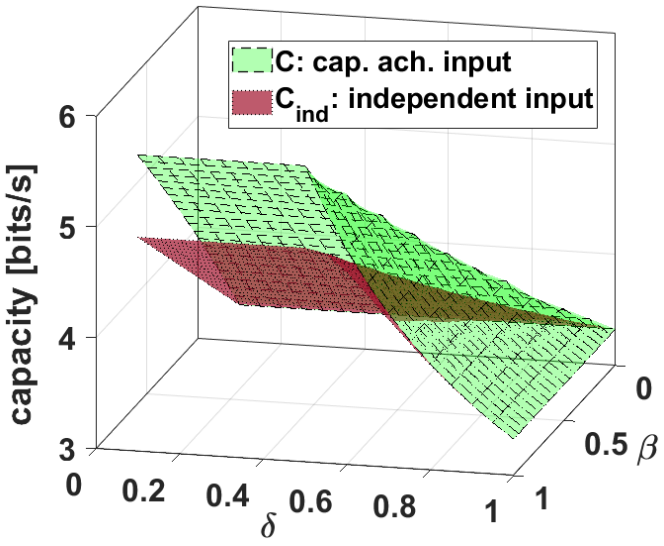


Figure 6.3: Comparison of C and C_{ind} for $P = 20\ dB$.

A comprehensive comparison of C and C_{ind} is illustrated in Fig. 6.3 for $P = 20\ dB$. Note that for the compression factor of $\frac{1}{1+\beta} < \delta < 1$, the excess bandwidth is exploited by FTN and the achievable rate is increased. However, as $\delta < \frac{1}{1+\beta}$, further compression provides no extra gain. Using FTN, the capacity-achieving input and the independent input can improve the performance of the Nyquist signaling up to 40% and 65%, respectively.

These results state that, in terms of capacity, Nyquist signaling is optimal only for the sinc pulse. For other pulse shapes, FTN can exploit the excess bandwidth and improve the capacity. Although FTN signaling with independent input can improve the achievable rate, the correlated input achieves the capacity and further boosts the advantage of FTN signaling. However, the capacity-achieving input imposes hardware-related problems, stability issues and OOB emission in practical settings. This is because the PSD of the capacity-achieving input is proportional to the inverse of the folded-scaled spectrum and can take very large values due to the small values of the folded-scaled spectrum. This results in OOB emission when time-limited pulse shapes are used. For further details on

practical aspects of FTN and hardware related issues refer to [96].

6.5 New Constrained Capacity

To solve the mentioned issue, we introduce an upperbound constraint on the input PSD, i.e., $S_a(f) \leq \theta P$ where $\theta \in [1, \infty)$ can be chosen based on the hardware capabilities and system sensitivity to OOB emission. Intuitively, looser upperbound results in higher capacity and on contrary more stringent upperbound results in lower capacity in exchange for less hardware complications and lower OOB emission. The constrained capacity for the FTN signaling with PSD upperbound-constraint can be stated as:

$$\begin{aligned} C_\theta(P, \delta T) = \sup_{S_a(f)} \frac{1}{2\delta T} \int_{-1/2}^{1/2} \log_2 \left(1 + \frac{1}{\sigma_n^2} G_{\delta T}\left(\frac{f}{\delta T}\right) S_a(f)\right) df. \\ s.t. \quad \frac{1}{\delta T} \int_{-1/2}^{1/2} S_a(f) G_{\delta T}\left(\frac{f}{\delta T}\right) df \leq P, \\ S_a(f) \leq \theta P. \end{aligned} \tag{6.14}$$

The KKT conditions can be written as:

$$\begin{aligned} &\lambda_1 \left(\frac{1}{\delta T} \int_{-1/2}^{1/2} S_a(f) G_{\delta T}\left(\frac{f}{\delta T}\right) df - P \right) = 0, \\ &\lambda_2(f) \left(S_a(f) - \theta P \right) = 0, \\ &\frac{G_{\delta T}\left(\frac{f}{\delta T}\right)}{\sigma_n^2 + G_{\delta T}\left(\frac{f}{\delta T}\right) S_a(f)} - 2\lambda_1 G_{\delta T}\left(\frac{f}{\delta T}\right) - 2\delta T \lambda_2(f) = 0, \end{aligned}$$

where $\lambda_1 \geq 0$ and $\lambda_2(f) \geq 0$ are the Lagrange multipliers. Therefore, $S_a(f)$ can be calculated as: $S_a(f) = \begin{cases} \frac{1}{2\lambda_1 G_{\delta T}(f/\delta T) + 2\delta T \lambda_2(f)} - \frac{\sigma_n^2}{G_{\delta T}(f/\delta T)} & G_{\delta T}(f/\delta T) \neq 0 \\ 0 & o.w. \end{cases}$. Considering

the KKT conditions and with some calculations, $S_a(f)$ can be further simplified as:

$$S_a(f) = \begin{cases} \theta P & \lambda > \theta P G_{\delta T}(f/\delta T) + \sigma_n^2 \\ \frac{\lambda - \sigma_n^2}{G_{\delta T}(f/\delta T)} & \theta P G_{\delta T}(f/\delta T) + \sigma_n^2 \geq \lambda > \sigma_n^2 \\ 0 & o.w. \end{cases}, \tag{6.15}$$

where $\lambda = 1/2\lambda_1$. The procedure to find λ is similar to the well-known water-filling algorithm. Defining $\mathcal{U} = \{f \in [-1/2, 1/2] | \lambda > \theta P G_{\delta T}(f/\delta T) + \sigma_n^2\}$ and $\mathcal{V} = \{f \in [-1/2, 1/2] | \theta P G_{\delta T}(f/\delta T) + \sigma_n^2 \geq \lambda\}$, then the power constraint can be written as:

$$T(\lambda) = \frac{\theta P}{\delta T} \int_{f \in \mathcal{U}} G_{\delta T}\left(\frac{f}{\delta T}\right) df + (\lambda - \sigma_n^2)\frac{\mu_{\mathcal{V}}}{\delta T} - P = 0,$$

where $T(\lambda) \leq (\theta - 1)P$ is a monotonically increasing function in λ. Therefore, the

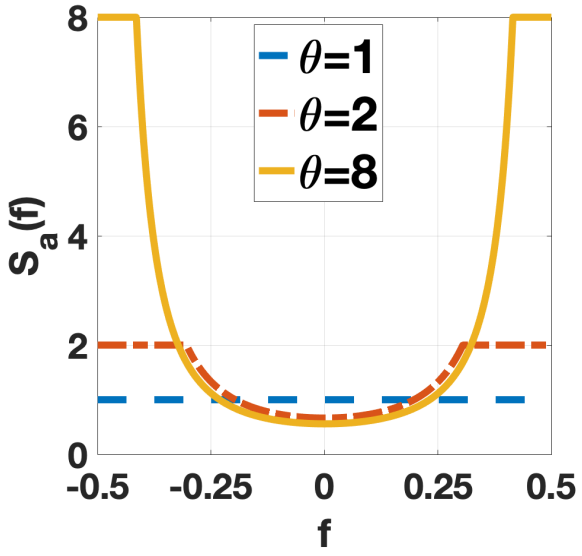


Figure 6.4: θ-limited PSD for different values of θ

optimal λ can be found by a simple bisection search. The optimal PSD can be calculated by inserting the optimal λ in (6.15). Using the proposed algorithm, the optimal PSDs are illustrated in Fig. 6.4 for $\beta = 1, \delta = 0.5, P = 1$ and different values of θ. In summary, the proposed method limits the very large values of the original PSD to θP and then adjusts

the rest of the PSD to satisfy the power constraint.

6.6 Fading Channels

Denoting the discrete-channel frequency response as $H_d(f)$, the capacity of FTN signaling can be formulated as:

$$C(P,\delta T) = \sup_{S_a(f)} \frac{1}{2\delta T} \int_{\frac{-1}{2}}^{\frac{1}{2}} \log_2 \left(1 + \frac{1}{\sigma_n^2} G_{\delta T}\left(\frac{f}{\delta T}\right) S_a(f)|H_d(f)|^2\right) df.$$
$$s.t. \;\; \frac{1}{\delta T} \int_{-1/2}^{1/2} S_a(f) G_{\delta T}\left(\frac{f}{\delta T}\right) df \leq P. \tag{6.16}$$

Considering the KKT conditions, the capacity is achieved by:

$$S_a(f) = \begin{cases} \frac{X(f)}{G_{\delta T}(f/\delta T)} & G_{\delta T}(f/\delta T) \neq 0 \\ 0 & o.w. \end{cases}, \tag{6.17}$$

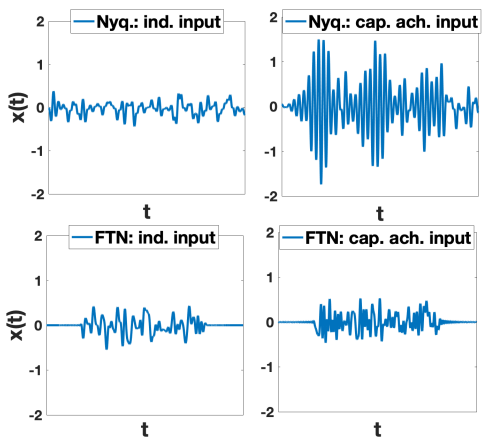
where $X(f) = [\lambda - \frac{\sigma_n^2}{H(f)}]^+$, $H(f) = \begin{cases} |H_d(f)|^2 & G_{\delta T}(f/\delta T), \neq 0 \\ 0 & o.w. \end{cases}$, and the water level, λ, is found by satisfying the power constraint, i.e., $\frac{1}{\delta T}\int_{\frac{-1}{2}}^{\frac{1}{2}} [\lambda - \frac{\sigma_n^2}{H(f)}]^+ df = P$. If the channel is time varying, then the capacity of FTN signaling equals:

$$C(P,\delta T) = E_H \left\{ \frac{1}{2\delta T} \int_{-1/2}^{1/2} \log_2 \left(1 + \frac{1}{\sigma_n^2} G_{\delta T}\left(\frac{f}{\delta T}\right) S_a^H(f)|H_d(f)|^2\right) df \right\}, \tag{6.18}$$

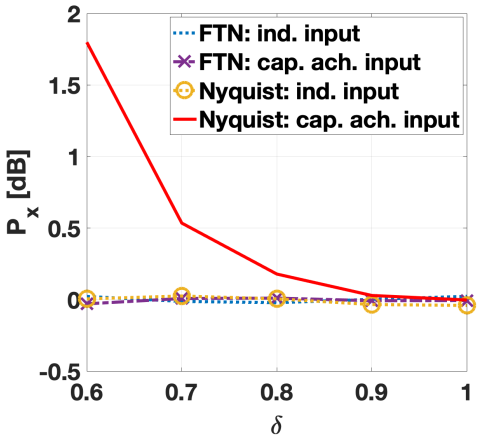
where $S_a^H(f)$ is found for each realization of channel similar to (6.17). The same calculation can be applied to the constrained capacity which is omitted due to lack of space. The effect of FTN in fading channels is further evaluated in the numerical results section.

6.7 Numerical Results

In this section, we analyze the derived results using the r.c. pulse. First, we examine the time-domain characteristics and the average transmit power of Nyquist and FTN signaling. We assume that the inputs are drawn from either an independent Gaussian process or a Gaussian process with the capacity-achieving PSD. To generate the capacity-achieving symbols, a random realization of white noise is convolved by a sequence with PSD of $S_a^{1/2}(f)$. Next, the symbols are modulated by a r.r.c. filter. Realizations of such sequences are shown in Fig. 6.5a for independent and capacity-achieving symbols with Nyquist rate, i.e., $\delta = 1$ and a FTN rate, i.e., $\delta = 0.6$. Then, the average power $P_x = \mathcal{E}\left\{\int_t |x(t)|^2 dt\right\}$ is shown in Fig. 6.5b for different values of δ.



(a) The time domain modulated signal for Nyquist and FTN signaling with $\delta = 0.6$

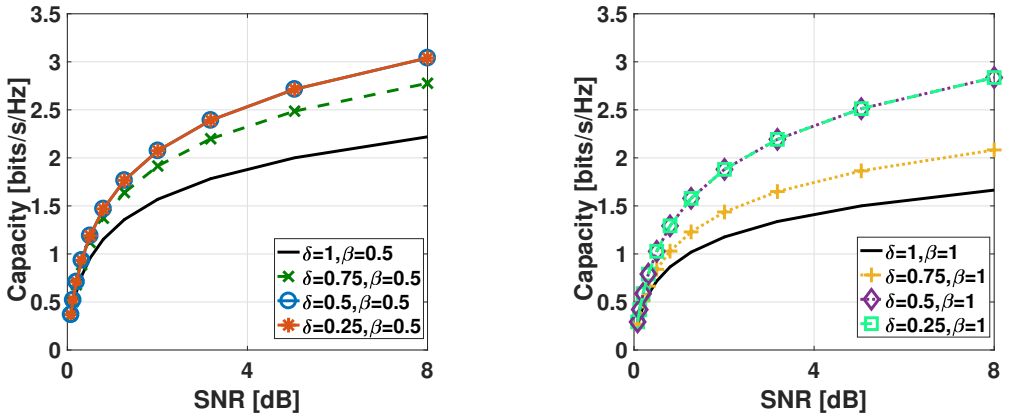


(b) P_x versus δ for Nyquist and FTN signaling

Figure 6.5: Illustration of the effect of FTN on transmit power.

Fig. 6.5 shows that the capacity-achieving PSD combined with FTN signaling satisfies the power constraint. On the other hand, the same PSD combined with Nyquist signaling, results in large instantaneous and average power. Also, note that independent PSD satisfies the power constraint and results in the same average power for all values of δ.

In Fig. 6.6, the effect of FTN signaling on the capacity is illustrated. As the signaling rate increases, the capacity increases. However, as δ drops below $\frac{1}{1+\beta}$, the capacity does not improve anymore. In addition, the capacity improvement for $\beta = 1$ is larger due to having more excess bandwidth.



(a) $\beta = 0.5$ (b) $\beta = 1$

Figure 6.6: Illustration of the effect of FTN on capacity.

As mentioned before, the capacity-achieving PSD imposes some practical issues such as OOB emission. The OOB emission is calculated as the total power leaked out of the allocated sub-band, and shown in Fig. 6.7 for $\beta = 1, \delta = 0.5, P = 10\ dB$. As θ increases, the OOB emission increases and for large values of θ, the OOB emission approaches an unaccepted level of 5 dB. On the other hand, by increasing θ, the achievable rate increases. In fact, we have $C_1 = C_{ind}$ and $C_\infty \to C$. Therefore, θ provides a flexible mechanism to capture the trade-off between an acceptable OOB emission and a desired rate. For example, $\theta = 10$ can result in an acceptable OOB emission of $\sim -30\ dB$ while it provides $\sim 2.5\ bits/s$ rate (i.e., 96% of the FTN capacity). In Fig. 6.8, the capacity of θ-limited PSD with $\theta = 10$ is compared with that of independent and capacity-achieving PSDs for $\beta = 0.5$ and $\delta = 0.5$. The channel follows a multi-path model with 5 taps where each tap is a Gaussian random variable and the average in (6.18) is taken over 10000 channel

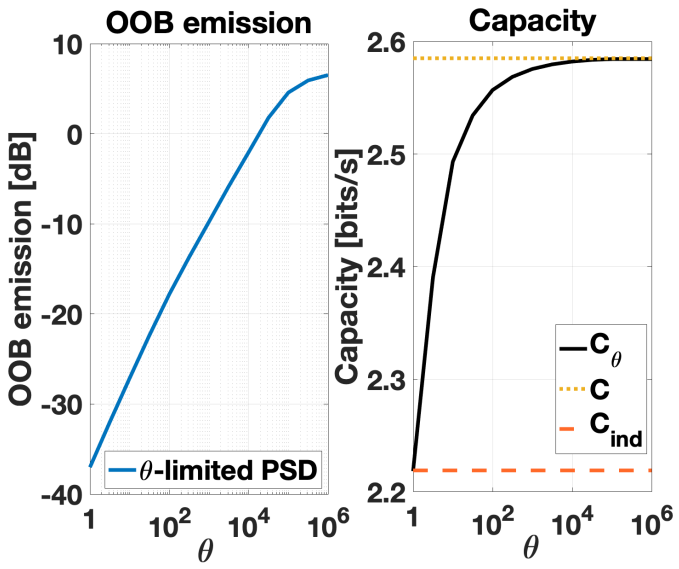


Figure 6.7: Illustration of the effect of θ on OOB emission and capacity.

realizations. The two step algorithm proposed in Section 6.6 is utilized which includes a water-filling step and a power adjustment step based on the threshold θ. The proposed PSD improves the performance compared with the independent PSD and its performance is very close to the capacity-achieving input.

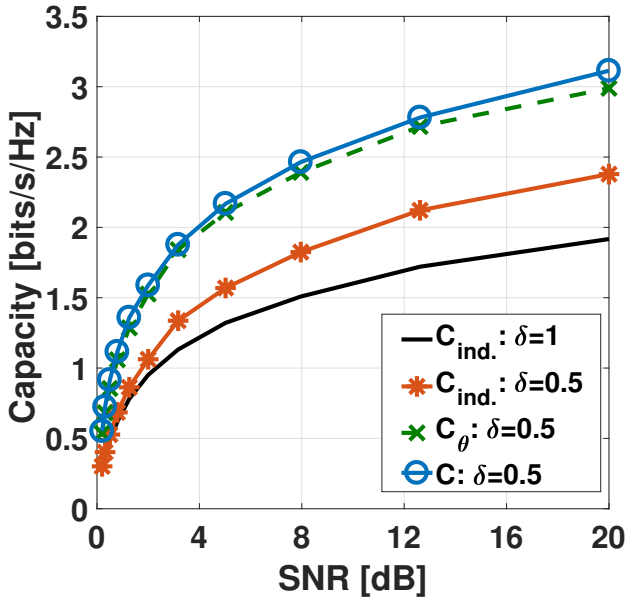


Figure 6.8: Comparison of C, C_θ and $C_{ind.}$.

6.8 Conclusion

In this chapter, we analyzed the capacity of FTN signaling. We showed that an independent PSD does not achieve the capacity of FTN signaling. Instead, the capacity-achieving PSD is derived, which is proportional to the inverse of the folded-scaled spectrum of the pulse shape. However, the capacity-achieving input causes practical issues such as OOB emission. To solve these issues, we introduce a new optimizing problem with an additional constraint on the input PSD. The introduced θ-limited input PSD can strike a trade-off between the independent input and the capacity-achieving input. The value of θ can be engineered to achieve an acceptable OOB emission and a desirable capacity.

Chapter 7

Exploiting Time Asynchrony in Multiuser Transmit Beamforming

In this chapter, we analyze the benefits of intentionally adding timing mismatch in the downlink transmit beamforming for wireless transmission. Transmit beamforming enables the so-called space-division multiple access (SDMA), where multiple spatially separated users are served simultaneously. The optimal beamforming vectors can be found to minimize the average transmit power under each user's Quality-of-Service (QoS) constraint. We show that intentionally adding timing offsets between the transmitted signals can significantly reduce the average transmission power compared with the conventional optimal beamforming method while providing the same QoSs for users. The frequency-selectivity in communication channels provides the opportunity to exploit intelligent design for performance improvement. The frequency-selectivity is limited in environments with line-of-sight links or little scattering. In such environments, we propose adding intentional time delays to induce frequency-selectivity that can be exploited. We provide three different methods exploiting the artificially induced frequency-selectivity, which improve the per-

formance with a computational complexity similar to that of the optimal synchronous beamforming. We derive the expressions for the achievable rates using the proposed methods and then provide efficient algorithms to solve the minimum power optimization. We show analytically and numerically that our proposed methods outperform the conventional optimal transmit beamforming.

7.1 Introduction

Transmit beamforming is a versatile technique for signal transmission to serve multiple users simultaneously in multiple-antenna systems [106]. With multiple antennas, the idea of transmit beamforming is to transmit directional beams to reduce the co-channel interference and, thus, enables serving several users using the same resource slot, which is called spatial division multiple access (SDMA). In contrast to the space-time coding methods [25], which can be designed in the time-space domain without the aid of channel state information (CSI), transmit beamforming exploits the CSI to combat the channel fading [22, 107, 108]. In theory, dirty paper coding (DPC), a multiuser encoding strategy based on the interference pre-subtraction [109], is the optimal (capacity-achieving) strategy in multiple-input multiple-output (MIMO) downlink channels from the base station to mobile users [110]. However, DPC is challenging to implement in practical systems due to the high computational burden of successive encoding and decoding, especially when the number of users is large.

Thus, in the practical scenarios, where users have limited computational capabilities and only employ a single antenna, the beamforming is proposed as an alternative solution of DPC [111, 112, 113, 114, 115, 116]. Beamforming has been shown to achieve a relatively large fraction of DPC capacity with a lower computational complexity when the base

station has multiple antennas, and each user has a single antenna [117]. Moreover, it has been shown that if the beamforming vectors are chosen optimally, the sum rate of BF approaches that of DPC as the number of users goes to infinity [118, 119]. As an example, in multibeam satellite communications [120], the burden of interference cancellation is carried by the transmission side at the gateway instead of the user terminals equipped with a single antenna and limited computational capabilities [121]. Therefore, beamforming techniques with low complexity are more favorable than more complex DPC or Superposition/ successive interference cancellation (SIC) methods.

Many different techniques have been proposed in the literature for the multiple-input-single-output (MISO) minimum power beamforming problem [122, 123, 124, 125] which generally assume symbol-level synchronization. However, in this chapter, we incorporate the idea of introducing timing offsets at the transmitter side to improve the performance. The idea of intentionally introducing timing offset was investigated in other contexts. It was shown in [83] that with time asynchrony, the achievable rate region could be improved for multiple access channels. By intentionally introducing symbol asynchrony in the transmitted signal, a higher diversity gain could be achieved by zero-forcing detection in spatial multiplexing [9, 4, 48]. The benefits of asynchronism in Code Division Multiple Access (CDMA) systems with random spreading were analyzed in [126], and it was shown that asynchronous transmission could indeed enhance the spectral efficiency. Besides, asynchronous NOMA (ANOMA) systems could achieve a better throughput performance than the conventional (synchronous) NOMA systems [127, 128, 89, 129]. Orthogonal differential decoding could be improved by utilizing the oversampling technique [47, 46] to achieve the sampling diversity gain. An asynchronous network coding (ANC) transmission strategy for multiuser cooperative networks was investigated in [130], where the received signals from multiple sources were asynchronous to each other. In [131], a novel interference cancellation (IC) technique was proposed for asynchronous NOMA systems, which exploited

a triangular pattern to perform the IC. It was shown that the NOMA with the proposed asynchronous IC technique could outperform the conventional synchronized orthogonal multiple access (OMA).

Traditionally, time asynchrony is viewed as a source of performance degradation that has to be mitigated. However, we intentionally add time delays to induce frequency-selectivity that can be exploited. In many applications, such as satellite communication, there is a line-of-sight path or little scattering in the environment, limiting the frequency-selectivity gains. In this chapter, we propose to exploit the artificially added frequency-selectivity, induced by adding time delays, when the environment has little scattering. In a related work [112], opportunistic beamforming is proposed, in which channel fluctuations are introduced in the time domain by changing the phase and amplitude of each antenna element even if the physical channel gains have minimal fluctuations. Opportunistic beamforming exploits the fluctuations to achieve multiuser diversity. Similarly, we induce the frequency-selectivity by intentionally adding timing offsets even if the physical channel is frequency-flat and exploit that to achieve better performance, i.e., reduce the transmission power.

We propose three methods to exploit induced frequency-selectivity. The first method requires no additional processing at the transmitter or the receiver beyond that in a conventional system, except for adding time delays among transmitted streams at the transmitter side. We prove that the average power of the induced frequency-selective channels can be reduced by proper choice of time delays. Thus, we will have reduced inter-user interference (IUI) and improved performance. The second method applies simple frequency transformations at the transmitter and receiver to decompose the effective frequency-selective channels into parallel multi-channel. It employs power allocation algorithms to reduce the transmit power further. The third method applies the concept of oversampling in conjunction with the asynchronous transmission to generate a system

model with a higher rank, enabling IUI cancellation.

The remainder of this chapter is organized as follows. In Section 7.2, the general system model and some introductory notes are presented to set the scene for introducing the proposed methods. In Section 7.3, Method A, its corresponding rate expressions and its performance are analyzed. In Section 7.4, Method B and its corresponding rate expressions are presented, and an efficient algorithm to solve the corresponding power minimization is provided. In Section 7.5, Method C with user scheduling and oversampling technique is presented. Next, the rate expressions with simplifying sub-channel/power assignments are provided to enable SDR. In Sections 7.6 and 7.7, the numerical results and final remarks are presented, respectively. The materials in this chapter are originally published in [132].

7.2 General System Model and Preliminaries

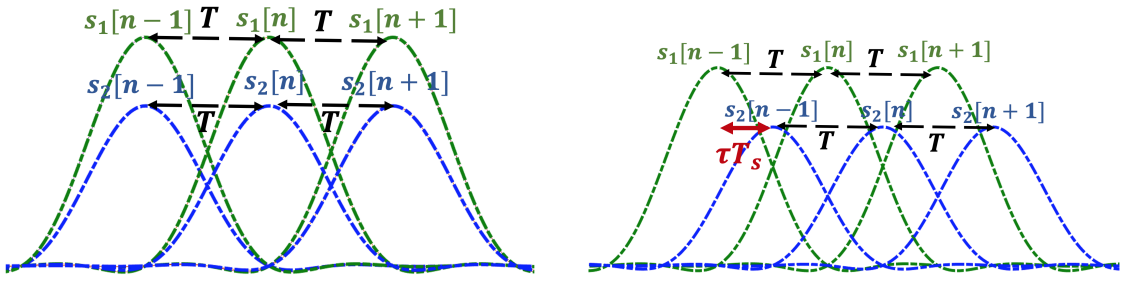
We consider a downlink wireless communication system consisting of one transmitter equipped with M transmit antennas and K single-antenna users where the environment has limited frequency-selectivity or more specifically is frequency-flat. The system operates over a multi-user fading channel and the transmitter sends a block of N complex symbols to each user. Channel coefficients change independently from one block to another and are perfectly known at the base station (BS) [123].

Let $\boldsymbol{d_k} = (d_k[1], \cdots, d_k[N])^T \in \mathbb{C}^{N\times 1}$ denote the random vector of symbols, where $d_k[n]$ is the symbol intended for User k at time instant n with average power of $\mathbb{E}[|d_k[n]|^2] = 1$. In practical systems, these symbols are selected from an arbitrary constellation; however, to derive the achievable rate expressions and compare the proposed methods, we assume that the symbols are drawn from a Gaussian distribution, as commonly assumed in the

related works in the literature [133]. The symbols can be time-precoded properly to yield precoded symbols of $\boldsymbol{s_k} = (s_k[1], \cdots, s_k[N])^T \in \mathbb{C}^{N\times 1}, k = 1, \cdots, K$. Then, the precoded symbols are linearly modulated by a unit-energy real-valued pulse as:

$$s_k(t) = \sum_{n=1}^{N} s_k[n] p(t - nT), \tag{7.1}$$

where T is the symbol interval and $p(t)$ is the pulse shaping filter which can be assumed to be a rectangular pulse shape (rect.) (a theoretical pulse shape most common in the literature) or a root-raised cosine (r.r.c.) pulse with roll-off factor of β (a common pulse shape in many communication standards like DVB-S2X standard [94] for high throughput satellite systems).



(a) Synchronous transmission (b) Asynchronous transmission

Figure 7.1: Demonstration of synchronous and asynchronous transmission.

To exploit the benefits of time asynchrony, the time delay of $\tau_k T$ is intentionally inserted between the transmitted streams where τ_k is the normalized timing offset intended for User k. Thus, the transmitted signal is denoted as $\boldsymbol{s}(t) = \sum_{k=1}^{K} \boldsymbol{w_k} s_k(t - \tau_k T)$, where $\boldsymbol{w_k} \in \mathbb{C}^{M\times 1}$ denotes the beamforming vector applied to the transmitting antenna elements to generate the spatial channel for transmission to User k. Putting $\tau_k = 0, k = 1, \cdots, K$ will result in the synchronous system model. In previous chapters, we have shown that the uniform time delays, i.e., $\tau_k = (k-1)/K$, result in optimal performance in different settings. Thus, in this chapter, we use the uniform time delays although any other choice

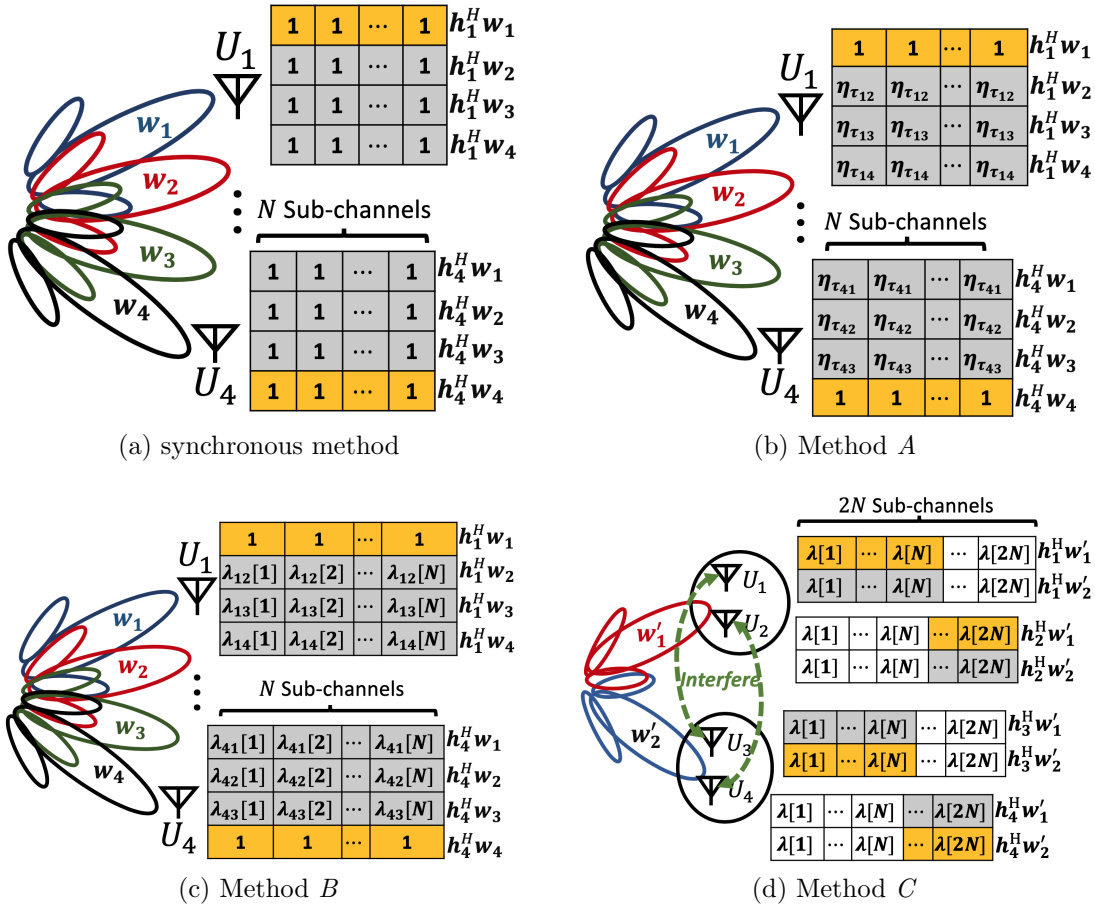
of time delays is still applicable. The difference between synchronous and asynchronous transmission is demonstrated in Fig. 7.1. The introduced timing offsets at the transmit side are known by the BS and can be sent to the users. The channel might also introduce additional time delay at the user side which needs to be estimated and compensated by the synchronization methods. The effect of the residual timing synchronization error is analyzed in [128]. We assume that channel-imposed time delays are compensated by the users.

The received signal at User k can be denoted as $y_k(t) = \sum_{l=1}^{K} \boldsymbol{h_k}^H \boldsymbol{w_l} s_l(t - \tau_l T) + n_k(t)$ where $\boldsymbol{h_k} \in \mathbb{C}^{M\times 1}$ denotes the channel vector for User k and $n_k(t)$ is the additive white Gaussian noise (AWGN) at the kth user with variance of σ_k^2. Applying the received signal at the kth user to a matched filter with the impulse response $p(t)$ and sampling the filter output at time instants $t_n = nT + \tau_k T$, $n = 1, 2, \cdots, N$ yields a set of statistics, $y_k[n]$, for detecting the transmitted symbol vectors. Denoting $\boldsymbol{y_k} = (y_k[1], \cdots, y_k[N])^T$, we will have $\boldsymbol{y_k} = \sum_{l=1}^{K} \boldsymbol{G_{kl}} \boldsymbol{h_k}^H \boldsymbol{w_l} \boldsymbol{s_l} + \boldsymbol{n_k}$. The matrix $\boldsymbol{G_{kl}}$, called the "*offset matrix*", is an $N \times N$ Toeplitz matrix whose elements depend on the pulse shape and the corresponding time delay and are denoted as:

$$[\boldsymbol{G_{kl}}]_{m,n} = g(\tau_{kl} T + (m-n)T) \triangleq g_{\tau_{kl}}[m-n], \quad m, n = 1, \cdots, N, \tag{7.2}$$

where $\tau_{kl} = \tau_k - \tau_l$ and $g(t) = p(t) * p(t)$. Denoting u as the number of significant (truncated) side-lobes in the Nyquist pulse shape, the offset matrix is a u-banded Toeplitz matrix. The vector $\boldsymbol{n_k}$ represents the noise vector at User k whose covariance matrix is $\boldsymbol{Q_k} = \mathbb{E}[\boldsymbol{n_k} \boldsymbol{n_k}^H] = \sigma_k^2 \boldsymbol{I_N}$. For any square-root Nyquist pulse, e.g., r.r.c., $\boldsymbol{G_{kk}} = \boldsymbol{I_N}$ and $\boldsymbol{G_{kl}^T} = \boldsymbol{G_{lk}}$. Also note that for the synchronous transmission, i.e., $\tau_k = 0, \forall k$, all the timing offset matrices become an identity matrix. Intuitively, by intentionally adding time delays, the interfering signals at the receiver side will appear as signals that have

passed through a channel with multiple taps, i.e., a frequency-selective channel. Method A exploits the reduced power of the interfering frequency-selective channels, and Method B exploits frequency-selectivity by using Discrete Fourier Transform (DFT) and proper power management, and Method C exploits oversampling to enable IUI cancellation which are illustrated in Fig. 7.2. The details of these methods will be presented in different sections as follows.



(a) synchronous method

(b) Method A

(c) Method B

(d) Method C

Figure 7.2: Simple illustration of the proposed methods: (a) the conventional synchronous method, (b) Method A: utilizes the "*reduced IUI*" caused by asynchronous transmission, (c) Method B: exploits the *"frequency-selectivity"* enabled by asynchronous transmission. (d) Method C: exploits *"additionally available rank"* introduced by asynchronous transmission and oversampling. Colors yellow, grey and white indicate the desired signal, the IUI, and no-interference, respectively.

7.3 Spatial Beamforming Design with No Time-domain Precoding

In this section, we present Method A which requires no additional processing at the transmitter or receiver side. Compared with a conventional system, the only change is the introduction of timing offsets at the transmitter. With no time-domain precoding, i.e., $\boldsymbol{s_k} = \boldsymbol{d_k}$, the received samples at User k can be written as $y_k[n] = \boldsymbol{h_k^H w_k} d_k[n] + \sum_{\substack{l=1 \\ l\neq k}}^{K} \boldsymbol{h_k^H w_l} \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m} d_l[m] + n_k[n], n = 1, \cdots, N$ where $\boldsymbol{G_{kl}}$ becomes an identity matrix for the synchronous transmission. Assuming Gaussian-distributed symbols, the variance of the effective noise, i.e., $\tilde{n}_k[n] = \sum_{\substack{l=1 \\ l\neq k}}^{K} \boldsymbol{h_k^H w_l} \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m} d_l[m] + n_k[n]$, can be calculated as $\tilde{\sigma}_k^2[n] = \sum_{\substack{l=1 \\ l\neq k}}^{K} |\boldsymbol{h_k^H w_l}|^2 \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m}^2 + \sigma_k^2$. As a result, the achievable rate at User k can be written as $r_k = \lim_{N\to\infty} \frac{1}{N} \sum_n \log_2(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\tilde{\sigma}_k^2[n]})$. Note that r_k provides a lower-bound for the achievable rate of non-Gaussian symbols. Due to the Toeplitz structure of the offset matrices, the expression for the achievable rate can be further simplified. For non-boundary sub-channels, i.e., $n = u, \cdots, N-u$, the variance of effective noise is independent of the sub-channel index, i.e., $\tilde{\sigma}_k^2 = \sum_{\substack{l=1 \\ l\neq k}}^{K} |\boldsymbol{h_k^H w_l}|^2 \sum_{n=-u}^{u} g_{\tau_{kl}}^2[n] + \sigma_k^2$. Therefore, as $N \to \infty$, the effect of boundary sub-channels vanishes and the achievable rate can be written as:

$$r_k^A = \log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} |\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2} \right), \tag{7.3}$$

where r_k^A denotes the achievable rate by asynchronous transmission and symbol-by-symbol detection at User k, $\eta_{\tau_{kl}} = \sum_{n=-u}^{u} g_{\tau_{kl}}^2[n]$ represents the "*offset coefficient*", and u is the number of significant (truncated) side-lobes in the Nyquist pulse shape. Note that, although $N \to \infty$ is required for rigorous derivation of the achievable rate, r_k^A approximates the achievable rate very well even for a moderate block length.

The offset coefficients $\eta_{\tau_{kl}}$ are the key factors to analyze the achievable rates of Method A. The offset coefficients represent the average power of the frequency-selective channels of interfering users. With synchronous transmission, i.e., $\tau_k = 0, \forall k$, the offset coefficients are equal to one and as a result, the achievable rates simplify to the conventional synchronous rate expressions $r_k^{synch} = \log_2\left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^{K} |\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right)$ [123]. However, for asynchronous transmission, i.e., $0 < \tau_k < 1$, the effective frequency-selective channel has a lower average power which is verified and quantified in the next lemma.

Lemma 7.1. *For any Nyquist filter denoted as $g(t)$, the offset coefficient defined as $\eta_\tau = \sum_{n=-u}^{u} g_\tau^2[n]$, for a given τ, has the following property:*

$$\eta_\tau \leq \eta_0 \tag{7.4}$$

In addition, η_τ can be calculated for the rectangular pulse and root raised cosine pulse with roll-off factor of β as:

$$\eta_\tau^{rect.} = \tau^2 + (1-\tau)^2, \quad \eta_\tau^{r.r.c.} = 1 - \beta/4 + \beta\cos(2\pi\tau)/4, \tag{7.5}$$

respectively.

Proof. The proof is presented in Appendix D.1. □

This lemma states that sampling a Nyquist pulse at a time instant different than its peak value leads to a reduced-power set of samples. Equivalently, it can be interpreted that the average power of the induced frequency-selective channels can be reduced by a proper choice of time asynchrony. The reduction in the power is caused by out-of-phase addition of folded spectrum and is proportional to the out-of-Nyquist-band spectrum of the used pulse shape, as explained in Appendix D.1. For example, for the perfect Nyquist pulse,

i.e., sinc function, the power of samples does not decrease with any choice of timing offsets in the sampling. In other words, when $\beta = 0$, $\eta_\tau^{r.r.c.} = 1$ for all values of τ. For other values of β, the choice of timing offset impacts the power of resulting samples. For example, $\tau = 0$ results in the maximum of $\eta_\tau = 1$ and $\tau = 1/2$ results in the minimum of $1 - \beta/2$ for the r.r.c. pulse shapes. In Fig. 7.3, the behaviour of the offset coefficient with respect to the normalized timing offsets is shown for a rect. pulse and truncated r.r.c. pulses with the roll-off factor $\beta = 0.1, 0.5, 1$. Note that $\tau = 0.5$ results in the smallest offset coefficient.

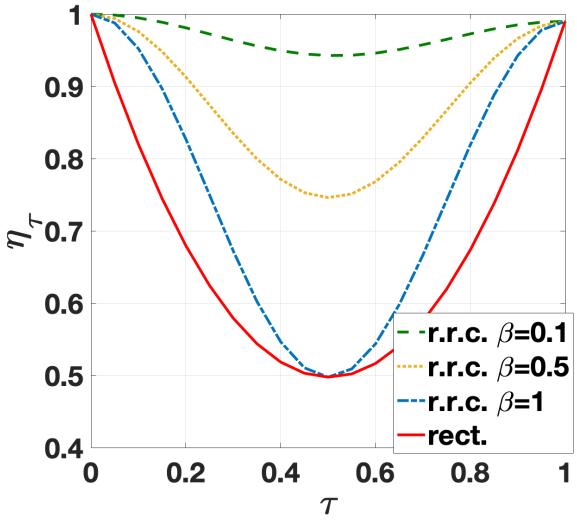


Figure 7.3: Values of η_τ for different timing offsets.

The power minimization problem can be expressed as [111, 125]:

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^K} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{7.6}$$
$$s.t. \ \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^K \eta_{\tau_{kl}} |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2} \geq \gamma_k^*, \ \ k = 1, \cdots, K,$$

where the parameters $\gamma_k^* = 2^{r_k^*} - 1$ describe the transmit power and the signal to interference plus noise ratios (SINRs) required by each user. The optimal solution can be obtained using the uplink-downlink duality and modifying the optimal algorithm pre-

sented in [111, 125]. For completeness, the highlight of the algorithm is presented in Appendix D.2.

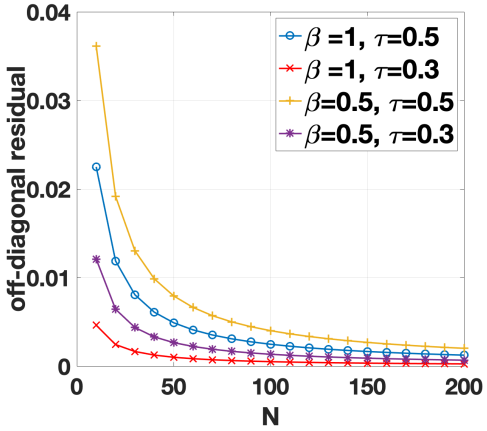
Proposition 7.1. *The average transmission power obtained from the asynchronous transmission in (7.6) is less than that of the synchronous method, i.e.,* $p^{A}_{avg,opt} < p^{synch}_{avg,opt}$.

Proof. The proof is trivial as the minimization problem for both the synchronous and asynchronous methods include the same objective function while having looser constraints for the latter due to the reduced IUI. □

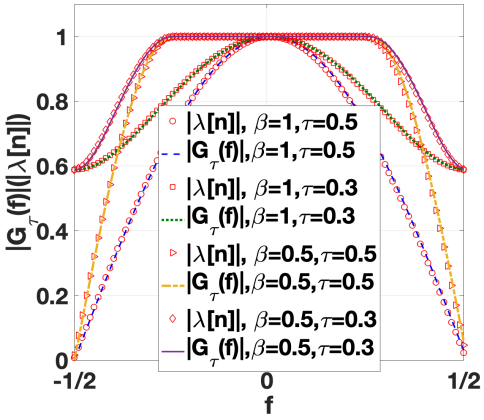
7.4 Spatial Beamforming design with Individual Time-domain Precoding

In this section, we present Method B which exploits DFT to decompose the induced frequency-selective channels into parallel multi-channels and applies power allocation to further reduce the transmit power. Recalling the received samples at User k, $\boldsymbol{y_k} = \sum_{l=1}^{K} \boldsymbol{G_{kl}} \boldsymbol{h_k}^H \boldsymbol{w_l} \boldsymbol{s_l} + \boldsymbol{n_k}$, note that the offset matrices are banded Toeplitz matrices. Banded Toeplitz matrices are asymptotically equivalent to circulant matrices as the matrix dimension goes to infinity [134, 135]. The first implication of the asymptotic equivalence of Banded Toeplitz matrices with circulate matrices is that a banded Toeplitz matrix can be diagonalized by DFT matrices as its size grows large. In other words, as $N \to \infty$, matrix $\boldsymbol{G_{kl}}$ can be denoted as $\boldsymbol{U_N} \boldsymbol{\Lambda_{kl}} \boldsymbol{U_N}^H$ where $\boldsymbol{U_N}$ denotes the $N \times N$ DFT matrix and $\boldsymbol{\Lambda_{kl}}$ is a diagonal matrix whose nth diagonal element is denoted as $[\boldsymbol{\Lambda_{kl}}]_{nn} = \lambda_{kl}[n]$. Diagonal structure of $\boldsymbol{\Lambda_{kl}} = \boldsymbol{U_N^H} \boldsymbol{G_{kl}} \boldsymbol{U_N}$ is also verified in Fig. 7.4a for a moderate number of block length. The off-diagonal residual, defined as sum of squared of off-diagonal elements is shown for the r.r.c. pulse and various values of N. It can be seen that, as N increases, the

off diagonal elements converge to zero and, thus, the offset matrices can be diagonalized by DFT matrix.



(a) Verification of diagonal structure of $\boldsymbol{\Lambda_{kl}} = \boldsymbol{U_N^H G_{kl} U_N}$ with r.r.c. pulse.



(b) Comparison of $G_{\tau_{kl}}(f)$ and $\lambda_{kl}(n)$ with r.r.c. pulse and $N = 100$.

Figure 7.4: This figure verifies that for sufficiently large N, Toeplitz matrix $\boldsymbol{G_{kl}}$ can be diagonalized with DFT matrices where the diagonal elements ($\lambda_{kl}[n]$) are the samples of the generating function of $\boldsymbol{G_{kl}}$, i.e., $G_{\tau_{kl}}(f)$. (a) shows that after diagonalizing matrix $\boldsymbol{G_{kl}}$ with DFT matrices, the off-diagonal residual goes to zero as N increases. (b) shows the equivalence of diagonal elements of $\boldsymbol{\Lambda_{kl}}$ and samples of $G_{\tau_{kl}}(f)$, i.e., $\lambda_{kl}[n] = G_{\tau_{kl}}(n/N)$

.

Therefore, to diagonalize the offset matrices, each data stream can be precoded as, $\boldsymbol{s_k} = \boldsymbol{U_N P_k}^{1/2} \boldsymbol{d_k}$ where $\boldsymbol{U_N}$ is the $N \times N$ DFT matrix and $\boldsymbol{P_k}$ is a diagonal matrix whose nth diagonal element is the power coefficient of User k's nth sub-channel denoted by $P_k[n]$. At User k, the received samples are multiplied by $\boldsymbol{U_N^H}$ to get:

$$\hat{\boldsymbol{y}}_{\boldsymbol{k}} = \boldsymbol{U_N^H y_k} = \boldsymbol{h_k^H w_k P_k^{1/2} d_k} + \sum_{\substack{l=1 \\ l \neq k}}^{K} \boldsymbol{\Lambda_{kl} h_k^H w_l P_l^{1/2} d_l} + \hat{\boldsymbol{n}}_{\boldsymbol{k}}, \tag{7.7}$$

where the covariance matrix of $\hat{\boldsymbol{n}}_{\boldsymbol{k}}$ is equal to $\sigma_k^2 \boldsymbol{I_N}$. Then, assuming Gaussian signaling, the achievable rate at User k can be written as:

$$r_k^B = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \log_2 \left(1 + \frac{P_k[n] |\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1, l \neq k}^{K} |\lambda_{kl}[n]|^2 P_l[n] |\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2} \right). \tag{7.8}$$

The next implication of the asymptotic equivalence of banded Toeplitz matrices with circular matrices is that the diagonal elements of $\boldsymbol{\Lambda_{kl}}$ are samples of the generating function of matrix $\boldsymbol{G_{kl}}$. In more details, defining the generating function of $\boldsymbol{G_{kl}}$ as $G_{\tau_{kl}}(f) = \sum_{n=-\infty}^{\infty} g_{\tau_{kl}}[n]e^{-j2\pi fn}, f \in [0,1]$, we have $\lambda_{kl}[n] = G_{\tau_{kl}}(n/N), n = 1, \cdots, N$ [53]. For example, for the offset matrix $\boldsymbol{G_{kl}}$ with r.r.c. pulse and $N = 100$, the absolute values of $\lambda_{kl}[n]$ and the function $|G_{\tau_{kl}}(f)|$ (denoted as $\lambda[n]$ and $G_\tau(f)$ in the legend for better presentation) are shown in Fig. 7.4b.

Defining $f_n = n/N$, $df_N = 1/N$ and $P_k(f_n) = P_k[n]$, we can rewrite the achievable rate as $r_k^B = \lim_{N\to\infty} \sum_{n=1}^{N} C(f_n)df_N$ where $C(f_n) = \log_2\left(1 + \frac{P_k(f_n)|\boldsymbol{h_k}^H\boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^{K}|G_{\tau_{kl}}(f_n)|^2 P_l(f_n)|\boldsymbol{h_k}^H\boldsymbol{w_l}|^2+\sigma_k^2}\right)$. Because $C(f_n)$ is bounded and almost everywhere continuous on the interval $[0,1]$, then it is Reimann integrable on the interval [136], and we get:

$$r_k^B = \int_0^1 \log_2\left(1 + \frac{P_k(f)|\boldsymbol{h_k}^H\boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^{K}\lambda_{kl}(f)P_l(f)|\boldsymbol{h_k}^H\boldsymbol{w_l}|^2 + \sigma_k^2}\right)df, \tag{7.9}$$

where $\lambda_{kl}(f) = |G_{\tau_{kl}}(f)|^2$, called the "*offset function*", depends on the pulse shape and the corresponding time delay τ_{kl}. The power distribution function of User k is denoted as $P_k(f), f \in [0,1]$. A similar approach is also used in [137] to show the capacity region of broadcast channels with inter-symbol interference (ISI) and colored Gaussian noise.

The offset function $\lambda_{kl}(f)$ is the deciding factor on the achievable rates of Method *B*. Various examples of the offset function is demonstrated in Fig. 7.5 for r.r.c. pulse shape and different time delays. The addition of time delays effectively transforms the flat fading channels into frequency-selective channels which can be exploited by proper power spectrum management.

The average transmission power can be calculated as $\sum_{k=1}^{K}||\boldsymbol{w_k}||^2\int_0^1 P_k(f)df$, hence, the

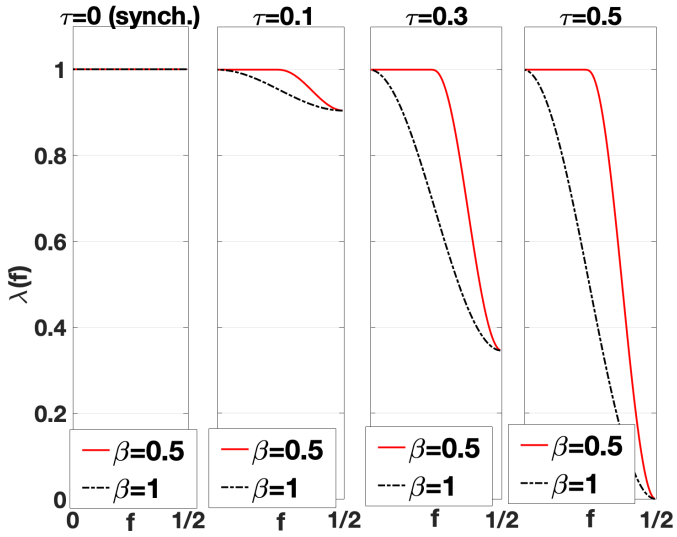


Figure 7.5: Demonstration of $\lambda_{kl}(f)$ for r.r.c. pulse with $\beta = 0.5, 1$ and different values of timing offsets.

power minimization can be written as:

$$\min_{\{P_k(f),\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \int_0^1 P_k(f)df \tag{7.10}$$

$$s.t. \quad \int_0^1 \log_2\left(1 + \frac{P_k(f)|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^{K} \lambda_{kl}(f)P_l(f)|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right)df \geq r_k^*, \quad k = 1, \cdots, K.$$

Proposition 7.2. *The optimal transmission power obtained from the power minimization in (7.10) is less than Method A, i.e.,* $p_{avg,opt}^B < p_{avg,opt}^A$.

Proof. The proof is presented in Appendix D.3. □

Due to non-convexity of the objective function in (10) and also the non-convex set of constraints, solving (10) leads to a computationally intractable problem. To reduce the complexity, a two-step sub-optimal algorithm is proposed to find the beamforming vectors and power distribution functions, subsequently. In the first step, the beamforming vectors are obtained by assuming uniform power distribution functions which reduces the

optimization problem to

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{7.11}$$
$$s.t. \quad \int_0^1 \log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^{K} \lambda_{kl}(f)|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right) df \geq r_k^*, \quad k = 1, \cdots, K.$$

To efficiently find the beamforming vectors under the rate constraints, the beamforming vectors can be found using the lower bounds of the actual rate expressions (refer to Appendix D.3 for more details). The optimization problem can be rewritten as

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{7.12}$$
$$s.t. \quad log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^{K} \int_0^1 \lambda_{kl}(f) df |\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right) \geq r_k^*, \quad k = 1, \cdots, K.$$

Since the rate constraints in (7.12) are tighter versions compared with those in (11), the solution of (7.12) is also a suboptimal solution for (11). The optimization problem in (7.12) is equivalent to the optimization problem in (7.6) because $\int_0^1 \lambda_{kl}(f) df = \eta_{\tau_{kl}}$. Hence, the optimal algorithm presented in Appendix D.2 can be used to find the optimal beamforming vectors for the above problem as well.

In the second step, the power distribution functions are found. Power spectrum management or Dynamic Spectrum Management (DSM) is a well-known and an effective method for reducing the effect of crosstalk in Digital Subscriber Line (DSL) systems [138]. Various DSM algorithms are proposed in the literature including the Optimal Spectrum Balancing (OSB), Iterative Spectrum Balancing (ISB), Iterative Water-Filling (IWF) and Successive Convex Approximation for Low complExity (SCALE) [139, 140, 141]. Denoting $||\boldsymbol{w_k^*}||^2$ and $\lambda_{kl}(f)|\boldsymbol{h_k^H w_l^*}|^2$ as ρ_k and $\hat{\lambda}_{kl}(f)$, respectively, where $\boldsymbol{w_k^*}$ is the beamforming vectors

obtained from the first step, the DSM problem can be formulated as:

$$\min_{\{P_k(f)\}_{k=1}^K} p_{avg} = \sum_{k=1}^{K} \rho_k \int_0^1 P_k(f) df \tag{7.13}$$
$$s.t. \quad \int_0^1 \log_2 \left(1 + \frac{P_k(f)\hat{\lambda}_{kk}(f)}{\sum_{l=1,l\neq k}^{K} \hat{\lambda}_{kl}(f) P_l(f) + \sigma_k^2} \right) df \geq r_k^*, \quad k = 1, \cdots, K,$$

which can be solved efficiently with SCALE algorithm [142]. For example, the power distribution functions for the case of $M = 4$, $K = 4$ using r.r.c. pulse shape with $\beta = 0.5$ and a random channel realization is shown in Fig. 7.6. Observe that the frequency-

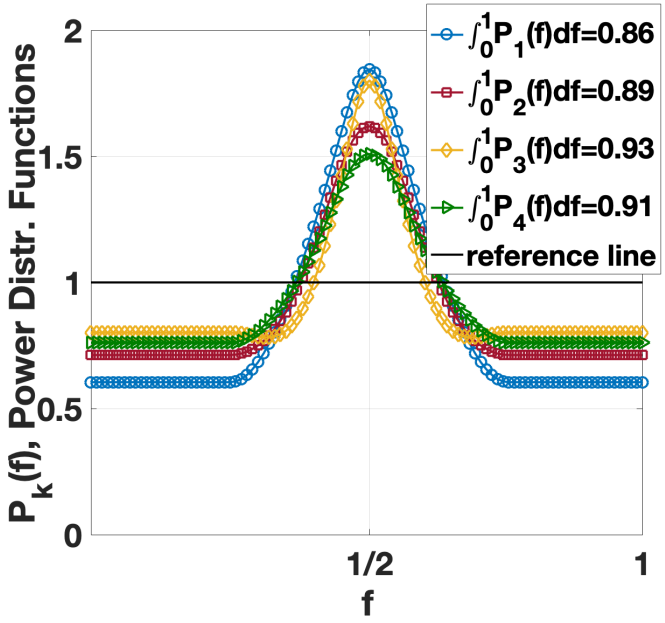
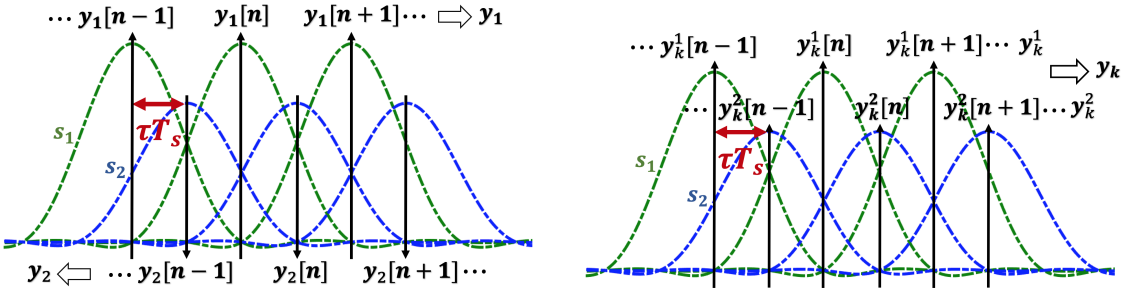


Figure 7.6: Demonstration of power distribution functions obtained by the proposed two-step algorithm for r.r.c. pulse with $\beta = 0.5$.

selectivity, imposed by asynchronous transmission, is exploited by the proposed algorithm and further reduction in the average transmit power is achieved, i.e., $\int_0^1 P_k(f)df < 1$, as shown in the legend of Fig. 7.6.

7.5 Spatial Beamforming Design with Joint Time-domain Precoding

In this section, we present Method C that exploits oversampling and timing asynchrony. Oversampling provides an independent set of sufficient statistics for the asynchronous transmission which can be exploited to cancel the IUI [83].



(a) Sampling at generic users 1 and 2 (b) Oversampling at User k

Figure 7.7: Demonstration of oversampling method.

To obtain the set of sufficient statistics, each user samples the received signal K times more than the previous methods, i.e., at time instants $t_n^j = nT + \tau_j T$, for $n = 1, 2, \cdots, N, j = 1, \cdots, K$ which yields a set of sufficient statistics $y_k^j[n]$ for detecting the transmitted symbol vectors [83]. The oversampling method is demonstrated in Fig. 7.7. Denoting $\boldsymbol{y_k^j} = (y_k^j[n], \cdots, y_k^j[N])^T$, we will have $\boldsymbol{y_k^j} = \sum_{l=1}^{K} \boldsymbol{h_k^H w_l G_{jl} s_l} + \boldsymbol{n_k^j}$ where $\boldsymbol{n_k^j}$ represents the noise vector of the jth set of samples at User k whose covariance matrix is $\boldsymbol{Q_k^j} = \mathbb{E}[\boldsymbol{n_k^j} \boldsymbol{n_k^j}^H] = \sigma_k^2 \boldsymbol{I_N}$. Putting all the samples together as $\boldsymbol{y_k} = (\boldsymbol{y_k^1}^T, \cdots, \boldsymbol{y_k^K}^T)^T$ results in:

$$\boldsymbol{y_k} = \begin{pmatrix} \boldsymbol{G_{11}} & \boldsymbol{G_{12}} & \cdots & \boldsymbol{G_{1K}} \\ \boldsymbol{G_{12}^T} & \boldsymbol{G_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{G_{K-1,K}} \\ \boldsymbol{G_{1K}^T} & \cdots & \boldsymbol{G_{K-1,K}^T} & \boldsymbol{G_{KK}} \end{pmatrix} \begin{pmatrix} \boldsymbol{h_k^H w_1 I_N} & \boldsymbol{0_N} & \cdots & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{h_k^H w_2 I_N} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \cdots & \boldsymbol{0_N} & \boldsymbol{h_k^H w_K I_N} \end{pmatrix} \begin{pmatrix} \boldsymbol{s_1} \\ \boldsymbol{s_2} \\ \vdots \\ \boldsymbol{s_K} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_k^1} \\ \boldsymbol{n_k^2} \\ \vdots \\ \boldsymbol{n_k^K} \end{pmatrix}$$

$$= \boldsymbol{G H_k s} + \boldsymbol{n_k}, \tag{7.14}$$

where $\mathbf{0_N}$ represents the $N \times N$ all-zero matrix. Matrix $\boldsymbol{G}$ contains the offset matrices as its constructive blocks. Matrix $\boldsymbol{H_k}$ represents the effective channel coefficients at User k which depends on the actual channel and the choice of beamforming vectors at the antenna elements. The vectors $\boldsymbol{s}$ and $\boldsymbol{n_k}$ denote the precoded symbols and the noise vector at User k whose covariance matrix equals $\boldsymbol{G}\sigma_k^2$ due to oversampling. It is shown that Matrix $\boldsymbol{G}$ with distinctive time delays is positive definite for any *time-limited* pulse shape (which encompasses all the pulse shapes in practice) [7]. Therefore, unlike the synchronous transmission where matrix $\boldsymbol{G}$ becomes rank-deficient, in the asynchronous transmission with distinctive time delays, matrix $\boldsymbol{G}$ is full-rank [8]. Because matrix $\boldsymbol{G}$ is full-rank, NK sub-channels are available to be exploited. In addition, matrix $\boldsymbol{G}$ is a block Toeplitz matrix which by using uniform time delays and proper ordering in the received samples become banded Toeplitz [93] and, thus can be diagonalized by DFT matrix, as explained in Section 7.4.

In order to use the additionally available rank, and be able to cancel the co-channel interference, the users should be divided into multiple groups with a common beamforming vector assigned to each group. In this way, the effective channel matrix will be the same for all the users in a group and the offset matrix can be diagonalized by DFT to remove the co-channel interference. The information symbols of users within each group are precoded jointly and transmitted by the same beamforming vector. Spatial-domain beamfroming is used to avoid inter-group interference and time-domain precoding is used to avoid intra-group interference as shown in Fig. 7.8.

Assume K/q groups, $\{\mathcal{G}_1, \cdots, \mathcal{G}_{K/q}\}$, where each of them includes q users, i.e., $|\mathcal{G}_g| = q,\ g = 1, \cdots, K/q$ and define a user-grouping function that assigns each user to a group, i.e., $\pi : \mathcal{K} \to \mathcal{G}$, where $\mathcal{K} = \{1, \cdots, K\}$ and $\mathcal{G} = \{1, \cdots, K/q\}$ are the set of user and group indices, respectively. Each user, k, is assigned to a group, g, $\pi(k) = g$, and the user-grouping policy will be discussed later. Unlike Method B with individual precoding,

where each user's data is precoded separately, here, the intended symbols of all users in a group are precoded together. The intended symbols for the users in Group g are precoded as $\left(\boldsymbol{s}^T_{\boldsymbol{q(g-1)+1}}, \cdots, \boldsymbol{s}^T_{\boldsymbol{qg}}\right)^T = \boldsymbol{U_{Nq}} \boldsymbol{P'_g}^{1/2} \left(\boldsymbol{d}^T_{\boldsymbol{q(g-1)+1}}, \cdots, \boldsymbol{d}^T_{\boldsymbol{qg}}\right)^T$ where $\boldsymbol{U_{Nq}}$ and $\boldsymbol{P'_g}$ are the $Nq \times Nq$ DFT matrix and $Nq \times Nq$ diagonal power allocation matrix, respectively. After time-precoding, power allocation and pulse shaping, the signal for Group g can be written as: $S_g(t) = \sum_{i=1}^{q} \sum_{n=1}^{N} s_{q(g-1)+i}[n] p(t - nT - \tau_i T)$. Then, the signal for each group is beamformed by $\boldsymbol{w'_g} \in \mathbb{C}^{M \times 1}$, and hence, the transmitted signal can be written as $\boldsymbol{S}(t) = \sum_{g=1}^{K/q} \boldsymbol{w'_g} S_g(t)$.

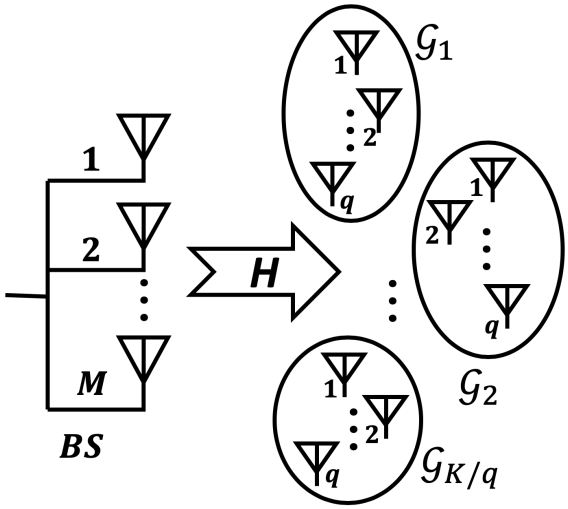


Figure 7.8: Method C's System Model.

The received signal at the kth user can be denoted as $y_k(t) = \sum_{g=1}^{K/q} \boldsymbol{h_k}^H \boldsymbol{w'_g} S_g(t) + n_k(t)$. By employing the oversampling technique, explained before, we can have:

$$\begin{aligned} \boldsymbol{y'_k} &= \begin{pmatrix} \boldsymbol{G_{11}} & \boldsymbol{G_{12}} & \cdots & \boldsymbol{G_{1,q}} \\ \boldsymbol{G^T_{12}} & \boldsymbol{G_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{G_{q-1,q}} \\ \boldsymbol{G^T_{1,q}} & \cdots & \boldsymbol{G^T_{q-1,q}} & \boldsymbol{G_{q,q}} \end{pmatrix} \sum_{g=1}^{K/q} \boldsymbol{h_k}^H \boldsymbol{w'_g} \boldsymbol{U_{Nq}} \boldsymbol{P'_g}^{1/2} \begin{pmatrix} \boldsymbol{d_{q(g-1)+1}} \\ \boldsymbol{d_{q(g-1)+2}} \\ \vdots \\ \boldsymbol{d_{qg}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_k^1} \\ \boldsymbol{n_k^2} \\ \vdots \\ \boldsymbol{n_k^q} \end{pmatrix} \\ &= \boldsymbol{G'} \sum_{g=1}^{K/q} \boldsymbol{h_k}^H \boldsymbol{w'_g} \boldsymbol{U_{Nq}} \boldsymbol{P'_g}^{1/2} \boldsymbol{d'_g} + \boldsymbol{n'_k}. \end{aligned} \tag{7.15}$$

As explained before, the banded Toeplitz matrix $\boldsymbol{G'}$ can be diagonalized by DFT matrix,

i.e., $\boldsymbol{G'} = \boldsymbol{U_{Nq}}\boldsymbol{\Lambda'}\boldsymbol{U_{Nq}^H}$, as $N \to \infty$ [135]. After multiplying $\boldsymbol{y'_k}$ with $\boldsymbol{U_{Nq}^H}$, the resulting samples can be written as $\hat{\boldsymbol{y}}'_{\boldsymbol{k}} = \boldsymbol{\Lambda'} \sum_{g=1}^{K/q} \boldsymbol{h_k^T w'}_{\boldsymbol{g}} \boldsymbol{P'}_{\boldsymbol{g}}^{1/2} \boldsymbol{d'_g} + \hat{\boldsymbol{n}}'_{\boldsymbol{k}}$ where the covariance matrix of the effective noise vector equals $\boldsymbol{\Lambda'}\sigma_k^2$. Denoting the nth diagonal elements of the $\boldsymbol{\Lambda'}$ and $\boldsymbol{P'_g}$, as $\lambda'[n]$ and $P'_g[n]$, respectively, and assuming Gaussian signaling, the rate at the kth user can be written as $r_k^C = \lim_{N\to\infty} \frac{1}{N} \sum_{n\in\mathcal{I}_k} \log_2 \left(1 + \frac{P'_{\pi(k)}[n]\lambda'[n]|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P'_j[n]\lambda'[n]|\boldsymbol{h_k^H w'_j}|^2+\sigma_k^2}\right)$ where $\mathcal{I}_k$ represents the set of sub-channels indices which are assigned to User k. The average transmit power equals $p_{avg} = \lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{\Lambda' P'_g})||\boldsymbol{w'_g}||^2$, thus, the optimization problem can be formulated as

$$\min_{\{\boldsymbol{w'_g},\boldsymbol{P'_g}\}_{g=1}^{K/q},\{\mathcal{I}_k\}_{k=1}^{K},\pi(.)} p_{avg} = \lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{\Lambda' P'_g})||\boldsymbol{w'_g}||^2 \tag{7.16}$$
$$s.t. \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n\in\mathcal{I}_k} \log_2 \left(1 + \frac{P'_{\pi(k)}[n]\lambda'[n]|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P'_j[n]\lambda'[n]|\boldsymbol{h_k^H w'_j}|^2+\sigma_k^2}\right) \geq r_k^*, \quad k = 1, \cdots, K.$$

The sub-channel and group assignment polices make the above optimization problem overly complicated and intractable. To solve the optimization problem efficiently, we use a sub-optimal method of sub-channel and power assignment rule to simplify the optimization problem. We assume the following simplifying assumptions: (I) a simple power assignment rule for sub-channels which is, indeed, optimal for an AWGN channel, and (II) the same q available sub-channel configurations for each group. These assumptions are explained in Appendix D.4 in more details. After considering the above simplifying assumptions, the power minimization problem can be formulated as:

$$\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q},\{P_k\}_{k=1}^{K},(\pi(.),\phi(.))} p_{avg} = \sum_{k=1}^{K} P_k||\boldsymbol{w'_{\pi(k)}}||^2 \tag{7.17}$$
$$s.t. \quad \frac{P_k|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P_{c(k,j)}|\boldsymbol{h_k^H w'_j}|^2+\sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K,$$

where $\pi(.), \phi(.)$ are the group assignment and sub-channel assignment functions, respec-

tively, assigning each user to a group index, i.e., ($\pi : \mathcal{K} \rightarrow \mathcal{G} = \{1, \cdots, K/q\}$) and a sub-channel assignment index ($\phi : \mathcal{K} \rightarrow \mathcal{S} = \{1, \cdots, q\}$). The index $c(k, j)$ denotes the index of the user in Group j which has the same sub-channel assignment as the kth user, i.e., $c(k, j) = \{l \in \mathcal{G}_j | \phi(l) = \phi(k)\}$.

A key step in the proposed method lies in sub-channel/user-grouping. User scheduling and sub-channel assignment is studied in various context in the literature including the NOMA and Multicast Mutigroup beamforming [143, 121, 144, 145]. The underlying intuition for user-grouping in Method C is that users assigned to the same group should have co-linear (i.e., similar) channels since they need to use the same beamforming vector. On the contrary, interfering users, assigned to other groups, should be orthogonal to minimize the interference [119]. Therefore, inspired by the multigroup multicast nature of Method C, we use the low-complexity user-grouping method detailed in [121]. Note that the focus in this thesis is analyzing the benefits of time asynchrony and not the user-grouping algorithm. Thus, any other alternative user grouping algorithm that fits our system model can be employed.

The well-known SDR technique [146] can be applied by assuming that all the power adjustment coefficients are equal to one. Thus, the original optimization problem which is a quadratically constrained quadratic programming (QCQP) problem with non-convex constraints can be relaxed to the following standard semi-definite programming (SDP).

$$\min_{\{\boldsymbol{W'_g}\}_{g:1}^{K/q}} p_{avg} = \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{W'_g}) \tag{7.18}$$

$$s.t.\ \gamma_k^* \sum_{j=1, j\neq\pi(k)}^{K/q} \mathrm{tr}(\boldsymbol{H_k W'_j}) + \gamma_k^* \sigma_k^2 - \mathrm{tr}(\boldsymbol{H_k W'_{\pi(k)}}) \leq 0$$

$$\boldsymbol{W'_g} \geq 0,\ \forall g \in \{1, \cdots, K/q\}, \tag{7.19}$$

where $\boldsymbol{H_k} = \boldsymbol{h_k}\boldsymbol{h_k}^H$, $\boldsymbol{W'_g} = \boldsymbol{w'_g}\boldsymbol{w'_g}^H$ and the non-convex rank-one condition is dropped. Thus, the solution can be found by SDP solvers like CVX [147]. However, due to the relaxation, the obtained solution will not, in general, consist of rank-one matrices. Hence, an approximate solution to the original problem can be found using a randomization technique like Gaussian randomization method [148]. We apply the following step to find the power adjustment coefficients.

Let denote $||\boldsymbol{w'^*_j}||^2$ and $|\boldsymbol{h}_k^H\boldsymbol{w'^*_j}|^2$ as ρ'_j and α_{kj}, respectively, where $\hat{\boldsymbol{w}}'^*_j$ is a beamforming vector candidate obtained from the Gaussian randomization method. Then, to find the power adjustment coefficients, we solve the following optimization problem:

$$\begin{aligned} &\min_{\{P_k\}_{k=1}^K} \; p_{avg} = \sum_{k=1}^{K} P_k \rho'_{\pi(k)} \\ s.t. \;\; &\frac{P_k \alpha_{k,\pi(k)}}{\sum_{j=1, j\neq\pi(k)}^{K/q} P_{c(k,j)}\alpha_{kj} + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K, \end{aligned} \tag{7.20}$$

which is a linear program (LP) and can be solved easily by matrix inversion as the inequalities are active at the optimal solution (see Appendix D.2 for more details). After feeding N_{rand} beamforming vector candidates to the power control step, the one with lowest objective function value is chosen as the final solution. In summary, the proposed algorithm includes solving the SDP problem once and solving the LP problem N_{rand} times [143].

7.6 Comparison and Numerical Results

7.6.1 Comparison

In Table 7.1, various properties of the proposed methods, including the complexity and delay, are compared. Due to the limited computational capabilities of the users in a beam-

Table 7.1: Comparison of the Proposed Methods

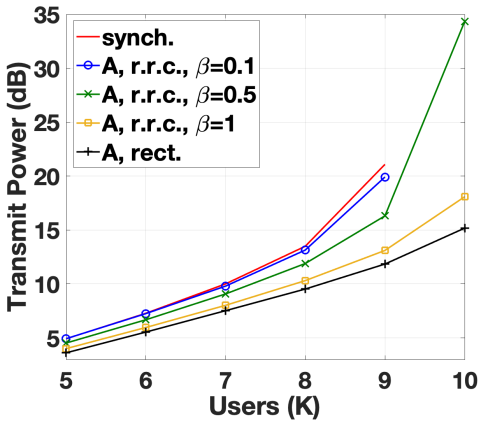
Methods	Complexity at Tx	Det. Delay	Uplink	Knowledge of Delays	Oversampling	Algorithm
synch.	STE	SS	yes	own delay	no	optimal
A	STE	SS	yes	own delay	no	optimal
B	N-IFFT+STE	FF	yes	own delay	no	sub-optimal
C	Nq-IFFT+STE	FF	no	all delays in own group	yes	sub-optimal

forming application, we only consider the complexity at the user side. The synchronous method and Method A enjoy symbol-by-symbol (SS) detection with a possible single-tap equalizer (STE) while Methods B and C perform detection on a frame-by-frame (FF) basis and have additional complexity of $N \times N$ IDFT matrix multiplication and $Nq \times Nq$ IDFT matrix multiplication, respectively. However, the $N \times N$ and $Nq \times Nq$ IDFT matrix multiplications can be effectively perforemd by N-point and Nq-point Inverse Fast Fourier Transform (IFFT). The synchronous method, Methods A and B are applicable to uplink with minor modifications, however, Method C is only applicable to downlink due to the joint precoding which requires the collective information of all users in a group. In addition, in Method C, the users require the knowledge of the user's time delays in their own group to perform oversampling, however, in other methods, each user only requires its own time delay. The suggested algorithms for the synchronous and Method A are optimal while the multi-step algorithms for Methods B and C are sub-optimal. Providing improved algorithms for Methods B and C is an interesting topic for future work.

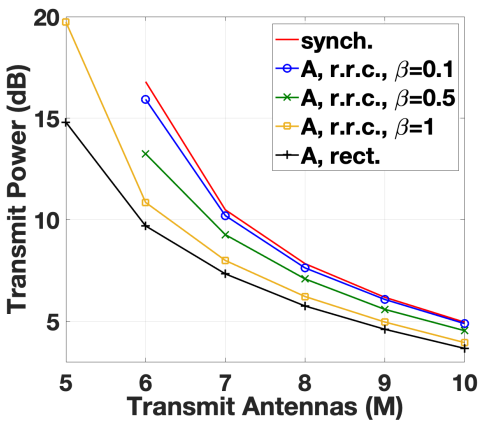
7.6.2 Numerical Results

In this section, numerical simulations are performed to verify our presented results. The Monte Carlo simulations are performed over 1000 realizations of random channel coefficients where the channel coefficients follow a Rayliegh fading model $\mathcal{CN}(0, 1)$. For each such configuration, the same rate constraint is assumed for all users and the noise variance is set to $\sigma_k^2 = 0.1$ for all users. Different configurations of transmit antennas (M), number of users (K) and rate constraints (r_k^*) are considered to show the advantages of the asyn-

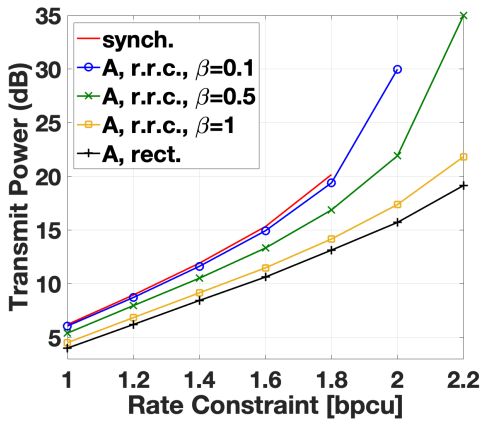
chronous methods. In Fig. 7.9a, the performance of the synchronous method is compared



(a) $M = 6$, $r_k^* = 1.5(bpcu)$



(b) $K = 10$, $r_k^* = 1.2(bpcu)$



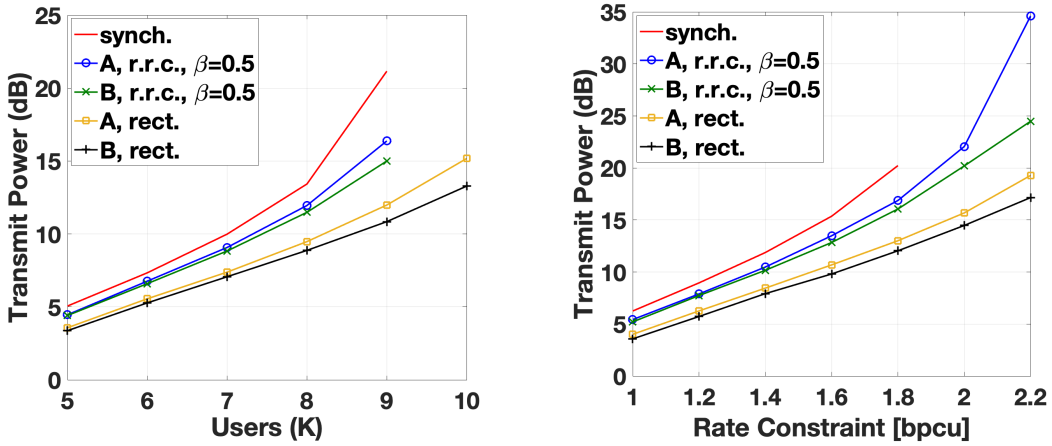
(c) $M = 6$, $K = 8$

Figure 7.9: Performance of Method A: (a) verifies that Method A needs less transmit power to serve the same number of users and is able to serve more number of users. (b) verifies that Method A needs less transmit power with a given number of transmit antennas and is able to serve the same number of users with fewer transmit antennas. (c) verifies that Method A needs less transmit power to satisfy a given rate constraint and is able to satisfy a wider range of rate constraints.

with Method A with $M = 6$ transmit antennas and the rate constraint of $r_k^* = 1.5(bpcu)$ for each user. Different pulse shapes are included to show the effect of pulse shape in the performance of beamforming methods. For fair comparison, the synchronous method is assumed to use the same pulse shape as its counterpart. As explained in Proposition 1, due to asynchrony, the choice of pulse shape has significant effect on the performance improvement of Method A. As the roll-off factor of the r.r.c. pulse shape increases, the

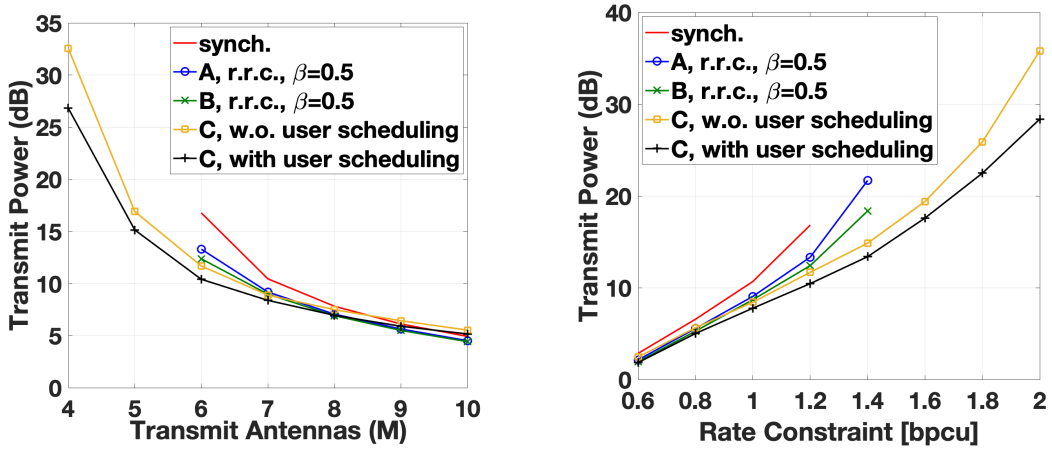
decrease in the transmit power increases. $\beta = 1$ provides the highest improvement while $\beta \to 0$ provides no reduction in power as proved in Proposition 1. Beside decreasing the average transmit power, Method A can also support 10 users which is not possible with the synchronous beamforming. In Fig. 7.9b, the comparison is performed for $K = 10$ and $r_k^* = 1.2(bpcu)$ with respect to different number of transmit antennas. By increasing the number of transmit antennas, the average transmit power decreases and the performance of both synchronous and asynchronous methods converge. With large number of antennas, the IUI can be completely removed by spatial beamforming and asynchrony loses its benefits. However, as the number of transmit antennas decreases and the system becomes overloaded, the reduction in the average transmit power, obtained by asynchrony, increases. For example, with $M = 6$ and r.r.c. pulse with $\beta = 0.5$, around $3dB$ power reduction is achieved. In addition, using Method A, the required rate constraints can be supported by 5 transmit antennas which is impossible with the synchronous beamforming. In Fig. 7.9c, the comparison is presented for $M = 6$ and $K = 8$ with respect to various rate constraints. For small rate constraints the gain provided by Method A is not noticeable, however, as the rate constraints increase the reduction in the average transmit power increases. In addition, the largest rate constraint that can be provided by the synchronous method is $r_k^* = 1.8(bpcu)$ while Method A can support up to $r_k^* = 2.2(bpcu)$.

In Fig. 7.10, the performance of Method B using the 2-step algorithm proposed in Section 7.4 with SCALE DSM algorithm [142] and $N = 100$ is shown. In Fig. 7.10a, $M = 6$, $r_k^* = 1.5[bpcu]$ and different number of users (K) are considered. As the number of users increases, the reduction in the average transmit power increases. For example, for $K = 9$, around $1dB$ power reduction is achieved. In Fig. 7.10b, considering $M = 6$ and $K = 8$, the power reduction of around $10dB$ and $2dB$ is achieved at $r_k^* = 2.2$ compared with Method A by r.r.c. ($\beta = 0.5$) and rect. pulses, respectively.



(a) $M = 6$, $r_k^* = 1.5(bpcu)$ (b) $M = 6$, $K = 8$

Figure 7.10: Performance of Method B: Method B improves the performance of the conventional synchronous method and Method A. (a) verifies that Method B reduces the transmit power to serve a given number of users. (b) verifies that Method B reduces the transmit power to satisfy a given rate constraint.



(a) $K = 10$, $r_k^* = 1.2(bpcu)$ (b) $M = 6$, $K = 10$

Figure 7.11: Performance of Method C: Method C improves the performance of the conventional synchronous method and Methods A and B (a) verifies that Method C is able to serve the same number of users with fewer transmit antennas. (b) verifies that Method C is able to satisfy a wider range of rate constraints.

In Fig. 7.11, the performance of Method C using the proposed algorithm in Section 7.5 with $N_{rand} = 300$, $q = 2$, and with/without user scheduling is shown. In Fig. 7.11a, $K = 10$, $r_k^* = 1.2(bpcu)$ and different number of transmit antennas (M) are considered. As the number of transmit antennas increases, the spatial domain becomes sufficient

to cancel the IUI and Method C is not helpful. However, with low number of transmit antennas, the spatial domain is unable to effectively cancel the IUI and user-grouping and the precoding/oversampling technique greatly improves the performance. Furthermore, by using Method C, the required rate constraints can be satisfied by $M = 4$ transmit antennas. In Fig. 7.11b, with $M = 6$, $K = 10$, the power reduction of around $6dB$ and $4dB$ are achieved at $r_k^* = 1.4(bpcu)$ compared with Methods A and B, respectively. In addition, for the chosen parameters, Method C can support up to $r_k^* = 2(bpcu)$ while the synchronous method and Methods A/B can support up to $r_k^* = 1.2(bpcu)$ and $r_k^* = 1.4(bpcu)$, respectively. Note that the heuristic user scheduling method further reduces the average transmit power. While we have presented the results for a few sets of parameters, we have done extensive simulations with other choices of parameters and a similar trend has been observed.

7.6.3 Frequency-Selective Channel Analysis

The main advantage of intentionally adding time delays is to transform flat-fading channels into frequency-selective channels and exploit the frequency-selectivity to improve the performance. However, in some applications, the channel is inherently frequency-selective which might impact the effectiveness of adding time delays at the transmitter. To analyze the effect of timing offsets in such scenarios, consider a frequency-selective environment where the channels follow a J-tap multipath channel. Then, the received signal at User k can be denoted by $y_k(t) = \sum_{j=1}^{J} \sum_{l=1}^{K} \boldsymbol{h}_{\boldsymbol{k,j}}^H \boldsymbol{w_l} s_l(t - \tau_l T - \tau_{k,j}) + n_k(t)$ where $\boldsymbol{h_{k,j}} \in \mathbb{C}^{M\times 1}$ denotes the jth channel tap vector for User k and $\tau_{k,j}$ is the jth delay tap to User k. Applying proper post-processing at the users, as explained in Section II, the received samples denoted by $\boldsymbol{y_k} = (y_k[1], \cdots, y_k[N])^T$, can be written as: $\boldsymbol{y_k} = \sum_{l=1}^{K} \boldsymbol{G_{kl}^{fs}} \boldsymbol{s_l} + \boldsymbol{n_k}$, where $\boldsymbol{G_{kl}^{fs}}$ denotes the effective frequency-selective channel ma-

trix calculated as $\boldsymbol{G_{kl}^{fs}} = \sum_{j=1}^{J} \boldsymbol{h_{k,j}^H w_l G_{kl,j}}$ and the offset matrix $\boldsymbol{G_{kl,j}}$ is defined similarly as:

$$[\boldsymbol{G_{kl,j}}]_{m,n} = g(\tau_l T + \tau_{k,j} + (m-n)T) \triangleq g_{\tau_{kl,j}}[m-n], \quad m,n = 1, \cdots, N. \quad (7.21)$$

By considering the above effective frequency-selective channel matrix, all proposed methods, including Methods A, B and C can be applied to frequency-selective channels. Due to limited space, the corresponding analysis is not included. For completeness, however, the effective SINR metric is used to show the general impact of time delays in frequency-selective channels. The effective SINR at User k can be calculated as:

$$\text{SINR}_k \approx \frac{\sum_{j=1}^{J} |\boldsymbol{h_{k,j}^H w_k}|^2 g^2_{\tau_{kk,j}}(0)}{\sum_{j=1}^{J} |\boldsymbol{h_{k,j}^H w_k}|^2 \sum_{\substack{n=-u \\ n\neq 0}}^{n=u} g^2_{\tau_{kk,j}}(n) + \sum_{\substack{l=1 \\ l\neq k}}^{K} \sum_{j=1}^{L} |\boldsymbol{h_{k,j}^H w_k}|^2 \sum_{n=-u}^{n=u} g^2_{\tau_{kl,j}}(n) + \sigma_n^2}, \quad (7.22)$$

where the first, second and third components in the denominator represent, ISI, IUI and noise, respectively. Because the channel is inherently frequency-selective, the effect of intentionally adding time delays , i.e., $\tau_l \neq 0, \forall l$ is not significant. Assuming a two-user scenario with r.r.c. pulse shape (β=0.5), MRT beamforming with respect to the dominant channel path, transmit power of 10 dB and $\sigma_n^2 = 1$ in a 5-tap channel model, the change of the averaged SINR for asynchronous transmission compared with synchronous transmission is shown in Fig. 7.12.

The average SINR over 10,000 realizations of the channel is considered. For frequency-selective channels, the average power ratio of the scattering paths and the dominant path is assumed to be $0, -3$ and -9 dB. For high scattering environments, the effect of asynchronous transmission is negligible. However, as the power of the dominant path compared with the scattering paths is increased, the effect of asynchronous transmission

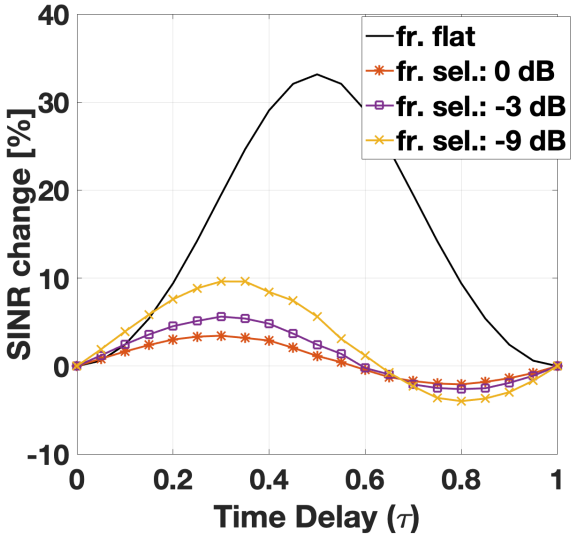


Figure 7.12: Effect of asynchronous transmission in frequency-selective channels

is intensified. In addition, due to multi-path effect, adding time delays can decrease the performance in some scenarios, unlike the flat fading case, where adding time delays always improve the effective SINR. In summary, in an inherently frequency-selective channel, intentionally adding time delays is less effective. As an alternative, frequency offsets can be intentionally added to induce time-selectivity to improve the performance [129].

7.7 Conclusion

In this chapter, we investigated the benefits of adding intentional time delays in the performance of downlink transmit beamforming. We compared our proposed methods with the conventional synchronous transmit beamforming method and showed that our proposed methods improve the performance by decreasing the average transmit power. Method A, which uses no time-domain precoding, exploits the reduction in the co-channel interference power by proper choice of time delays. In Method B, the induced frequency-selective channels are decomposed by DFT, and a further reduction in the average transmit power is realized by applying power spectrum management algorithms. Method C employs joint

time-domain precoding and an oversampling technique to obtain a set of sufficient statistics, which results in a system model with higher dimensional full-rank matrices. The additionally available rank helps to cancel the intra-group co-channel interference, while the spatial beamforming reduces the inter-group IUI. In a nutshell, adding intentional timing offsets provide powerful tools to manage the co-channel interference in downlink beamforming scenarios.

Chapter 8

Rate-region Analysis of the Asynchronous Transmission for Multiple Access Channels

In this chapter, we thoroughly analyze the rate-region provided by the asynchronous transmission in multiple access channels (MACs). We derive the theoretical capacity regions which apply to a wide range of pulse shaping methods. We analytically prove that asynchronous transmission enlarges the capacity region of MACs. We show that although successive interference cancellation (SIC) is sum-rate achieving for the conventional uplink Non-orthogonal Multiple Access (NOMA) methods, it is unable to achieve the boundary of the capacity region for the asynchronous transmission. We also demonstrate that for the asynchronous transmission, the optimal SIC decoding order to achieve the maximum sum-rate is based on the channel strengths of the users. This phenomenon is in contrast to the conventional uplink NOMA, where the order of decoding does not affect the sum-rate.

8.1 Introduction

With the increase of mobile users and the higher data demand, the use of Non-orthogonal methods in wireless networks is inevitable. Notably, in the uplink, where multiple users attempt to connect the base station (BS), orthogonal multiple access (OMA) incurs inefficiency in resource utilization and requires more overhead signals. The capacity region of the MAC channel is derived and expressed in [149, 150], which is conventionally named the Cover-Wyner pentagon. However, recently, the system design and performance analysis of MAC channels have received new attention from academia and industry under the category of uplink Non-orthogonal-multiple-access (NOMA). For example, in the power-domain NOMA scheme, the signals for multiple users are superposed at different power levels using superposition coding, and a multiuser detection method, such as successive interference cancellation (SIC), is employed at the receiver [151]. The advantages of the NOMA over the OMA have been extensively studied in [152] and the references therein, e.g., providing higher system throughput compared with OMA and supporting massive connectivity.

In this chapter, the results of [83] on the capacity of asynchronous MACs are generalized to a wide range of pulse shapes, including the prevailing r.c. pulse shape. We further analyze and compare the SIC detection method for the conventional synchronous transmission and the asynchronous transmission. We prove that while SIC is optimal for synchronous transmission in the capacity-achieving sense, it provides a sub-optimal rate region for the asynchronous transmission. We analytically show the optimal decoding order of SIC detection for the asynchronous transmission.

The rest of this chapter is organized as follows: In Section 8.2, the general system model, characteristics, and input-output relation is outlined. In Section 8.3, the capacity region is analytically derived for general types of pulse shapes, and the superiority of asyn-

chronous transmission is proved. Finally, in Section 8.4, the main points of this chapter are summarized.

8.2 System Model

In many practical situations, it is reasonable to assume that frame synchronism is achievable with a means of channel feedback or cooperation among transmitters. However, symbol anachronism is difficult to achieve due to the much smaller time scale involved as explained in previous chapters. For a channel with K single-antenna users and one common receiver, assuming frame synchronism, we can write the channel output as

$$y(t) = \sum_{k=1}^{K} h_k \sum_{n=1}^{N} s_k[n] p(t - nT - \tau_k T) + n(t), \tag{8.1}$$

User k transmits a codeword $(s_k[1], \cdots, s_k[N]) \in \mathbb{S}_k^N$ where $\mathbb{S}_k$ represents the input domain, $p(t)$ is the normalized pulse shaping function, i.e., $\int |p(t)|^2 dt = 1$, with the support of T_p, T denotes the symbol interval and $n(t)$ is white Gaussian noise with power spectral density equal to σ_n^2. The $\tau_k \in [0, 1)$ accounts for the timing offset of User k, and h_k represents the kth user channel coefficient and are known to the receiver to decode reliably each of the transmitted messages. For capacity analysis, the symbols are assumed to be chosen from Gaussian processes whose PSD are found optimally. It is assumed that the pulse shape occupies frequency bandwidth of B. For example, for a r.r.c. pulse shape with roll-off factor of β, the occupied bandwidth is $B = \frac{1+\beta}{T}$. T is assumed to be normalized to 1 if not specified.

If the signals transmitted by the users are not aligned at the receiver, then the channel is symbol asynchronous. For completeness of the material in this chapter, we include the

system model derivation which is indeed similar to the one in the previous chapters. The received signal is applied to a matched filter with the impulse response $p(t)$ and its output is samples at time instants $t_k^m = mT + \tau_k T$, $m = 1, 2, \cdots, N$, $k = 1, \cdots, K$ which results in the samples of $y_k[m]$. One can arrange the samples and define $\boldsymbol{y_k} = (y_k[1], \cdots, y_k[N])^T$. Thus, we can have the input output relation of $\boldsymbol{y_k} = \sum_{l=1}^{K} h_l \boldsymbol{R_{kl}} \boldsymbol{s_l} + \boldsymbol{n_k}$. Matrix $\boldsymbol{R_{kl}}$ is an $N \times N$ Toeplitz matrix whose elements depend on the pulse shape, the corresponding time delay and are denoted as:

$$\begin{aligned} [\boldsymbol{R_{kl}}]_{m,n} =& g(\tau_{kl} T + (m-n)T) \triangleq g_{\tau_{kl}}[m-n], \\ & m, n = 1, \cdots, N \end{aligned} \tag{8.2}$$

where $\tau_{kl} = \tau_k - \tau_l$ and $g(t) = p(t) * p(t)$. The vector $\boldsymbol{n_k}$ represents the noise vector whose co-variance matrix is defined as $\boldsymbol{Q_{kl}} = \mathbb{E}[\boldsymbol{n_k} \boldsymbol{n_l}^H] = \sigma_n^2 \boldsymbol{R_{kl}}$. For any square-root Nyquist pulse, e.g., r.r.c., $\boldsymbol{R_{kk}} = \boldsymbol{I_N}$ and $\boldsymbol{R_{kl}^T} = \boldsymbol{R_{lk}}$. Also note that for the synchronous transmission, i.e., $\tau_k = 0 \forall k$, $\boldsymbol{R_{lk}} = \boldsymbol{I_N}$. We can put all the received samples together and define $\boldsymbol{y} = \left(\boldsymbol{y_1}^T, \cdots, \boldsymbol{y_K}^T\right)^T$ to get the system model in a matrix form as:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{R_{11}} & \cdots & \boldsymbol{R_{1K}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R_{K1}} & \cdots & \boldsymbol{R_{KK}} \end{pmatrix} \begin{pmatrix} h_1 \boldsymbol{I_N} & \cdots & \boldsymbol{0_N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0_N} & \cdots & h_K \boldsymbol{I_N} \end{pmatrix} \begin{pmatrix} \boldsymbol{s_1} \\ \vdots \\ \boldsymbol{s_K} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_1} \\ \vdots \\ \boldsymbol{n_K} \end{pmatrix} = \boldsymbol{R H s} + \boldsymbol{n} \tag{8.3}$$

where $\boldsymbol{R}$ is the $NK \times NK$ matrix which is called the "*offset matrix*". Note that $\boldsymbol{R}$ is constructed by Toeplitz blocks of $\boldsymbol{R_{lk}}$.

8.3 Capacity Region Analysis

To analyze the capacity region of the resulting asynchronous system model, we assume two users while generalization to more than two users is straightforward. Let us define $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ as stationary Gaussian processes with power spectral density (PSD) of $\{S_1(f), S_2(f)\ f \in [0,1]\}$ for User 1 and User 2, respectively.

Then, the capacity region of the asynchronous MAC channel can be denoted as:

$$C = \bigcup_{\substack{S_k(f)\geq 0, f\in[0,1] \\ \int_0^1 S_k(f)df \leq P_k \\ k=1,2}} \left\{ (R_1, R_2), \begin{array}{r} 0 \leq R_1 \leq \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{y}; \boldsymbol{s_1}|\boldsymbol{s_2}) \\ 0 \leq R_2 \leq \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{y}; \boldsymbol{s_2}|\boldsymbol{s_1}) \\ 0 \leq R_1 + R_2 \leq \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{y}; \boldsymbol{s_1}, \boldsymbol{s_2}) \end{array} \right\} \tag{8.4}$$

where R_k and P_k represents the achievable rate and the available power of User k, respectively, and $\boldsymbol{y}$ represents the $2N$ received samples obtained as explained in Eq. 8.3. In the next theorem, the capacity region is derived:

Theorem 8.1. *For a two user MAC channel, the capacity region can be denoted as:*

$$C = \bigcup_{\substack{S_k(f)\geq 0, k=1,2 \\ \int_0^1 S_k(f)df \leq P_k}} \left\{ (R_1, R_2), \begin{array}{r} 0 \leq R_1 \leq \frac{1}{2}\int_0^1 \log_2\left(1 + \frac{S_1(f)}{\sigma_1^2}\right) \\ 0 \leq R_2 \leq \frac{1}{2}\int_0^1 \log_2\left(1 + \frac{S_2(f)}{\sigma_2^2}\right) \\ 0 \leq R_1 + R_2 \leq \frac{1}{2}\int_0^1 \log\left(1 + \frac{S_1(f)}{\sigma_1^2} + \frac{S_2(f)}{\sigma_2^2} + \frac{S_1(f)S_2(f)(1-G_\tau^2(f))}{\sigma_1^2\sigma_2^2}\right) df \end{array} \right\} \tag{8.5}$$

where $\sigma_1^2 = \frac{\sigma_n^2}{|h_1|^2}$, $\sigma_2^2 = \frac{\sigma_n^2}{|h_2|^2}$ *and* $G_\tau(f)$ *depends on the pulse shape and the timing offset and is defined as*

$$G_\tau(f) = \left| \frac{1}{T} \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}\left(\frac{f+i}{T}\right) \right|, \tag{8.6}$$

where $\hat{g}(f)$ *is the Fourier transform of the matched filter pulse* $g(t)$.

Proof. The proof is presented in Appendix E.1. □

Note that with no timing offset, $G_\tau(f)$ becomes the conventional folded spectrum which is constant and is equal to 1 for Normalized Nyquist no-ISI satisfying pulse shapes. Then, the capacity region of the MAC channel turns into the conventional Cover-Wyner region:

$$C_{synch} = \left\{ \begin{array}{c} \\ (R_1, R_2), \\ \\ \end{array} \begin{array}{c} 0 \leq R_1 \leq \frac{1}{2}\log_2\left(1+\frac{P_1}{\sigma_1^2}\right) \\ 0 \leq R_2 \leq \frac{1}{2}\log_2\left(1+\frac{P_2}{\sigma_2^2}\right) \\ 0 \leq R_1+R_2 \leq \frac{1}{2}\log\left(1+\frac{P_1}{\sigma_1^2}+\frac{P_2}{\sigma_2^2}\right) df \end{array} \right\} \tag{8.7}$$

We present our first conclusion in the next proposition. Although this is a straightforward outcome of the derived capacity region, it is of great importance from practical and theoretical perspective.

Proposition 8.1. *In a 2-users MAC channel, asynchronous transmission enlarges the capacity region compared with the synchronous transmission.*

Proof. To prove the proposition, it is enough to prove that for the asynchronous transmission the function $G_\tau(f)$ is less than or equal to that of the synchronous transmission in every frequency. This can be easily proved by using the fact that $\hat{g}(f)$ is a positive real-valued function which is satisfied because of matched filtering at the receiver side. Thus,

$$\begin{aligned} G_{asynch}(f) = \left|\frac{1}{T}\sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)}\hat{g}\left(\frac{f+i}{T}\right)\right| &\leq \frac{1}{T}\sum_{i=-\infty}^{\infty}\left|e^{-j2\pi\tau(f+i)}\right|\left|\hat{g}\left(\frac{f+i}{T}\right)\right|, \\ &= \frac{1}{T}\sum_{i=-\infty}^{\infty}\hat{g}\left(\frac{f+i}{T}\right) = G_{synch}(f) \end{aligned}$$

□

The capacity region enlargement by the asynchronous transmission depends on the pulse shape and the timing offset which is explained in more details next.

Example 8.1. *In this example, the well-known and practically common pulse shape of r.r.c is considered. After matched filtering, the effective pulse shape is raised cosine whose frequency spectrum is denoted as:*

$$\hat{g}(f) = \begin{cases} T & |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2}\left[1+\cos\left(\frac{\pi T}{\beta}\left(|f|-\frac{1-\beta}{2T}\right)\right)\right] & \frac{1-\beta}{2T} < |f| \leq \frac{1+\beta}{2T} \\ 0 & o.w. \end{cases}. \tag{8.8}$$

where β is the roll-off factor. The phase-shifted folded spectrum, $G_\tau(f) = \left|\frac{1}{T}\sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)}\hat{g}\left(\frac{f+i}{T}\right)\right|$ *is periodic with period of 1 and based on the definition of $\hat{g}(f)$, can be derived as:*

$$G_\tau(f) = \begin{cases} e^{-j2\pi\tau f}A(-f) + e^{-j2\pi\tau(f+1)}A(f+1) & \frac{-1}{2} < f \leq \frac{-1+\beta}{2} \\ e^{-j2\pi\tau f} & |f| \leq \frac{1-\beta}{2} \\ e^{-j2\pi\tau f}A(f) + e^{-j2\pi\tau(f-1)}A(-f+1) & \frac{1-\beta}{2} < f \leq \frac{1}{2} \end{cases}, \tag{8.9}$$

where $A(f) = \frac{1}{2}\left[1+\cos\left(\frac{\pi T}{\beta}\left(\frac{f}{T}-\frac{1-\beta}{2T}\right)\right)\right]$. *To illustrate the effect of roll-off factor and timing offset, $G_\tau^2(f)$ is shown with various parameters in Fig. 8.1.*

The next step to specify the capacity region is to find the optimum power allocation. However, it is worth describing the achievable rate region with the constant power PSDs. Assuming constant power for each user, i.e., $S_k(f) = P_k$, the achievable rate region can

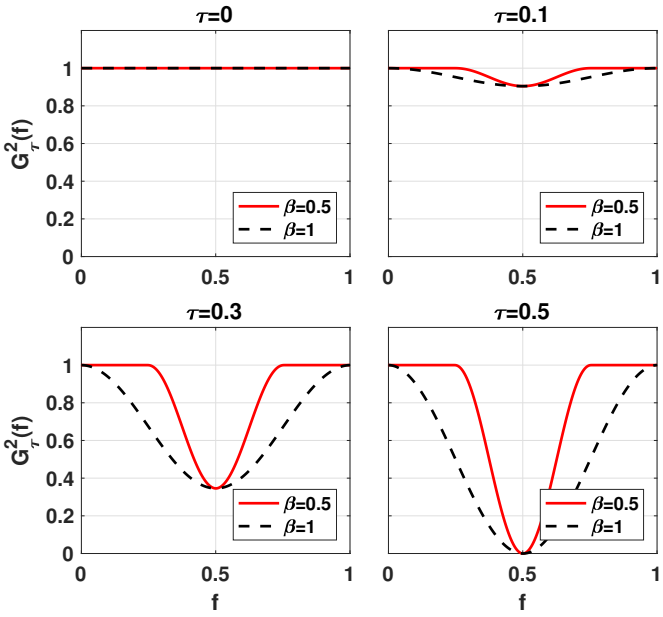


Figure 8.1: Demonstration of $G_\tau^2(f)$ for r.r.c. pulse with $\beta = 0.5, 1$.

be denoted as:

$$C = \bigcup \left\{ (R_1, R_2), \begin{array}{c} 0 \le R_1 \le \frac{1}{2}\log_2\left(1+\frac{P_1}{\sigma_1^2}\right) \\ 0 \le R_2 \le \frac{1}{2}\log_2\left(1+\frac{P_2}{\sigma_2^2}\right) \\ 0 \le R_1+R_2 \le \frac{1}{2}\log\left(1+\frac{P_1}{\sigma_1^2}+\frac{P_2}{\sigma_2^2}+\frac{P_1P_2(1-G_\tau^2(f))}{\sigma_1^2\sigma_2^2}\right) df \end{array} \right\} \tag{8.10}$$

Using Proposition 1, it is obvious that even with constant PSDs, the asynchronous transmission enlarges the capacity region, however, with optimizing the inputs PSDs, even further enlargement is possible. Following the steps in [83], the optimization problem to find the optimal PSD can be formulated as:

$$\begin{aligned} &\arg\max_{S_1(f),S_2(f)} \max_{R_1,R_2} \alpha R_1 + \alpha R_2 \\ &s.t. \;\; R_k \in C \; in \; (8.5), \\ &\int_0^1 S_k(f)df = P_k \end{aligned} \tag{8.11}$$

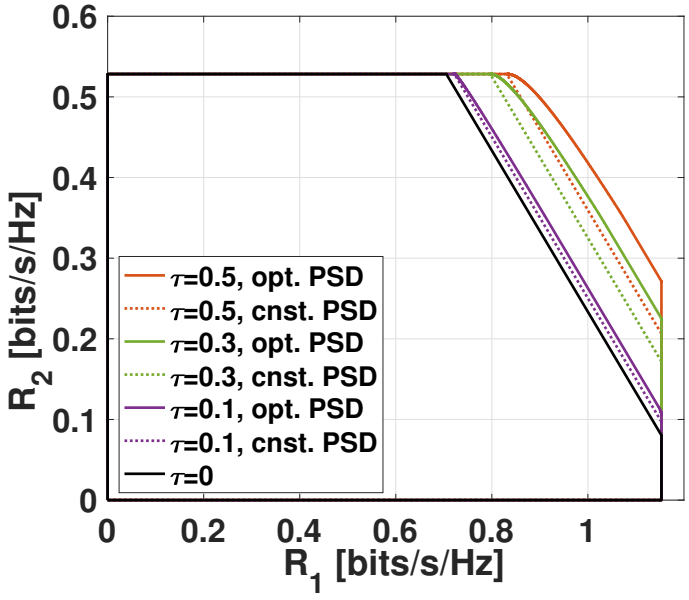
for every $0 \leq \alpha \leq 1$. The inner maximization which describes the pareto points of the region can be simplified as:

$$\begin{aligned} &\arg\max_{S_1(f),S_2(f)} \max \{\alpha F(S_1(f),0) + (1-\alpha)[F(S_1(f),S_2(f)) - F(S_1(f),0)] \,, \\ &\qquad \alpha[F(S_1(f),S_2(f)) - F(0,S_2(f))] + (1-\alpha)F(0,S_2(f))\} \qquad (8.12) \\ &= \arg\max_{S_1(f),S_2(f)} \begin{cases} (2\alpha-1)F(S_1(f),0) + (1-\alpha)F(S_1(f),S_2(f)) & 1/2 \leq \alpha \leq 1 \\ (1-2\alpha)F(S_1(f),0) + \alpha F(S_1(f),S_2(f)) & 0 \leq \alpha \leq 1/2 \end{cases} \qquad (8.13) \end{aligned}$$

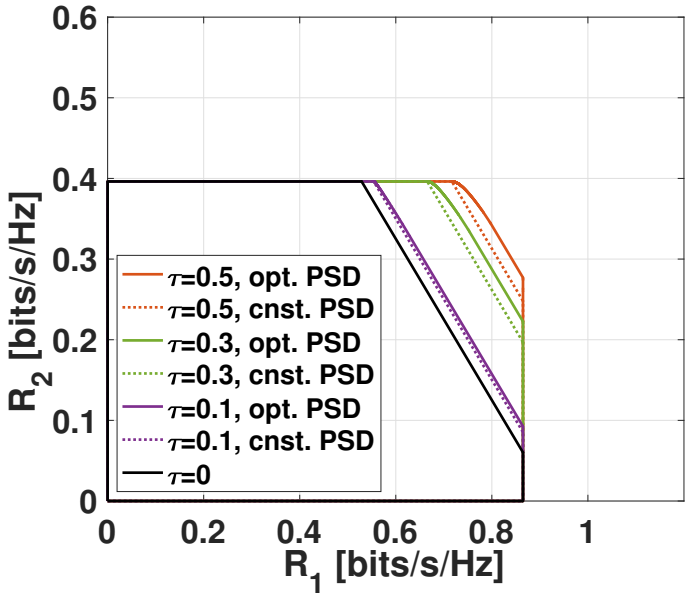
where $F(S_1(f),S_2(f)) = \frac{1}{2}\int_0^1 \log\left(1 + \frac{S_1(f)}{\sigma_1^2} + \frac{S_2(f)}{\sigma_2^2} + \frac{S_1(f)S_2(f)(1-|G_\tau(f)|^2)}{\sigma_1^2\sigma_2^2}\right) df$. Unlike the synchronous case, where $S_k(f) = P_k$ can achieve all the boundary points, in the asynchronous case, not a single optimal PSD pair can achieve all the boundary points. After solving the nonlinear optimization problem for various values of α, the result is shown in Fig. 8.2.

It is shown that the asynchronous transmission improves the capacity region compared with the synchronous transmission. In addition, $\tau = 0.5$ provides the largest capacity improvement. Although constant PSDs are optimal for synchronous transmission, constant PSDs can not achieve the capacity for asynchronous transmission. However, even with constant PSDs, the asynchronous transmission provides a significant improvement in achievable rates compared with the synchronous transmission.

Note that by increasing the roll-off factor, β, the gain provided by asynchronous transmission is increased. For example, with constant PSDs, $P_1 = P_2 = 10, \sigma_1^2 = 1, \sigma_2^2 = 5, \tau = 0.5$, and using r.r.c. pulse shape with $\beta = 0.5$ and $\beta = 1$, asynchronous transmission provides 10% (R_{sum} : 1.23 $\rightarrow$ 1.36 $[bits/s/Hz]$) and 20% (R_{sum} : 0.92 $\rightarrow$ 1.11 $[bits/s/Hz]$) improvement in the sum-rate, respectively, compared with the synchronous transmission. Optimizing PSDs enables additional $3-4\%$ improvement. This is supported by the



(a) $\beta = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 5$



(b) $\beta = 1$, $\sigma_1^2 = 1, \sigma_2^2 = 5$

Figure 8.2: Capacity region of the asynchronous MAC channel for two users with $P_1 = P_2 = 10$ and r.r.c. pulse shape.

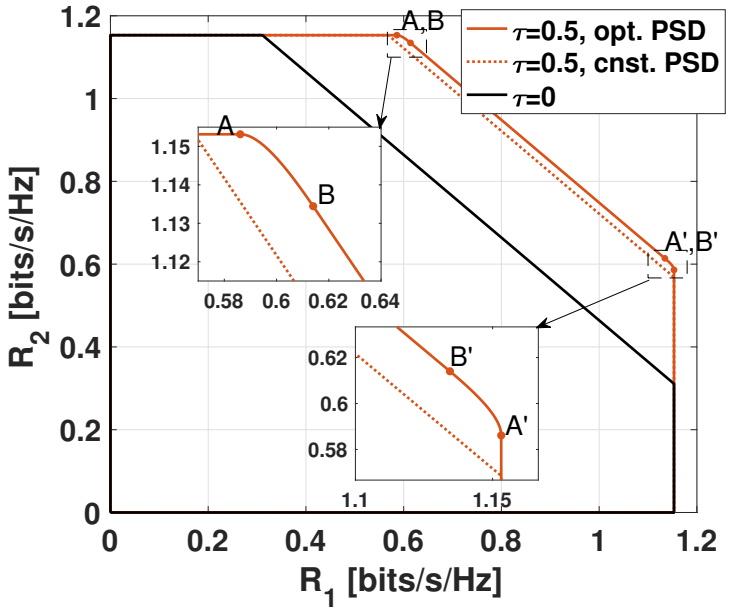
derivation of $G_\tau(f)$ function, as $\beta = 1$ results in more reduction in $G_\tau(f)$ function, shown in Fig. 8.1. The underlying reason is because of the out-of-Nyquist-band portion of r.c.

spectrum which increases as the roll-off factor increases. Increased out-of-Nyquist-band portion causes more destructive addition in the phase-shifted folded spectrum which results in smaller values for $G_\tau(f)$. Note that with no out-of-Nyquist-band spectrum, i.e., $\beta = 0$, we have $G_\tau(f) = 1, f \in [0, 1]$, and thus, the capacity region is reduced to the conventional Wyner-Cover pentagon. In summary, asynchronous transmission exploits the out-of-Nyquist-band spectrum, and is of no use without it.

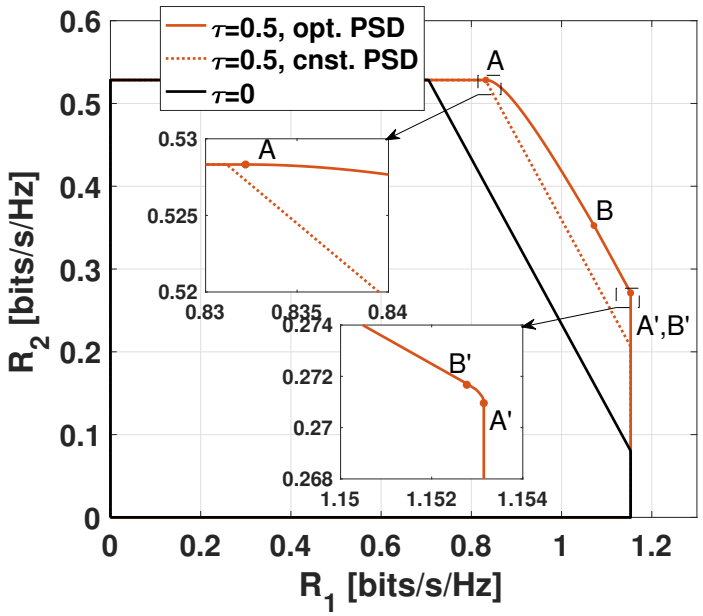
In the capacity region, 4 points are of great importance denoted as A, A', B, B' in Fig. 8.3. Points A and A' are obtained by maximizing the sum-rate upper-bound over one user's PSD while the other user's PSD is assumed to be constant. By formulating the optimization problem, considering the KKT conditions and simple simplifications, we have:

$$S_k^*(f) = \left[\lambda - \frac{\sigma_{\bar{k}}^2 + P_{\bar{k}}}{\frac{\sigma_{\bar{k}}^2}{\sigma_k^2} + \frac{P_{\bar{k}}}{\sigma_k^2}(1 - G_\tau^2(f))}\right]^+, \ k = 1, 2 \tag{8.14}$$

where $\bar{k} = \{1, 2\} - k$ and $[x]^+ = \max\{0, x\}$. It shows that the optimal PSD for points A and A' is not constant and depends on $G_\tau(f)$ as also stated in [83]. In addition, it depends on the ratio of channel strengths, i.e., $\frac{\sigma_{\bar{k}}^2}{\sigma_k^2}$. Roughly speaking, if one of the users has a much stronger channel strength, the dependence of stronger user's PSD on $G_\tau(f)$ reduces and it approaches a constant PSD. On the other hand, the weaker user's PSD greatly depends on $G_\tau(f)$, and no power is assigned to frequencies that $G_\tau(f) = 1$. These observations are illustrated in Fig. 8.4, where the optimized PSD for the stronger user in Fig. 8.4b (row A) is almost constant with insignificant variations while the weaker user selectively assigns its available power depending on $G_\tau(f)$. This fact is the underlying reason for proximity of point A to the constant PSD corner and the large gap between A' and the constant corner in Fig. 8.3b.



(a) $\beta = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 1$



(b) $\beta = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 5$

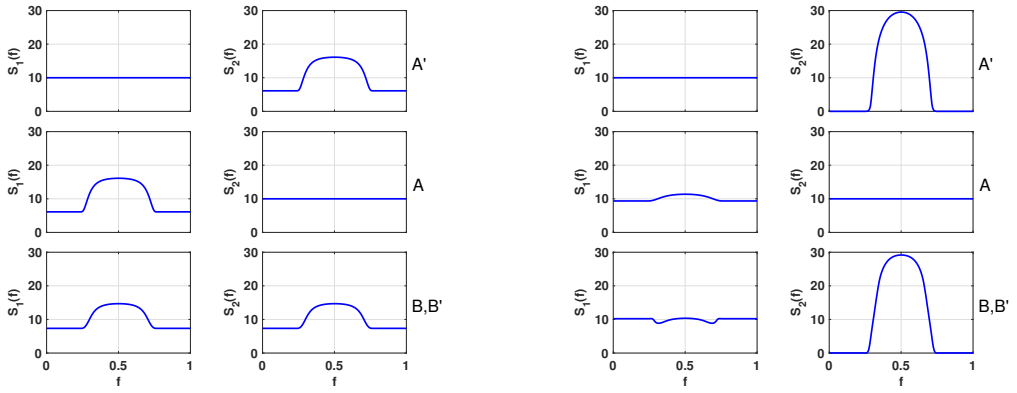
Figure 8.3: Capacity region analysis of MAC channel with $P_1 = P_2 = 10$ and r.c. pulse shape while the channels have equal or unequal strength

Points B and B' are obtained by maximizing the sum-rate upper-bound over both users'

PSDs. Similarly, the optimized PSDs can be obtained as:

$$S_k^*(f) = \left[\lambda - \frac{\sigma_{\bar{k}}^2 + S_{\bar{k}}^*(f)}{\frac{\sigma_{\bar{k}}^2}{\sigma_k^2} + \frac{S_{\bar{k}}^*(f)}{\sigma_k^2}(1 - G_\tau^2(f))}\right]^+, \quad k = 1, 2 \tag{8.15}$$

Again, if the difference between channel strengths is large, then, the stronger user's optimal PSD approaches a constant one as shown in Fig. 8.4b (row B, B'). This is also the reason that the point B' is very close to A' in Fig. 8.3b. For point B', the stronger user achieves the rate $R_1 = \int_0^1 \log 2 \left(1 + \frac{S_1^*(f)}{\sigma_1^2}\right) df$ where $S_1^*(f)$ is derived in (8.15) and is almost a constant spectrum (also shown in Fig. 8.4b). The weaker user utilizes the residual available rate from the optimized sum-rate which is $R_2 = R_{sum} - R_1$. On the other hand, for point B, weaker user achieves the rate $R_2 = \int_0^1 \log 2 \left(1 + \frac{S_2^*(f)}{\sigma_2^2}\right) df$ where $S_2^*(f)$ is derived in (8.15) (also shown in Fig. 8.4b), which is far less than the rate achieved by the constant PSD. In exchange, the sum-rate is maximized and the stronger user achieves $R_1 = R_{sum} - R_2$. In a nutshell, with asynchronous transmission, in a near-far scenario, the power optimization is more critical for the weaker user as the stronger user's optimal PSD approaches a constant PSD.



(a) $\beta = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 1$ (b) $\beta = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 5$

Figure 8.4

For a synchronous scenario, the pentagon's corner point can be achieved by the SIC

method, and the rest of Pareto points on the connecting line are achieved by time-sharing. This optimality is the reason that SIC is widely used in NOMA uplink literature. Due to the SIC method's importance, the SIC achievable rates are analyzed in the next section, and few critical comparisons are drawn between the synchronous and asynchronous transmissions.

8.3.1 SIC achievable rate region

In the SIC method, the information symbols of one user are decoded while the other user's signal is considered as noise. After the decoded symbols are removed, the symbols of the other user are decoded with reduced interference. With Gaussian signaling, the achievable rate tuples for the synchronous transmission can be denoted as:

$$R_1 = \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{P_2|h_2|^2+\sigma_n^2}\right), R_2 = \frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}\right) \tag{8.16}$$

$$R_2 = \frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{P_1|h_1|^2+\sigma_n^2}\right), R_1 = \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\sigma_n^2}\right) \tag{8.17}$$

which are achieved by the decoding order of $\{1,2\}$ and $\{2,1\}$, respectively. The mentioned rate tuples coincide with the corner points of the well-known pentagon capacity region and the sum-rate is calculated as:

$$R_{sum} = \frac{1}{2}\log_2\left(1+\frac{P_j|h_j|^2}{P_k|h_k|^2+\sigma_n^2}\right)+\frac{1}{2}\log_2\left(1+\frac{P_k|h_k|^2}{\sigma_n^2}\right) \ (k,j)=\{(1,2),(2,1)\} \tag{8.18}$$

$$= \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\sigma_n^2}+\frac{P_2|h_2|^2}{\sigma_n^2}\right) \tag{8.19}$$

which is equal to the maximum sum-rate in capacity region Eq. (8.7). Recall the asynchronous system model in Eq. (8.3), and rewrite the received samples as $y_k[m] =$

$h_k s_k[m] + h_{\bar{k}} \sum_{n=1}^{N} g_\tau[m-n] s_{\bar{k}}[m-n] + n_k[m], k = 1, 2$. Note that due to the time delay, the IUI is caused by multiple interfering symbols rather than the one in the synchronous case. Assuming Gaussian signaling, the achievable rate tuples for the asynchronous transmission is denoted as:

$$R_1 = \frac{1}{2}\log_2\left(1 + \frac{P_1|h_1|^2}{\eta_\tau P_2|h_2|^2 + \sigma_n^2}\right), R_2 = \frac{1}{2}\log_2\left(1 + \frac{P_2|h_2|^2}{\sigma_n^2}\right) \tag{8.20}$$

$$R_1 = \frac{1}{2}\log_2\left(1 + \frac{P_2|h_2|^2}{\eta_\tau P_1|h_1|^2 + \sigma_n^2}\right), R_2 = \frac{1}{2}\log_2\left(1 + \frac{P_1|h_1|^2}{\sigma_n^2}\right) \tag{8.21}$$

where $\eta_\tau = \sum_{n=-\infty}^{n=\infty} g_\tau^2(n)$ and the corresponding decoding orders are $\{1, 2\}$ and $\{2, 1\}$, respectively. Equivalently, η_τ can be denoted as $\eta_\tau = \int_{-1/2}^{1/2} G_\tau^2(f) df$ and thus, $\eta_\tau \leq 1$ with equality achieved by $\tau = 0$, as shown in Chapters 5. Particularly, for the r.r.c. pulse shape $\eta_\tau = 1 - \beta/4 + \beta/4\cos(2\pi\tau)$, which can be minimized by $\tau = 0.5$ and has been derived in Chapter 7. Therefore, for every set of power and channel coefficients, the asynchronous transmission enjoys less inter-user interference and greater achievable rates. Unlike the synchronous transmission where the decoding order has no effect on the sum rate, decoding the stronger user first results in a larger sum-rate for the asynchronous transmission. In more details, assuming $|h_1|^2 > |h_2|^2$, and $P_1, P_2 \neq 0$, results in $\frac{1}{2}\log_2\left(1 + \frac{P_1|h_1|^2}{\eta_\tau P_2|h_2|^2 + \sigma_n^2}\right) + \frac{1}{2}\log_2\left(1 + \frac{P_2|h_2|^2}{\sigma_n^2}\right) > \frac{1}{2}\log_2\left(1 + \frac{P_2|h_2|^2}{\eta_\tau P_1|h_1|^2 + \sigma_n^2}\right) + \frac{1}{2}\log_2\left(1 + \frac{P_1|h_1|^2}{\sigma_n^2}\right)$. Summary of our results are listed below:

Proposition 8.2. • *The SIC rate tuples achieved by asynchronous transmission are greater than that of the synchronous transmission.*

Proof. The proof is immediately obtained by comparing the rate tuples expressed previously and considering the fact that $\eta_\tau \leq 1$. □

• *For synchronous transmission, the order of decoding in the SIC method does not affect the sum-rate. On the contrary, for the asynchronous transmission, decoding*

the stronger user first and the weaker user next results in a larger sum-rate.

Proof. The proof is simply obtained by algebraic manipulations. □

- *Unlike the synchronous transmission, in which SIC is capacity-achieving optimal, the SIC method can not achieve the optimal sum-rate for the asynchronous transmission.*

  Proof. The optimallity of SIC method in the synchronous transmission was expressed, previously. For the asynchronous transmission, the maximal sum-rate achieved by SIC, assuming $|h_1|^2 \geq |h_2|^2$, equals $R_{sum}^{SIC} = \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\eta_\tau P_2|h_2|^2+\sigma_n^2}\right) + \frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}\right)$. However, the capacity region boundary with constant PSDs is denoted as $R_{sum}^{const.} = \frac{1}{2}\int_0^1 \log\left(1+\frac{P_1|h_1|^2}{\sigma_n^2}+\frac{P_2|h_2|^2}{\sigma_n^2}+\frac{P_1P_2|h_1|^2|h_2|^2(1-G_\tau^2(f))}{\sigma_n^2\sigma_n^2}\right)df$. We can prove that:

$$
\begin{aligned}
R_{sum}^{SIC} &= \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\eta_\tau P_2|h_2|^2+\sigma_n^2}\right)+\frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}\right) \qquad (8.22)\\
&=^1 \frac{1}{2}\log_2\left(1+\frac{P_1|h_1|^2}{\int_0^1 G_\tau(f)df P_2|h_2|^2+\sigma_n^2}\right)+\frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}\right)\\
&\leq^2 \frac{1}{2}\int_0^1 \log_2\left(1+\frac{P_1|h_1|^2}{G_\tau(f)P_2|h_2|^2+\sigma_n^2}\right)df+\frac{1}{2}\log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}\right)\\
&= \frac{1}{2}\int_0^1 \log_2\left(1+\frac{P_2|h_2|^2}{\sigma_n^2}+\frac{P_1|h_1|^2}{G_\tau(f)P_2|h_2|^2+\sigma_n^2}+\frac{P_1P_2|h_1|^2|h_2|^2}{G_\tau(f)P_2|h_2|^2\sigma_n^2+\sigma^4}\right)df\\
&\leq^3 \frac{1}{2}\int_0^1 \log\left(1+\frac{P_1|h_1|^2}{\sigma_n^2}+\frac{P_2|h_2|^2}{\sigma_n^2}+\frac{P_1P_2|h_1|^2|h_2|^2(1-G_\tau^2(f))}{\sigma_n^2\sigma_n^2}\right)df\\
&= R_{sum}^{const.}
\end{aligned}
$$

  where (1) is obtained by substituting $\eta = \int_0^1 G_\tau(f)df$, (2) is the result of applying the Jenson's inequality to the convex function of $\log_2(1+\frac{1}{x})$, and finally, (3) can be achieved by simple calculations assuming $G_\tau(f) \leq 1, \forall f$. □

The mentioned results are illustrated in Fig. 8.5 where the capacity region with the optimized PSDs, the optimal rate regions with constant PSDs and the achievable rate

region by SIC method are included for $P_1 = P_2 = 10, \sigma_1^2 = 1, \sigma_2^2 = 5$, and using r.r.c. pulse shape with $\beta = 0.5$.

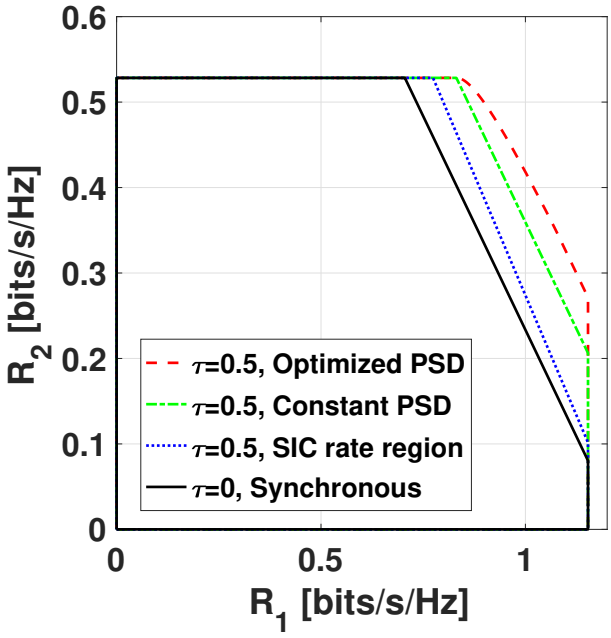


Figure 8.5: Rate region analysis of SIC method

8.4 Conclusion

In this chapter, we analyzed the capacity region of the uplink MAC channels. The results apply to a wide range of pulse shapes, including the well-known r.r.c. pulse shape. The asynchronous MAC channel provides a larger capacity region compared with the synchronous MAC channel. Besides, we proved that the capacity region boundary of the asynchronous MAC channel could not be achieved by the SIC method. Furthermore, the sum-capacity-achieving SIC decoding order for asynchronous MAC is expressed.

Chapter 9

Discussions and Future work

In this thesis, we analyzed the opportunities and challenges that time asynchrony provides in a multiuser network. We showed that conventional system designs are incapable of removing the harmful effects of time asynchrony on system performance. As an example, we demonstrated that in a distributed massive MIMO system, neglecting the existing time delays deteriorate the available rates and hinder the power scaling law. On the other hand, we showed that the inherent time delays in a multiuser system could indeed be useful and improve the system performance by proper architecture design. To this end, we introduced proper sampling methods that generated sufficient statistics to detect the transmitted symbols and analyzed the resulting system models. Favorably, the effective system model is turned into an ISI channel with memory, which enables the use of MLSD methods such as the Viterbi algorithm. As a result, a superior BEP performance is obtained compared with the conventional synchronous system.

Next, we proposed to add timing offsets in downlink NOMA schemes intentionally. The receiver architecture includes oversampling and a SIC scheme similar to the conventional NOMA schemes, however, asynchrony reduces IUI and improves the overall performance.

Besides, using novel precoding and simple post-processing at users, we exploited the additional degrees of freedom introduced by time asynchrony. In other words, we decomposed the channel into independent sub-channels and eliminated the interference. We showed that the proposed methods could improve the achievable rate-regions. We also analyzed the capacity of FTN signaling, which is similar to the asynchronous downlink transmission. We showed that the capacity-achieving input causes practical issues such as OOB emission. To solve these issues, we proposed a simple solution that can achieve an acceptable OOB emission and a desirable capacity.

Next, we investigated the benefits of adding intentional time delays in the performance of downlink transmit beamforming. We compared our proposed methods with the conventional synchronous transmit beamforming method and showed that our proposed methods improve the performance by decreasing the average transmit power. In summary, adding intentional timing offsets transform the flat fading channel into frequency selective channels, and by using proper power allocation method and precoding, reduced transmit power can be achieved compared with the conventional synchronous beamforming methods. Lastly, we thoroughly analyzed the capacity region of the uplink MAC channel for a wide range of pulse shapes, including the well-known r.c. pulse shape.

In this thesis, the opportunities and fundamental performance bounds provided by time asynchrony are discussed along with their underlying reasons. However, in this thesis, the information about time delays was assumed to be known. The knowledge of time delays plays a significant role in the system design and performance improvement brought by asynchronous transmission. Therefore, it is essential to analyze and design a composite receiver structure, including time delay estimation and symbols detection. The interplay between time delay estimation and symbol detection and the effect of estimation error in the system performance is interesting to analyze. It is interesting to consider asynchronous transmission in more practical scenarios that encounter issues such as quantization error,

phase offset, and analyze the overall system performance. Finally, as the last step to validate the analytical results, it is vital to practically build such a system and evaluate its performance with real data.

Bibliography

[1] Ali A Nasir, Salman Durrani, Hani Mehrpouyan, Steven D Blostein, and Rodney A Kennedy. Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years. *EURASIP Journal on Wireless Communications and Networking*, (1):180, Dec. 2016.

[2] Heinrich Meyr, Marc Moeneclaey, and Stefan A Fechtel. *Digital communication receivers: synchronization, channel estimation, and signal processing*. Wiley Online Library, 1998.

[3] Erik Dahlman, Stefan Parkvall, and Johan Skold. *4G:LTE/LTE-advanced for mobile broadband*. Academic Press, 2013.

[4] Aniruddha Das and Bhaskar D Rao. MIMO systems with intentional timing offset. *EURASIP Journal on Advances in Signal Processing*, (1):1–14, Dec. 2011.

[5] S. Barghi and H. Jafarkhani. Exploiting asynchronous Amplify-and-Forward relays to enhance the performance of IEEE 802.11 networks. *IEEE/ACM Transactions on Networking*, 23(2):479–490, Feb. 2014.

[6] S. Barghi, H. Jafarkhani, and H. Yousefi'zadeh. MIMO-assisted MPR-aware MAC design for asynchronous WLANs. *IEEE/ACM Transactions on Networking*, 19(6):1652–1665, Dec. 2011.

[7] Mehdi Torbatian. *Communication over Asynchronous Networks: Signaling and Rate-Reliability Analysis*. PhD thesis, University of Waterloo, 2011.

[8] Kishor Barman and Onkar Dabeer. Capacity of MIMO systems with asynchronous PAM. *IEEE Transactions on Communications*, 57(11):3366–3375, Nov. 2009.

[9] Shihai Shao, Youxi Tang, Ting Kong, Kai Deng, and Ying Shen. Performance analysis of a modified V-BLAST system with delay offsets using zero-forcing detection. *IEEE Transactions on Vehicular Technology*, 56(6):3827–3837, Nov. 2007.

[10] Michael L Honig. *Advances in multiuser detection*. Wiley Online Library, 2009.

[11] Sergio Verdu. *Multiuser detection*. Cambridge University Press, 1998.

[12] Javad Kazemitabar and Hamid Jafarkhani. Performance analysis of multiple antenna multi-user detection. In *IEEE Information Theory and Applications Workshop*, pages 150–159, Feb. 2009.

[13] Viveck R Cadambe and Syed Ali Jafar. Interference alignment and degrees of freedom of the k-user interference channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, Jul. 2008.

[14] Shimon Moshavi. Multi-user detection for DS-CDMA communications. *IEEE Communications Magazine*, 34(10):124–136, Oct. 1996.

[15] A. F. Naguib, N. Seshadri, and A. R. Calderbank. Applications of space-time block codes and interference suppression for high capacity and high data rate wireless systems. In *Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1803–1810, Nov. 1998.

[16] Javad Kazemitabar and Hamid Jafarkhani. Multiuser interference cancellation and detection for users with more than two transmit antennas. *IEEE Transactions on Communications*, 56(4):574–583, Apr. 2008.

[17] Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, and Hiroyuki Ishii. Future steps of LTE-A: evolution toward integration of local area and wide area systems. *IEEE Wireless Communications*, 20(1):12–18, Mar. 2013.

[18] Sundaram Vanka, Sunil Srinivasa, Zhenhua Gong, Peter Vizi, Kostas Stamatiou, and Martin Haenggi. Superposition coding strategies: Design and experimental evaluation. *IEEE Transactions on Wireless Communications*, 11(7):2628–2639, Jul. 2012.

[19] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, Anxin Li, and Kenichi Higuchi. Non-orthogonal multiple access (NOMA) for cellular future radio access. In *IEEE Vehicular Technology Conference (VTC)*, pages 1–5, Jun. 2013.

[20] Yuya Saito, Anass Benjebbour, Yoshihisa Kishiyama, and Takehiro Nakamura. System-level performance evaluation of downlink non-orthogonal multiple access (NOMA). In *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 611–615, Sep. 2013.

[21] Thomas Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, Jan. 1972.

[22] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[23] James E Mazo. Faster-than-Nyquist signaling. *The Bell System Technical Journal*, 54(8):1451–1462, 1975.

[24] Helmut Bölcskei. *Space-time wireless systems: from array processing to MIMO communications*. Cambridge University Press, 2006.

[25] Hamid Jafarkhani. *Space-time coding: theory and practice*. Cambridge University Press, 2005.

[26] J. J. van de Beek, P. O. Borjesson, M. L. Boucheret, D. Landstrom, J. M. Arenas, P. Odling, C. Ostberg, M. Wahlqvist, and S. K. Wilson. A time and frequency synchronization scheme for multiuser OFDM. *IEEE Journal on Selected Areas in Communications*, 17(11):1900–1914, Nov. 1999.

[27] En Zhou, Xing Zhang, Hui Zhao, and Wenbo Wang. Synchronization algorithms for MIMO OFDM systems. In *IEEE Wireless Communications and Networking Conference*, volume 1, pages 18–22, Mar. 2005.

[28] A. A. Nasir, S. Durrani, and R. A. Kennedy. Blind timing and carrier synchronization in distributed multiple input multiple output communication systems. *IET Communications*, 5(7):1028–1037, May 2011.

[29] T. L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600, Nov. 2010.

[30] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, Feb. 2014.

[31] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. Energy and spectral efficiency of very large multiuser MIMO systems. *IEEE Transactions on Communications*, 61(4):1436–1449, Apr. 2013.

[32] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang. An overview of massive MIMO: Benefits and challenges. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):742–758, Oct. 2014.

[33] Ang Yang, Yindi Jing, Chengwen Xing, Zesong Fei, and Jingming Kuang. Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas. *IEEE Transactions on Wireless Communications*, 14(10):5659–5671, Jun. 2015.

[34] Howard H Xia, Angel B Herrera, Steve Kim, and Fernando S Rico. A CDMA-distributed antenna system for in-building personal communications services. *IEEE Journal on Selected Areas in Communications*, 14(4):644–650, May 1996.

[35] Ralf E Schuh and Magnus Sommer. W-CDMA coverage and capacity analysis for active and passive distributed antenna systems. In *IEEE 55th Vehicular Technology Conference, VTC Spring*, volume 1, pages 434–438, May 2002.

[36] Upamanyu Madhow, Donald Richard Brown, Soura Dasgupta, and Raghuraman Mudumbai. Distributed massive MIMO: algorithms, architectures and concept systems. In *IEEE Information Theory and Applications Workshop (ITA)*, pages 1–7, Feb. 2014.

[37] Kien T Truong and Robert W Heath. The viability of distributed antennas for massive MIMO systems. In *Asilomar Conference on Signals, Systems and Computers*, pages 1318–1323, Nov. 2013.

[38] H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and S. H. Dai. Asynchronous interference mitigation in cooperative base station systems. *IEEE Transactions on Wireless Communications*, 7(1):155–165, Jan. 2008.

[39] Q. Zhang, S. Jin, K. K. Wong, H. Zhu, and M. Matthaiou. Power scaling of uplink massive mimo systems with arbitrary-rank channel means. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):966–981, Oct. 2014.

[40] Mehdi Ganji and Hamid Jafarkhani. On the performance of MRC receiver with unknown timing mismatch-a large scale analysis. In *IEEE International Conference on Communications (ICC)*, pages 1–6, May 2018.

[41] Mehdi Ganji and Hamid Jafarkhani. On the performance of MRC receiver with unknown timing mismatch-a large scale analysis. 2018. available at https://arxiv.org/abs/1703.10422.

[42] Haifan Yin, David Gesbert, and Laura Cottatellucci. Dealing with interference in distributed large-scale MIMO systems: A statistical approach. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):942–953, May 2014.

[43] Jingon Joung, Yeow Khiang Chia, and Sumei Sun. Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):954–965, Mar. 2014.

[44] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, Apr. 2003.

[45] Steven M Kay. *Fundamentals of statistical signal processing: Practical algorithm development*, volume 3. Pearson Education, 2013.

[46] Sina Poorkasmaei and Hamid Jafarkhani. Asynchronous orthogonal differential decoding for multiple access channels. *IEEE Transactions on Wireless Communications*, 14(1):481–493, Jan. 2015.

[47] MR Avendi and Hamid Jafarkhani. Differential distributed space-time coding with imperfect synchronization in frequency-selective channels. *IEEE Transactions on Wireless Communications*, 14(4):1811–1822, Nov. 2014.

[48] Mehdi Ganji and Hamid Jafarkhani. Interference mitigation using asynchronous transmission and sampling diversity. In *EEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec. 2016.

[49] Jerry M Mendel. *Lessons in estimation theory for signal processing, communications, and control*. Pearson Education, 1995.

[50] Pouya Pezeshkpour, Carlos Guestrin, and Sameer Singh. Compact factorization of matrices using generalized round-rank. *arXiv preprint arXiv:1805.00184*, 2018.

[51] Ulf Grenander and Gabor Szego. *Toeplitz forms and their applications*. University of California Press, 1958.

[52] Robert Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Transactions on Information Theory*, 18(6):725–730, Nov. 1972.

[53] Robert M Gray et al. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, Jan. 2006.

[54] Jesús Gutiérrez-Gutiérrez and Pedro M Crespo. Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems. *IEEE Transactions on Information Theory*, 54(12):5671–5680, Dec. 2008.

[55] Erik G Larsson. MIMO detection methods: How they work [lecture notes]. *IEEE signal processing magazine*, 26(3):91–95, Apr. 2009.

[56] Shaoshi Yang and Lajos Hanzo. Fifty years of MIMO detection: The road to large-scale MIMOs. *IEEE Communications Surveys and Tutorials*, 17(4):1941–1988, Sep. 2015.

[57] Harold Artés, Dominik Seethaler, and Franz Hlawatsch. Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ML detection. *IEEE transactions on signal processing*, 51(11):2808–2820, Dec. 2003.

[58] GDJR Forney. Maximum-likelihood sequence estimation of digital sequences in the presence of inter-symbol interference. *IEEE Transactions on Information theory*, 18(3):363–378, May 1972.

[59] John G Proakis. *Intersymbol interference in digital communication systems*. Wiley Online Library, 2001.

[60] Lin Bai, Jinho Choi, and Quan Yu. *Low complexity MIMO receivers*. Springer, 2014.

[61] Ahmadreza Hedayat and Aria Nosratinia. Outage and diversity of linear receivers in flat-fading MIMO channels. *IEEE Transactions on Signal Processing*, 55(12):5868–5873, Nov. 2007.

[62] Ersin Sengul, Enis Akay, and Ender Ayanoglu. Diversity analysis of single and multiple beamforming. *IEEE Transactions on Communications*, 54(6):990–993, Jun. 2006.

[63] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 2012.

[64] R Xu and FCM Lau. Performance analysis for MIMO systems using zero forcing detector over fading channels. *IEEE Proceedings-Communications*, 153(1):74–80, Feb. 2006.

[65] Lizhong Zheng and David N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on information theory*, 49(5):1073–1096, May 2003.

[66] Linglong Dai, Bichai Wang, Yifei Yuan, Shuangfeng Han, I Chih-Lin, and Zhaocheng Wang. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. *IEEE Communications Magazine*, 53(9):74–81, Sep. 2015.

[67] Pouya Pezeshkpour and Hamid Behroozi. Optimal tradeoff between source and state distortions over a gaussian channel using single and hybrid digital analog codes. In *7'th International Symposium on Telecommunications (IST'2014)*, pages 619–622, Sep. 2014.

[68] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley and Sons, 2012.

[69] Giuseppe Caire and Shlomo Shamai. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7):1691–1706, Jul. 2003.

[70] Pramod Viswanath and David N. C. Tse. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Transactions on Information Theory*, 49(8):1912–1921, Aug. 2003.

[71] Sriram Vishwanath, Nihar Jindal, and Andrea Goldsmith. Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels. *IEEE Transactions on Information Theory*, 49(10):2658–2668, Oct. 2003.

[72] Wei Yu and John M Cioffi. Sum capacity of Gaussian vector broadcast channels. *IEEE Transactions on Information Theory*, 50(9):1875–1892, Sep. 2004.

[73] Hanan Weingarten, Yossef Steinberg, and Shlomo Shitz Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Transactions on Information Theory*, 52(9):3936–3964, Sep. 2006.

[74] Aaron D Wyner. Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Transactions on Information Theory*, 40(6):1713–1727, Nov. 1994.

[75] Peng Xu, Zhiguo Ding, Xuchu Dai, and H Vincent Poor. A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks. *IEEE Access*, 3:1633–1639, Sep. 2015.

[76] Boya Di, Siavash Bayat, Lingyang Song, and Yonghui Li. Radio resource allocation for downlink non-orthogonal multiple access (NOMA) networks using matching theory. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec. 2015.

[77] Shimei Liu, Chao Zhang, and Gangming Lyu. User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system. In *IEEE International Conference on Communication Workshop (ICCW)*, pages 2561–2565, Jun. 2015.

[78] Zhiguo Ding, Mugen Peng, and H Vincent Poor. Cooperative non-orthogonal multiple access in 5G systems. *IEEE Communications Letters*, 19(8):1462–1465, Aug. 2015.

[79] Xun Zou, Mehdi Ganji, and Hamid Jafarkhani. Trellis-coded non-orthogonal multiple access. *IEEE Wireless Communications Letters*, 9(4):538–542, Dec. 2019.

[80] Mehdi Ganji, Xun Zou, and Hamid Jafarkhani. A block-based non-orthogonal multicarrier scheme. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec. 2019.

[81] Jinho Choi. On the power allocation for MIMO-NOMA systems with layered transmissions. *IEEE Transactions on Wireless Communications*, 15(5):3226–3237, May 2016.

[82] Zhiguo Ding, Fumiyuki Adachi, and H Vincent Poor. The application of MIMO to non-orthogonal multiple access. *IEEE Transactions on Wireless Communications*, 15(1):537–552, Jan. 2016.

[83] Sergio Verdu. The capacity region of the symbol-asynchronous Gaussian multiple-access channel. *IEEE Transactions on Information Theory*, 35(4):733–751, Jul. 1989.

[84] Jianxiong Cui, Guangliang Dong, Shengli Zhang, Haitao Li, and Guinian Feng. Asynchronous NOMA for downlink transmissions. *IEEE Communications Letters*, 21(2):402–405, Feb. 2017.

[85] Xun Zou, Biao He, and Hamid Jafarkhani. On uplink asynchronous non-orthogonal multiple access systems with timing error. In *IEEE International Conference on Communications (ICC)*, pages 1–6, May 2018.

[86] Xun Zou, Mehdi Ganji, and Hamid Jafarkhani. Cooperative asynchronous non-orthogonal multiple access with power minimization under qos constraints. *IEEE Transactions on Wireless Communications*, Nov. 2019.

[87] Behnam Pourghassemi, Ardalan Amiri Sani, and Aparna Chandramowlishwaran. What-if analysis of page load time in web browsers using causal profiling. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–23, Jun. 2019.

[88] Mehdi Ganji and Hamid Jafarkhani. Novel time asynchronous NOMA schemes for downlink transmissions. 2018. available at https://arxiv.org/abs/1808.08665.

[89] Mehdi Ganji and Hamid Jafarkhani. Time asynchronous NOMA for downlink transmission. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, Apr. 2019.

[90] Thomas M Cover. Comments on broadcast channels. *IEEE Transactions on Information Theory*, 44(6):2524–2530, Oct. 1998.

[91] Lokenath Debnath and Piotr Mikusiński. *Hilbert spaces with applications*. Academic Press, 2005.

[92] Robert G Gallager. *Principles of digital communication*, volume 1. Cambridge University Press Cambridge, UK, 2008.

[93] YJD Kim. *Faster than Nyquist Transmission over Continuous-Time Channels: Capacity Analysis and Coding*. PhD thesis, McGill University, Montreal, Canada, 2013.

[94] Pantelis-Daniel Arapoglou, Alberto Ginesi, Stefano Cioni, Stefan Erl, Federico Clazzer, Stefano Andrenacci, and Alessandro Vanelli-Coralli. DVB-S2X-enabled precoding for high throughput satellite systems. *International Journal of Satellite Communications and Networking*, 34(3):439–455, May 2016.

[95] Fredrik Rusek and John B Anderson. Constrained capacities for faster-than-Nyquist signaling. *IEEE Transactions on Information Theory*, 55(2):764–775, Feb. 2009.

[96] Deepak Dasalukunte, V Owall, Fredrik Rusek, and John B Anderson. *Faster than Nyquist Signaling*. Springer, 2016.

[97] Young Geon Yoo and Joon Ho Cho. Asymptotic optimality of binary faster-than-Nyquist signaling. *IEEE Communications Letters*, 14(9):788–790, Aug. 2010.

[98] Fredrik Rusek and John B Anderson. Non binary and precoded faster than Nyquist signaling. *IEEE Transactions on Communications*, 56(5):808–817, May 2008.

[99] Ebrahim Bedeer, Mohamed Hossam Ahmed, and Halim Yanikomeroglu. A very low complexity successive symbol-by-symbol sequence estimator for faster-than-Nyquist signaling. *IEEE Access*, 5:7414–7422, Mar. 2017.

[100] Yong Jin Daniel Kim, Jan Bajcsy, and David Vargas. Faster-than-Nyquist broadcasting in Gaussian channels: Achievable rate regions and coding. *IEEE Transactions on Communication*, 64(3):1016–1030, Jan. 2016.

[101] Ather Gattami, Emil Ringh, and Johan Karlsson. Time localization and capacity of faster-than-nyquist signaling. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, Dec. 2015.

[102] Ji Zhou, Yaojun Qiao, Zhanyu Yang, Qixiang Cheng, Qi Wang, Mengqi Guo, and Xizi Tang. Capacity limit for faster-than-nyquist non-orthogonal frequency-division multiplexing signaling. *Scientific reports*, 7(1):3380, Jun. 2017.

[103] Dzevdan Kapetanovic and Fredrik Rusek. The effect of signaling rate on information rate for single carrier linear transmission systems. *IEEE Transactions on Communications*, 60(2):421–428, Jan. 2012.

[104] Mehdi Ganji, Xun Zou, and Hamid Jafarkhani. On the capacity of faster than nyquist signaling. *IEEE Communications Letters, Early Access*, Mar. 2020.

[105] Walter Hirt and James L Massey. Capacity of the discrete-time Gaussian channel with intersymbol interference. *IEEE Transactions on Information Theory*, 34(3):380–388, May 1988.

[106] Lal Chand Godara. *Handbook of antennas in wireless communications*, volume 4. CRC press, 2018.

[107] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[108] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H Vincent Poor. *MIMO wireless communications*. Cambridge university press, 2007.

[109] Max Costa. Writing on dirty paper (corresp.). *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.

[110] Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai. The capacity region of the Gaussian MIMO broadcast channel. In *Proceedings of IEEE International Symposium on Information Theory, ISIT*, page 174, Jun. 2004.

[111] Farrokh Rashid-Farrokhi, KJ Ray Liu, and Leandros Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, Oct. 1998.

[112] P. Viswanath, D. N. C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, Jun. 2002.

[113] M. Schubert and H. Boche. Solution of the multiuser downlink beamforming problem with individual SINR constraints. *IEEE Transactions on Vehicular Technology*, 53(1):18–28, Jan. 2004.

[114] Ami Wiesel, Yonina C Eldar, and Shlomo Shamai. Linear precoding via conic optimization for fixed MIMO receivers. *IEEE Transactions on Signal Processing*, 54(1):161–176, Dec. 2005.

[115] Li Liu and Hamid Jafarkhani. Novel transmit beamforming schemes for time-selective fading multiantenna systems. *IEEE Transactions on Signal Processing*, 54(12):4767–4781, Dec. 2006.

[116] Wei Yu and Tian Lan. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Transactions on Signal Processing*, 55(6):2646–2660, May 2007.

[117] Harish Viswanathan, Sivarama Venkatesan, and Howard Huang. Downlink capacity evaluation of cellular networks with known-interference cancellation. *IEEE Journal on Selected Areas in Communications*, 21(5):802–811, Jun. 2003.

[118] Masoud Sharif and Babak Hassibi. A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users. *IEEE Transactions on Communications*, 55(1):11–15, Jan. 2007.

[119] Taesang Yoo and Andrea Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE Journal on Selected Areas in Communications*, 24(3):528–541, Mar. 2006.

[120] Miguel Angel Vazquez, Ana Perez-Neira, Dimitrios Christopoulos, Symeon Chatzinotas, Bjorn Ottersten, Pantelis-Daniel Arapoglou, Alberto Ginesi, and Giorgio Tarocco. Precoding in multibeam satellite communications: Present and future challenges. *IEEE Transactions on Wireless Communications*, 23(6):88–95, Dec. 2016.

[121] Dimitrios Christopoulos, Symeon Chatzinotas, and Björn Ottersten. Multicast multigroup precoding and user scheduling for frame-based satellite communications. *IEEE Transactions on Wireless Communications*, 14(9):4695–4707, Apr. 2015.

[122] Martin Schubert and Holger Boche. Solution of the multiuser downlink beamforming problem with individual SINR constraints. *IEEE Transactions on Vehicular Technology*, 53(1):18–28, Jan. 2004.

[123] Emil Björnson, Mats Bengtsson, and Björn Ottersten. Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]. *IEEE Signal Processing Magazine*, 31(4):142–148, Jul. 2014.

[124] Erdem Koyuncu and Hamid Jafarkhani. Variable-length limited feedback beamforming in multiple-antenna fading channels. *IEEE Transactions on Information Theory*, 60(11):7140–7165, Nov. 2014.

[125] Mats Bengtsson and Björn Ottersten. Optimal and suboptimal transmit beamforming. *in Handbook of Antennas in Wireless Communications, L. C. Godara, volume 4, CRC press*, 2018.

[126] Laura Cottatellucci, Ralf R Muller, and Merouane Debbah. Asynchronous CDMA systems with random spreading—Part I: Fundamental limits. *IEEE Transactions on Information Theory*, 56(4):1477–1497, Mar. 2010.

[127] Jianxiong Cui, Guangliang Dong, Shengli Zhang, Haitao Li, and Guinian Feng. Asynchronous NOMA for downlink transmissions. *IEEE Communications Letters*, 21(2):402–405, Oct. 2016.

[128] Xun Zou, Biao He, and Hamid Jafarkhani. An analysis of two-user uplink asynchronous non-orthogonal multiple access systems. *IEEE Transactions on Wireless Communications*, 18(2):1404–1418, Jan. 2019.

[129] Mehdi Ganji and Hamid Jafarkhani. Improving NOMA multi-carrier systems with intentional frequency offsets. *IEEE Wireless Communications Letters*, 8(4):1060–1063, Aug. 2019.

[130] Xuehua Zhang, Mehdi Ganji, and Hamid Jafarkhani. Exploiting asynchronous signaling for multiuser cooperative networks with analog network coding. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, Mar. 2017.

[131] Huseyin Haci, Huiling Zhu, and Jiangzhou Wang. Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique. *IEEE Transactions on Communications*, 65(3):1319–1335, Jan. 2017.

[132] Mehdi Ganji, Xun Zou, and Hamid Jafarkhani. Exploiting time asynchrony in multi-user transmit beamforming. *IEEE Transactions on Wireless Communications*, 19(5):3156–3169, Feb. 2020.

[133] Shuying Shi, Martin Schubert, and Holger Boche. Rate optimization for multiuser mimo systems with linear processing. *IEEE Transactions on Signal Processing*, 56(8):4020–4030, May 2008.

[134] Evgenij E Tyrtyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear algebra and its applications*, 232:1–43, 1996.

[135] Zhihui Zhu and Michael B Wakin. On the asymptotic equivalence of circulant and Toeplitz matrices. *IEEE Transactions on Information Theory*, 63(5):2975–2992, Mar. 2017.

[136] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[137] Andrea J Goldsmith and Michelle Effros. The capacity region of broadcast channels with intersymbol interference and colored gaussian noise. *IEEE Transactions on Information Theory*, 47(1):219–240, Jan. 2001.

[138] Wei Yu, George Ginis, and John M Cioffi. Distributed multiuser power control for digital subscriber lines. *IEEE Journal on Selected Areas in Communications*, 20(5):1105–1115, Aug. 2002.

[139] Sean Huberman, Christopher Leung, and Tho Le-Ngoc. Dynamic spectrum management (DSM) algorithms for multi-user xDSL. *IEEE Communications Surveys and Tutorials*, 14(1):109–130, Oct. 2010.

[140] Navid Rezazadeh and Solmaz S Kia. Privacy preservation in a continuous-time static average consensus algorithm over directed graphs. In *2018 Annual American Control Conference (ACC)*, pages 5890–5895. IEEE, Jun. 2018.

[141] Navid Rezazadeh and Solmaz S Kia. A sub-modular receding horizon approach to persistent monitoring for a group of mobile agents over an urban area. *IFAC-PapersOnLine*, 52(20):217–222, Jan. 2019.

[142] John Papandriopoulos and Jamie S Evans. SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks. *IEEE Transaction on Information Theory*, 55(8):3711–3724, Jul. 2009.

[143] Eleftherios Karipidis, Nicholas D Sidiropoulos, and Zhi-Quan Luo. Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups. *IEEE Transactions on Signal Processing*, 56(3):1268–1279, Feb. 2008.

[144] M. S. Ali, H. Tabassum, and E. Hossain. Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems. *IEEE Access*, 4:6325–6343, 2016.

[145] J. Seo, Y. Sung, and H. Jafarkhani. A High-Diversity Transceiver Design for MISO Broadcast Channels. *IEEE Transactions on Wireless Communications*, 18(5):2591–2606, May 2019.

[146] Wing-Kin Ken Ma. Semidefinite relaxation of quadratic optimization problems and applications. *IEEE Signal Processing Magazine*, 1053(5888/10), May 2010.

[147] Michael Grant and Stephen Boyd. Cvx: Matlab software for disciplined convex programming, version 2.1, 2014.

[148] Shuzhong Zhang and Yongwei Huang. Complex quadratic optimization and semidefinite programming. *SIAM Journal on Optimization*, 16(3):871–890, Jul. 2006.

[149] Aaron Wyner. Recent results in the shannon theory. *IEEE Transactions on information Theory*, 20(1):2–10, Jan. 1974.

[150] Thomas M Cover. Some advances in broadcast channels. In *Advances in communication systems*, volume 4, pages 229–260. Elsevier, Jan. 1975.

[151] Kenichi Higuchi and Anass Benjebbour. Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access. *IEICE Transactions on Communications*, 98(3):403–414, Mar. 2015.

[152] Zhiguo Ding, Xianfu Lei, George K Karagiannidis, Robert Schober, Jinhong Yuan, and Vijay K Bhargava. A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends. *IEEE Journal on Selected Areas in Communications*, 35(10):2181–2195, Jul. 2017.

[153] R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober. Linear massive MIMO precoders in the presence of phase noise-a large-scale analysis. *IEEE Transactions on Vehicular Technology*, 65(5):3057–3071, Jun. 2016.

[154] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson. On the optimality of single-carrier transmission in large-scale antenna systems. *IEEE Wireless Communications Letters*, 1(4):276–279, Aug. 2012.

[155] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson. Effect of oscillator phase noise on uplink performance of large MU-MIMO systems. In *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1190–1197, Oct. 2012.

[156] Thomas Eng and Laurence B Milstein. Coherent DS-CDMA performance in Nakagami multipath fading. *IEEE Transactions on Communications*, 43(2):1134–1143, Feb. 1995.

[157] Murray Dow. Explicit inverses of Toeplitz and associated matrices. *ANZIAM Journal*, 44(E):185–215, Jan. 2008.

[158] Dennis S Bernstein. *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.

Appendix A

Supplementary Proofs for Chapter 2

A.1 Proof of Theorem 2.1

We analyze four different receiver structures, including: MRC with perfect and imperfect CSI and MRC-ZF with perfect and imperfect CSI. The output sample of either of these receivers for detection of the ath symbol of the lth user can be written in a general framework as:

$$\boldsymbol{y}_{\boldsymbol{l,p/ip}}^{\boldsymbol{mrc/mrc-zf}}(a) = \sqrt{\rho_d}\sum_{k=1}^{K}\sum_{n=1}^{N}\boldsymbol{T}_{\boldsymbol{lk,p/ip}}^{\boldsymbol{mrc/mrc-zf}}(a,n)\boldsymbol{s_k}(n) + \boldsymbol{n}_{\boldsymbol{l,p/ip}}^{\boldsymbol{mrc/mrc-zf}}(a)$$

Discarding the subscripts for different receivers and based on the assumption that the coefficients are known perfectly at the receiver, the corresponding achievable rate can be calculated as follows[31]:

$$R_l(a) = E\left\{\log_2\left(1+\frac{\rho_d(\boldsymbol{T_{ll}}(a,a))^2}{\rho_d\sum\sum_{\substack{k,n\\(k,n)\neq(l,a)}}(\boldsymbol{T_{lk}}(a,n))^2+\sigma^2_{n_l(a)}}\right)\right\} \tag{A.1}$$

Because each of the coefficients is the arithmetic mean of M uncorrelated random variables, as M increases, their randomness decreases and they approach to their expected value. As a result, the above expression can be approximated in the Massive MIMO context as follows: [39, 153]

$$R_l(a) = \log_2 \left(1 + \frac{\rho_d E\left\{(\boldsymbol{T_{ll}}(a,a))^2\right\}}{\rho_d \sum\sum\limits_{\substack{k,n \\ (k,n)\neq(l,a)}} E\left\{(\boldsymbol{T_{lk}}(a,n))^2\right\} + \sigma^2_{n_l(a)}} \right) \tag{A.2}$$

However, due to the existence of unknown time delays, the assumption of perfect knowledge of coefficients at the receiver is not valid anymore. Because the expected value of the coefficients is known by the receiver, in order to find the achievable rates, we rewrite Eq. (A.1), as follows (the subscripts are discarded):

$$\boldsymbol{y_l}(a) = \sqrt{\rho_d} \sum_{k=1}^{K} \sum_{n=1}^{N} E\{\boldsymbol{T_{lk}}(a,n)\} \boldsymbol{s_k}(n) + \tilde{\boldsymbol{n}}_{\boldsymbol{l}}(a) \tag{A.3}$$

where $\tilde{\boldsymbol{n}}_{\boldsymbol{l}}(a) = \sqrt{\rho_d} \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\boldsymbol{T_{lk}}(a,n) - E\{\boldsymbol{T_{lk}}(a,n)\}\right) \boldsymbol{s_k}(n) + \boldsymbol{n_l}(a)$. In this new system model, all the effective channel coefficients are known by the receiver and the effective noise is denoted as $\tilde{\boldsymbol{n}}_{\boldsymbol{l}}$. It is mentioned in [154] that the achievable information rate for the system model in Eq. (A.3) is given by considering the worst case uncorrelated additive noise having the same variance as $\tilde{\boldsymbol{n}}_{\boldsymbol{l}}(a)$. Variance of $\tilde{\boldsymbol{n}}_{\boldsymbol{l}}(a)$ can be easily calculated as $\sigma^2_{\tilde{n}_l(a)} = \rho_d \sum_{k=1}^{K} \sum_{n=1}^{N} Var\{\boldsymbol{T_{lk}}(a,n)\} + \sigma^2_{n_l(a)}$. As a result, the achievable rate can be written as:

$$R_l(a) = E\left\{ \log_2 \left(1 + \frac{\rho_d E^2\{\boldsymbol{T_{ll}}(a,a)\}}{\rho_d \sum\sum\limits_{\substack{k,n \\ (k,n)\neq(l,a)}} E^2\left\{\boldsymbol{T_{lk}}(a,n)\right\} + \rho_d \sum_{k=1}^{K} \sum_{n=1}^{N} Var\{\boldsymbol{T_{lk}}(a,n)\} + \sigma^2_{n_l(a)}} \right) \right\}$$

The expectation can be discarded because the randomness of the coefficients is combined with the noise and no randomness is left except the noise realizations where the worst

case of Gaussian uncorrelated additive noise is considered [155]. Then,

$$R_l(a) = \log_2 \left(1 + \frac{\rho_d E^2\{\boldsymbol{T_{ll}}(a,a)\}}{\rho_d \sum\sum_{\substack{k,n \\ (k,n)\neq(l,a)}} E^2\{\boldsymbol{T_{lk}}(a,n)\} + \rho_d \sum_{k=1}^{K}\sum_{n=1}^{N} Var\{\boldsymbol{T_{lk}}(a,n)\} + \sigma^2_{n_l(a)}} \right)$$

By using the fact that $Var\{x\} = E\{x^2\} - E^2\{x\}$, the formula for the achievable rate can be further simplified to:

$$R_l(a) = \log_2 \left(1 + \frac{\rho_d E^2\{\boldsymbol{T_{ll}}(a,a)\}}{\rho_d \sum_{k=1}^{K}\sum_{n=1}^{N} E\{(\boldsymbol{T_{lk}}(a,n))^2\} - \rho_d E^2\{\boldsymbol{T_{ll}}(a,a)\} + \sigma^2_{n_l(a)}} \right) \quad (A.4)$$

Hence, for finding the achievable rates for different receivers, we just need to calculate the values of $E\{(\boldsymbol{T_{lk}}(a,n))^2\}$ and the variance of effective noise vector for different receivers. Note that due to the structure of the system, the achievable rate for different symbols of the frame except the I-boundary ones (negligible with respect to the frame length) is the same, thus the index of a can be discarded. Note that, throughout the paper, for element-wise power we use $(\boldsymbol{T}(p,q))^2$ and for elements of powered matrices we use $\boldsymbol{T}^2(p,q)$.
We show the step by step derivation for the MRC receiver with perfect CSI, and the corresponding formulas for other scenarios can be derived similarly.

For the MRC receiver with perfect CSI the effective channel matrix is denoted as $\boldsymbol{T^{mrc}_{lk,p}} = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\beta_{ml}\beta_{mk}}h^*_{ml}h_{mk}\boldsymbol{G_{mk}}$ and the values of $E\{|\boldsymbol{T^{mrc}_{lk,p}}(a,n)|^2\}$ are calculated in the next lemma.

Lemma A.1. *The expected value of $|\boldsymbol{T^{mrc}_{lk,p}}(a,n)|^2$ can be calculated as follows:*

$$E\left\{|\boldsymbol{T^{mrc}_{lk,p}}(a,n)|^2\right\} = \begin{cases} \frac{1}{M}\left(\overline{\beta_{ll}}\left(2g''(a-n) - g'(a-n)\right) + M\overline{\beta_l}^2 g'(a-n)\right) & k = l \\ \frac{\overline{\beta_{lk}}}{M} g''(a-n) & k \neq l \end{cases} \quad (A.5)$$

where

$$g'[i] = E^2_{f(\tau)}\{g(iT-\tau)\} = \left[\int_{-\infty}^{\infty} g(iT-\tau)f(\tau)d\tau\right]^2$$

$$g''[i] = E_{f(\tau)}\{g^2(iT-\tau)\} = \int_{-\infty}^{\infty} g^2(iT-\tau)f(\tau)d\tau$$

Proof. Case $k = l$:

$$E\left\{|\boldsymbol{T}_{\boldsymbol{ll,p}}^{\boldsymbol{mrc}}(a,n)|^2\right\} = \frac{1}{M^2}E\left\{\left(\sum_{m=1}^{M}\beta_{ml}|h_{ml}|^2\boldsymbol{G_{ml}}(a,n)\right)\left(\sum_{m=1}^{M}\beta_{ml}|h_{ml}|^2\boldsymbol{G_{ml}}(a,n)\right)\right\} \tag{A.6}$$
$$= \frac{1}{M^2}\left(\sum_{m=1}^{M}\beta_{ml}^2E\left\{|h_{ml}|^4(\boldsymbol{G_{ml}}(a,n))^2\right\} + \sum_{m_1=1}^{M}\sum_{\substack{m_2=1\\ m_2\neq m_1}}^{M}\beta_{m_1l}\beta_{m_2l}E\left\{|h_{m_1l}|^2\boldsymbol{G_{m_1l}}(a,n)|h_{m_2l}|^2\boldsymbol{G_{m_2l}}(a,n)\right\}\right)$$

where the expectation is taken over different realizations in time. Expectations of the elements of matrices $\boldsymbol{G_{ml}}$ only depend on the pulse shape and the joint distribution of time delays which are known by the receiver. For example, $E\left\{\boldsymbol{G_{m_1l}}(a,n)\boldsymbol{G_{m_2l}}(a,n)\right\}$ can be expressed as:

$$E\left\{\boldsymbol{G_{m_1l}}(a,n)\boldsymbol{G_{m_2l}}(a,n)\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t'-\tau_1)g(t'-\tau_2)f_{\tau_{m_1l},\tau_{m_2l}}(\tau_1,\tau_2)d\tau_1d\tau_2$$

where $t' = (a-n)T$. Note that based on the system characteristics, the time delays can follow any given correlated distribution or even they might be the same which is more suited for collocated receive antennas , i.e., $f_{\tau_{ml_1},\tau_{ml_2}}(\tau_1,\tau_2) = f(\tau_1)\delta(\tau_1-\tau_2)$, or independent for distributed receive antennas, i.e., $f_{\tau_{ml_1},\tau_{ml_2}}(\tau_1,\tau_2) = f_{\tau_{ml_1}}(\tau_1)f_{\tau_{ml_2}}(\tau_2)$. Although no assumption is required for the joint distribution of delays, we consider the distributed scenario and assume that all the time delays are independent and follow the same distribution, i.e., $f_{\tau_{ml}}(\tau) = f(\tau)$. Therefore, we will have:

$$E\left\{\boldsymbol{G_{m_1l}}(a,n)\boldsymbol{G_{m_2l}}(a,n)\right\} = \left[\int_{-\infty}^{\infty} g(t'-\tau)f(\tau)d\tau\right]^2,$$
$$E\left\{(\boldsymbol{G_{ml}}(a,n))^2\right\} = \int_{-\infty}^{\infty} g^2(t'-\tau)f(\tau)d\tau \tag{A.7}$$

Note that, $E^2\left[\boldsymbol{G_{ml}}(a,n)\right]$ and $E\left[(\boldsymbol{G_{ml}}(a,n))^2\right]$ depend on a and n thorough their difference, therefore we denote them as $g'(a-n)$ and $g''(a-n)$, respectively. As a result, Eq.

(A.6) can be presented as:

$$E\left\{|\boldsymbol{T}_{\boldsymbol{ll,p}}^{\boldsymbol{mrc}}(a,n)|^2\right\} = \frac{1}{M^2}\left(2g''(a-n)\sum_{m=1}^{M}\beta_{ml}^2 + g'(a-n)\sum_{m_1=1}^{M}\sum_{\substack{m_2=1\\ m_2\neq m_1}}^{M}\beta_{m_1 l}\beta_{m_2 l}\right)$$
$$= \frac{1}{M^2}\left(2g''(a-n)M\overline{\beta_{ll}} + g'(a-n)(M^2\overline{\beta_l}^2 - M\overline{\beta_{ll}})\right)$$
$$= \frac{1}{M}\left(\overline{\beta_{ll}}\left(2g''(a-n) - g'(a-n)\right) + M\overline{\beta_l}^2 g'(a-n)\right) \quad \text{(A.8)}$$

where $\overline{\beta_k}$ and $\overline{\beta_{lk}}$ are defined as follows, respectively:

$$\overline{\beta_k} = \frac{1}{M}\sum_{m=1}^{M}\beta_{mk}, \quad \overline{\beta_{lk}} = \frac{1}{M}\sum_{m=1}^{M}\beta_{ml}\beta_{mk} \quad \text{(A.9)}$$

Note that these values are bounded, i.e., $\min_m \beta_{mk} < \overline{\beta_k} < \max_m \beta_{mk}$ and $\min_m \beta_{ml}\beta_{mk} < \overline{\beta_{lk}} < \max_m \beta_{ml}\beta_{mk}$.
Case $k \neq l$: By taking similar steps, we can show that:

$$E\left\{|\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc}}(a,n)|^2\right\} = \frac{1}{M^2}E\left\{\left(\sum_{m=1}^{M}\sqrt{\beta_{ml}\beta_{mk}}h_{ml}^* h_{mk}\boldsymbol{G}_{\boldsymbol{mk}}(a,n)\right)\left(\sum_{m=1}^{M}\sqrt{\beta_{ml}\beta_{mk}}h_{ml}h_{mk}^*\boldsymbol{G}_{\boldsymbol{mk}}(a,n)\right)\right\}$$
$$= \frac{1}{M^2}\left(\sum_{m=1}^{M}\beta_{ml}\beta_{mk}E\left\{|h_{ml}|^2|h_{mk}|^2(\boldsymbol{G}_{\boldsymbol{mk}}(a,n))^2\right\}\right) = \frac{1}{M}\left(g''(a-n)\overline{\beta_{lk}}\right) \quad \text{(A.10)}$$

which completes the derivation of expectations. □

Covariance matrix of the effective noise vector is also calculated as:

$$COV\{\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}}\} = E\left\{\left(\frac{1}{M}\sum_{m=1}^{M}\sqrt{\beta_{ml}}h_{ml}^*\boldsymbol{n_m}\right)\left(\frac{1}{M}\sum_{m=1}^{M}\sqrt{\beta_{ml}}h_{ml}\boldsymbol{n}_{\boldsymbol{m}}^H\right)\right\}$$
$$= \frac{1}{M^2}\sum_{m=1}^{M}\beta_{ml}E\left\{|h_{ml}|^2\right\}E\left\{\boldsymbol{n_m}\boldsymbol{n}_{\boldsymbol{m}}^H\right\} = \frac{\overline{\beta_l}}{M}\boldsymbol{I_N} \quad \text{(A.11)}$$

By inserting the expected values of $|\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc}}(a,n)|^2$ and also the variance of the effective

noise vector into Eq. (A.4), we have:

$$\tilde{R}_{l,p}^{mrc} \approx \log_2 \left(1 + \frac{\rho_d M \overline{\beta_l}^2 g'[0]}{\rho_d \left(g'' \sum_{k=1}^{K} \overline{\beta_{lk}} + (g''-g')\overline{\beta_{ll}} + M\overline{\beta_l}^2 g' \right) - \rho_d M \overline{\beta_l}^2 g'[0] + \overline{\beta_l}} \right)$$

where $g' = \sum_{i=-I}^{I} g'[i]$, $g'' = \sum_{i=-I}^{I} g''[i]$ and I is the number of significant side lobes of the pulse shape.

A.2 Proof of Theorem 2.2

By taking similar steps as Appendix A.1, we can derive the expectations of the elements of the effective matrix presented in Eq. (2.38), as follows:

$$E\left\{ |\boldsymbol{T}_{\boldsymbol{lk,ip}}^{\boldsymbol{mrc}}(a,n)|^2 \right\} = \frac{1}{M} \left[\overline{\beta_{kk}} \left(\gamma''_{lkk}[a-n] - \gamma'_{lkk}[a-n] \right) + M\overline{\beta_k}^2 \gamma'_{lkk}[a-n] + \sum_{j=1}^{K} \overline{\beta_{jk}} \gamma''_{ljk}[a-n] \right]$$

where $\gamma'_{ljk}[i] = E^2_{f(\tau_j,\tau_k)} \{\lambda_{lmj} g(iT - \tau_{mk})\}$ and $\gamma''_{ljk}[i] = E_{f(\tau_j,\tau_k)} \left\{ \lambda^2_{lmj} g^2(iT - \tau_{mk}) \right\}$ are the expectations over the time delay distributions. Assuming the same distribution for all the time delays, the receive antenna index is discarded. The covariance matrix of the effective noise vector presented in Eq. (2.39) can be calculated similarly as:

$$COV\left\{ \boldsymbol{n}_{\boldsymbol{l,ip}}^{\boldsymbol{mrc}} \right\} = \left(\frac{\rho_d}{M\rho_p} g'' \sum_{k=1}^{K} \overline{\beta_k} + \frac{1}{M} \left(\sum_{j=1}^{K} \overline{\beta_j} \lambda''_{lj} + \frac{1}{\rho_p} \right) \right) \boldsymbol{I_N} \tag{A.12}$$

where $\lambda''_{lk} = E_{f(\tau)}\{\lambda^2_{lmk}\}$. By inserting these values into Eq. (A.4), the proof will be complete.

A.3 Details of MRC-ZF Receiver with Imperfect CSI

In this section, we provide some details on MRC-ZF structure, particularly for the imperfect CSI case. As explained before, when the channel coefficients are estimated by using orthogonal pilot sequences, the effective channel matrices for all the K users are nonzero. Therefore, we need at least K sets of samples to cancel them. Denoting $\boldsymbol{y_m^t},\ t = 1, \cdots, K$ as the set of N samples, obtained at sampling times of $e_tT + iT, i = 1, \cdots, N$, we collect all the samples obtained at receive antenna m in a vector $\boldsymbol{y_m^{os}} = [(\boldsymbol{y_m^1})^T, \cdots, (\boldsymbol{y_m^K})^T]^T$ to derive:

$$\boldsymbol{y_m^{os}} = \sqrt{\rho_d}\boldsymbol{T_m}\boldsymbol{s} + \boldsymbol{n_m^{os}} \tag{A.13}$$

where $\boldsymbol{s} = [\boldsymbol{s_1}^T, \cdots, \boldsymbol{s_K}^T]^T$ includes transmitted vectors of all users and $\boldsymbol{T_m}$ is defined as:

$$\boldsymbol{T_m} = \begin{pmatrix} \boldsymbol{T_{m1}^1} & \boldsymbol{T_{m2}^1} & \cdots & \boldsymbol{T_{mk}^1} \\ \boldsymbol{T_{m1}^2} & \boldsymbol{T_{m2}^2} & \cdots & \boldsymbol{T_{mk}^2} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{T_{m1}^K} & \boldsymbol{T_{m2}^K} & \cdots & \boldsymbol{T_{mk}^K} \end{pmatrix} \tag{A.14}$$

where $\boldsymbol{T_{mk}^t}$ represents the channel matrix of User k to receive antenna m in the tth set of samples, i.e., $\boldsymbol{T_{mk}^t} = \sqrt{\beta_{mk}}h_{mk}\boldsymbol{G_{mk}^t}$ where $\boldsymbol{G_{mk}^t}$ is defined by Eq. (2.21) with $e = e_t$. The noise vector also includes all the noise vectors obtained from different sampling times, i.e., $\boldsymbol{n_m^{os}} = \left((\boldsymbol{n_m^1})^T, \cdots, (\boldsymbol{n_m^K})^T\right)^T$ and its covariance matrix is calculated by:

$$\boldsymbol{\Sigma_{n^{os}}} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} & \dots & \boldsymbol{\Sigma_{1K}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} & \dots & \boldsymbol{\Sigma_{2K}} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\Sigma_{K1}} & \boldsymbol{\Sigma_{K2}} & \dots & \boldsymbol{\Sigma_{KK}} \end{pmatrix} \tag{A.15}$$

where $\boldsymbol{\Sigma_{t_1 t_2}}$ is the covariance matrix between the noise samples obtained at times t_1 and t_2, and its elements are defined as $\boldsymbol{\Sigma_{t_1 t_2}}(i,j) = g((i-j)T + (e_{t_1} - e_{t_2})T)$. The receive antenna index is discarded because the noise covariance matrix is the same at all receive antennas. The User l's channel estimation to the mth receive antenna, i.e., $\tilde{c}^s_{ml}$ is equal to:

$$\tilde{c}^s_{ml} = \sum_{j=1}^{K} \lambda^s_{lmj} c_{mj} + \tilde{n}^s_{ml} \tag{A.16}$$

where λ^s_{lmj} is defined the same as Eq. (2.37) with e_s. Because oversampling is only performed in the data detection phase, we use index s for the set of sampling times in the channel estimation phase to differentiate it from the sets of sampling times in the data detection phase. After performing MRC for the lth user, the resulting system of equations is:

$$\boldsymbol{y^{os-mrc}_{l,ip}} = \sqrt{\rho_d}\boldsymbol{\hat{T}_l s} + \boldsymbol{n^{os-mrc}_{l,ip}} \tag{A.17}$$

where $\boldsymbol{\hat{T}_l}$ is the effective channel matrix with constructive blocks of $\boldsymbol{\hat{T}^t_{lk}}$ where each sub-block is defined as:

$$\boldsymbol{\hat{T}^t_{lk}} = \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{j=1}^{K} \lambda^s_{lmj} c^*_{mj} \right) \boldsymbol{T^t_{mk}} \tag{A.18}$$

The effective noise is also calculated as:

$$\boldsymbol{n^{os-mrc}_{l,ip}} = \frac{\sqrt{\rho_d}}{M} \sum_{m=1}^{M} (\tilde{n}^s_{ml})^* \boldsymbol{T_m b} + \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{j=1}^{K} \lambda^s_{lmj} c^*_{mj} + (\tilde{n}^s_{ml})^* \right) \boldsymbol{n^{os}_m} \tag{A.19}$$

The expected value of $\hat{\boldsymbol{T}}_{\boldsymbol{l}}$ is equal to:

$$E\{\hat{\boldsymbol{T}}_{\boldsymbol{l}}\} = \begin{pmatrix} \boldsymbol{\Gamma}_{l1}^{1} & \boldsymbol{\Gamma}_{l2}^{1} & \cdots & \boldsymbol{\Gamma}_{lK}^{1} \\ \boldsymbol{\Gamma}_{l1}^{2} & \boldsymbol{\Gamma}_{l2}^{2} & \cdots & \boldsymbol{\Gamma}_{lK}^{2} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{l1}^{K} & \boldsymbol{\Gamma}_{l2}^{K} & \cdots & \boldsymbol{\Gamma}_{lK}^{K} \end{pmatrix} \begin{pmatrix} \overline{\beta_1}\boldsymbol{I_N} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\beta_2}\boldsymbol{I_N} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \overline{\beta_K}\boldsymbol{I_N} \end{pmatrix}$$
$$= \boldsymbol{\Gamma_l} \begin{pmatrix} \overline{\beta_1}\boldsymbol{I_N} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\beta_2}\boldsymbol{I_N} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \overline{\beta_K}\boldsymbol{I_N} \end{pmatrix} \tag{A.20}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{lk}}^{\boldsymbol{t}}$ is defined the same as Eq. (2.48); however, because of oversampling, there is an extra index of t which represents the sampling origin index, i.e., t in e_t. Matrix $\boldsymbol{\Gamma_l}$ in Eq. (A.20) is only related to sampling origins, i.e., e_ts, pilot sequences, the pulse shape and delay distributions and is known at the receiver. To resolve the problem of ISI and IUI, we calculate the inverse of $\boldsymbol{\Gamma_l}$ and denote it as $\boldsymbol{W_l}$, which is constructed by sub-blocks of $\boldsymbol{W_{lk}}$, i.e., $\boldsymbol{W_l} = (\boldsymbol{W_{l1}}^T, \cdots, \boldsymbol{W_{lK}}^T)^T$. Then, in order to detect the transmitted symbols of the lth user, we multiply the output of the MRC receiver by the lth sub-block of $\boldsymbol{W_l}$, i.e., $\boldsymbol{W_{ll}}$.

A.4 Proof of Theorem 2.3

The expected values of $|\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc-zf}}(a,n)|^2$ defined in Eq. (2.49) can be calculated similar to Appendix A.1:

$$E\left\{|\boldsymbol{T}_{\boldsymbol{lk,p}}^{\boldsymbol{mrc-zf}}(a,n)|^2\right\} = \begin{cases} \frac{1}{M}\left(\overline{\beta_{ll}}(2\hat{g}''[a-n] - \delta[a-n]) + M\overline{\beta_l}^2\delta[a-n]\right) & k = l \\ \frac{\overline{\beta_{lk}}}{M}\hat{g}''[a-n] & k \neq l \end{cases}$$

where $\hat{g}''[a-n] = E_{f(\tau)}\left\{(\hat{\boldsymbol{G}}(a,n))^2\right\}$ with respect to the distribution of time delays and $\hat{\boldsymbol{G}} = \boldsymbol{ZG}$. Similarly, $\hat{g}''$ is defined as the summation of nonzero elements of $\hat{g}''[i]$.

Covariance of the effective noise vector in Eq. (2.49) is also calculated as:

$$COV\{\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}}\} = E\left\{\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}}\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc-zf}^H}\right\} \tag{A.21}$$

$$= E\left\{\boldsymbol{Z}^{-1}\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}}\boldsymbol{n}_{\boldsymbol{l,p}}^{\boldsymbol{mrc}H}\boldsymbol{Z}^{-H}\right\} \tag{A.22}$$

$$= \frac{\overline{\beta_l}}{M}\left(\boldsymbol{Z}^H\boldsymbol{Z}\right)^{-1} \tag{A.23}$$

By inserting these values into Eq. (A.4), we can conclude the proof.

A.5 Proof of Theorem 2.4

Similar to Appendix A.1, the expected values of $|\boldsymbol{T}_{\boldsymbol{lk,ip}}^{\boldsymbol{mrc-zf}}(a,n)|^2$ and the covariance matrix of the effective noise vector defined in Eq. (2.52) can be calculated as:

$$E\left\{|\boldsymbol{T}_{\boldsymbol{lk,ip}}^{\boldsymbol{mrc-zf}}(a,n)|^2\right\} = \frac{1}{M}\left\{\overline{\beta_{kk}}\left(\hat{\gamma}''_{lkk}[a-n]-\delta[a-n]\delta[l-k]\right)+M\overline{\beta_k}^2\delta[a-n]\delta[l-k]+\sum_{j=1}^{K}\overline{\beta_{jk}}\hat{\gamma}''_{ljk}[a-n]\right\}$$

$$COV\{\boldsymbol{n}_{\boldsymbol{l,ip}}^{\boldsymbol{mrc-zf}}\} = \frac{\rho_d}{M\rho_p}\sum_{k=1}^{K}\overline{\beta_k}\boldsymbol{U_l}+\frac{1}{M}\left(\sum_{j=1}^{K}\overline{\beta_j}\lambda''_{lj}+\frac{1}{\rho_p}\right)\boldsymbol{V_l} \tag{A.24}$$

where

$$\hat{\gamma}''_{ljk}[a-n] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\lambda_{lj}^{s}\hat{\boldsymbol{G}}_{\boldsymbol{lk}}(a,n)\right)^2 f(\tau_j)f(\tau_k)d\tau_j d\tau_k$$

$$\lambda''_{lk} = \int_{-\infty}^{\infty}\left(\lambda_{lk}^{s}\right)^2 f(\tau_k)d\tau_k$$

where $\hat{\boldsymbol{G}}_{\boldsymbol{lk}} = \boldsymbol{W}_{ll}\hat{\boldsymbol{G}}_{\boldsymbol{k}}$ and $\hat{\boldsymbol{G}}_{\boldsymbol{k}} = ((\boldsymbol{G}_{\boldsymbol{k}}^{\boldsymbol{1}})^T,\cdots,(\boldsymbol{G}_{\boldsymbol{k}}^{\boldsymbol{K}})^T)^T$. Assuming the same distribution for all the time delays, the receive antenna index is discarded. Similarly, $\hat{\gamma}''_{ljk}$ is defined as the summation of nonzero elements of $\hat{\gamma}''_{ljk}[i]$. Also $\boldsymbol{U_l} = \boldsymbol{W_{ll}}E_{f(\tau)}\{\hat{\boldsymbol{G}}_{\boldsymbol{k}}\hat{\boldsymbol{G}}_{\boldsymbol{k}}^{H}\}\boldsymbol{W}_{\boldsymbol{ll}}^{H}$ and

$\boldsymbol{V_l} = \boldsymbol{W_{ll}}\boldsymbol{\Sigma_{n^{os}}}\boldsymbol{W_{ll}^H}$ whose diagonal elements are denoted as u_{l0} and v_{l0}, respectively. Inserting these values into Eq. (A.4), will conclude the proof.

Appendix B

Supplementary Proofs for Chapter 4

B.1 Derivation of the Upper-bound for the Bit Error Probability (BEP) Expression

The post SNR of the ith sub-channel at the receiver can be expressed as:

$$\delta_i = \frac{\delta_0 \sum_{m=1}^{M} |h_{\pi(i),m}|^2}{\boldsymbol{R'}^{-1}(i,i)} \tag{B.1}$$

where $\delta_0 = \frac{E[|s_k[i]|^2]}{\sigma_n^2}$. The function $\pi(.)$ is the assigning function that assign each sub-channel to a specific user, i.e., $\pi : \mathcal{I} \rightarrow \mathcal{K}$, where $\mathcal{I} = \{1, \cdots, NK\}$ and $\mathcal{K} = \{1, \cdots, K\}$ are the set of sub-channel and user indices, respectively. Also, $h_{k,m}$ represents the channel coefficient between User k and Receive Antenna m. We know that $|h_{k,m}|^2$ follows a chi-squared distribution with two degrees of freedom for all is and js. Therefore, $\sum_{m=1}^{M} |h_{k,m}|^2$ is chi-squared distributed with $2M$ degrees of freedom. As a result, the distribution of δ_i

can be calculated as follows:

$$P_{\delta_i}(\delta) = \frac{\boldsymbol{R'}^{-1}(i,i)}{\delta_0} \frac{\left(\frac{\boldsymbol{R'}^{-1}(i,i)}{\delta_0}\delta\right)^{M-1} \exp\left(-\frac{\boldsymbol{R'}^{-1}(i,i)}{2\delta_0}\delta\right)}{2^M \Gamma(M)} \tag{B.2}$$

where $\Gamma(.)$ is the Gamma function. For a specific value of SNR, BER varies according to the modulation. We assume that BPSK is used, however, extension to other modulations is straightforward. Based on this assumption, the BER for a given value of SNR, e.g., δ is equal to $Q(\sqrt{2\delta})$. The next step is to calculate the following integral:

$$BEP_{u,i} = \int_0^\infty Q(\sqrt{2\delta}) P_{\delta_i}(\delta) d\delta$$

The integral of $\frac{a^m}{\Gamma(m)} \int_0^\infty \exp(-az) z^{m-1} Q(\sqrt{bz}) dz$ has a closed-form of:

$$\frac{\sqrt{b/2\pi a}}{2\left(1+\frac{b}{2a}\right)^{m+1/2}} \frac{\Gamma(m+1/2)}{\Gamma(m+1)} {}_2F_1(1, m+\frac{1}{2}; m+1; \frac{1}{1+\frac{b}{2a}})$$

where ${}_2F_1(q, w; e; r)$ is the hypergeometric function [156]. Therefore, the upper-bound for the actual BEP is equal to:

$$BEP_{u,i} = \frac{\sqrt{\frac{\delta_0 2}{\pi \boldsymbol{R'}^{-1}(i,i)}}}{2\left(1+\frac{\delta_0 2}{\boldsymbol{R'}^{-1}(i,i)}\right)^{M+\frac{1}{2}}} \times \frac{\Gamma(M+\frac{1}{2})}{\Gamma(M+1)} \times {}_2F_1(1, M+\frac{1}{2}; M+1; \frac{1}{1+\frac{\delta_0 2}{\boldsymbol{R'}^{-1}(i,i)}}) \tag{B.3}$$

B.2 Average BEP and Its Approximation at High SNR

In Eq. (B.3), BEP_i depends on $\boldsymbol{R'}^{-1}(i,i)$ which varies for different values of i, and therefore each sub-channel has a different BEP. This is unlike the synchronous ZF, where all

resulting sub-channels have the same performance. In order to evaluate the performance of the entire system, we define the average BER performance as follows:

$$BEP_{avg} = \frac{\sum_{i=1}^{NK} BEP_i}{NK} \tag{B.4}$$

Since BEP_{avg} is not tractable, we approximate it at high SNR, using the fact that ${}_2F_1(1, m+\frac{1}{2}; m+1; \frac{1}{1+c})$ converges to one as c grows large [64]. Hence, at high SNR, BEP_{avg} can be approximated as follows:

$$\widetilde{BEP}_{avg} = \text{Const } \times \frac{\sum_{i=1}^{NK} (\boldsymbol{R'}^{-1}(i,i))^M}{\delta_0^M} \tag{B.5}$$

where the constant value is equal to $\frac{1}{2^{(M+1)}NK\sqrt{\pi}} \frac{\Gamma(M+\frac{1}{2})}{\Gamma(M+1)}$.

B.3 Proof of Lemma 4.2

When the frame length is N, we denote $\boldsymbol{R'}$ by $\boldsymbol{R^N}$. Note that the prime sign is discarded in this section for notational simplicity. However, the same sampling structure as in Eq. (4.5) is used. Then, we prove by induction that, for all $N \in \mathbb{Z}_+$,

$$\text{trace}((\boldsymbol{R^N})^{-1}) = \frac{(N-1)(N+1)}{3(1+\tau_1-\tau_K)} + \frac{2N+1}{3(N+1+\tau_1-\tau_K)} + \frac{N(N+2)}{3} \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1}-\tau_i}$$

Base case: When $N = 1$, $\boldsymbol{R^1}$ is equal to $\boldsymbol{R_0}$ which can be written as a generalized

Fiedler's matrix whose inverse is given by [157]:

$$\boldsymbol{R_0}^{-1} = -\frac{1}{2}\begin{pmatrix} d_1 & \frac{1}{\tau_2-\tau_1} & ... & 0 & f \\ \frac{1}{\tau_2-\tau_1} & d_2 & \frac{1}{\tau_3-\tau_2} & ... & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & ... & \frac{1}{\tau_{K-1}-\tau_{K-2}} & d_{K-1} & \frac{1}{\tau_K-\tau_{K-1}} \\ f & 0 & ... & \frac{1}{\tau_K-\tau_{K-1}} & d_K \end{pmatrix}$$

where f and d_is are defined as:

$$f = \frac{1}{\tau_K - \tau_1 - 2} \tag{B.6}$$

$$d_1 = \frac{1}{\tau_1 - \tau_2} - \frac{1}{\tau_1 - \tau_K + 2} \tag{B.7}$$

$$d_K = \frac{1}{\tau_{K-1} - \tau_K} - \frac{1}{\tau_1 - \tau_K + 2} \tag{B.8}$$

$$d_i = \frac{1}{\tau_{i-1} - \tau_i} + \frac{1}{\tau_i - \tau_{i+1}} \quad 2 \leq i \leq K-1 \tag{B.9}$$

Then, trace($\boldsymbol{R_0^{-1}}$) is equal to $\left(-\frac{1}{2}\sum_{i=1}^{K} d_i\right)$, which can be calculated using the above equations:

$$\text{trace}(\boldsymbol{R_0^{-1}}) = \frac{1}{(2 + \tau_1 - \tau_K)} + \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1} - \tau_i} \tag{B.10}$$

Therefore, Eq. (B.6) is true for $N = 1$.

Induction step: Suppose Eq. (B.6) is true for N. We need to show that it also holds for $N + 1$, i.e.,

$$\text{trace}((\boldsymbol{R^{(N+1)}})^{-1}) = \frac{(N)(N+2)}{3(1 + \tau_1 - \tau_K)} + \frac{2N+3}{3(N + 2 + \tau_1 - \tau_K)} + \frac{(N+1)(N+3)}{3} \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1} - \tau_i}$$

Because matrix $\boldsymbol{R}$ follows a recursive structure, $\boldsymbol{R^{N+1}}$can be presented as follows:

$$\boldsymbol{R^{N+1}} = \begin{bmatrix} (\boldsymbol{R^N})_{NK \times NK} & (\boldsymbol{L})_{NK \times K} \\ (\boldsymbol{L}^T)_{K \times NK} & (\boldsymbol{R_0})_{K \times K} \end{bmatrix} \tag{B.11}$$

where $\boldsymbol{L}^T = [\mathbf{0}_K, \ldots, \mathbf{0}_K, (\boldsymbol{R_1})_K]$. For calculating the inverse of $\boldsymbol{R^{N+1}}$, we use the following lemma for matrix inversion in block form.

Lemma B.1. *Let na $(m+n) \times (m+n)$ matrix $\boldsymbol{T}$ be partitioned into a block form:*

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$$

where the $m \times m$ matrix $\boldsymbol{A}$ and $n \times n$ matrix $\boldsymbol{D}$ are invertible. Then, we have:

$$\boldsymbol{T}^{-1} = \begin{bmatrix} \boldsymbol{M}^{-1} & -\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{M}^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{bmatrix}$$

where $\boldsymbol{M} = \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$ [158].

Here, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are equal to $\boldsymbol{R^N}$, $\boldsymbol{L}$, $\boldsymbol{L}^T$ and $\boldsymbol{R_0}$, respectively. Therefore, $\boldsymbol{M}$ is equal to:

$$\boldsymbol{M} = \boldsymbol{R^N} - \boldsymbol{L}(\boldsymbol{R_0})^{-1}\boldsymbol{L}^T \tag{B.12}$$

Now, we need to find the inverse of $\boldsymbol{M}$. By defining $\boldsymbol{Z}$ as $(\boldsymbol{R^N})^{-1}$, the inverse of $\boldsymbol{M}$ can be presented as:

$$\boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{I}_K & \ldots & \mathbf{0}_K & \boldsymbol{Z_{1N}}\boldsymbol{Q}(\boldsymbol{I}_K - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1} \\ \mathbf{0}_K & \ddots & \vdots & \vdots \\ \vdots & \vdots & \boldsymbol{I}_K & \boldsymbol{Z_{(N-1)N}}\boldsymbol{Q}(\boldsymbol{I}_K - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1} \\ \mathbf{0}_K & \ldots & \mathbf{0}_K & (\boldsymbol{I}_K - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1} \end{bmatrix} \boldsymbol{Z}$$

where $\boldsymbol{Q} = \boldsymbol{R_{-1}}{\boldsymbol{R_0}}^{-1}\boldsymbol{R_1}$ and $\boldsymbol{Z_{ij}}$s are $K \times K$ partitioning blocks of $\boldsymbol{Z}$. Also, $\boldsymbol{I}_k$ and $\mathbf{0}_k$ are a $k \times k$ identity and zero matrices , respectively.

To show the correctness of Eq. (B.13), we need to take the following steps:

Step 1: By some calculations, it can be shown that $\boldsymbol{L}(\boldsymbol{R_0})^{-1}\boldsymbol{L}^T$ is equal to $\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q} \end{bmatrix}$. As a result, we have:

$$\boldsymbol{M} = \boldsymbol{R^N} - \begin{bmatrix} \mathbf{0}_{(N-1)K\times(N-1)K} & \mathbf{0}_{(N-1)K\times K} \\ \mathbf{0}_{K\times(N-1)K} & \boldsymbol{Q} \end{bmatrix} \tag{B.13}$$

Step 2: If we multiply both sides by $\boldsymbol{Z}$, we will have:

$$\boldsymbol{ZM} = \boldsymbol{I}_{NK} - \begin{bmatrix} \mathbf{0}_K & \dots & \mathbf{0}_K & \boldsymbol{Z_{1N}Q} \\ \mathbf{0}_K & \dots & \mathbf{0}_K & \boldsymbol{Z_{2N}Q} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_K & \dots & \mathbf{0}_K & \boldsymbol{Z_{NN}Q} \end{bmatrix} \tag{B.14}$$

Step 3: We denote the right hand side of Eq. (B.14) by $\boldsymbol{X}$, then, we can conclude that the inverse of $\boldsymbol{M}$ is equal to:

$$\boldsymbol{M}^{-1} = \boldsymbol{X}^{-1}\boldsymbol{Z} \tag{B.15}$$

Step 4: $\boldsymbol{X}^{-1}$ can be calculated as follows:

$$\boldsymbol{X}^{-1} = \begin{bmatrix} \boldsymbol{I}_K & \dots & \mathbf{0}_K & \boldsymbol{Z_{1N}Q}(\boldsymbol{I}_K - \boldsymbol{Z_{NN}Q})^{-1} \\ \mathbf{0}_K & \ddots & \vdots & \vdots \\ \vdots & \vdots & \boldsymbol{I}_K & \boldsymbol{Z_{(N-1)N}Q}(\boldsymbol{I}_K - \boldsymbol{Z_{NN}Q})^{-1} \\ \mathbf{0}_K & \dots & \mathbf{0}_K & (\boldsymbol{I}_K - \boldsymbol{Z_{NN}Q})^{-1} \end{bmatrix} \tag{B.16}$$

Step 5: Finally, if we plug $\boldsymbol{X}^{-1}$ in Eq. (B.15), we will reach Eq. (B.13).

If we denote $K \times K$ diagonal blocks of $\boldsymbol{M}^{-1}$ as $[\boldsymbol{M}^{-1}]_{i,i}$ $1 \leq i \leq N$, then, by use of Lemma B.1, trace$((\boldsymbol{R^{N+1}})^{-1})$ can be written as:

$$\text{trace}((\boldsymbol{R^{N+1}})^{-1}) = \sum_{i=1}^{N} \text{trace}([\boldsymbol{M}^{-1}]_{i,i}) + \text{trace}(\boldsymbol{R_0}^{-1} + \boldsymbol{R_0}^{-1}\boldsymbol{R_1}[\boldsymbol{M}^{-1}]_{N,N}\boldsymbol{R_{-1}}\boldsymbol{R_0}^{-1}) \tag{B.17}$$

By simplifying Eq. (B.13), diagonal blocks of $\boldsymbol{M}^{-1}$ can be presented as follows:

$$1 \leq i \leq N-1 :$$

$$[\boldsymbol{M}^{-1}]_{i,i} = \boldsymbol{Z_{ii}} + \boldsymbol{Z_{iN}}\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1}\boldsymbol{Z_{Ni}} \tag{B.18}$$

$$i = N :$$

$$[\boldsymbol{M}^{-1}]_{i,i} = (\boldsymbol{I} - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1}\boldsymbol{Z_{NN}} \tag{B.19}$$

In Eq. (B.17), we set the diagonal blocks of $\boldsymbol{M}^{-1}$ as Eqs. (B.18) and (B.19). Then, by some manipulations, trace$((\boldsymbol{R^{N+1}})^{-1})$ can be presented as:

$$\begin{aligned}\text{trace}((\boldsymbol{R^{N+1}})^{-1}) = \text{trace}((\boldsymbol{R^{N}})^{-1}) + \text{trace}(\boldsymbol{R_0}^{-1}) \\ + \sum_{i=1}^{N-1} \text{trace}(\boldsymbol{Z_{iN}}\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1}\boldsymbol{Z_{Ni}}) \qquad \text{(B.20)} \\ + \text{trace}((\boldsymbol{I} - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1}\boldsymbol{Z_{NN}}) - \text{trace}(\boldsymbol{Z_{NN}}) \\ + \text{trace}(\boldsymbol{R_0}^{-1}\boldsymbol{R_1}(\boldsymbol{I} - \boldsymbol{Z_{NN}}\boldsymbol{Q})^{-1}\boldsymbol{Z_{NN}}\boldsymbol{R_{-1}}\boldsymbol{R_0}^{-1}) \qquad \text{(B.21)}\end{aligned}$$

The first and second terms in Eq. (B.21) can be calculated by induction hypothesis and induction base, respectively. Calculating other terms in Eq. (B.21) is tedious but similar for different values of K. Therefore, we only calculate it for $K = 2$ and skip the rest. For

$K = 2$, $\boldsymbol{Q}$ is equal to:

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1-\tau}{1+\tau} \end{bmatrix} \tag{B.22}$$

where $\tau = \tau_2 - \tau_1$.

If we plug $\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1-\tau}{1+\tau} \end{bmatrix}$ in Eq. (B.21), after some calculations we will have:

$$\begin{aligned} \text{trace}((\boldsymbol{R^{N+1}})^{-1}) = \text{trace}((\boldsymbol{R^N})^{-1}) &+ \frac{2}{1-(1-\tau)^2} \\ &+ \frac{1-\tau}{(1+\tau)-(1-\tau)r(2N,2N)} \sum_{i=1}^{2N} (r(2N,i))^2 \\ &+ \frac{(1+\tau)(1+(\tau-1)^2)}{(2-\tau)^2[(1+\tau)-(1-\tau)r(2N,2N)]} r(2N,2N) \end{aligned} \tag{B.23}$$

where $r(i,j)$ is the (i,j)th element of matrix $(\boldsymbol{R^N})^{-1}$. By induction hypothesis, the first term in Eq. (B.23) is equal to $\frac{(N-1)(N+1)}{3(1-\tau_2+\tau_1)} + \frac{N(N+2)}{3(\tau_2-\tau_1)} + \frac{2N+1}{3(N+1-\tau_2+\tau_1)}$. For calculating Eq. (B.23), we also need values of $r(2N,i)$, $1 \le i \le 2N$, which are elements of the last row of $(\boldsymbol{R^N})^{-1}$. Due to the special structure of matrix $\boldsymbol{R}$, values of $r(2N,i)$ can be calculated as follows:

$$\begin{cases} r(2N,2i-1) = \frac{\tau-i}{\tau(N+1-\tau)} \\ r(2N,2i) = \frac{i}{\tau(N+1-\tau)} \end{cases} 1 \le i \le N \tag{B.24}$$

To verify Eq. (B.24), we can multiply the last row of $(\boldsymbol{R^N})^{-1}$, i.e., $[r(2N,1), r(2N,2), \ldots, r(2N,2N)]$, by different columns of $\boldsymbol{R^N}$ as follows:

$$\text{1st column:} \quad \frac{(\tau-1)1}{\tau(N+1-\tau)} + \frac{(1)(1-\tau)}{\tau(N+1-\tau)} = 0$$

$$(2i)_{th} \text{ column:} \quad 1 \le i \le N-1$$

$$\frac{(\tau-i)(1-\tau)}{\tau(N+1-\tau)}+\frac{(i)1}{\tau(N+1-\tau)}+\frac{(\tau-(i+1))\tau}{\tau(N+1-\tau)}=0$$

$(2i-1)_{th}$ column: $\quad 2\leq i\leq N$

$$\frac{((i-1))(\tau)}{\tau(N+1-\tau)}+\frac{(\tau-i)1}{\tau(N+1-\tau)}+\frac{(i)(1-\tau)}{\tau(N+1-\tau)}=0$$

$2N$th column: $\quad \frac{(\tau-N)(1-\tau)}{\tau(N+1-\tau)}+\frac{(N)1}{\tau(N+1-\tau)}=1$

These results verify that the last row of $(\boldsymbol{R^N})^{-1}$ follows the pattern in Eq. (B.24).

The last step is to plug Eq. (B.24) into Eq. (B.23). As a result, trace$((\boldsymbol{R^{N+1}})^{-1})$ is equal to $\frac{(N)(N+2)}{3(1-\tau_2+\tau_1)}+\frac{(N+1)(N+3)}{3(\tau_2-\tau_1)}+\frac{2N+3}{3(N+2-\tau_2+\tau_1)}$, which verifies the induction step and completes the proof.

Appendix C

Supplementary Proofs for Chapter 5

C.1 Proof of Lemma 5.1

Denoting g(t) as a pulse shape with real spectrum, we show that $IUI(\tau)$ calculated as:

$$IUI(\tau) = \sum_{i=-\infty}^{\infty} |g(\tau T + iT)|^2 \tag{C.1}$$

is maximized at $\tau = 0$. In other words, $IUI(\tau) < IUI(0), \quad \tau \neq 0$. Assume that $\hat{g}(f)$ is the Fourier transform of the pulse shape $g(t)$. Then, the Fourier transform of the shifted version of $g(t)$, i.e., $g(\tau T + t)$, will be $\hat{g}(f)e^{j2\pi f\tau}$. The DTFT of the samples of $g(\tau T + t)$, i.e., $g(\tau T + iT), i \in Z$ can be expressed as:

$$G_\tau(f) = \frac{1}{T} \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}(\frac{f+i}{T}) \tag{C.2}$$

Note that $G_\tau(f)$ is periodic with period of 1. Based on the Parseval's theorem, the IUI energy, i.e., $IUI(\tau) = \sum_{i=-\infty}^{\infty} |g(\tau T + iT)|^2$, will be equal to:

$$IUI(\tau) = \int_{-1/2}^{1/2} |G_\tau(f)|^2 df \tag{C.3}$$

Then, with the assumption of having real spectrum, we will have:

$$\begin{aligned} IUI(\tau) &= \frac{1}{T} \int_{-1/2}^{1/2} \left| \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}\left(\frac{f+i}{T}\right) \right|^2 df \\ &\leq \frac{1}{T} \int_{-1/2}^{1/2} \left| \sum_{i=-\infty}^{\infty} \hat{g}\left(\frac{f+i}{T}\right) \right|^2 df = IUI(0) \end{aligned} \tag{C.4}$$

which concludes the proof.

C.2 Proof of Lemma 5.2

The power of the asynchronous signal can be written as:

$$\begin{aligned} P_{asynch} &= E\left\{ \int_{-\infty}^{\infty} s_{asynch}(t) s_{asynch}^*(t) dt \right\} \\ &= E\left\{ \int_{-\infty}^{\infty} \left(\sum_{k=1}^{K} s_k(t - \tau_k T) \right) \left(\sum_{k=1}^{K} s_k(t - \tau_k T) \right)^* dt \right\} \end{aligned} \tag{C.5}$$

Then, we have

$$\begin{aligned} P_{asynch} &= E\left\{ \int_{-\infty}^{\infty} \left(\sum_{k=1}^{K} \sum_{n=1}^{N} s_k[n] p(t - nT - \tau_k T) \right) \left(\sum_{k=1}^{K} \sum_{n=1}^{N} s_k[n] p(t - nT - \tau_k T) \right)^* dt \right\} \\ &= \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} E\left[s_{k_1}[n_1] s_{k_2}[n_2] \right] \int_{-\infty}^{\infty} p(t - n_1 T - \tau_{k_1} T) p(t - n_2 T - \tau_{k_2} T) dt \\ &= \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} trace(\boldsymbol{R}_{\boldsymbol{k_1 k_2}} COV\{\boldsymbol{s_{k_1}}, \boldsymbol{s_{k_2}}\}) = trace(\boldsymbol{R}\ COV\{\boldsymbol{s}\}) \end{aligned} \tag{C.6}$$

which concludes the proof.

C.3 Proof of Lemma 5.3

The achievable rate for each user is the sum of the achievable rates for the corresponding sub-channels:

$$R_r = \frac{1}{2N} \sum_{i \in \mathcal{I}_r} log_2 \left(1 + \frac{p_i \lambda_i |h_r|^2}{\sigma_n^2} \right)$$

where $\mathcal{I}_r$ represents the set of sub-channels indices that are assigned to User r. with the power constraint of:

$$\sum_{i=1}^{N} p_i \lambda_i \leq NP \tag{C.7}$$

Denoting P_r as $\sum_{i \in \mathcal{I}_r} p_i \lambda_i / N$, then the achievable rates can be re-written as:

$$R_r = \frac{1}{2N} \sum_{i \in \mathcal{I}_r} log_2 \left(1 + \frac{p_i \lambda_i |h_r|^2}{\sigma_n^2} \right)$$

$$s.t. \quad \sum_{i \in \mathcal{I}_k} p_i \lambda_i / N = P_r \tag{C.8}$$

$$\sum_{k=1}^{K} P_k \leq P \tag{C.9}$$

It can be easily shown that the power assignment that maximizes R_r is such that:

$$p_i \lambda_i = P_r \quad \forall i \in \mathcal{I}_r \tag{C.10}$$

Therefore, by simple substitution, we can conclude that the achievable rate for each user is:

$$R_r = \frac{1}{2} log_2 \left(1 + \frac{P_r |h_r|^2}{\sigma_n^2}\right)$$

such that $\sum_{r=1}^{K} P_r \leq P$. Note that the different assignments of sub-channels to users only change the power assignment, otherwise, the final result remains the same.

Appendix D

Supplementary Proofs for Chapter 7

D.1 Proof of Lemma 7.1

Assuming a rectangular pulse shape, due to its finite time support, it can be easily shown that $\eta_\tau^{rect.} = \sum_{i=-\infty}^{\infty} g^2(\tau T + iT) = \tau^2 + (1-\tau)^2$. Therefore, we focus on deriving the results for the r.r.c. pulse shape which has infinite time support. Denoting $g(t)$ as a raised cosine pulse shape with symbol period of T and the roll-off factor of β, the goal is to find $\eta_\tau = \sum_{i=-u}^{u} g^2(\tau T + iT)$. Assuming truncation of the raised cosine pulse shape with large number of side lobes, we can approximate $\eta_\tau \approx \eta_\tau^{\infty} = \sum_{i=-\infty}^{\infty} g^2(\tau T + iT)$. The frequency spectrum of the discrete sequence of $\{g(\tau T + iT)\}_{i=-\infty}^{\infty}$ can be denoted as:

$$G_\tau(f) = \frac{1}{T} \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}\left(\frac{f+i}{T}\right), \quad f \in [-1/2, 1/2], \tag{D.1}$$

where $\hat{g}(f)$ is the Fourier transform of the raised cosine pulse shape $g(t)$ and is denoted as:

$$\hat{g}(f) = \begin{cases} T & |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2}\left[1+\cos\left(\frac{\pi T}{\beta}\left(|f|-\frac{1-\beta}{2T}\right)\right)\right] & \frac{1-\beta}{2T} < |f| \leq \frac{1+\beta}{2T} \\ 0 & o.w. \end{cases}. \tag{D.2}$$

The spectrum function $G_\tau(f)$ is periodic with period of 1 and based on the Parseval's theorem, we have $\eta_\tau^\infty = \int_{-1/2}^{1/2} |G_\tau(f)|^2 df$. Based on the definition of $\hat{g}(f)$, the spectrum function $G_\tau(f)$ for $f \in [-1/2, 1/2]$ can be derived as:

$$G_\tau(f) = \begin{cases} e^{-j2\pi\tau f}A(-f) + e^{-j2\pi\tau(f+1)}A(f+1) & \frac{-1}{2} < f \leq \frac{-1+\beta}{2} \\ e^{-j2\pi\tau f} & |f| \leq \frac{1-\beta}{2} \\ e^{-j2\pi\tau f}A(f) + e^{-j2\pi\tau(f-1)}A(-f+1) & \frac{1-\beta}{2} < f \leq \frac{1}{2} \end{cases}, \tag{D.3}$$

where $A(f) = \frac{1}{2}\left[1+\cos\left(\frac{\pi T}{\beta}\left(\frac{f}{T}-\frac{1-\beta}{2T}\right)\right)\right]$. Thus, η_τ^∞ can be calculated as follows:

$$\eta_\tau^\infty = \underbrace{\int_{\frac{-1}{2}}^{\frac{-1+\beta}{2}} \left|A(-f) + e^{-j2\pi\tau}A(f+1)\right|^2 df}_{b} + \underbrace{\int_{\frac{1-\beta}{2}}^{\frac{1}{2}} \left|A(f) + e^{-j2\pi\tau}A(-f+1)\right|^2 df}_{a} + (1-\beta). \tag{D.4}$$

In general, for the pulse shapes that satisfy the Nyquist no-ISI condition, the frequency-shifted replicas of the spectrum add up to a constant value, here, e.g., $A(f)+A(-f+1) = 1,\ f \in [\frac{1-\beta}{2}, \frac{1}{2}]$. Thus, putting $\tau = 0$ yields $\eta_0^\infty = 1$, which is in fact the maximum value of η_τ^∞. However, with asynchronous transmission, the effective frequency-selective channel attains less power which is the result of the out-of-phase addition of folded spectrum. In more details, for the non-zero values of timing offset, the phase rotation of the frequency

replicas, due to the timing offset, results in the out-of-phase addition of replicas and hence, $\eta_\tau^\infty < \eta_0^\infty, \tau \in (0,1)$, as shown in Fig. D.1.

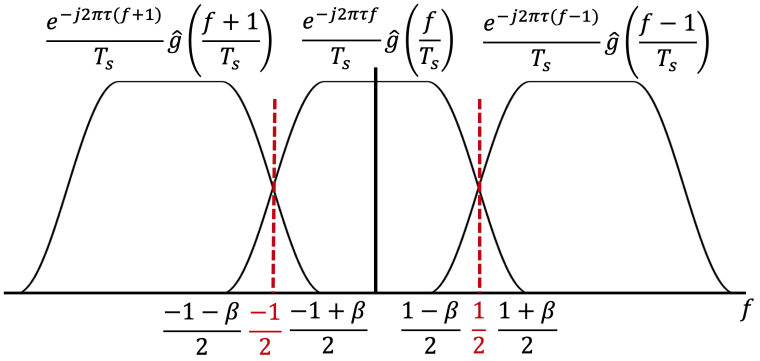


Figure D.1: Schematic illustration of the folded spectrum with phase rotation of the frequency replicas.

To further simplify η_τ^∞, note that $a = b$, and we calculate one of them as:

$$a = \int_{\frac{1-\beta}{2}}^{\frac{1}{2}} A^2(f) + A^2(-f+1) + 2A(f)A(-f+1)\cos(2\pi\tau)df. \tag{D.5}$$

Because the function $A(f)$ is always positive, thus, $\tau = 0$ maximizes a as explained before, and $\tau = 1/2$ minimizes a. By some calculations, we can show that $a = 3\beta/8 + \beta\cos(2\pi\tau)/8$. As a result, η_τ^∞ can be calculated as $\eta_\tau^\infty = 1 - \beta/4 + \beta\cos(2\pi\tau)/4$ which concludes the proof.

D.2 Summary of the Optimal Algorithm for Downlink Beamforming

Denoting $\boldsymbol{w_k} = \sqrt{\rho_k}\boldsymbol{u_k}$ where $||\boldsymbol{u_k}||^2 = 1$, the power optimization problem can be written as:

$$\min_{\{\rho_k, \boldsymbol{u_k}\}_{k=1}^K} p_{avg} = \sum_{k=1}^{K} \rho_k \tag{D.6}$$
$$s.t.\ \frac{\rho_k \boldsymbol{u_k}^H \boldsymbol{h_k} \boldsymbol{h_k}^H \boldsymbol{u_k}}{\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} \rho_l \boldsymbol{u_l}^H \boldsymbol{h_k} \boldsymbol{h_k}^H \boldsymbol{u_l} + \sigma_k^2} \geq \gamma_k^*,\ \ k = 1, \cdots, K.$$

By exploiting the virtual uplink duality, the power minimization can be equivalently stated as [111]:

$$\min_{\{\rho_k, \boldsymbol{u_k}\}_{k=1}^K} p_{avg} = \sum_{k=1}^{K} \rho_k \tag{D.7}$$
$$s.t.\ \frac{\rho_k \boldsymbol{u_k}^H \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H \boldsymbol{u_k}}{\boldsymbol{u_l}^H \left(\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} \rho_l \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M} \right) \boldsymbol{u_l}} \geq \gamma_k^*,\ \ k = 1, \cdots, K.$$

where $\tilde{\boldsymbol{h}}_{\boldsymbol{k}} = \boldsymbol{h_k}/\sigma_k$. Then, the beamforming direction $\boldsymbol{u_k}$ and beamforming amplitude ρ_k can be recursively updated to find the optimal answer, as follows:

- Update beamforming direction:
  $\boldsymbol{u_k}(t+1) = \left(\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} \rho_l(t) \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M} \right)^{-1} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}$.
- Update beamforming amplitude:
  $\rho_k(t+1) = \frac{\gamma_k^*}{\mu_k(t)} \rho_k(t)$ where $\mu_k(t) = \frac{\rho_k(t) \boldsymbol{u}^H(t)_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H \boldsymbol{u_k}(t)}{\boldsymbol{u_l}^H(t) \left(\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} \rho_l(t) \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M} \right) \boldsymbol{u_l}(t)}$.

The rate constraint inequalities are active at the optimal point, thus, to satisfy the rate constraint inequalities, the power control procedure is adopted as follows $\boldsymbol{\rho} = \boldsymbol{F}^{-1} \boldsymbol{\gamma^*}$

where $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_K)$, $\boldsymbol{\gamma^*} = (\gamma_1^*, \cdots, \gamma_K^*)$ and $[\boldsymbol{F}]_{i,j} = \begin{cases} \boldsymbol{u_i^H \tilde{h}_i \tilde{h}_i^H u_i} & i = j \\ -\eta_{\tau_{ij}} \gamma_i^* \boldsymbol{u_j^H h_i h_i^H u_j} & i \neq j \end{cases}$.

D.3 Proof of Proposition 7.2

To show the superiority of Method B compared with Method A, we simply assume uniform power distribution function in Method B which simplifies the achievable rate expression for User k to $r_k^B = \int_0^1 \log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^K \lambda_{kl}(f)|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right) df$. Then, to prove that the asynchronous method with individual precoding can reduce the power transmission, we show that for the set of optimal beamforming vectors obtained for Method A, we have $r_k^B > r_k^A$. Then, the beamforming vectors' amplitudes can be accordingly reduced which results in power reduction. To show that, we use the Jensen's inequality for the convex function of $\log_2(1+1/x)$, which results in $\sum_i \log_2(1+1/x_i) \geq \log_2(1+1/\sum_i x_i)$. Therefore, we have:

$$r_k^B = \int_0^1 log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^K \lambda_{kl}(f)|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right) df \tag{D.8}$$

$$\geq \log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^K \int_0^1 \lambda_{k,l}(f) df \boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right) = r_k^A. \tag{D.9}$$

Besides, by applying the Parseval's theorem, we can show that $\eta_{\tau_{kl}} = \sum_{n=-\infty}^{\infty} g_{\tau_{kl}}^2(n) = \int_0^1 |G_{\tau_{kl}}(f)|^2 df = \int_0^1 \lambda_{kl}(f) df$ (refer to Appendix D.1 for more details). The right hand side of the inequality is equal to $r_k^A = \log_2 \left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^K \eta_{\tau_{k,l}}|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right)$ which verifies that $r_k^B > r_k^A$ and concludes the proof.

D.4 Derivation of the Rate Expression for Method *C* with Simplifying Assumptions

The term $\frac{1}{N}\text{tr}(\boldsymbol{\Lambda' P'_g})$ in the average transmit power expression can be written as $\frac{1}{N}\sum_{n=1}^{Nq}\lambda'[n]P'_g[n] = \sum_{k\in\mathcal{G}_g}P_k$ where $P_k = \frac{1}{N}\sum_{n\in\mathcal{I}_k}\lambda'[n]P'_g[n]$. To simplify the optimization problem, we use a sub-optimal power allocation which is, in fact, optimal for AWGN channels [93]. Assuming an AWGN channel, the optimal capacity-achieving power allocation simplifies to $\lambda'[n]P'_g[n] = P_k,\ \forall n \in \mathcal{I}_k$ which is followed by the concavity of $\log_2$ function and Jensen's inequality. Therefore, by substitution, the optimization problem simplifies to:

$$\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q},\{\mathcal{I}_k,P_k\}_{k=1}^{K},\pi(.)} p_{avg} = \sum_{k=1}^{K} P_k||\boldsymbol{w'_{\pi(k)}}||^2 \tag{D.10}$$
$$s.t.\quad \lim_{N\to\infty}\frac{1}{N}\sum_{n\in\mathcal{I}_k}\log_2\left(1+\frac{P_k|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q}P_{c(n,j)}|\boldsymbol{h_k^H w'_j}|^2+\sigma_k^2}\right)\geq r_k^*$$

where $c(n,j) = \{l \in \mathcal{G}_j | n \in \mathcal{I}_l\}$. In simple words, $P_{c(n,j)}$ is the power adjustment coefficient of the user in Group j which Sub-channel n is assigned to it. In the above optimization, the summation is over assigned sub-channels, however, with a simple sub-channel assignment rule, we can get rid of the summation and the optimization problem can be further simplified. We assume that there are q different sub-channel configurations,

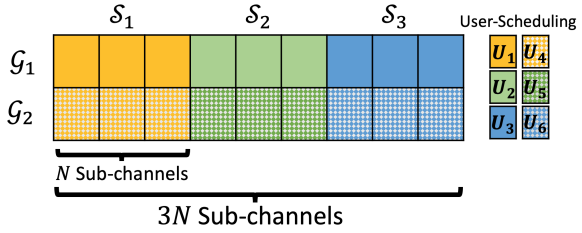


Figure D.2: Schematic representation of the simplified sub-channel assignment.

$\{\mathcal{S}_1, \cdots, \mathcal{S}_q\}$, where each of them includes K/q different users in K/q different groups,

i.e., $|\mathcal{S}_s| = K/q,\ s = 1, \cdots, q$. Define a sub-channel-assignment function that assigns each user to a sub-channel configuration, i.e., $\phi : \mathcal{K} \to \mathcal{S}$, where $\mathcal{K} = \{1, \cdots, K\}$ and $\mathcal{S} = \{1, \cdots, q\}$ are the set of user and sub-channel configuration indices, respectively. Each user, k, is assigned to a sub-channel configuration, s, $\phi(k) = s$. An example of the simplified sub-channel assignment is shown in Fig. D.2 for $K = 6$ and $q = 3$, where $\mathcal{G}_1 = \{1, 2, 3\}$, $\mathcal{G}_2 = \{4, 5, 6\}$, $\mathcal{S}_1 = \{1, 4\}$, $\mathcal{S}_2 = \{2, 5\}$ and $\mathcal{S}_3 = \{3, 6\}$. Note that the users with the same color (same sub-channel configuration) interfere with each other. Therefore, the optimization problem is simplified to

$$\begin{aligned} &\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q}, \{P_k\}_{k=1}^{K}, (\pi(.), \phi(.))} p_{avg} = \sum_{k=1}^{K} P_k ||\boldsymbol{w'_{\pi(k)}}||^2 \\ s.t.\ \ &\frac{P_k |\boldsymbol{h_k}^H \boldsymbol{w'_{\pi(k)}}|^2}{\sum_{j=1, j\neq\pi(k)}^{K/q} P_{c(k,j)} |\boldsymbol{h_k}^H \boldsymbol{w'_j}|^2 + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K. \end{aligned} \tag{D.11}$$

where $c(k, j) = \{l \in \mathcal{G}_j | \phi(k) = \phi(l)\}$.

Appendix E

Supplementary Proofs for Chapter 8

E.1 Proof of Theorem 8.1

The system model in Eq. (8.3) can be re-written for a two-user scenario as:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{R_{11}} & \boldsymbol{R_{12}} \\ \boldsymbol{R_{21}} & \boldsymbol{R_{22}} \end{pmatrix} \begin{pmatrix} h_1 \boldsymbol{I_N} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & h_2 \boldsymbol{I_N} \end{pmatrix} \begin{pmatrix} \boldsymbol{s_1} \\ \boldsymbol{s_2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu_1} \\ \boldsymbol{\nu_2} \end{pmatrix} = \boldsymbol{RHs} + \boldsymbol{\nu} \tag{E.1}$$

The mutual information $I(\boldsymbol{y}; \boldsymbol{s_1}, \boldsymbol{s_2})$ can be upper-bounded by:

$$I(\boldsymbol{y}; \boldsymbol{s_1}, \boldsymbol{s_2}) \leq \frac{1}{2} \log \det[cov(\boldsymbol{y})] - \frac{1}{2} \log \det[cov(\boldsymbol{\nu})] \tag{E.2}$$

where $cov(\boldsymbol{y}) = \boldsymbol{RH}\mathbb{E}[\boldsymbol{ss}^H]\boldsymbol{H}^H\boldsymbol{R} + \boldsymbol{R}\sigma^2$ and $cov(\boldsymbol{n}) = \boldsymbol{R}\sigma^2$. Therefore,

$$I(\boldsymbol{y}; \boldsymbol{s_1}, \boldsymbol{s_2}) \leq \frac{1}{2} \log \det \left[\boldsymbol{I_{2N}} + \frac{1}{\sigma^2} \begin{pmatrix} |h_1|^2 \boldsymbol{Q}_{\boldsymbol{s_1}} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & |h_2|^2 \boldsymbol{Q}_{\boldsymbol{s_2}} \end{pmatrix} \begin{pmatrix} \boldsymbol{R_{11}} & \boldsymbol{R_{12}} \\ \boldsymbol{R_{21}} & \boldsymbol{R_{22}} \end{pmatrix} \right] \tag{E.3}$$

where $\boldsymbol{Q_{s_k}}$ is the co-variance matrix of the input stationary process for User k. The upper-bound is achieved by Gaussian processes. Matrices $\boldsymbol{R_{11}}$ and $\boldsymbol{R_{22}}$ are equal to identity matrices and $\boldsymbol{R_{12}}$ and $\boldsymbol{R_{21}}$ are Toeplitz matrices which can be decomposed using the singular value decomposition, i.e., $\boldsymbol{R_{12}} = \boldsymbol{R_{21}^T} = \boldsymbol{U G_\tau V}^H$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and $\boldsymbol{G_\tau}$ is an $N \times N$ diagonal matrix containing singular values of $\boldsymbol{R_{12}}$ on its diagonal and depends on the pulse shape and the timing offset. Hence, we can simplify the upper-bound using:

$$\det\left[\boldsymbol{I_{2N}} + \frac{1}{\sigma_n^2}\begin{pmatrix} |h_1|^2\boldsymbol{Q_{s_1}} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & |h_2|^2\boldsymbol{Q_{s_2}}\end{pmatrix}\begin{pmatrix}\boldsymbol{R_{11}} & \boldsymbol{R_{12}} \\ \boldsymbol{R_{21}} & \boldsymbol{R_{22}}\end{pmatrix}\right] = \tag{E.4}$$

$$\det\left[\begin{pmatrix}\boldsymbol{U}^H & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{V}^H\end{pmatrix}\right]\det\left[\boldsymbol{I_{2N}} + \frac{1}{\sigma_n^2}\begin{pmatrix} |h_1|^2\boldsymbol{Q_{s_1}} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & |h_2|^2\boldsymbol{Q_{s_2}}\end{pmatrix}\begin{pmatrix}\boldsymbol{I_N} & \boldsymbol{R_{12}} \\ \boldsymbol{R_{21}} & \boldsymbol{I_N}\end{pmatrix}\right]\det\left[\begin{pmatrix}\boldsymbol{U} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{V}\end{pmatrix}\right] =$$

$$\det\left[\boldsymbol{I_{2N}} + \frac{1}{\sigma_n^2}\begin{pmatrix} |h_1|^2\boldsymbol{U}^H\boldsymbol{Q_{s_1}} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{V}^H|h_2|^2\boldsymbol{Q_{s_2}}\end{pmatrix}\begin{pmatrix}\boldsymbol{U} & \boldsymbol{R_{12}V} \\ \boldsymbol{R_{21}U} & \boldsymbol{V}\end{pmatrix}\right] =$$

$$\det\left[\boldsymbol{I_{2N}} + \frac{1}{\sigma_n^2}\begin{pmatrix} |h_1|^2\boldsymbol{U}^H\boldsymbol{Q_{s_1}U} & |h_1|^2\boldsymbol{U}^H\boldsymbol{Q_{s_1}UG_\tau} \\ |h_2|^2\boldsymbol{V}^H\boldsymbol{Q_{s_2}G_\tau} & |h_2|^2\boldsymbol{V}^H\boldsymbol{Q_{s_2}V}\end{pmatrix}\right] =$$

$$\det\left[\boldsymbol{I_{2N}} + \frac{1}{\sigma_n^2}\begin{pmatrix} |h_1|^2\boldsymbol{\mathcal{S}_1} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & |h_2|^2\boldsymbol{\mathcal{S}_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{I_N} & \boldsymbol{G_\tau} \\ \boldsymbol{G_\tau} & \boldsymbol{I_N}\end{pmatrix}\right] \tag{E.5}$$

where $\boldsymbol{\mathcal{S}_1} = \boldsymbol{U}^H\boldsymbol{Q_{s_1}U}$ and $\boldsymbol{\mathcal{S}_2} = \boldsymbol{V}^H\boldsymbol{Q_{s_2}V}$ are diagonal matrices whose diagonal elements can be denoted as s_{1n} and s_{2n}, respectively. To further simplify the upper-bound, we use the Lemma 2 in [83], which states:

Lemma E.1. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $N \times N$ non-negative-definite matrices, and let $\boldsymbol{G} = diag[g_1, \cdots, g_N]$ where $|g_n| \leq 1, n = 1, \cdots, N$. Then,*

$$\det\left[\boldsymbol{I_{2N}} + \begin{pmatrix}\boldsymbol{A} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{B}\end{pmatrix}\begin{pmatrix}\boldsymbol{I_N} & \boldsymbol{G} \\ \boldsymbol{G} & \boldsymbol{I_N}\end{pmatrix}\right] \leq \prod_{n=1}^{N}\left(1 + a_{nn} + b_{nn} + a_{nn}b_{nn}(1 - |g_n|^2)\right) \tag{E.6}$$

where a_{nn} and b_{nn} are the diagonal elements of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The equality is achieved with $\boldsymbol{A}$ and $\boldsymbol{B}$ being diagonal.

As a result, we will have:

$$R_1 + R_2 \leq \frac{1}{2} \lim_{N\to\infty} \sum_{n=1}^{N} \log\left(1 + \frac{|h_1|^2}{\sigma^2} s_{1n} + \frac{|h_2|^2}{\sigma^2} s_{2n} + \frac{|h_1|^2|h_2|^2}{\sigma^4} s_{1n}s_{2n}(1-|g_n|^2)\right) \frac{1}{N} \tag{E.7}$$

Toeplitz matrices are asymptotically equivalent to circulant matrices as the matrix dimension goes to infinity [134, 135]. The implication of the asymptotic equivalence of Toeplitz matrices with circular matrices is that the values of the singular values of Toeplitz matrices are asymptotically equal to samples of their generating function. In more details, considering a Toeplitz matrix, $\boldsymbol{R}$, its generating function, $R(f), f \in [0,1]$, and its singular values, $r_n, n = 1, \cdots, N$, we have $r_n = R(n/N), n = 1, \cdots, N$ [53].

Defining $f_n = n/N$, $df_N = 1/N$, $S_k(f_n) = s_{kn}$, and $G_\tau(f_n) = |g_n|$ we can rewrite the sum-rate upper-bound as $C = \frac{1}{2}\lim_{N\to\infty}\sum_{n=1}^{N} C(f_n)df_N$ where
$C(f_n) = \log\left(1 + \frac{|h_1|^2}{\sigma^2}S_1(f_n) + \frac{|h_2|^2}{\sigma^2}S_2(f_n) + \frac{|h_1|^2|h_2|^2}{\sigma^4}S_1(f_n)S_2(f_n)(1-G^2(f_n))\right)$. Because $C(f_n)$ is bounded and almost everywhere continuous on the interval $[0,1]$, then it is Reimann integrable on the interval [136], and we get:

$$R_1 + R_2 \leq \frac{1}{2}\int_0^1 \log\left(1 + \frac{|h_1|^2}{\sigma^2}S_1(f) + \frac{|h_2|^2}{\sigma^2}S_2(f) + \frac{|h_1|^2|h_2|^2}{\sigma^4}S_1(f)S_2(f)(1-G_\tau^2(f))\right) df \tag{E.8}$$

where $S_1(f)$ and $S_2(f)$ are the generating functions of $\boldsymbol{Q_{s_1}}$ and $\boldsymbol{Q_{s_2}}$, or equivalently, PSDs of User 1 and User 2, respectively. In addition, $G_\tau(f) = |R_{12}(f)| = |R_{12}(f)|$ where $R_{12}(f)$ and $R_{21}(f)$ are the generating functions of Toeplitz matrices $\boldsymbol{R_{12}}$ and $\boldsymbol{R_{21}}$, respectively.

To analyze $G_\tau(f)$, recall the structure of $\boldsymbol{R_{21}}$ as:

$$\boldsymbol{R_{21}} = \begin{pmatrix} g_\tau[0] & g_\tau[-1] & \cdots & g_\tau[1-n] \\ g_\tau[1] & g_\tau[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & g[-1] \\ g_\tau[N-1] & \cdots & g_\tau[1] & g_\tau(0) \end{pmatrix} \tag{E.9}$$

where τ is the timing offset between two users and $g(t)$ is the matched filter pulse. Recall that $g_\tau[n] = g(\tau T + nT)$. The generating function of $\boldsymbol{R_{21}}$ is defined as $R_{21}(f) = \sum_{n=1-N}^{N-1} g_\tau[n] e^{-j2\pi nf}$ and is periodic with period of 1. Equivalently, $R_{21}(f)$ can be denoted as:

$$R_{21}(f) = \frac{1}{T} \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}(\frac{f+i}{T}), \tag{E.10}$$

where $\hat{g}(f)$ is the Fourier transform of $g(t)$. Note that $R_{21}(f)$ can be interpreted as the folded-spectrum. The only difference is that each replica is phase shifted due to the timing offset between users.