UNIVERSITY OF CALIFORNIA,
IRVINE

Analysis and Design of Asynchronous Non-Orthogonal Multiple Access

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical Engineering

by

Xun Zou

Dissertation Committee:
Professor Hamid Jafarkhani, Chair
Professor Ender Ayanoglu
Assistant Professor Zhiying Wang

2020

DEDICATION

To my wife and my parents, for their unwavering support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ALGORITHMS

ACKNOWLEDGMENTS

This thesis would never have been possible without the support and guidance from my advisor, Professor Hamid Jafarkhani. It was a precious experience working with Prof. Jafarkhani in the past five years. He always gave me valuable advises and encouraged me to fulfill my studies. His dedication to research sets a role model in my future life.

I would like to thank other members in our research group over these years: Mehdi Ganji, Biao He, Jun Guo, Xiaoyi Liu, Lisi Jiang, Weiqi Li, Erdem Koyuncu, Saeed Karimi Bidhendi, Carles Diaz. Especially, it is my honor to conduct researches together with two excellent researchers, Mehdi Ganji and Biao He, from whom I acquired a great amount of knowledge and pleasure.

Another very special gratitude goes out to 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.

I would like to thank my parents for their selfless love. They always show a great understanding of the decision I made and the dilemma I faced. Last but not least, I wouldn't have done any of this without the endless support from my wife Yan, who was still my girlfriend when I just started the journey of PhD study. I sincerely appreciate her trust and companionship.

CURRICULUM VITAE

Xun Zou

EDUCATION

Doctor of Philosophy in Electrical Engineering **2020**
University of California, Irvine *Irvine, California*

Master of Science in Electronics Science and Technology **2015**
Beijing University of Posts and Telecommunications *China*

Bachelor of Science in Electronic Information Engineering **2012**
Harbin Institute of Technology *China*

RESEARCH EXPERIENCE

Graduate Research Assistant **2015 – 2020**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **Jan. 2020 – Mar. 2020**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Downlink Asynchronous Non-Orthogonal Multiple Access with Quantizer Optimization **2020**
submitted to IEEE Wireless Communications Letters

Trellis-Coded Non-Orthogonal Multiple Access **2020**
IEEE Wireless Communications Letters

Cooperative Asynchronous Non-Orthogonal Multiple Access with Power Minimization Under QoS Constraints **2020**
IEEE Transactions on Wireless Communications

Exploiting Time Asynchrony in Multi-user Transmit Beamforming **2020**
IEEE Transactions on Wireless Communications

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

An Analysis of Two-User Uplink Asynchronous Non-Orthogonal Multiple Access Systems **2019**
IEEE Transactions on Wireless Communications

Interleaving Channel Estimation and Limited Feedback for Point-to-point Systems with a Large Number of Transmit Antennas **2018**
IEEE Transactions on Wireless Communications

REFEREED CONFERENCE PUBLICATIONS

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

A Block-Based Non-Orthogonal Multicarrier Scheme **Dec. 2019**
IEEE Global Communications Conference

On Uplink Asynchronous Non-Orthogonal Multiple Access Systems with Timing Error **May 2018**
IEEE International Conference on Communications

Asynchronous Channel Training in Massive MIMO Systems **Dec. 2016**
IEEE Global Communications Conference

ABSTRACT OF THE DISSERTATION

Analysis and Design of Asynchronous Non-Orthogonal Multiple Access

By

Xun Zou

Doctor of Philosophy in Electrical Engineering

University of California, Irvine, 2020

Professor Hamid Jafarkhani, Chair

In this dissertation, the analysis and design of asynchronous non-orthogonal multiple access (ANOMA) are presented and the advantages of ANOMA over the conventional non-orthogonal multiple access (NOMA) are detailed. The gain of ANOMA over NOMA, which is called "sampling diversity", results from intentionally introducing the symbol asynchrony at the transmitter and applying the oversampling technique at the receiver. In this dissertation, first, the conventional NOMA is introduced, including the superposition coding and the successive interference cancellation (SIC), and the trellis code is designed and optimized specifically for NOMA systems with the aid of the tensor product of trellises. Second, we analyze the performance of an uplink ANOMA system and the effects of timing errors on the system performance. It is revealed that ANOMA achieves a higher sum throughput compared with NOMA and the synchronization timing error causes a more severe performance loss compared with the coordination timing error. Third, a downlink cooperative ANOMA system with user relaying is studied, including the throughput performance and the power consumption. It is demonstrated that ANOMA is more power efficient compared with NOMA. Finally, we further investigate the impact of imperfect channel state information (CSI) on ANOMA systems. It is shown that under the same channel estimation error, the users in ANOMA systems achieve a lower outage probability compared with those in NOMA systems. Furthermore, the limited feedback scheme for the downlink ANOMA

systems is proposed and optimized. It is manifested that ANOMA can achieve the same or even higher average max-min rate as NOMA with a lower feedback rate.

Chapter 1

Introduction

Non-orthogonal multiple access (NOMA) is envisaged as a promising technique for future radio access [1]. Traditional orthogonal multiple access (OMA) techniques allocate orthogonal resources to different users, e.g., orthogonal time resources in the time division multiple access (TDMA) scheme. Differently, the NOMA provides the multiuser access by allocating non-orthogonal resources to users [2]. For example, in the power-domain NOMA scheme, the signals for multiple users are superposed at different power levels using superposition coding [3], and the multiuser detection method, such as successive interference cancellation (SIC) [4], is employed at the receiver. The advantages of the NOMA over the OMA have been extensively studied in [1] and the references therein, e.g., providing higher system throughput compared with OMA and supporting massive connectivity.

Another line of research is to study the effects of asynchronous transmission on the performance of the wireless communication systems. Asynchronous transmission refers to the case where the symbol epochs of the signals transmitted by the users are not aligned at the receiver [5]. In particular, [5] first pointed out the potential advantages of symbol-asynchronous communications in terms of increasing the capacity of a multiple-access channel. The work in

[6] applied the symbol-asynchronous channel estimation method to tackle the pilot contamination problem in massive multiple-input multiple-output (MIMO) systems. Asynchronous transmission was studied in [7, 8] as a tool to mitigate or cancel the inter-user interference. In addition, the nonzero symbol offset was used to reduce the inter-antenna interference in MIMO systems in [9]. Moreover, an asynchronous analog network coding scheme for multiuser cooperative communications was proposed in [10] to provide a greater diversity order compared with that of synchronous analog network coding. Also, adding intentional timing mismatch was proposed in [11] to improve the performance of a relay network. The authors of [12, 13] further proposed several differential decoding schemes for asynchronous multiuser MIMO systems based on orthogonal space-time block codes (OSTBCs) and for differential distributed space-time coding systems with imperfect synchronization.

Asynchronous NOMA (ANOMA) is proposed by combining the asynchronous transmission with NOMA. It can be proved that ANOMA outperforms NOMA in terms of, e.g., the throughput, the outage probability, and the power consumption, by intentionally introducing a timing mismatch at the transmitter and applying the oversampling technique at the receiver. The main contribution of this dissertation is summarized as follows:

1. We study the design of trellis code in NOMA systems. The signals for different users are produced by trellis coded modulation (TCM) and then superimposed on different power levels. By interpreting the encoding process via the tensor product of trellises, we introduce a joint detection method based on the Viterbi algorithm. Then, we determine the optimal power allocation between the two users by maximizing the free distance of the tensor product trellis. Finally, we manifest that the trellis-coded NOMA outperforms the uncoded NOMA at high signal-to-noise ratio (SNR).

2. We investigate the uplink ANOMA system. Focusing on a two-user uplink system, for the first time, we analytically prove that the ANOMA with a sufficiently large frame length can always outperform the NOMA in terms of the sum throughput.

To this end, we derive the expression for the sum throughput of the ANOMA as a function of SNR, frame length, and normalized timing mismatch. Based on the derived expression, we find that users should transmit at full powers to maximize the sum throughput. In addition, we obtain the optimal timing mismatch as the frame length goes to infinity. Moreover, we comprehensively study the impact of timing error on the ANOMA throughput performance. Two types of timing error, i.e., the synchronization timing error and the coordination timing error, are considered. We derive the throughput loss incurred by both types of timing error and find that the synchronization timing error has a greater impact on the throughput performance compared with the coordination timing error.

3. We propose a novel half-duplex cooperative ANOMA (C-ANOMA) framework with user relaying, where a timing mismatch is intentionally added in the broadcast signal. We derive the expressions for the throughputs of the strong user (acts as relay) which employs the block-wise SIC and the weak user which combines the symbol-asynchronous signal with the interference-free signal. We analytically prove that in the C-ANOMA systems with a sufficiently large block length, the strong user attains the same throughput to decode its own message while both users can achieve a higher throughput to decode the weak user's message compared with those in the cooperative NOMA (C-NOMA) systems. Besides, we obtain the optimal timing mismatch when the block length goes to infinity. Furthermore, to exploit the tradeoff between the power consumption of the base station and that of the relay user, we solve a weighted sum power minimization problem under quality of services (QoS) constraints. Numerical results show that the C-ANOMA system can consume less power compared with the C-NOMA system to satisfy the same QoS requirements.

4. We investigate a downlink ANOMA system with imperfect CSI. It is analytically proved that the ANOMA system with a relatively large frame length outperforms the NOMA

system in terms of the outage probability. To this end, we derive the analytical expressions for the individual throughput of each user and simplify them in the asymptotic case of infinite frame length. Besides, we show that with channel estimation error, the optimal timing mismatch converges to half of a single symbol length as the frame length goes to infinity.

5. We study a two-user downlink ANOMA with limited feedback. We employ the max-min criterion for the power allocation and derive the closed-form expressions for the upper and lower bounds of the max-min rate. It is demonstrated that ANOMA can achieve the same or even higher average max-min rate with a lower feedback rate compared with NOMA. Moreover, we propose a quantizer optimization algorithm which applies to both NOMA and ANOMA. Simulation results show that the optimized quantizer significantly improves the average max-min rate compared with the conventional uniform quantizer, especially in the scenario with a low feedback rate.

Chapter 2

Non-Orthogonal Multiple Access and Trellis Code Design

2.1 Introduction

From a unified perspective, NOMA consists of code-domain NOMA and power-domain NOMA [14]. Both code-domain and power-domain NOMA have been extensively studied in the existing literature.

2.1.1 Power-domain NOMA

In the power-domain NOMA systems, the signals of different users are assigned different powers. Then, one major challenge is the optimal power allocation as discussed, for example, in [2, 15]. The optimal power can be determined according to the channel conditions to maximize users' achievable rates. Superposition coding and successive interference cancellation (SIC) techniques are utilized at the transmitter and the receiver, respectively. Again,

there are many studies on how to perform these techniques efficiently, for example [16, 17].

For simplicity, we consider a two-user scenario to illustrate the NOMA technique. In an uplink system, two users transmit signals to the base station (BS) simultaneously using the same bandwidth. The received signal at the BS is given by

$$y = h_1\sqrt{P_1}s_1 + h_2\sqrt{P_2}s_2 + n, \tag{2.1}$$

where h_i, P_i, and s_i are the channel coefficient, the transmit power, and the transmitted symbol of User i, respectively, and n is the additive Gaussian noise, i.e., $n \sim \mathcal{CN}(0, \sigma_n^2)$. SIC is employed at the BS. Assume that User 1 is decoded first. The rate of User 1 is given by

$$R_1 = \log\left(1 + \frac{P_1|h_1|^2}{P_2|h_2|^2 + \sigma_n^2}\right). \tag{2.2}$$

Under the assumption of perfect SIC, after removing the signal from User 1, the rate of User 2 is given by

$$R_2 = \log\left(1 + \frac{P_2|h_2|^2}{\sigma_n^2}\right). \tag{2.3}$$

In the downlink NOMA system, the superposition coding is adopted at the BS. The transmitted signal is given by

$$s = \sqrt{P_1}s_1 + \sqrt{P_2}s_2, \tag{2.4}$$

where P_i and s_i is the power allocated to User i and the symbol for User i, respectively. The received signal at User i is given by

$$y_i = h_i s + n_i = h_i\sqrt{P_1}s_1 + h_i\sqrt{P_2}s_2 + n_i. \tag{2.5}$$

Assume that User 1 is the stronger user with a better channel. User 1 can employ the SIC, which first detects User 2's messages, removes it, and then detects its own messages. The rate of User 1 to detect User 2's message is given by

$$R_{2\to 1} = \log\left(1 + \frac{P_2|h_1|^2}{P_1|h_1|^2 + \sigma_{n_1}^2}\right). \tag{2.6}$$

With perfect SIC, the rate of User 1 is given by

$$R_1 = \log\left(1 + \frac{P_1|h_1|^2}{\sigma_{n_1}^2}\right). \tag{2.7}$$

Similarly, the rate of User 2 is given by

$$R_2 = \log\left(1 + \frac{P_2|h_2|^2}{P_1|h_2|^2 + \sigma_{n_2}^2}\right). \tag{2.8}$$

2.1.2 Code-domain NOMA

The code-domain NOMA has its origin in code division multiple access (CDMA), including sparse code multiple access (SCMA) [18] and trellis coded multiple access (TCMA) [19]. The signals of multiple users are separated by user-specific features, e.g., the uniquely assigned codeword of each user. In the code-domain NOMA, the main efforts are devoted to the multi-user detection, for example, the design of multidimensional constellations [18, 20]. To the best of our knowledge, the joint design of the code-domain and power-domain NOMA has never been studied.

In this chapter, we apply trellis coded modulation (TCM) to the power-domain NOMA, taking advantages of the coding gain and the power optimization. Utilizing superposition coding, the signals for multiple users are superimposed on different power levels. Compared

with [20], the main contribution of this work is introducing the power allocation to code-domain NOMA. The performance can be improved by allocating proper powers to the signals of different users. Instead of utilizing TCM purely for codeword design in [20], TCM is employed in this work to jointly optimize the error control coding and modulation. Therefore, the Viterbi algorithm can be directly applied to the proposed scheme. By interpreting the modulating process via the tensor product of trellises [21, 22], we implement the maximum likelihood sequence detection (MLSD) based on the Viterbi algorithm [23]. Furthermore, we derive the optimal power allocation between the two users by maximizing the free distance of the tensor product trellis.

The key difference between the trellis-coded NOMA and the traditional TCMA lies in the multiple access scheme. In TCMA, the signals of multiple users are differentiated by their unique features, for example, convolutional encoder, constellation, or interleaver [24]. However, in the trellis-coded NOMA, the signals are differentiated only by the power levels. Furthermore, for the first time, we provide insight into the power optimization for the superimposed TCM signals.

2.2 System Model

In this chapter, we consider a downlink NOMA system consisting of one base station (BS) and two users. Superposition coding is employed at the transmitter. The power allocated to User i's signal is denoted as P_i, $i = 1, 2$. The channel coefficient between the BS and User i is represented by h_i. We adopt the block fading channel model, i.e., the channel remains static within each block and changes independently from one block to another [2, 15]. We assume that the channel state information is perfectly known by the BS and users. Without loss of generality, we assume that $|h_1|^2 > |h_2|^2$. To stipulate the user fairness, we set $P_2 > P_1$. In what follows, the 8-phase-shift keying (PSK) 4-state TCM serves as an example

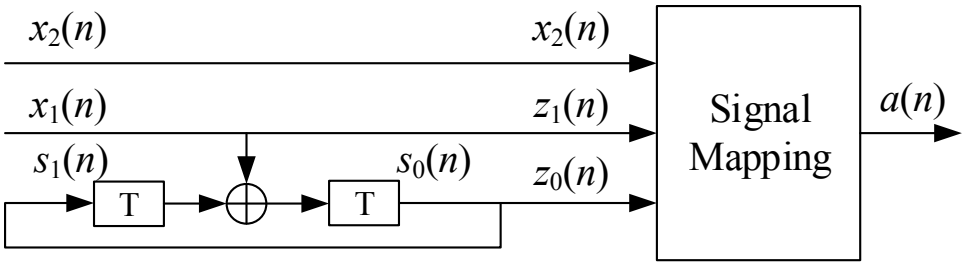

Figure 2.1: Illustration of an 8-PSK 4-state TCM encoder.

of TCM [25], which is depicted in Fig. 2.1. The trellis diagram and the 8-PSK mapping are shown in Figs. 2.2 (a) and (b), respectively. In Figs. 2.1 and 2.2, x_1 and x_2 represent the uncoded bits while z_0 and z_1 denote the coded bits via the convolutional encoder. For the sake of brevity, we employ the signal constellation with unit signal power, i.e., $\mathrm{E_b} = 1$. Note that the proposed scheme can be applied to the case where two users employ different modulations/trellises and also the case of more than two users.

In the proposed trellis-coded NOMA, the signals for Users 1 and 2 are first modulated by TCM, as shown in Figs. 2.1 and 2.2, and then superimposed on different power levels. Using superposition coding, the nth transmitted symbol at the BS is given by $\sqrt{P_1}a_1(n)+\sqrt{P_2}a_2(n)$ where $a_i(n)$ is the nth symbol for User i after TCM. Then, the nth received sample at User i is given by

$$y_i(n) = h_i \left[\sqrt{P_1}a_1(n) + \sqrt{P_2}a_2(n)\right] + w_i(n), \tag{2.9}$$

where $w_i(n) \sim \mathcal{CN}(0, \sigma_i^2)$ is the additive noise. At users, the modulated symbols are detected and then the binary information bits are recovered from the modulated symbols, which will be explained in the next section.

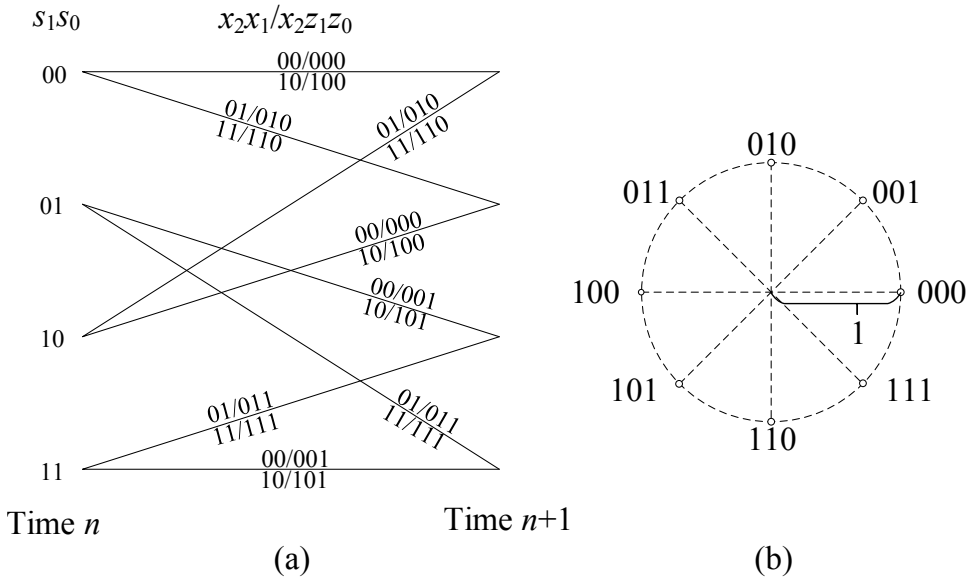

Figure 2.2: (a) Trellis representation of 8-PSK 4-state TCM. (b) The mapping of 8-PSK constellation.

2.3 Tensor Product of Trellises and Detection Design

In this section, we first present the separate detection method with SIC. Then, we propose the joint detection method based on a novel trellis structure known as "tensor product of trellises".

2.3.1 Separate Detection with SIC

In the separate detection scheme, the signals for Users 1 and 2 are detected separately. User 2 (the weak user) detects its own signal by considering User 1's signal as noise. User 1 (the strong user) utilizes SIC, i.e., first detects User 2's signal, removes it from the superimposed signal, and then detects its own signal. The Viterbi algorithm [23] can be employed to determine the sequence with the minimum Euclidean distance from the received sequence using the 4-state trellis in Fig. 2.2 (a).

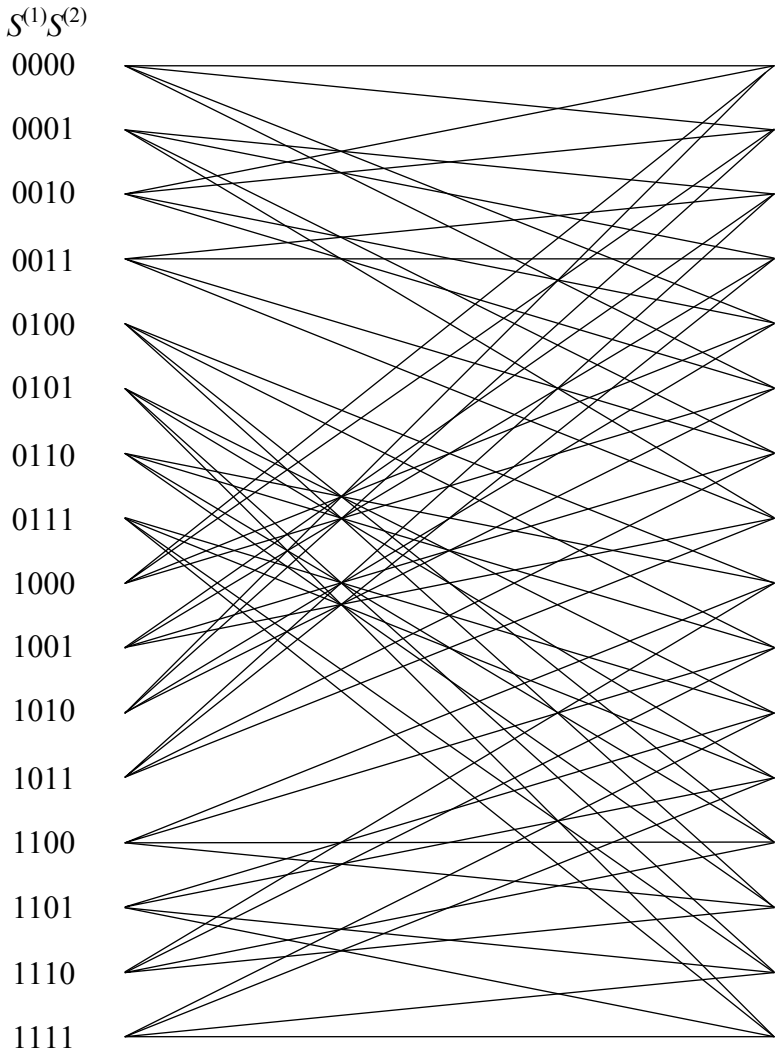

Figure 2.3: Underlying tensor product of trellises.

2.3.2 Joint Detection with Tensor Product of Trellises

First, we review the concept of the tensor product of trellises [22, 21]. Let us consider trellises T_1 and T_2 with r_1 and r_2 states, respectively, and $S_i^{(l)}$, $i = 1, \cdots, r_l$, denotes the ith state of T_l. The tensor product of T_1 and T_2, denoted as $T_1 \otimes T_2$, can be represented as a trellis with $r_1 \times r_2$ states. Each state in $T_1 \otimes T_2$ is given by $S_i^{(1)} S_j^{(2)}$, $i = 1, \cdots, r_1$, $j = 1, \cdots, r_2$. The state transition from $S_i^{(1)} S_j^{(2)}$ to $S_k^{(1)} S_l^{(2)}$ exists if and only if there exist transitions from $S_i^{(1)}$ to $S_k^{(1)}$ in T_1 and from $S_j^{(2)}$ to $S_l^{(2)}$ in T_2. One can easily extend the definition of the tensor product trellis to the case of more than two trellises.

Let us revisit the modulating process of two users' signals in Section 2.2. The symbols for Users 1 and 2 are modulated independently through the 4-state trellis, shown in Fig. 2.2 (a). Let T_1 and T_2 stand for the trellises employed to modulate the symbols for Users 1 and 2, respectively. The tensor product trellis $T_1 \otimes T_2$ is the 16-state trellis in Fig. 2.3. Every pair of state transitions in T_1 and T_2 can be represented by a unique transition path in $T_1 \otimes T_2$. For example, let us assume that the state of T_1 transits from $S_i^{(1)}$ to $S_k^{(1)}$ producing the modulated symbol a_1 and the state of T_2 transits from $S_j^{(2)}$ to $S_l^{(2)}$ generating the modulated symbol a_2. From the perspective of $T_1 \otimes T_2$, the state transits from $S_i^{(1)}S_j^{(2)}$ to $S_k^{(1)}S_l^{(2)}$ and the superimposed symbol $\sqrt{P_1}a_1 + \sqrt{P_2}a_2$ is produced. Since every state transition can be realized by two parallel paths in T_1 and T_2, as shown in Fig. 2.2 (a), every state transition in $T_1 \otimes T_2$ includes $2 \times 2 = 4$ parallel paths.

The description of the tensor product trellis demonstrates the equivalence of the trellis-coded NOMA and the TCM using the tensor product trellis. The joint detection is to detect both users' signals jointly by treating the trellis-coded NOMA as a regular TCM with the tensor product trellis. In the joint detection, the Viterbi algorithm is implemented using the tensor product trellis. It is worth mentioning that there is no necessity to modulate the signals for Users 1 and 2 jointly using the tensor product trellis at the transmitter. The transmitted symbols for each user can be modulated independently according to its own trellis by applying an appropriate power allocation scheme to ensure a good decoding performance (as shown in Section V).

Since the Viterbi algorithm can be employed in joint decoding, the computational complexity increases linearly with the number of decoded symbols, N. More specifically, if the number of states in T_i $(i = 1, 2)$ is K_i and the total number of edges in T_i is L_i, the computational complexity of the joint detection method is given by $O(N(K_1K_2 + L_1L_2))$ while that of the separate detection method with SIC is $O(N(K_1 + K_2 + L_1 + L_2))$.

2.4 Power Optimization

In this section, we study the power allocation to optimize the performance of the joint detection scheme. The power allocation is optimized under two power constraints. One is the sum power constraint, i.e., $P_1 + P_2 \leq P$ where P is the total transmit power. The other constraint is $P_1 < P_2$ which is added with no loss of generality. We adopt the free distance of the tensor product trellis, d_{free}, to measure the performance, which is widely used in the existing TCM studies, for example [26]. A larger free distance results in a better performance at high signal-to-noise ratio (SNR). As will be illustrated later, the free distance is a function of the power coefficients P_1 and P_2. We obtain the optimal powers by maximizing the free distance.

The free distance is defined as the minimum Euclidean distance between any pair of valid and distinct sequences produced by a given trellis, i.e., $d_{\text{free}} = \arg\min_{\mathbf{a}_1,\mathbf{a}_2 \in V, \mathbf{a}_1 \neq \mathbf{a}_2} ||\mathbf{a}_1 - \mathbf{a}_2||$ where V is the set of all valid sequences. The free distance can be determined by choosing the minimum of two candidates: the minimum Euclidean distance between the symbols produced by the parallel paths, i.e., d_{parallel}, and that between the sequences which diverge from the same state and then merge at the same state, i.e., $d_{\text{D\&M}}$. The subscript D&M is the acronym for "diverging and merging". In what follows, we analyze these two distances separately. Assume that there are two different paths in $T_1 \otimes T_2$ producing $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where u_1 and u_2 are the modulated symbols of T_1 and v_1 and v_2 are those of T_2.

2.4.1 Parallel Paths

First, we study the case where $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$ are produced by the parallel paths in $T_1 \otimes T_2$. Fig. 2.4 illustrates the possible positions of $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$ in the superimposed constellation when v_1 and v_2 are chosen from

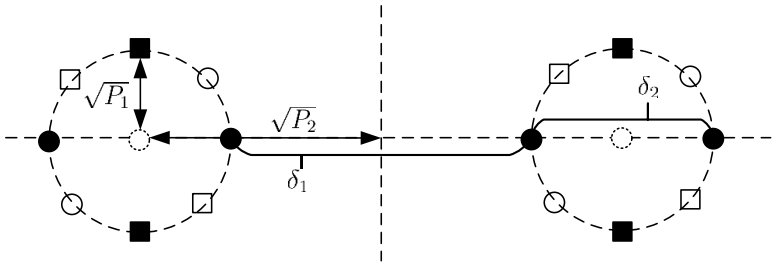

Figure 2.4: Illustration of the minimum Euclidean distance in the superimposed constellation.

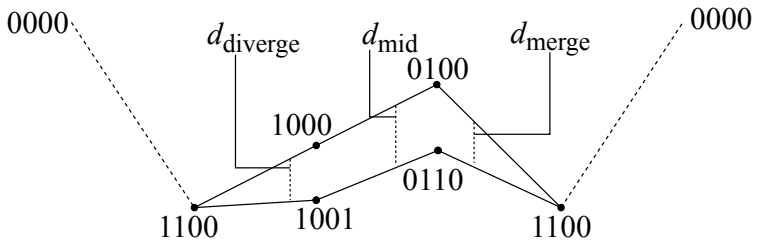

Figure 2.5: Illustration of the diverging-and-merging paths with the minimum Euclidean distance.

$\{1, -1\}$. Because of symmetry, the minimum Euclidean distance for all the other choices will be the same. In Fig. 2.4, there are four different markers, hollow/solid square/circle. The superimposed symbols depicted by the same marker are the symbols produced by the parallel paths for a specific state transition in $T_1 \otimes T_2$. Every state transition in $T_1 \otimes T_2$ can be realized by four parallel paths. Therefore, there are four positions for every marker. The minimum Euclidean distance between parallel paths can be found by calculating the Euclidean distance between the points sharing the same marker. It is clear from Fig. 2.4 that the minimum Euclidean distance is either δ_1 or δ_2. Thus,

$$d_{\text{parallel}} = \min\{\delta_1, \delta_2\} = \min\left\{2\sqrt{P_2} - 2\sqrt{P_1}, 2\sqrt{P_1}\right\}. \tag{2.10}$$

2.4.2 Diverging-and-Merging Paths

Second, we study the Euclidean distance between the sequences which diverge from the same state and then merge at the same state. It can be shown that if two sequences diverge from

any state, it takes at least three transitions to merge at the same state. We utilize the exhaustive search to find a pair of sequences with the minimum Euclidean distance among all pairs of distinct sequences, which is shown in Fig. 2.5. Note that all valid codewords start and end at state zero. However, any common sub-sequence will not contribute to d_{free}. Therefore, to calculate d_{free} in Fig. 2.5, we need to consider the state transitions $1100 \to 1000 \to 0100 \to 1100$ and $1100 \to 1001 \to 0110 \to 1100$. As shown in Fig. 2.5, the squared Euclidean distance between the diverging-and-merging paths is given by

$$d^2_{\text{D\&M}} = d^2_{\text{diverge}} + d^2_{\text{mid}} + d^2_{\text{merge}}. \tag{2.11}$$

First, let us focus on the diverging paths in Fig. 2.5. According to Fig. 2.2, the superimposed symbol produced by the path $1100 \to 1000$ is given by $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$, where $u_1 \in \{e^{j3\pi/4}, e^{j7\pi/4}\}$ and $v_1 \in \{1, -1\}$. Similarly, the superimposed symbol produced by $1100 \to 1001$ is given by $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where $u_2 \in \{e^{j3\pi/4}, e^{j7\pi/4}\}$ and $v_2 \in \{e^{j\pi/2}, e^{j3\pi/2}\}$. The positions of the superimposed symbols can be shown in Fig. 2.6. The minimum Euclidean distance between the diverging paths is given by

$$d_{\text{diverge}} = \delta_3 = |\sqrt{2P_2} - 2\sqrt{P_1}|.$$

One can employ the same approach to derive the minimum Euclidean distance between the merging paths and find that $d_{\text{merge}} = d_{\text{diverge}}$.

Second, we investigate the Euclidean distance d_{mid} in Fig. 2.5. The superimposed symbol produced by the path $1000 \to 0100$ is given by $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$, where $u_1 \in \{1, -1\}$ and $v_1 \in \{1, -1\}$. Similarly, the superimposed symbol produced by the path $1001 \to 0110$ is given by $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where $u_2 \in \{1, -1\}$ and $v_2 \in \{e^{j\pi/4}, e^{j5\pi/4}\}$. The positions of the superimposed symbols can be shown in Fig. 2.7. According to Fig. 2.7, the minimum

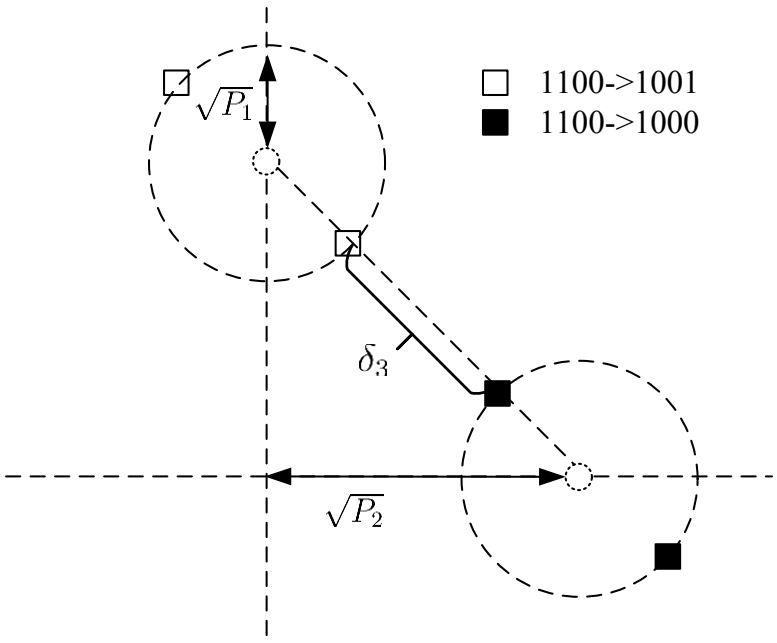

Figure 2.6: Illustration of the minimum Euclidean distance between the symbols produced by diverging paths.

Euclidean distance d_{mid} is given by

$$\begin{aligned} d_{\text{mid}}^2 &= \min\{\delta_4^2, \delta_5^2\} \\ &= (2-\sqrt{2})P_2 + \min\left\{0, 4P_1 + 2\sqrt{P_1P_2}\left(\sqrt{2}-2\right)\right\}. \end{aligned}$$

To summarize, the minimum Euclidean distance between the diverging-and-merging paths is given by

$$\begin{aligned} d_{\text{D\&M}}^2 =& d_{\text{diverge}}^2 + d_{\text{mid}}^2 + d_{\text{merge}}^2 \\ &= \left(6-\sqrt{2}\right)P_2 + 8P_1 - 8\sqrt{2P_1P_2} + \min\left\{0, 4P_1 + 2\sqrt{P_1P_2}\left(\sqrt{2}-2\right)\right\}. \quad (2.12) \end{aligned}$$

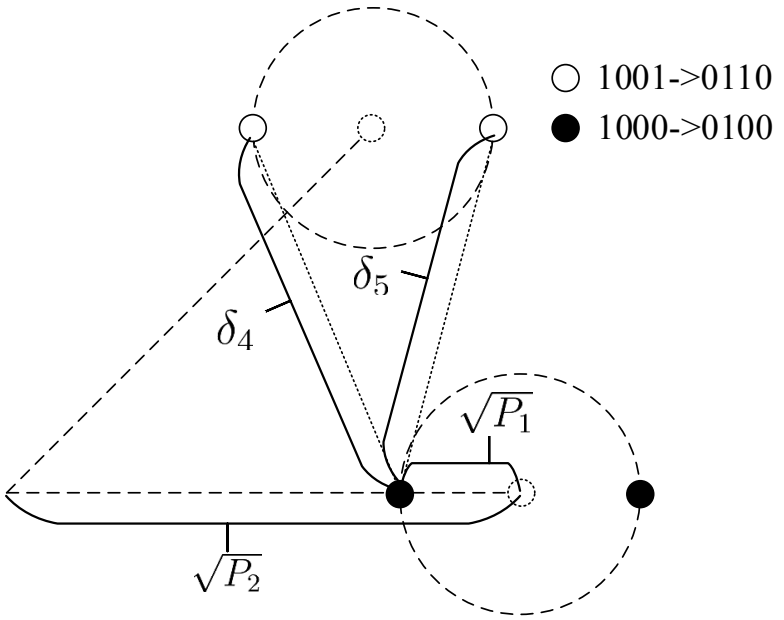

Figure 2.7: Illustration of the minimum Euclidean distance between the symbols in the intermediate stage of the diverging-and-merging paths.

2.4.3 Free Distance

The free distance of $T_1 \otimes T_2$ is determined by finding the minimum of d_{parallel} and $d_{\text{D\&M}}$, i.e.,

$$\begin{aligned} d_{\text{free}}^2 &= \min\{d_{\text{parallel}}^2, d_{\text{D\&M}}^2\} \\ &= \min\Big\{4P_1, 4\left(\sqrt{P_2}-\sqrt{P_1}\right)^2, \left(6-\sqrt{2}\right)P_2+8P_1 \\ &\quad -8\sqrt{2P_1P_2}+\min\Big\{0, 4P_1+2\sqrt{P_1P_2}\left(\sqrt{2}-2\right)\Big\}\Big\}. \end{aligned} \tag{2.13}$$

The optimal powers can be derived by maximizing the free distance, i.e.,

$$[P_1^*, P_2^*] = \arg\max_{P_1,P_2} d_{\text{free}}^2, \text{ s.t. } P_1+P_2 \le P, \tag{2.14}$$

where P is the total transmit power. According to (2.13), one can derive that d_{free} is maximized when $4P_1 = \left(6-\sqrt{2}\right)P_2+8P_1-8\sqrt{2P_1P_2}$, which then results in $\frac{P_1^*}{P_2^*} = \left(\frac{2\sqrt{2}-\sqrt{2+\sqrt{2}}}{2}\right)^2 \approx 0.2404$. Besides, to combat the channel noise, P_1+P_2 should be maximized. As a result,

$P_1^* = \frac{0.2404}{1+0.2404} P \approx 0.1938P$ and $P_2^* \approx 0.8062P$.

While we presented the results for a two-user scenario with 8-PSK 4-state TCM, our approach can be generalized to any TCM.

2.5 Simulation Results

In this section, we present the simulation results of the 8-PSK 4-state trellis-coded NOMA (TC-NOMA), TCMA, and the uncoded NOMA (UC-NOMA) with 4-PSK. We ensure a fair comparison among these schemes since the TCM is implemented without consuming extra bandwidth compared with the uncoded modulation [25]. In our simulation, we employ bit error ratio (BER) as the measure of performance. In the uncoded NOMA, the maximum likelihood detection is employed. We also present the results for the TCMA where the signals for Users 1 and 2 are modulated by the identical trellis shown in Fig. 2.2 but differentiated by constellation [19]. In TCMA, the constellation used by one user is the other user's constellation rotated by $\pi/8$. In contrast to the trellis-coded NOMA, the transmitted signal in TCMA is given by $\sqrt{(P_1 + P_2)/2}[a_1(n) + a_2(n)]$, which ensures a fair comparison by using the same sum transmit power.

First, we show the BER as a function of SNR for NOMA and TCMA schemes in Figs. 2.8 and 2.9 when $P_2 = 1$ and $P_1 = 0.1$ or 0.3, respectively. SNR is given by $\frac{1}{\sigma^2}$ where σ^2 is the variance of noise. For $P_1 = 0.1$ or 0.3, it is manifested that at high-SNR, similar to conventional TCM [25], the trellis-coded NOMA using the joint detection outperforms the uncoded NOMA. Besides, the trellis-coded NOMA using the separate detection achieves a similar performance to that using the joint detection when $P_1 = 0.1$. In contrast, there is a huge gap between the BER curves of the separate detection and those of the joint detection when $P_1 = 0.3$. This is because of the severe inter-user interference when detecting two

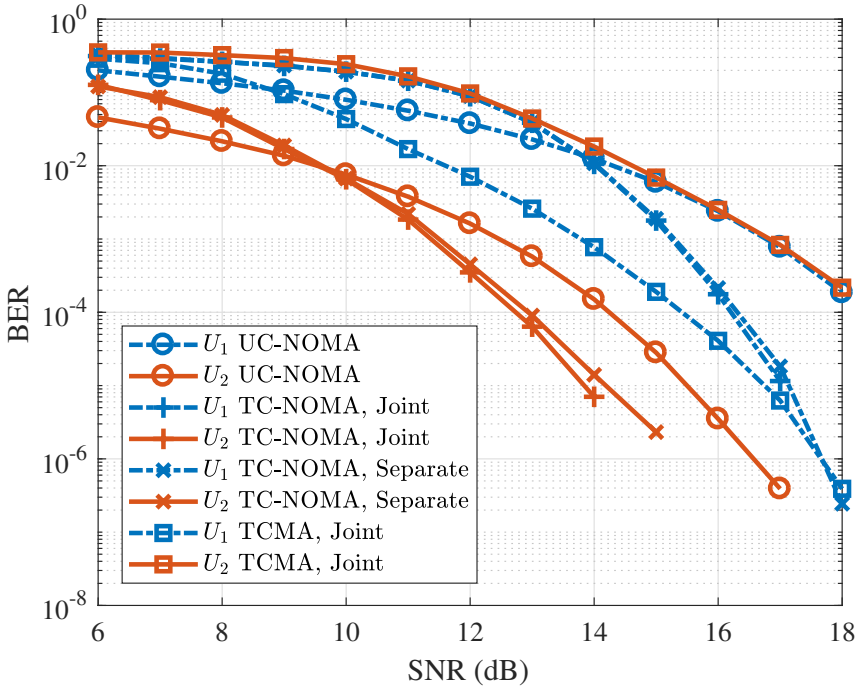

Figure 2.8: BER vs. SNR for TCMA, uncoded and trellis-coded NOMA when $P_1 = 0.1$, $P_2 = 1$, $|h_1|^2 = 2$, $|h_2|^2 = 1$.

user's signals separately and the error propagation problem in SIC. Furthermore, using the joint detection, the trellis-coded NOMA outperforms TCMA at high-SNR in Figs. 2.8 and 2.9. Moreover, in the trellis-coded NOMA, the signals of different users can also employ different constellations. The curves with "TC-NOMA, Joint, Rotate" in Fig. 2.9 are for the case where the constellation used by one user is the other user's constellation rotated by $\pi/8$. The trellis-coded NOMA with constellation rotation achieves a better performance compared with the trellis-coded or uncoded NOMA without constellation rotation and TCMA. It can be explained intuitively by considering how the constellation rotation affects the Euclidean distance between superimposed symbols. According to Figs. 2.4 and 2.6, the minimum Euclidean distance may increase if the constellation of User 1's signal rotates by $\pi/8$, which then improves the performance.

Fig. 2.10 shows how the average BER changes with the power ratio P_1/P_2 using the joint detection at SNRs 16dB and 18dB for the 4-state trellis in Fig. 2.2 and the 8-state trellis

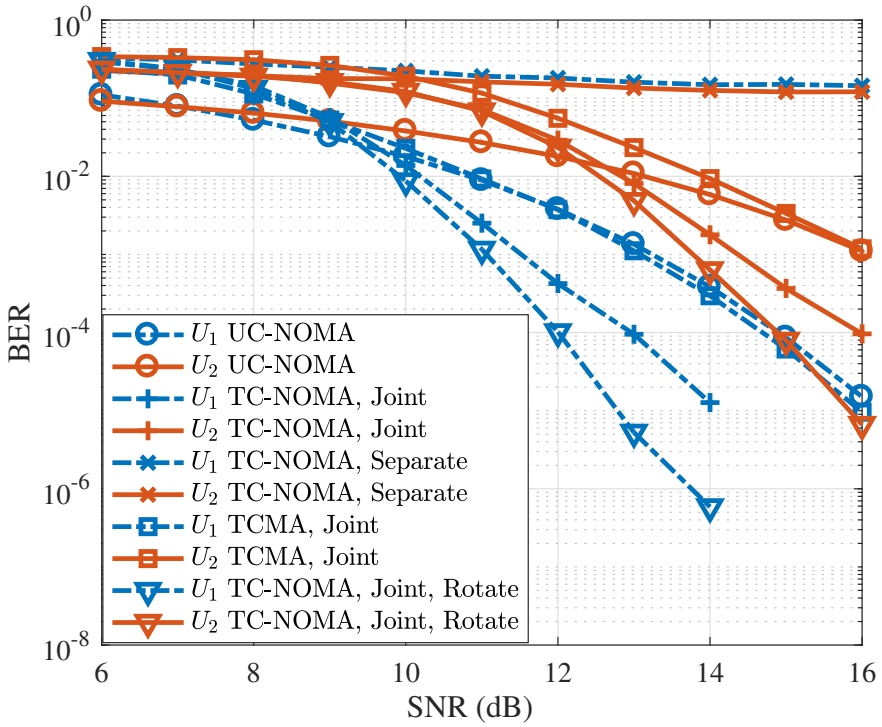

Figure 2.9: BER vs. SNR for TCMA, uncoded and trellis-coded NOMA when $P_1 = 0.3$, $P_2 = 1$, $|h_1|^2 = 2$, $|h_2|^2 = 1$.

in Fig. 12.8 of [26]. It is shown that the minimum BER is achieved when $P_1/P_2 \approx 0.25$ for the uncoded NOMA and the 8-state trellis-coded NOMA. The optimal power ratio for the 4-state trellis-coded NOMA is 0.24 for SNR=16dB and 0.22 for SNR=18dB, which are close to the optimal power ratio of 0.2404 derived in Section 2.4. Moreover, the trellis-coded NOMA in its best case scenario outperforms the uncoded NOMA in its best case scenario. Besides, the performance of TCMA does not change with P_1/P_2. By choosing the proper powers, the trellis-coded NOMA outperforms TCMA.

2.6 Conclusions

In this chapter, we study the trellis-coded NOMA and propose a joint detection method based on the tensor product of trellises. Besides, we derive the optimal power allocation between the

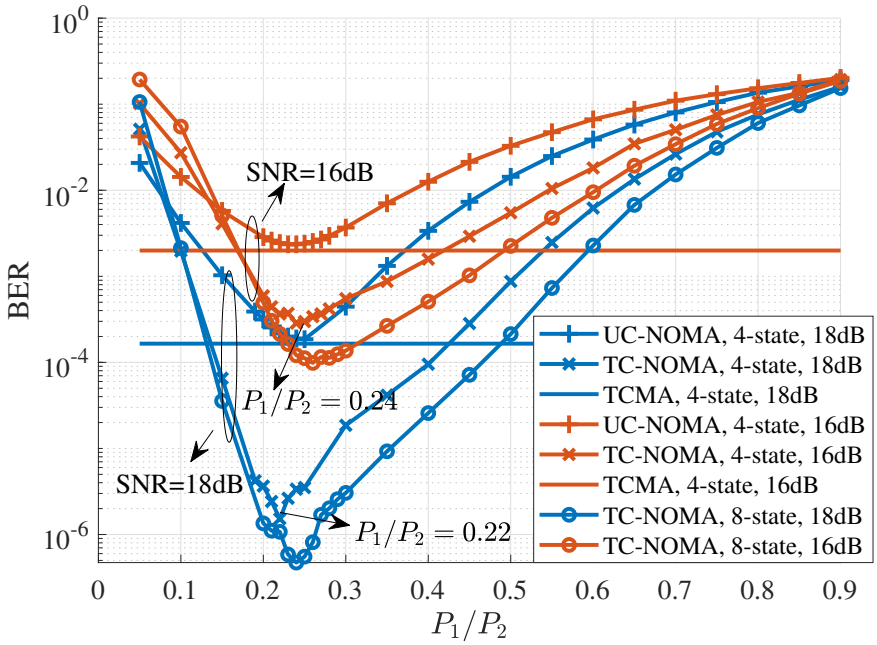

Figure 2.10: BER vs. P_1/P_2 for TCMA, uncoded and trellis-coded NOMA schemes at users employing the joint detection when $P_1 + P_2 = 1$.

two users by maximizing the free distance of the tensor product trellis. Simulation results demonstrate that the trellis-coded NOMA outperforms the uncoded NOMA and TCMA using an appropriate power allocation. The study of the trellis-coded NOMA systems with more than two users is our future work.

Chapter 3

Uplink Asynchronous Non-Orthogonal Multiple Access

3.1 Introduction

Applying the symbol-asynchronous transmission to NOMA, a scheme named asynchronous NOMA (ANOMA) was studied in [27]. In fact, an idea similar to the ANOMA in [27], i.e., applying asynchronous transmission for multiple access, has also been proposed and investigated in, e.g., [11, 7, 8]. Specially, a timing mismatch between signals for different users is intentionally added as an additional resource to address the problem of inter-user interference. It has been shown using the numerical simulation in [27] that the ANOMA outperforms the conventional (synchronized) NOMA by achieving a larger throughput.

However, the work in [27] has several limitations. While addressing those limitations is important to understand the ANOMA systems, to the best of our knowledge, no existing paper has tackled the following issues. First, there is no analytical result on the comparison between the performance of the ANOMA and that of the NOMA in terms of the throughput,

although numerically it is shown that ANOMA outperforms NOMA in certain scenarios. This is probably because the existing expression for the throughput of the ANOMA system is given as a function of the channel matrix but not the signal-to-noise ratio (SNR). The lack of such an expression in terms of SNR makes the analytical comparison between NOMA and ANOMA almost intractable. Second, the optimal design of ANOMA has not been investigated. Despite the fact that the performance of ANOMA is directly affected by important design parameters such as the transmit power and the timing mismatch, existing papers mainly focused on the performance demonstration only. Third, the impact of timing error has not been studied on the ANOMA systems. In the ANOMA systems, existing studies ideally assumed that the timing information was perfectly known. However, the timing information in practice cannot always be perfectly obtained, and the timing error is often inevitable. The timing information plays a vital role in ANOMA systems, since oversampling is designed using the timing information [6, 7]. Note that the impact of timing error has been widely studied on different communication systems, such as [28] on the direct-sequence code division multiple-access (DS-CDMA) system, [29] on the multi-carrier code division multiple-access (MC-CDMA) system, and [30] on the multiple-input single-output (MISO) system using distributed OSTBC to name a few.

In this chapter, we comprehensively investigate the ANOMA in a two-user uplink system. The primary contributions of the chapter are summarized as follows:

1. For the first time, we analytically prove that the ANOMA with a sufficiently large frame length can always outperform the NOMA in terms of the system sum throughput. To this end, we derive the expression for the sum throughput of the ANOMA system as a function of SNR, frame length, and normalized timing mismatch. A simplified throughput expression is further obtained for the asymptotic case of infinite frame length.

2. We investigate the optimal design of the two-user uplink ANOMA system aiming at

maximizing the sum throughput. We find that each user should transmit at full power despite the negative effect of inter-user interference. In addition, we prove that the optimal timing mismatch converges to one half of a symbol time as the frame length goes to infinity.

3. We analyze the impact of timing error on the performance of the uplink ANOMA system. Two types of timing error are taken into consideration, i.e., the synchronization timing error and the coordination timing error, which account for the timing error caused in signal synchronization and the coordination of the timing mismatch between asynchronous signals, respectively. We derive the expressions for the throughput loss of the ANOMA system with respect to both types of timing error, and analyze how the synchronization timing error and the coordination timing error individually and jointly affect the system performance.

Compared with our conference version [31] which briefly analyzed the impact of timing error in the ANOMA systems, the new analyses incorporated in this chapter include, e.g., the sum throughput analysis of the ANOMA, the analytical comparisons between ANOMA and NOMA, the asymptotic analysis for a large frame length, and the optimal transmit power and timing mismatch designs. The remainder of the chapter is organized as follows. The two-user uplink system model is presented in Section 3.2. The performance of the ANOMA system is analyzed in Section 3.3. We discuss the optimal design of the ANOMA system in Section 3.4. We analyze the outputs of ANOMA matched filters with timing error and the throughput loss incurred by timing error in Section 3.5. Numerical results are presented in Section 3.6. Finally, we draw the conclusions in Section 3.7.

Notations: $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^T$ denotes the transpose, $\mathrm{Tr}(\cdot)$ denotes the trace operation, $(\cdot)^{-1}$ denotes the inverse operation, $|x|$ denotes the absolute value of x, $\mathbb{E}[\cdot]$ denotes the expectation operation, $\mathcal{CN}(0,1)$ denotes the complex normal distribution with zero mean and unit variance, and $\mathbf{1}(\cdot)$ denotes the unit step function whose value is

zero for negative arguments and one for positive arguments.

3.2 System Model

In this chapter, we consider an uplink system which consists of a single base station (BS) and two users, as shown in Fig. 3.1. The two users share the same frequency-time resource to transmit signals to the BS. We assume that perfect channel state information (CSI) is known at the BS via the uplink channel training.

3.2.1 ANOMA System

For the ANOMA, a timing mismatch is intentionally introduced between the symbols from two users. By intentionally introducing the timing mismatch, the oversampling technique can be adopted at the receiver, so that extra linearly independent samples can be obtained to have sampling diversity. Then, the performance of the ANOMA can be improved by utilizing the sampling diversity. In contrast, one cannot get sampling diversity by oversampling the synchronized signals in conventional NOMA schemes. As shown in Fig. 3.2, the intended timing mismatch between the symbols of Users 1 and 2 is denoted by τT, where T is the duration of each symbol and τ, $0 \leq \tau < 1$, is the normalized timing mismatch. Note that the ANOMA system becomes a synchronous NOMA system when $\tau = 0$. In this section, we assume that τ is perfectly known at BS via timing offset estimation and uplink timing control techniques, such as the timing advance [32]. We will study the ANOMA system with timing error in Section 3.5.

Let $a_1[i] = h_1\sqrt{P_1}s_1[i]$ and $a_2[i] = h_2\sqrt{P_2}s_2[i]$, where the subscripts 1 and 2 denote the parameters for Users 1 and 2, respectively, $s_j[i]$ $(j = 1, 2)$ denotes the ith normalized transmitted symbol, h_j denotes the channel coefficient in the block of transmission, and P_j denotes

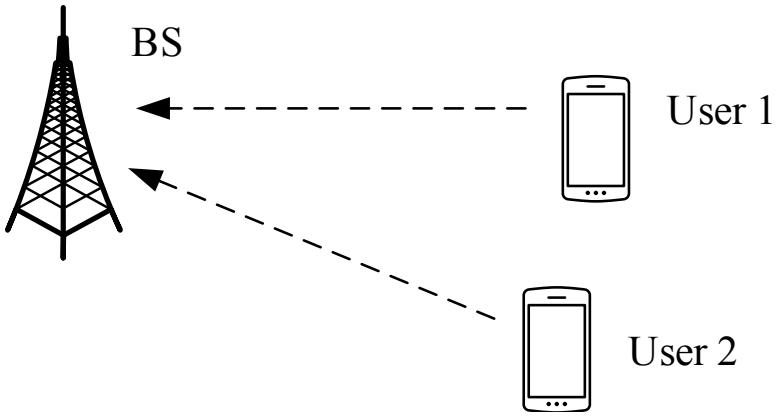

Figure 3.1: Illustration of a two-user uplink system.

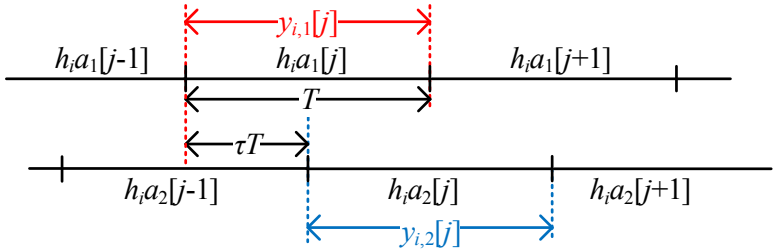

Figure 3.2: Illustration of the sampling for ANOMA.

the transmit power. The BS's received signal at time t is then given by

$$y(t) = \sum_{i=0}^{N-1} a_1[i]p(t - iT) + \sum_{i=0}^{N-1} a_2[i]p(t - iT - \tau T) + n(t), \tag{3.1}$$

where N denotes the number of symbols in a frame, i.e., the frame length, T denotes the time duration of one symbol, $p(\cdot)$ denotes the pulse-shaping filter, and $n(t) \sim \mathcal{CN}(0, 1)$ denotes the normalized additive white Gaussian noise (AWGN). Without loss of generality, the rectangular pulse shape is adopted, i.e., $p(t) = 1/\sqrt{T}$ when $t \in [0, T]$ and $p(t) = 0$ when $t \notin [0, T]$. With the block fading model, we assume that the channels remain static during the transmission of N consecutive symbols. As an initial study on the ANOMA systems, our work adopts a basic flat fading channel model without the consideration of a frequency selective channel and OFDM. The assumption of flat fading has been widely used in the existing literature; see, e.g., [2, 9, 33, 27], and the references therein.

The BS uses the oversampling technique to take advantage of sampling diversity in the

asynchronous systems [6, 31]. Oversampling uses the matched filter to sample signals at instants iT and $(i+\tau)T$, $i = 1, \cdots, N$, which produces $2N$ samples without doubling the sampling rate. It has been shown in [5] that the samples obtained by oversampling are sufficient statistics for the transmitted messages in the symbol-asynchronous scenario.

As shown in Fig. 3.2, the BS obtains two sample vectors, denoted by $[y_1[1], \cdots, y_1[N]]^T$ and $[y_2[1], \cdots, y_2[N]]^T$. Specifically, the ith element in the first sample vector is given by

$$\begin{aligned} y_1[i] &= \int_0^\infty y(t)p(t-iT)dt \\ &= \int_0^\infty a_1[i]p(t-iT)p(t-iT)dt + \int_0^\infty \{a_2[i-1]p(t-(i+1+\tau)T) \\ &+a_2[i]p(t-(i+\tau)T)\}\, p(t-iT)dt + n_1[i] \\ &= a_1[i] + \tau a_2[i-1] + (1-\tau)a_2[i] + n_1[i], \end{aligned} \tag{3.2}$$

where $n_1[i] = \int_0^\infty n(t)p(t-iT)dt$ denotes the additive noise in the first sampled vector. The ith element in the second sample vector is given by

$$y_2[i] = \int_0^\infty y(t)p(t-iT-\tau T)dt = a_2[i] + \tau a_1[i+1] + (1-\tau)a_1[i] + n_2[i], \tag{3.3}$$

where $n_2[i] = \int_0^\infty n(t)p(t-iT-\tau T)dt$ denotes the additive noise in the second sampled vector. From (3.2) and (3.3), we note that the inter-user interference exists, since the symbols for Users 1 and 2, i.e., $a_1[i]$ and $a_2[i]$, are added together to interfere with each other in the received samples.

We can write the outputs at the BS in a matrix form as

$$\mathbf{Y} = \mathbf{RHX} + \mathbf{N}, \tag{3.4}$$

where

$$\mathbf{Y} = [y_1[1]\ y_2[1]\ y_1[2]\ y_2[2]\ \cdots\ y_1[N]\ y_2[N]]^T, \tag{3.5}$$

$$\mathbf{X} = [s_1[1]\ s_2[1]\ s_1[2]\ s_2[2]\ \cdots\ s_1[N]\ s_2[N]]^T, \tag{3.6}$$

$$\mathbf{N} = [n_1[1]\ n_2[1]\ n_1[2]\ n_2[2]\ \cdots\ n_1[N]\ n_2[N]]^T, \tag{3.7}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1 & \tau & 0 & \cdots & 0 \\ 0 & \tau & 1 & 1-\tau & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \tau & 1 & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1 \end{bmatrix}, \tag{3.8}$$

and

$$\mathbf{H} = \begin{bmatrix} h_1\sqrt{P_1} & & & & \\ & h_2\sqrt{P_2} & & & \\ & & \ddots & & \\ & & & h_1\sqrt{P_1} & \\ & & & & h_2\sqrt{P_2} \end{bmatrix}. \tag{3.9}$$

We note that the inter-user interference is represented in matrix $\mathbf{R}$. If there is no inter-user interference, $\mathbf{R}$ becomes a diagonal matrix. With inter-user interference, $\mathbf{R}$ is given as (3.8) which is a tridiagonal matrix, but not a diagonal matrix.

We assume that the transmitted symbols are independent, such that $\mathbb{E}\left[\mathbf{X}\mathbf{X}^H\right] = \mathbf{I}$. Note that the noise terms in (3.4) are colored due to the oversampling, and we have

$$\mathbb{E}\left\{n_1[i]n_2^H[i]\right\} = \int_0^\infty \int_0^\infty \mathbb{E}\left\{n(t)n^H(s)\right\} p\left(t - iT\right) p\left(s - iT - \tau T\right) dtds = 1 - \tau. \tag{3.10}$$

Thus, the covariance matrix of $\mathbf{N}$ is given by

$$\mathbf{R}_\mathbf{N} = \mathbb{E}\left\{\mathbf{N}\mathbf{N}^H\right\} = \mathbf{R}. \tag{3.11}$$

As an initial study on ANOMA, the analysis in this chapter focuses on the rectangular

pulse shape. It is worth mentioning that the ANOMA with oversampling is also applicable to other pulse shapes, such as the raised cosine pulse shape. Since the raised cosine pulse shape spans more than T, it causes more severe inter-user interference compared with the rectangular pulse shape. In terms of the mathematical expression, the matrix $\mathbf{R}$ in (3.8) changes accordingly while (3.11) still holds. The (i,j)th entry of $\mathbf{R}$ for the raised cosine pulse shape is given as

$$\mathbf{R}_{i,j} = \begin{cases} \rho\left(\tau T + \lfloor\frac{j-i}{2}\rfloor T\right), & i \text{ is odd, } j \text{ is even} \\ \rho\left(\tau T + \lfloor\frac{i-j}{2}\rfloor T\right), & i \text{ is even, } j \text{ is odd} \\ \rho\left(\frac{j-i}{2}T\right), & \text{otherwise,} \end{cases} \tag{3.12}$$

where ρ is the auto-correlation function of the raised cosine function. The rest of the throughput formulas remain the same.

3.2.2 Benchmark System – NOMA

By setting $\tau = 0$, the ANOMA system becomes the synchronous NOMA system. For the NOMA, the BS uses the typical matched filter instead of the oversampling technique, and the ith sample at the BS can be written as

$$y[i] = a_1[i] + a_2[i] + n[i], \tag{3.13}$$

where $n[i] = \int_0^\infty n(t)p(t - iT)dt$. Note that (3.13) can be derived from either (3.2) or (3.3) by letting $\tau = 0$.

It is worth mentioning that the OMA systems also have the asynchrony issue in practice. For OMA, different users transmit signals using orthogonal resources. For TDMA, orthogonal time resources are allocated to different users. The asynchrony in time domain will destroy

such orthogonality, which will then degrade the system performance. For orthogonal frequency division multiple access (OFDMA), orthogonal frequency resources are allocated to different users. The asynchrony in frequency domain, for example in the form of frequency offsets, will destroy such orthogonality, which will then degrade the system performance. In general, the asynchrony does not benefit the OMA schemes.

3.3 Performance Analysis of ANOMA Systems

In this chapter, we employ the widely used performance metric, i.e., the sum throughput, to investigate the rate performance limit of the ANOMA systems. From (3.4), the sum throughput of the two-user uplink ANOMA system can be written as

$$R^{\text{ANOMA}} = \frac{1}{N+\tau} \log\det\left(\mathbf{I}_{2N} + \mathbf{H}\mathbf{H}^H\mathbf{R}\right). \tag{3.14}$$

Some existing papers, e.g., [27], define the throughput of ANOMA as

$$R^{\text{ANOMA}}_{\text{exist}} = \frac{1}{N} \log\det\left(\mathbf{I}_{2N} + \mathbf{H}\mathbf{H}^H\mathbf{R}\right), \tag{3.15}$$

which is slightly different from (3.14). Although (3.14) and (3.15) converge to the same expression as $N \to \infty$, we highlight that our adopted expression in (3.14) is more accurate than (3.15) for evaluating the throughput of ANOMA with finite frame length N, since the system actually spends $N + \tau$ instead of N symbol times to transmit N symbols for each user.

It is worth mentioning that the practical transmission scheme may even simply allocate $N + 1$ instead of $N + \tau$ symbol times for the transmission, and the throughput becomes $R^{\text{ANOMA}}_{N+1} = \frac{1}{N+1} \log\det\left(\mathbf{I}_{2N} + \mathbf{H}\mathbf{H}^H\mathbf{R}\right)$. Our analysis is still applicable to that case, since

one can simply revise (most) results according to $R_{N+1}^{\text{ANOMA}} = \frac{N+\tau}{N+1} R^{\text{ANOMA}}$. It is also worth mentioning that (3.14) and (3.15) are based on two main assumptions. The first assumption is the symbol-level asynchrony. That is, there is a timing mismatch τT ($\tau \in [0, 1)$) between symbols from different users. The second assumption is that the timing mismatch is perfectly known at BS, which is a common assumption in the existing literature, e.g., [7, 9, 27]. We will analyze the impact of timing error in the case that the timing information is not perfectly known in Section 3.5.

In the following theorem, we derive the sum throughput of the two-user uplink ANOMA system in terms of the receive SNRs, $\mu_1 = P_1|h_1|^2$ and $\mu_2 = P_2|h_2|^2$, the normalized timing mismatch, τ, and the frame length, N.

Theorem 3.1. *The sum throughput of the two-user uplink ANOMA system is derived as*

$$R^{\text{ANOMA}} = \frac{N}{N+\tau} \log(\mu_1 \mu_2) + \frac{1}{N+\tau} \log \frac{(r_1^{N+1} - r_2^{N+1}) + \tau^2 (r_1^N - r_2^N)}{r_1 - r_2}, \tag{3.16}$$

where

$$r_1 = \frac{\mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1} + 2\tau(1-\tau)}{2} + \frac{\sqrt{\left[\mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1} + 2\tau(1-\tau)\right]^2 - 4\tau^2(1-\tau)^2}}{2}, \tag{3.17}$$

$$r_2 = \frac{\mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1} + 2\tau(1-\tau)}{2} - \frac{\sqrt{\left[\mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1} + 2\tau(1-\tau)\right]^2 - 4\tau^2(1-\tau)^2}}{2}. \tag{3.18}$$

Proof. See Appendix A.1. □

Based on Theorem 3.1, we present the throughput of the two-user uplink ANOMA system for the asymptotic case of $N \to \infty$ in the following corollary, which characterizes the limiting performance of the system when the frame length N is large.

Corollary 3.1. *The throughput of the two-user uplink ANOMA system in the asymptotic case of $N \to \infty$ is given by*

$$\lim_{N\to\infty} R^{\text{ANOMA}} = \log\left(\mu_1\mu_2 r_1\right). \tag{3.19}$$

Proof. See Appendix A.2. □

3.3.1 Comparison with NOMA

In NOMA systems, SIC is adopted at BS to decode transmitted symbols, which works as follows. The BS first decodes the message from User 1 (stronger user) while treating the codeword from User 2 (weaker user) as an extra source of interference or noise. Then, the BS subtracts the decoded message from the received signal, and decodes the message from User 2.

According to (3.13), with perfect SIC at BS, the sum throughput of the two users in the uplink NOMA system can be written as [34]

$$R^{\text{NOMA}} = \log(1 + P_1|h_1|^2 + P_2|h_2|^2) = \log(1 + \mu_1 + \mu_2), \tag{3.20}$$

which can also be obtained from (3.16) by setting $\tau = 0$.

Due to the complicated expression for the throughput of ANOMA in (3.16), it is difficult to analytically compare the NOMA with the ANOMA for a general value of N. Instead, we provide numerical results in Section 3.6 and consider the asymptotic case of $N \to \infty$ for an analytical comparison in the following theorem. Our asymptotic analysis here aims to provide useful insights for ANOMA in the scenario where the frame length, N, is relatively large. Most results in the chapter are valid for any arbitrary value of N, e.g., the throughput analysis, the optimal transmit power design, and the impact of timing error.

Theorem 3.2. *The two-user uplink ANOMA system as $N \to \infty$ achieves an equal or higher throughput compared with the NOMA system, i.e.,*

$$\lim_{N\to\infty} R^{\mathrm{ANOMA}} \geq R^{\mathrm{NOMA}}, \tag{3.21}$$

where $\lim_{N\to\infty} R^{\mathrm{ANOMA}} = R^{\mathrm{NOMA}}$ *if and only if the normalized timing mismatch* $\tau = 0$.

Proof. See Appendix A.3. □

We note that the oversampling in ANOMA enables the sampling diversity, which leads to the possible performance advantage of the ANOMA compared with the NOMA. On the other hand, an extra τT time resource is used to transmit N symbols in the ANOMA, which has a negative effect on the performance of the ANOMA compared with that of the NOMA. When N is small, the negative effect of the extra transmission time dominates the sum throughput of ANOMA. As N grows, the effect of the extra transmission time becomes negligible and the sampling diversity dominants, which results in Theorem 3.2. Thus, one can expect a better ANOMA throughput performance compared with the NOMA when N is larger than a certain value.

The physical meaning of (3.21) is further clarified by the following corollary.

Corollary 3.2. *With a sufficiently large frame length, the ANOMA outperforms the NOMA for the two-user uplink system in terms of the sum throughput.*

In general, the ANOMA system requires a higher detection complexity compared with the NOMA system. ANOMA adopts the oversampling technique to obtain more samples of the signal compared with the conventional NOMA system. Thus, the decoding process of the ANOMA system involves a larger number of samples compared with that of the NOMA system. Maximum-likelihood sequence detection, with relatively high complexity, can be

used to decode the transmitted messages. For the low-complexity decoding methods in the ANOMA systems, one can employ the methods in the existing literature with some modifications, e.g., SIC with hard decision passing and the forward backward belief propagation detection proposed in [7] or the low complexity decoder using dynamic programming in [12].

3.4 Design of ANOMA Systems

From the analysis in Section 3.3, we note that the throughput performance of the uplink ANOMA system is directly determined by the transmit powers and the normalized timing mismatch, i.e., P_1, P_2, and τ. In this section, we investigate the optimal P_1, P_2, and τ that maximize the throughput of the system.

The design problem is formulated as follows:

$$\begin{aligned} &\arg\max_{P_1,P_2,\tau} \quad R^{\text{ANOMA}}, \\ &\quad s.t. \quad 0 \leq \tau < 1,\ 0 \leq P_1 \leq P_{1,\max},\ 0 \leq P_2 \leq P_{2,\max}, \end{aligned} \tag{3.22}$$

where $P_{1,\max}$ and $P_{2,\max}$ are the maximum available powers at which Users 1 and 2 can transmit, respectively. Note that the transmit powers are coupled together in a complicated way in the expression for the throughput of the ANOMA system in (3.16), which is different from the case of NOMA in (3.20). Thus, the optimal transmit powers for the ANOMA system are not easy to determine, while it is easy to find that we shall use the maximum available transmit powers at users for the NOMA system.

It is worth mentioning that the performance of the uplink ANOMA system is also affected by the frame length, N. However, the frame length is constrained by the channel condition, i.e., the length of each block of the block fading channel, and the acceptable transceiver

complexity. Hence, we do not investigate the design of N in this work and assume that it is fixed based on the channel conditions and the overall system design.

3.4.1 Optimal Transmit Power

We first obtain the optimal transmit power scheme. We summarize the optimal transmit powers for the ANOMA system as follows.

Theorem 3.3. *For the two-user uplink ANOMA system with any frame length, N, and the normalized timing mismatch, τ, the optimal transmit powers at Users 1 and 2, P_1^* and P_2^*, are equal to the maximum available powers at which Users 1 and 2 can transmit, $P_{1,\max}$ and $P_{2,\max}$, i.e., $P_1^* = P_{1,\max}$ and $P_2^* = P_{2,\max}$*

Proof. See Appendix A.4. □

From Theorem 3.3, we find that the optimal design of transmit powers for the two-user uplink ANOMA system is the same as that for the NOMA system.

3.4.2 Optimal Normalized Timing Mismatch

We now study the optimal normalized timing mismatch, τ^*. The optimal normalized timing mismatch, τ^*, is analytically intractable for a general finite frame length N, while we can numerically obtain τ^* for a given finite N by simply searching in the range of $0 \leq \tau < 1$. In addition, we present τ^* in the asymptotic case of $N \to \infty$ in the following theorem.

Theorem 3.4. *For the two-user uplink ANOMA system with the frame length $N \to \infty$, the optimal normalized timing mismatch to maximize the sum throughput is given by $\tau^* = 0.5$.*

Proof. See Appendix A.5. □

3.5 Impact of Timing Error on ANOMA systems

The analysis in the previous sections is based on the assumption that the BS perfectly knows the timing information. However, the timing information cannot always be perfectly obtained in practice, and the timing error is often inevitable. In this section, we study the impact of timing error on the ANOMA system.

3.5.1 Timing Error

We consider two types of timing error for the ANOMA system, i.e., the synchronization timing error and the coordination timing error.

3.5.1.1 Synchronization Timing Error

To synchronize the signals, we need a reference signal. Without loss of generality, we use the signal from User 1 as the timing reference (the timing offset is 0). This requires a symbol-level timing synchronization with User 1 at the BS, as it is also done in NOMA. The normalized synchronization timing error, denoted by ϵ_1 in Fig. 3.3, is due to the imperfect timing synchronization. Without loss of generality, we assume that $\epsilon_1 \in (\tau - 1, \tau)$. With the synchronization timing error, $y_1[i]$ is taken from the time $(i-1)T + \epsilon_1 T$ to $iT + \epsilon_1 T$ and $y_2[i]$ is taken from the time $(i-1)T + (\tau + \epsilon_1)T$ to $iT + (\tau + \epsilon_1)T$, although the BS intends to take $y_1[i]$ from the time $(i-1)T$ to iT and $y_2[i]$ from the time $(i-1)T + \tau T$ to $iT + \tau T$. We will study the effect of this timing error later.

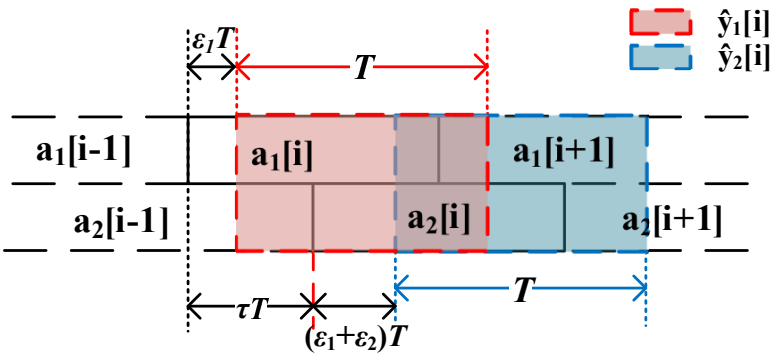

Figure 3.3: Illustration of the sampling for ANOMA with timing error.

3.5.1.2 Coordination Timing Error

In order to achieve the desired timing mismatch between the two signals, the BS coordinates the uplink transmission timing of the two users to add the intended timing offsets at each transmitter. For example, the timing advance is the technique employed in long term evolution (LTE) systems to estimate and adjust the timing offsets among uplink signals at BS [35, 32]. The normalized coordination timing error, denoted by ϵ_2 in Fig. 3.3, results from the imperfect coordination between the users. With the coordination timing error, the actual timing mismatch becomes $(\tau+\epsilon_2)T$, while the intended timing mismatch is τT. In addition to the synchronization timing error $\epsilon_1 T$, the sample $y_2[i]$ is taken from $(i-1)T+(\tau+\epsilon_1+\epsilon_2)T$ to $iT+(\tau+\epsilon_1+\epsilon_2)T$, although the BS intends to take $y_2[i]$ from $(i-1)T+\tau T$ to $iT+\tau T$. Without loss of generality, we assume that $\epsilon_1+\epsilon_2 \in (-\tau, 1-\tau)$.

Fig. 3.3 illustrates the sampling for an ANOMA system with timing error. It is worth mentioning that the sign of the timing error stands for the direction in which the function of the matched filter is shifted. For example, as shown in Fig. 3.3, the matched filter is shifted to the right by $\epsilon_1 T$ if $\epsilon_1 > 0$ compared with the matched filter designed with no timing error in Fig. 3.2. Fig. 3.3 only presents the case when $\epsilon_1 > 0$ and $\epsilon_1+\epsilon_2 > 0$, while our analysis works for any values of ϵ_1 and ϵ_2.

3.5.2 Outputs of ANOMA Matched Filters with Timing Error

In the presence of timing error, the ith element of the first sample vector is given by

$$\begin{aligned}
\hat{y}_1[i] &= \int_0^\infty y(t)p(t-iT-\epsilon_1 T)dt \\
&= \int_0^\infty a_1[i]p(t-iT)p(t-iT-\epsilon_1 T)dt \\
&+ \mathbf{1}(-\epsilon_1)\int_0^\infty a_1[i-1]p(t-(i-1)T)p(t-iT-\epsilon_1 T)dt \\
&+ \mathbf{1}(\epsilon_1)\int_0^\infty a_1[i+1]p(t-(i+1)T)p(t-iT-\epsilon_1 T)dt \\
&+ \int_0^\infty a_2[i-1]p(t-\tau T-(i-1)T)p(t-iT-\epsilon_1 T)dt \\
&+ \int_0^\infty a_2[i]p(t-\tau T-iT)p(t-iT-\epsilon_1 T)dt \\
&+ \int_0^\infty n(t)p(t-iT-\epsilon_1 T)dt \\
&= a_1[i](1-|\epsilon_1|)+a_1[i-1]\mathbf{1}(-\epsilon_1)(-\epsilon_1)+a_1[i+1]\mathbf{1}(\epsilon_1)\epsilon_1 \\
&+ a_2[i-1](\tau-\epsilon_1)+a_2[i](1-\tau+\epsilon_1)+\hat{n}_1[i], \qquad (3.23)
\end{aligned}$$

and the ith element of the second sample vector is given by

$$\begin{aligned}
\hat{y}_2[i] &= \int_0^\infty y(t)p(t-(i+\tau+\epsilon_1+\epsilon_2)T)dt \\
&= a_2[i](1-|\epsilon_1+\epsilon_2|)+a_2[i-1]\mathbf{1}(-\epsilon_1-\epsilon_2)(-\epsilon_1-\epsilon_2) \\
&+ a_2[i+1]\mathbf{1}(\epsilon_1+\epsilon_2)(\epsilon_1+\epsilon_2)+a_1[i](\tau-\epsilon_1-\epsilon_2) \\
&+ a_1[i+1](1-\tau+\epsilon_1+\epsilon_2)+\hat{n}_2[i], \qquad (3.24)
\end{aligned}$$

where $\hat{n}_1[i] = \int_0^\infty n(t)p(t-iT-\epsilon_1 T)dt$ and $\hat{n}_2[i] = \int_0^\infty n(t)p(t-(i+\tau+\epsilon_1+\epsilon_2)T)dt$.

We note from (3.23) and (3.24) that the first sample vector is affected by the normalized synchronization timing error ϵ_1 only, while the second sample vector is affected by the sum of the normalized synchronization timing error ϵ_1 and the normalized coordination timing

error ϵ_2.

With (3.23) and (3.24), we obtain the outputs of the two matched filters at the BS subject to the timing error in the matrix form as

$$\hat{\mathbf{Y}} = \hat{\mathbf{R}}\mathbf{H}\mathbf{X} + \hat{\mathbf{N}}, \tag{3.25}$$

where $\hat{\mathbf{Y}} = [\hat{y}_1[1]\ \hat{y}_2[1]\ \hat{y}_1[2]\ \hat{y}_2[2]\ \cdots\ \hat{y}_1[N]\ \hat{y}_2[N]]^T$, $\hat{\mathbf{N}} = [\hat{n}_1[1]\ \hat{n}_2[1]\ \hat{n}_1[2]\ \hat{n}_2[2]\ \cdots\ \hat{n}_1[N]\ \hat{n}_2[N]]^T$, and $\hat{\mathbf{R}}$ is given by

$$\begin{aligned}
\hat{\mathbf{R}} &= \begin{bmatrix}
1-|\epsilon_1| & 1-\tau+\epsilon_1 & \mathbf{1}(\epsilon_1)\epsilon_1 & 0 & \cdots & \cdots & 0 \\
1-\tau-\epsilon_1-\epsilon_2 & 1-|\epsilon_1+\epsilon_2| & \tau+\epsilon_1+\epsilon_2 & \mathbf{1}(\epsilon_1+\epsilon_2)(\epsilon_1+\epsilon_2) & 0 & \cdots & 0 \\
\mathbf{1}(-\epsilon_1)(-\epsilon_1) & \tau-\epsilon_1 & 1-|\epsilon_1| & 1-\tau+\epsilon_1 & \mathbf{1}(\epsilon_1)\epsilon_1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \mathbf{1}(-\epsilon_1-\epsilon_2)(-\epsilon_1-\epsilon_2) & 1-\tau-\epsilon_1-\epsilon_2 & 1-|\epsilon_1+\epsilon_2| & \tau+\epsilon_1+\epsilon_2 & \mathbf{1}(\epsilon_1+\epsilon_2)(\epsilon_1+\epsilon_2) \\
0 & \cdots & 0 & \mathbf{1}(-\epsilon_1)(-\epsilon_1) & \tau-\epsilon_1 & 1-|\epsilon_1| & 1-\tau+\epsilon_1 \\
0 & \cdots & \cdots & 0 & \mathbf{1}(-\epsilon_1-\epsilon_2)(-\epsilon_1-\epsilon_2) & 1-\tau-\epsilon_1-\epsilon_2 & 1-|\epsilon_1+\epsilon_2|
\end{bmatrix} \\
&= \mathbf{R} \\
&+ \underbrace{\begin{bmatrix}
-|\epsilon_1| & \epsilon_1 & \mathbf{1}(\epsilon_1)\epsilon_1 & 0 & \cdots & \cdots & 0 \\
-\epsilon_1-\epsilon_2 & -|\epsilon_1+\epsilon_2| & \epsilon_1+\epsilon_2 & \mathbf{1}(\epsilon_1+\epsilon_2)(\epsilon_1+\epsilon_2) & 0 & \cdots & 0 \\
\mathbf{1}(-\epsilon_1)(-\epsilon_1) & -\epsilon_1 & -|\epsilon_1| & \epsilon_1 & \mathbf{1}(\epsilon_1)\epsilon_1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \mathbf{1}(-\epsilon_1-\epsilon_2)(-\epsilon_1-\epsilon_2) & -\epsilon_1-\epsilon_2 & -|\epsilon_1+\epsilon_2| & \epsilon_1+\epsilon_2 & \mathbf{1}(\epsilon_1+\epsilon_2)(\epsilon_1+\epsilon_2) \\
0 & \cdots & 0 & \mathbf{1}(-\epsilon_1)(-\epsilon_1) & -\epsilon_1 & -|\epsilon_1| & \epsilon_1 \\
0 & \cdots & \cdots & 0 & \mathbf{1}(-\epsilon_1-\epsilon_2)(-\epsilon_1-\epsilon_2) & -\epsilon_1-\epsilon_2 & -|\epsilon_1+\epsilon_2|
\end{bmatrix}}_{\mathbf{E}_1}.
\end{aligned} \tag{3.26}$$

We note from (3.26) that the expression for $\mathbf{E}_1$ is related to the signs of ϵ_1 and $\epsilon_1 + \epsilon_2$. For the sake of brevity, we present the analytical results for the case of $\epsilon_1 > 0$ and $\epsilon_1 + \epsilon_2 > 0$ in the rest of the chapter, while our analytical method and findings are applicable to all cases. In addition, we will present the numerical results in Section 3.6 for all possible cases of ϵ_1

and $\epsilon_1 + \epsilon_2$. With $\epsilon_1 > 0$ and $\epsilon_1 + \epsilon_2 > 0$, the expression for $\mathbf{E}_1$ is rewritten as

$$\mathbf{E}_1 = \epsilon_1 \underbrace{\begin{bmatrix} -1 & 1 & 1 & 0 & \cdots & \cdots & 0 \\ -1 & -1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & -1 & 1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & -1 & 1 & 1 \\ 0 & \cdots & 0 & 0 & -1 & -1 & 1 \\ 0 & \cdots & \cdots & 0 & 0 & -1 & -1 \end{bmatrix}}_{\mathbf{Z}_1} + \epsilon_2 \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & -1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & -1 & 1 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & \cdots & 0 & 0 & -1 & -1 \end{bmatrix}}_{\mathbf{Z}_2}. \tag{3.27}$$

The covariance matrix of $\hat{\mathbf{N}}$ is given by

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{N}} &= \mathbb{E}\left\{\hat{\mathbf{N}}\hat{\mathbf{N}}^H\right\} \\ &= \begin{bmatrix} 1 & 1-\tau-\epsilon_2 & 0 & \cdots & \cdots & 0 \\ 1-\tau-\epsilon_2 & 1 & \tau+\epsilon_2 & 0 & \cdots & 0 \\ 0 & \tau+\epsilon_2 & 1 & 1-\tau-\epsilon_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \tau+\epsilon_2 & 1 & 1-\tau-\epsilon_2 \\ 0 & \cdots & \cdots & 0 & 1-\tau-\epsilon_2 & 1 \end{bmatrix} \\ &= \mathbf{R} + \underbrace{\begin{bmatrix} 0 & -\epsilon_2 & 0 & \cdots & \cdots & 0 \\ -\epsilon_2 & 0 & \epsilon_2 & 0 & \cdots & 0 \\ 0 & \epsilon_2 & 0 & -\epsilon_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \epsilon_2 & 0 & -\epsilon_2 \\ 0 & \cdots & \cdots & 0 & -\epsilon_2 & 0 \end{bmatrix}}_{\mathbf{E}_2}, \end{aligned} \tag{3.28}$$

where $\mathbf{E}_2$ can be rewritten as

$$\mathbf{E}_2 = \epsilon_2 \underbrace{\begin{bmatrix} 0 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 0 \end{bmatrix}}_{\mathbf{Z}_3}. \tag{3.29}$$

We note from (3.29) that the covariance matrix of the noise terms is affected by the normalized coordination timing error ϵ_2, while it is not related to the normalized synchronization timing error ϵ_1.

3.5.3 Impact of Timing Error on Throughput Performance

According to (3.25), the throughput of the ANOMA system with timing error is given by

$$\begin{aligned} R_e^{\text{ANOMA}} &= \frac{1}{N+\tau} \log\det\left(\mathbf{I}_{2N} + \hat{\mathbf{R}}_{\mathbf{N}}^{-1}\hat{\mathbf{R}}\mathbf{H}\mathbf{H}^H\hat{\mathbf{R}}^H\right) \\ &= \frac{1}{N+\tau} \log\det\left(\mathbf{I}_{2N}+(\mathbf{R}+\mathbf{E}_2)^{-1}(\mathbf{R}+\mathbf{E}_1)\mathbf{H}\mathbf{H}^H(\mathbf{R}+\mathbf{E}_1^H)\right) \\ &= \frac{1}{N+\tau} \log\det\left(\mathbf{I}_{2N} + \left(\mathbf{I}_{2N} + (\mathbf{R}+\mathbf{E}_2)^{-1}(\mathbf{E}_1-\mathbf{E}_2)\right)\mathbf{H}\mathbf{H}^H(\mathbf{R}+\mathbf{E}_1^H)\right) \\ &= \frac{1}{N+\tau} \log\det\left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R} + \mathbf{H}\mathbf{H}^H\mathbf{E}_1^H + (\mathbf{R}+\mathbf{E}_2)^{-1}\right. \\ &\quad \left.\cdot(\mathbf{E}_1-\mathbf{E}_2)\mathbf{H}\mathbf{H}^H(\mathbf{R}+\mathbf{E}_1^H)\right). \end{aligned} \tag{3.30}$$

When there is no timing error, i.e., $\epsilon_1 = \epsilon_2 = 0$, we have $\mathbf{E}_1 = \mathbf{E}_2 = \mathbf{0}$. Hence, substituting $\mathbf{E}_1 = \mathbf{E}_2 = \mathbf{0}$ into (3.30), we obtain the throughput of the ANOMA system without timing error, which is the same as (3.14).

From (3.14) and (3.30), we derive the throughput loss incurred by the timing error as

$$\begin{aligned} \Delta &= R^{\text{ANOMA}} - R_e^{\text{ANOMA}} \\ &= -\frac{1}{N+\tau} \log\det\left\{\mathbf{I}_{2N} + \left(\mathbf{I}_{2N} + \mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1}\left[\mathbf{H}\mathbf{H}^H\mathbf{E}_1^H\right.\right. \\ &\quad \left.\left.+(\mathbf{R}+\mathbf{E}_2)^{-1}(\mathbf{E}_1-\mathbf{E}_2)\mathbf{H}\mathbf{H}^H(\mathbf{R}+\mathbf{E}_1^H)\right]\right\}. \end{aligned} \tag{3.31}$$

In what follows, we separately analyze the throughput loss incurred by the synchronization timing error and the coordination timing error with the practical consideration that these two types of timing error both are relatively small.

3.5.3.1 Impact of Synchronization Timing Error

We first investigate the impact of synchronization timing error on the throughput loss and consider the practical scenario where the error is relatively small, such that $\epsilon_2 = 0$ and $\epsilon_1 \ll 1$.

In this case, by omitting high-order terms of ϵ_1, we obtain the throughput loss incurred by the synchronization timing error from (3.31) as

$$\begin{aligned}\Delta_{\epsilon_1} &= -\frac{1}{N+\tau}\log\det\left\{\mathbf{I}_{2N}+\epsilon_1\left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1}\left[\mathbf{H}\mathbf{H}^H\mathbf{Z}_1^H+\mathbf{R}^{-1}\mathbf{Z}_1\mathbf{H}\mathbf{H}^H(\mathbf{R}+\epsilon_1\mathbf{Z}_1^H)\right]\right\}\\ &\overset{(a)}{\approx} -\frac{1}{N+\tau}\log\det\left\{\mathbf{I}_{2N}+\epsilon_1\left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1}\left[\mathbf{H}\mathbf{H}^H\mathbf{Z}_1^H+\mathbf{R}^{-1}\mathbf{Z}_1\mathbf{H}\mathbf{H}^H\mathbf{R}\right]\right\}\\ &\overset{(b)}{\approx} -\frac{1}{N+\tau}\log\left(1+\epsilon_1\mathrm{Tr}(\mathbf{F}_1)+O(\epsilon_1^2)\right)\overset{(c)}{\approx}\epsilon_1 c_1,\end{aligned} \tag{3.32}$$

where $\mathbf{F}_1 = \left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1}\left(\mathbf{H}\mathbf{H}^H\mathbf{Z}_1^H+\mathbf{R}^{-1}\mathbf{Z}_1\mathbf{H}\mathbf{H}^H\mathbf{R}\right)$, $c_1 = -\frac{1}{N+\tau}\mathrm{Tr}(\mathbf{F}_1)$, (a) is approximated by using $\mathbf{R}+\epsilon_1\mathbf{Z}\approx\mathbf{R}$ as $\epsilon_1\to 0$, (b) is derived using the special case of Jacobi's formula [36], i.e., $\det(\mathbf{I}+\epsilon\mathbf{A}) = 1+\epsilon\mathrm{Tr}(\mathbf{A})+O(\epsilon^2)$, and (c) is derived by omitting the high-order terms of ϵ_1 and applying the approximation $\log(1+x)\approx x$ when $x\ll 1$. From (3.32), we note that the throughput loss is approximately linear to ϵ_1 when $\epsilon_2 = 0$ and $\epsilon_1 \ll 1$.

3.5.3.2 Impact of Coordination Timing Error

We now investigate the impact of the coordination timing error on the throughput loss and still consider the practical scenario where the error is relatively small, such that $\epsilon_1 = 0$ and $\epsilon_2 \ll 1$.

By omitting high-order terms of ϵ_2, we obtain the throughput loss incurred by the coordina-

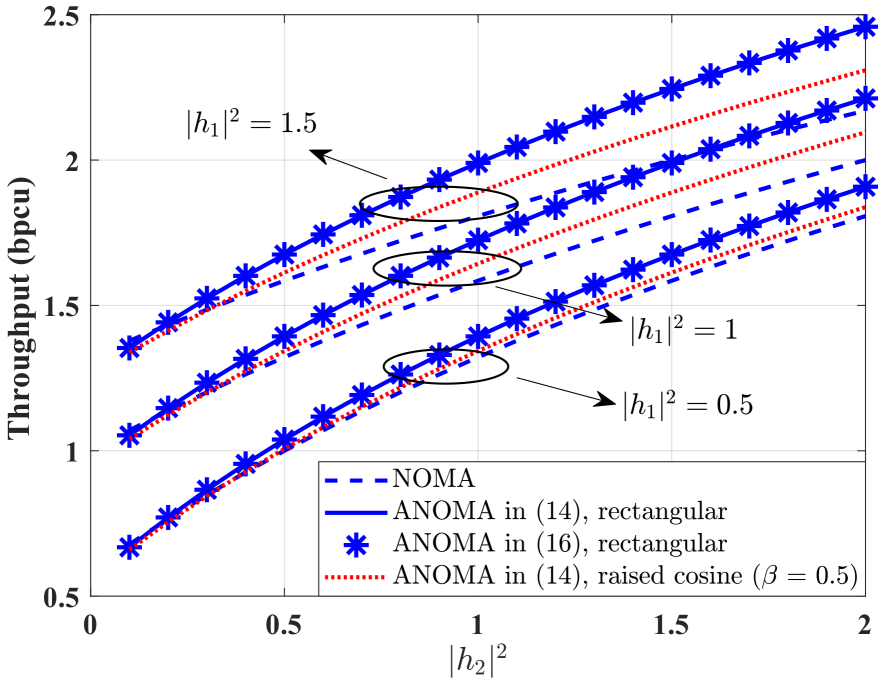

Figure 3.4: The sum throughput of two users as a function of channel gains for ANOMA and NOMA systems when $P_1 = 1$, $P_2 = 1$, $\tau = 0.5$, and $N = 10$.

tion timing error from (3.31) as

$$\begin{aligned}\Delta_{\epsilon_2} = &-\frac{1}{N+\tau}\log\det\Big\{\mathbf{I}_{2N} + \epsilon_2\left(\mathbf{I}_{2N} + \mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1} \\ &\cdot\left[\mathbf{H}\mathbf{H}^H\mathbf{Z}_2^H + (\mathbf{R}+\epsilon_2\mathbf{Z}_3)^{-1}(\mathbf{Z}_2-\mathbf{Z}_3)\mathbf{H}\mathbf{H}^H(\mathbf{R}+\epsilon_2\mathbf{Z}_2^H)\right]\Big\} \\ &\overset{(a)}{\approx} \epsilon_2 c_2,\end{aligned} \tag{3.33}$$

where $\mathbf{F}_2 = \left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right)^{-1}\left(\mathbf{H}\mathbf{H}^H\mathbf{Z}_2^H + \mathbf{R}^{-1}(\mathbf{Z}_2-\mathbf{Z}_3)\cdot\mathbf{H}\mathbf{H}^H\mathbf{R}\right)$, $c_2 = -\frac{1}{N+\tau}\mathrm{Tr}(\mathbf{F}_2)$, and (a) can be derived by following the same steps in the derivation of (3.32). From (3.33), we note that the throughput loss is approximately linear to ϵ_2 when $\epsilon_1 = 0$ and $\epsilon_2 \ll 1$.

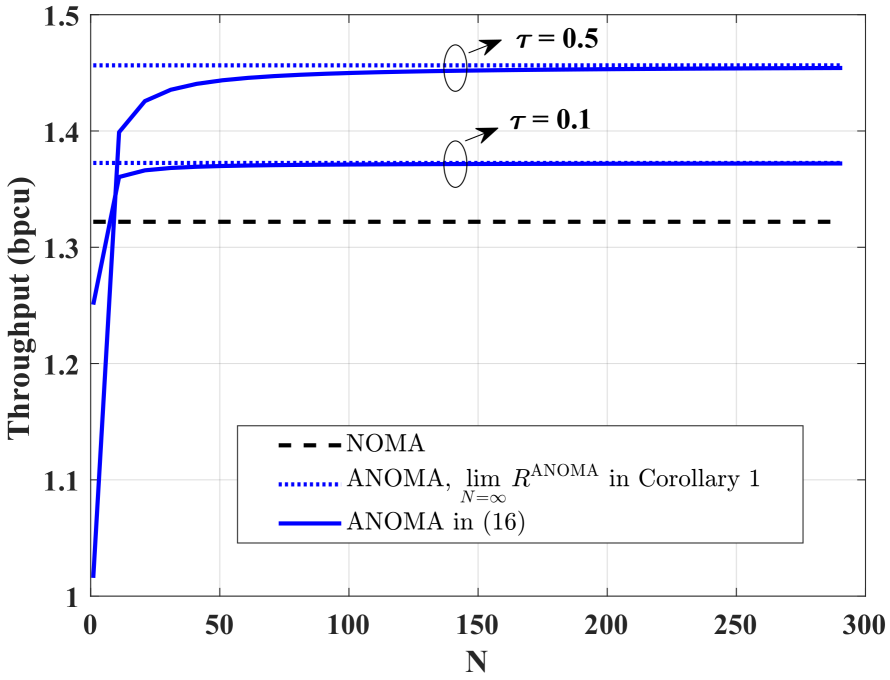

Figure 3.5: The sum throughput of two users as a function of the frame length N for ANOMA and NOMA systems when $P_1|h_1|^2 = 1$, $P_2|h_2|^2 = 0.5$, $\tau = 0.5$ or 0.1.

3.6 Numerical Results

In this section, we present numerical results to compare the throughput performances of NOMA and ANOMA systems and illustrate the impact of timing error on the performance of the ANOMA system. Figures 3.5, 3.6, and 3.7 show the ANOMA system without timing error while the other figures are for the impact of timing error. In our simulations, we set the symbol length $T = 1$ and the AWGN with unit power. If not specified, the normalized timing mismatch between the two signals τ is set to 0.5.

At first, we present the throughput performances of NOMA and ANOMA systems under different channel conditions and pulse shapes in Figure 3.4. In Fig. 3.4, the curves of "ANOMA in (3.14)" are derived directly from the definition in (3.14), and the curves of "ANOMA in (3.16)" are obtained from our result in Theorem 3.1. Note that the performance of NOMA is not affected by the adopted pulse shape if the pulse shape has unit power and

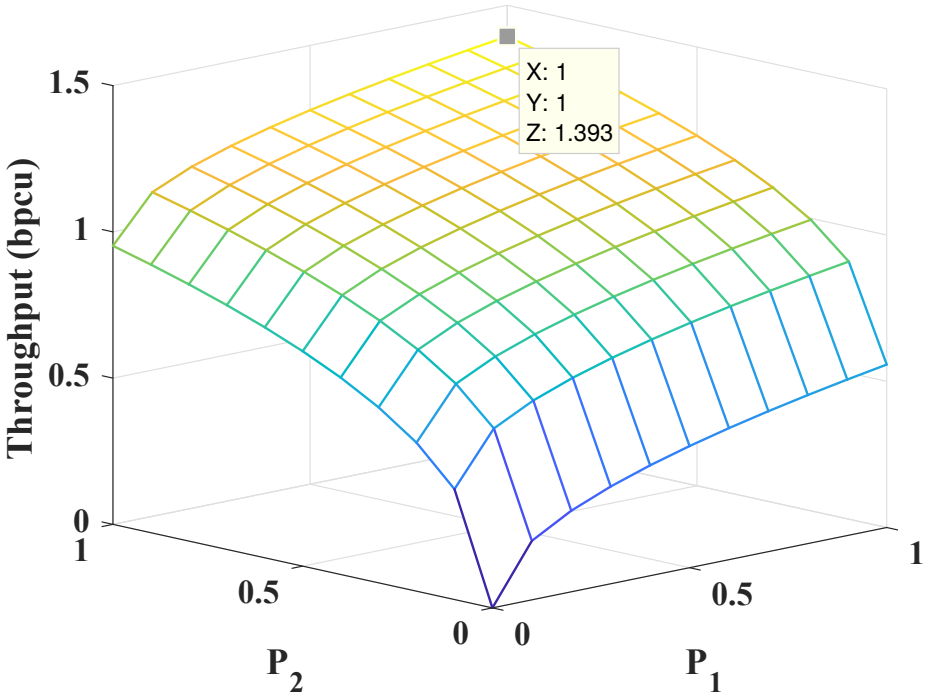

Figure 3.6: The sum throughput of two users as a function of transmit powers of Users 1 & 2 for ANOMA systems when $|h_1|^2 = 1$, $|h_2|^2 = 0.5$, $P_{1,\max} = P_{2,\max} = 1$, $\tau = 0.5$, and $N = 10$.

causes no inter-symbol interference. It is shown that, for the rectangular pulse shape, the throughput computed by Theorem 3.1 completely aligns with that calculated by (3.14) for different combinations of channel conditions, which confirms the correctness of Theorem 3.1. Besides, Fig. 3.4 demonstrates that the throughputs of ANOMA and NOMA systems increase with the channel gains $|h_1|^2$ and $|h_2|^2$ for both rectangular and raised cosine pulse shapes. Furthermore, the ANOMA systems using the rectangular pulse shape outperform those using the raised cosine pulse shape with roll-off factor $\beta = 0.5$. It is because the raised cosine pulse shape spans more than one symbol time, causing more severe interference compared with the rectangular pulse shape.

Then, we compare the throughput performances of NOMA and ANOMA systems. Figure 3.5 shows the throughput as a function of the frame length N. In Fig. 3.5, it is demonstrated that as N increases, the throughput of ANOMA systems converges to the result in Corollary

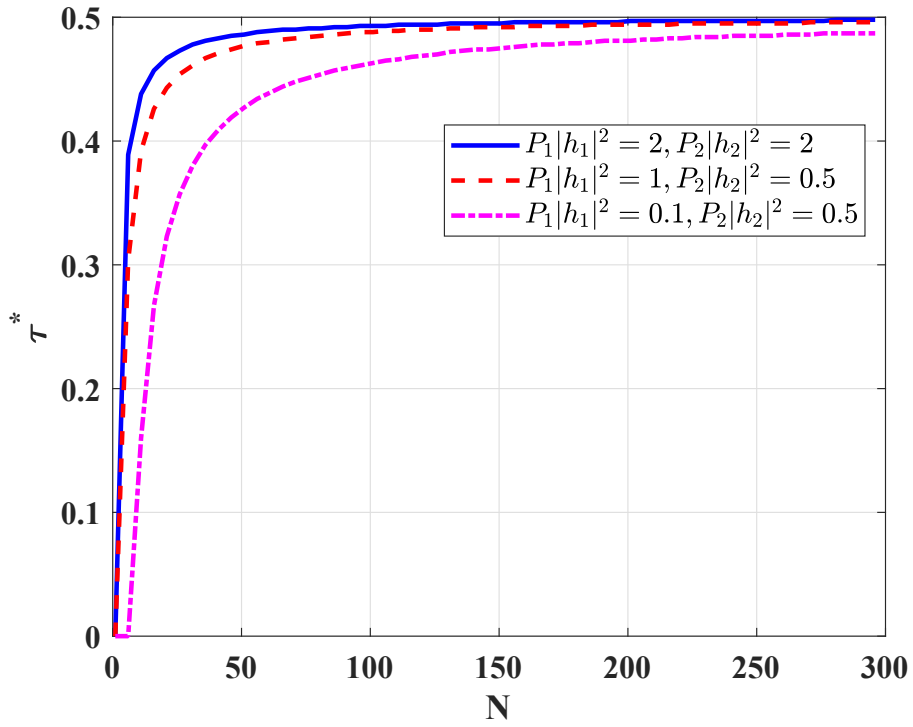

Figure 3.7: The optimal normalized timing mismatch τ^* to maximize the sum throughput of two users as a function of the frame length N for different channel conditions.

3.1, which provides the throughput in the asymptotic case of $N \to \infty$. Furthermore, we note from the figure that the throughputs of ANOMA systems for different τs as $N \to \infty$ are greater than that of the NOMA system, which is consistent with our analytical results in Theorem 3.2 and Corollary 3.2. In global system for mobile communications (GSM), there are approximately 156 symbols in a normal burst (a physical channel carrying information on traffic and control channels) [37]. In LTE, there are 140 symbols with normal cyclic prefix (CP) in a frame [32]. We find from Fig. 3.5 that the ANOMA outperforms the NOMA if N is greater than 20, which is much smaller than the number of symbols in a burst/frame of GSM/LTE. Since the needed frame length of ANOMA to outperform NOMA is much less than the burst/frame length in GSM/LTE systems, the detection delay is within a reasonable range.

In addition, we illustrate the optimal parameter design of the ANOMA system in Figures 3.6

and 3.7. Figure 3.6 demonstrates the sum throughput of two users as a function of their transmit powers. It is shown that the maximal sum throughput is reached when the transmit powers are equal to the maximum available powers, which aligns with Theorem 3.3. We present the optimal normalized timing mismatch τ^* found by exhaustive search to maximize the sum throughput of two users in Figure 3.7. As shown in Fig. 3.7, τ^* starts with 0, and then increases with N, finally converges to 0.5 as N grows, which verifies the correctness of Theorem 3.4.

Figure 3.5 demonstrates that ANOMA outperforms NOMA for $N \geq 20$ in the considered scenario. Also, from Fig. 3.7, we note that the optimal τ closely approaches 0.5 as $N \geq 50$. It is worth mentioning that having $N \geq 50$ is reasonable for practical communication systems. For a high-speed train traveling at 200 km/h using a 900 MHz carrier, the coherence time is approximately 3 ms [37]. In GSM (operating at 900 MHz), the symbol rate is approximately 271 ksymbols/second. As a result, it is reasonable to assume a static channel with flat fading if the frame length N does not exceed 3 ms $\times$ 271 ksymbols/second = 813, which is much greater than the threshold needed, i.e., 50.

In what follows, we evaluate the impact of timing error on the throughput of ANOMA systems. In the following figures, the throughput loss ratio is defined as the ratio of the throughput loss in (3.31) and the throughput of the ANOMA system without timing error in (3.14), i.e.,$\gamma = \frac{\Delta}{R^{\text{ANOMA}}}$.

In Fig. 3.8, we present the throughput loss ratio as a function of ϵ_1 and ϵ_2 ranging from -0.1 to 0.1. As shown in Fig. 3.8, the throughput loss ratio increases with both the synchronization timing error and the coordination timing error. We also find that the throughput loss ratio γ is a continuous function with respect to ϵ_1 and ϵ_2 but non-differentiable when $\epsilon_1 = 0$ or $\epsilon_1 + \epsilon_2 = 0$. This is because there are non-linear step functions in the expression for $\mathbf{E}_1$ in (3.26). We find from the figure that there still exists a considerable performance loss when $\epsilon_1 + \epsilon_2 = 0$, which can be explained as follows. Note that $\epsilon_1 + \epsilon_2 = 0$ does not necessarily

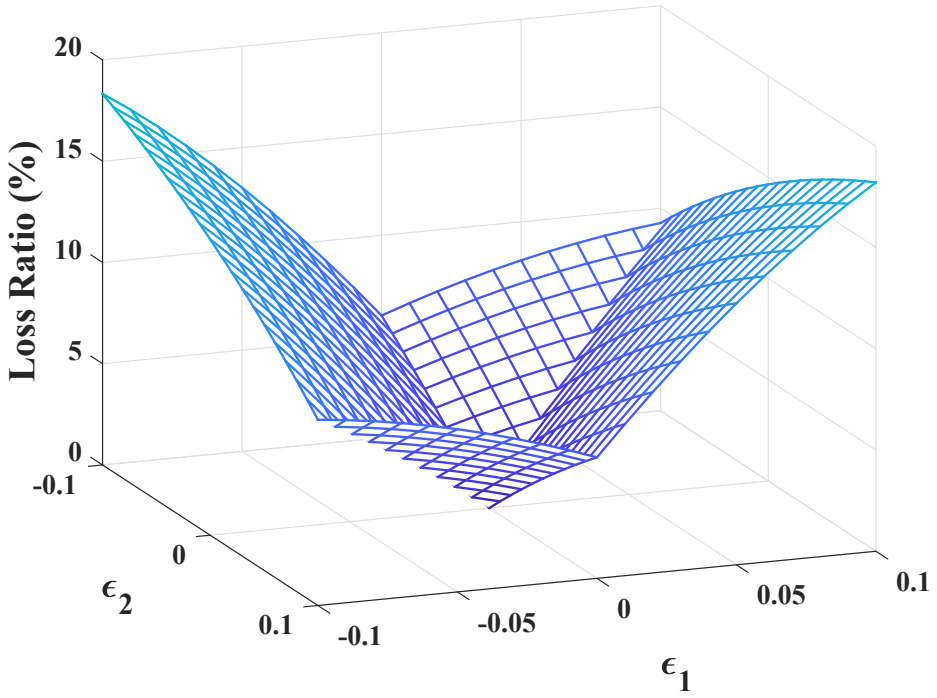

Figure 3.8: The throughput loss ratio as a function of the normalized synchronization timing error ϵ_1 and the normalized coordination timing error ϵ_2 when $P_1|h_1|^2 = 1$, $P_2|h_2|^2 = 0.5$, $\tau = 0.5$ and $N = 10$.

mean that $\epsilon_1 = 0$ and $\epsilon_2 = 0$, since ϵ_1 and ϵ_2 can be negative values. The system may be still affected by non-zero timing errors, even when $\epsilon_1 + \epsilon_2 = 0$. According to (3.23) and (3.24), when $\epsilon_1 + \epsilon_2 = 0$, the first sample vector is still affected by ϵ_1, although the second sample vector will have no timing error.

We also study the individual effects of the timing synchronization error and the coordination timing error on the throughput of ANOMA systems. In Fig. 3.9, we show the throughput loss ratio as a function of ϵ_1 when $\epsilon_2 = 0$ and ϵ_2 when $\epsilon_1 = 0$. Note that the curves of "impact of ϵ_1" and "impact of ϵ_2" are the slices of Fig. 3.8 when $\epsilon_2 = 0$ and $\epsilon_1 = 0$, respectively. The approximated results are calculated by $\Delta_{\epsilon_1}/R^{\text{ANOMA}}$ and $\Delta_{\epsilon_2}/R^{\text{ANOMA}}$ using (3.32) and (3.33). It is demonstrated that the expressions in (3.32) and (3.33) are good approximations of (3.31) when $|\epsilon_1| < 0.05$ and $|\epsilon_2| < 0.05$, respectively. Besides, ϵ_1 causes almost twice throughput loss compared with ϵ_2 for the same value of error. This phenomenon reveals that the synchronization timing error deteriorates the system performance more severely

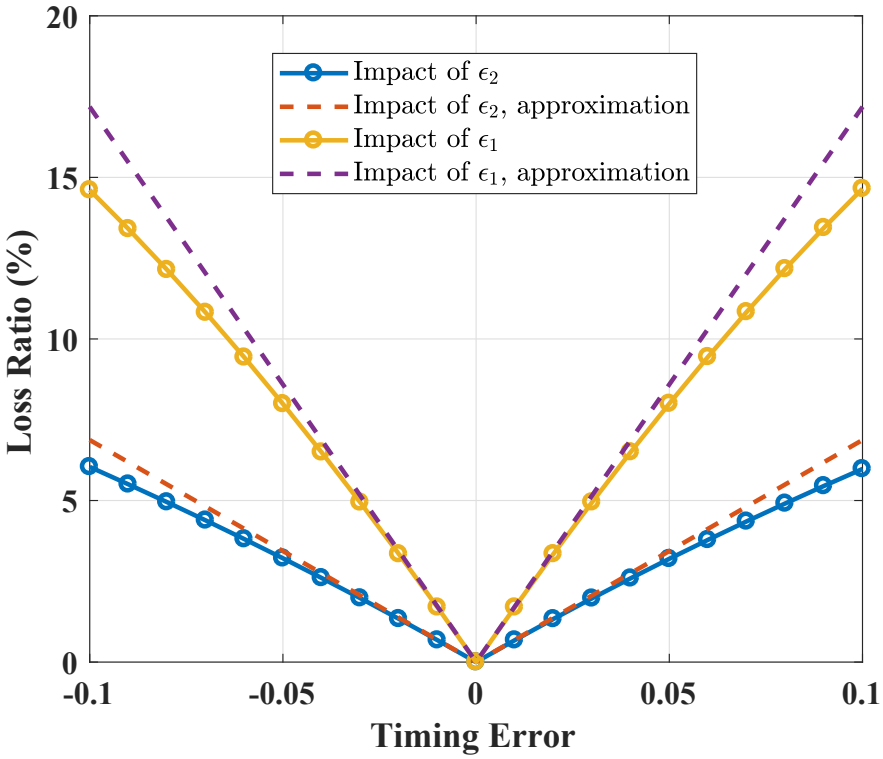

Figure 3.9: The individual impacts of the normalized synchronization timing error ϵ_1 and the normalized coordination timing error ϵ_2 on the throughput loss ratio when $P_1|h_1|^2 = 1$, $P_2|h_2|^2 = 0.5$, $\tau = 0.5$, and $N = 10$.

compared with the coordination timing error. This observation can be explained as follows. With the oversampling, the sampling instants are at $(i + \epsilon_1)T$ and $(i + \tau + \epsilon_1 + \epsilon_2)T$, $i = 1, \cdots, N$, which illustrates that the synchronization timing error affects all sampling instants while the coordination timing error only has impacts on half of the sampling instants.

Finally, we compare the performances of OMA, NOMA, ANOMA without and with timing error in Figure 3.10. In our simulation, the conventional TDMA is adopted as the OMA scheme. As shown in the figure, the throughput curve of OMA is a single point because it is not a function of timing error. The throughput of the NOMA system is calculated under the assumption that perfect SIC is realized at BS. It is demonstrated that the rate performance for ANOMA without timing error is better than that of NOMA with SIC which is further greater than that of OMA. Also, ANOMA always outperforms a perfectly

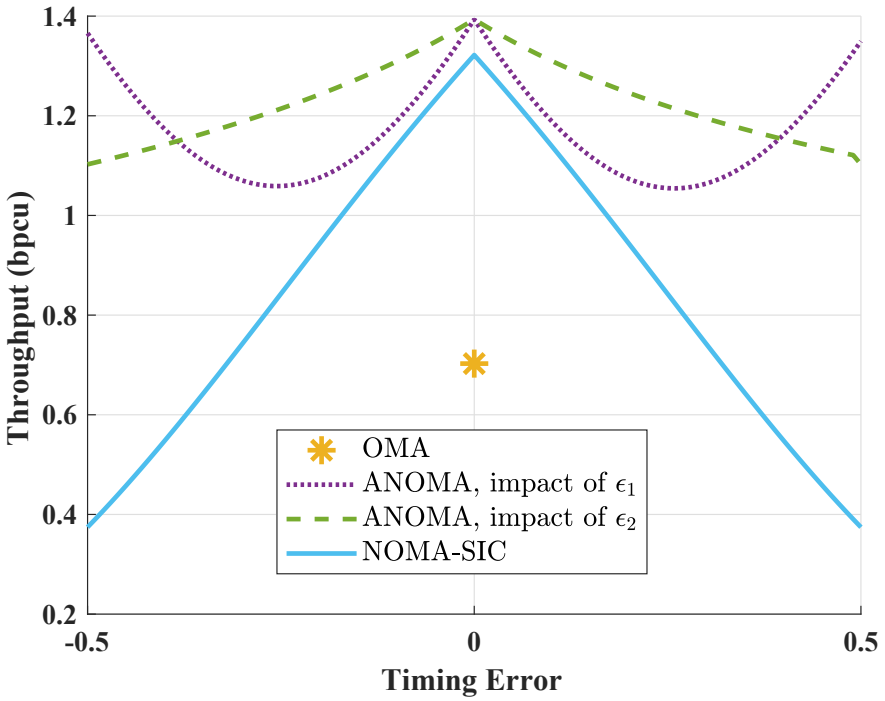

Figure 3.10: Comparison of throughputs among OMA, NOMA, and ANOMA when $P_1|h_1|^2 = 1$, $P_2|h_2|^2 = 0.5$, $\tau = 0.5$, and $N = 10$.

synchronized OMA. We note that for small values of timing error, ANOMA outperforms even a perfectly synchronized NOMA. For the same timing error, the performance of ANOMA is better than that of NOMA. Besides, as shown in Fig. 3.10, the throughput decreases with the absolute value of ϵ_2 monotonously, while the throughput decreases at the beginning and then increases as the absolute value of ϵ_1 increases. This phenomenon can be explained as follows: If there is no timing error ($\epsilon_1 = \epsilon_2 = 0$) and $\tau = 0.5$, the sampling moments are at iT and $(i + 0.5)T$, $i = 1, \cdots, N$. If $|\epsilon_1| = 0.5$ and $\epsilon_2 = 0$, the sampling moments are at $(i \pm 0.5)T$ and $(i + 1 \pm 0.5)T$, $i = 1, \cdots, N$, which are equivalent to advancing ($\epsilon_1 = -0.5$) or delaying ($\epsilon_1 = 0.5$) all sampling moments by $0.5T$. The sampling diversity can still be achieved except that there will be throughput loss due to the shift of sampling moments. For the case $\epsilon_1 = 0$ and $|\epsilon_2| = 0.5$, the second sample vector is a duplicate ($\epsilon_2 = -0.5$) or shifted version ($\epsilon_2 = 0.5$) of the first sample vector. Hence, the sampling diversity cannot be obtained and only the first sample vector can be used to recover the transmitted symbols.

3.7 Conclusion and Future Work

In this chapter, we have studied the performance of a two-user uplink ANOMA system and compared it with the NOMA system. We derive an analytical expression for the two-user sum throughput in the ANOMA system as a function of SNR, frame length, and normalized timing mismatch. We have demonstrated that the ANOMA outperforms the NOMA when the frame length is sufficiently large. Furthermore, we have shown that two users should transmit at full power to maximize the two-user sum throughput. The optimal timing mismatch to maximize the sum throughput converges to a half of one time slot as the frame length goes to infinity. Besides, we discuss the impact of timing error on the throughput performance of the ANOMA system, including the synchronization timing error and the coordination timing error. We have shown how these two types of timing error individually and jointly affect the throughput performance of the ANOMA system.

Chapter 4

Cooperative Asynchronous Non-Orthogonal Multiple Access

4.1 Introduction

Cooperative communication is an effective approach to exploit spatial diversity available through cooperating terminals' relaying signals for one another [38, 39, 40]. Cooperative relaying network with NOMA has been extensively studied in the literature, e.g., [41, 42, 43]. It has been shown that the cooperative NOMA (C-NOMA) systems outperform the cooperative OMA systems in terms of the spectral efficiency [41] and the outage probability [42]. Instead of dedicated relay nodes, users can also be adopted as relays in a cooperative network. A key feature of NOMA is that users with better channel conditions have prior information about the messages of other users. Ding et al. [44] proposed a C-NOMA scheme to fully exploit the prior knowledge at the strong user, where the users could cooperate with each other via short-range communication channels. Yue et al. [45] compared different operation modes of the relay user in a C-NOMA system. The half-duplex relay user receives and transmits in

separate time slots while the full-duplex relay user receives and transmits simultaneously. In [45], the outage probability, the ergodic rate, and the energy efficiency were analyzed in a NOMA user relaying system where the near user could switch between full-duplex and half-duplex modes to relay messages to the far user. Zhang et al. [46] studied an adaptive multiple access scheme to further improve the outage performance, which dynamically switched among the C-NOMA with user relaying, conventional NOMA, and OMA schemes, according to the level of residual self-interference and the quality of links. Wei et al. [47] solved the energy efficiency maximization problem of a full-duplex C-NOMA system under the constraint of successful SIC operation.

4.1.1 Motivations and Related Works

By intentionally introducing symbol asynchrony in the transmitted signal, asynchronous NOMA (ANOMA) systems can achieve a better throughput performance compared with the conventional (synchronous) NOMA systems [48, 27, 49, 50]. In ANOMA systems, the receiver utilizes the oversampling technique [12] to achieve the sampling diversity gain. It has been revealed that the cooperative communication systems can also benefit from the symbol-asynchronous transmission. Sodagari et al. [51] studied an asynchronous cognitive radio framework, where the primary user and the secondary user were not aligned in their timing. They conclude that not only can asynchronous cognitive radio reduce the interference to the primary user, but it also saves power at the secondary user compared with synchronous cognitive radio systems. An asynchronous network coding (ANC) transmission strategy for multiuser cooperative networks was investigated in [10, 11], where the received signals from multiple sources were asynchronous to each other. The proposed scheme achieves full diversity and outperforms the complex field network coding in terms of decoding complexity and bit error rate (BER).

In this chapter, we consider a half-duplex cooperative ANOMA (C-ANOMA) system with user relaying, including a base station (BS), a strong user (also acting as a relay), and a weak user. The half-duplex C-ANOMA system employs a transmission scheme similar to that of the conventional half-duplex C-NOMA system [45, 52, 44]: The BS transmits the superimposed signals to two users simultaneously in the first time block and then the relay user transmits the signal to the weak user at the second time block. Different from the conventional C-NOMA systems, a symbol asynchrony is intentionally added to the downlink superposed signal in the broadcast phase of C-ANOMA systems. The weak user receives two blocks of signals via the broadcast link and the relay link separately. The questions then arise: How to realize SIC based on the symbol-asynchronous signal and then evaluate the performance of the strong user in the C-ANOMA systems? How to evaluate the performance of the weak user which combines a symbol-asynchronous signal from the broadcast link with an interference-free signal from the relay link? Moreover, compared with the cooperative systems with dedicated relay nodes, the power control strategy plays a more critical role in the cooperative systems with user relaying because the power consumption of the relay user affects the lifetime of the cooperative network. We assume that the channel information is available at transmitters [47, 52] and the system works in the delay-tolerant transmission mode [45], such that the transmitters can dynamically adjust their transmit powers according to the channel states to avoid outage and save energy. On one hand, the relay user with very limited battery capacity is more sensitive to the power consumption compared with the BS. On the other hand, the relay user can transmit signals to the weak user more efficiently because the relay user is usually closer to the weak user. As a result, an effective power control strategy is of practical interest to make a trade-off between the transmit power of the BS and that of the relay user while satisfying the quality of service (QoS) constraints in the C-ANOMA/C-NOMA systems with user relaying. To reduce the energy consumption, the power minimization problem has been investigated in several systems, e.g., the downlink NOMA systems [53], the multicell NOMA systems [54], and the cooperative beamforming

networks [55]. Besides, Liu et al. [52] and Chen et al. [56] studied the power allocation problem for half-duplex and full-duplex C-NOMA systems, respectively, to maximize the minimum achievable user rate in a NOMA user pair. To the best of our knowledge, the power minimization problem has never been studied in the C-NOMA or C-ANOMA systems with user relaying.

4.1.2 Contributions

In this chapter, we comprehensively investigate a half-duplex C-ANOMA system with user relaying. The primary contributions of the chapter are summarized as follows:

- We introduce the block-wise SIC technique into C-ANOMA systems, which is employed at the strong (relay) user. We derive the analytical expressions for throughputs achieved by the strong user to decode both users' messages and study their asymptotic performances as the block length goes to infinity. We analytically show that in the C-ANOMA systems with a sufficiently large block length, the strong user can achieve a higher throughput to detect the weak user's message while attains the same throughput when detecting its own message compared with those in C-NOMA systems.

- We derive the expression for the combining throughput achieved by the weak user which combines the asynchronously superimposed signal from the broadcast link with the interference-free signal from the relay link. Based on the derived throughput expressions, we obtain the asymptotic throughput as the block length goes to infinity and its simple upper and lower bounds. We analytically prove that in the C-ANOMA systems with a sufficiently large block length, the combining throughput of the weak user is greater than that in the C-NOMA systems.

- We further study the optimal design of C-ANOMA systems. We analytically prove that the optimal timing mismatch to maximize the individual throughput converges

to half of the symbol interval as the block length increases. Besides, we solve the weighted sum power minimization problem under the QoS constraints for C-ANOMA and C-NOMA systems. The solution is given by the explicit expressions of the powers allocated to the strong and weak users at the BS and the transmit power of the relay (strong) user. It is demonstrated that for a relatively large block length, the C-ANOMA systems consume less power compared with the C-NOMA systems in order to satisfy the same QoS requirements. In other words, under the same transmit power limits, the C-ANOMA systems can provide a higher QoS for users compared with the C-NOMA systems.

4.1.3 Organization and Notation

The remainder of the chapter is organized as follows. The C-ANOMA system model is presented in Section 4.2. The throughput performance of the C-ANOMA system is analyzed in Section 4.3. We discuss the optimal design of the C-ANOMA system in Section 4.4 where we investigate the optimal timing mismatch and solve the weighted power minimization problem under QoS constraints. Numerical results are presented in Section 4.5. Finally, we draw the conclusions in Section 4.6.

Notations: $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^T$ denotes the transpose, $(\cdot)^{-1}$ denotes the inverse operation, $\otimes$ denotes the Kronecker product, $|x|$ denotes the absolute value of x, $\bar{x}$ denotes the complex conjugate of x, $\mathbb{E}[\cdot]$ denotes the expectation operation, $\mathcal{CN}(0,1)$ denotes the complex normal distribution with zero mean and unit variance. $\mathrm{diag}(\mathbf{x})$ stands for a diagonal matrix whose k-th diagonal element is equal to the k-th entry of vector $\mathbf{x}$.

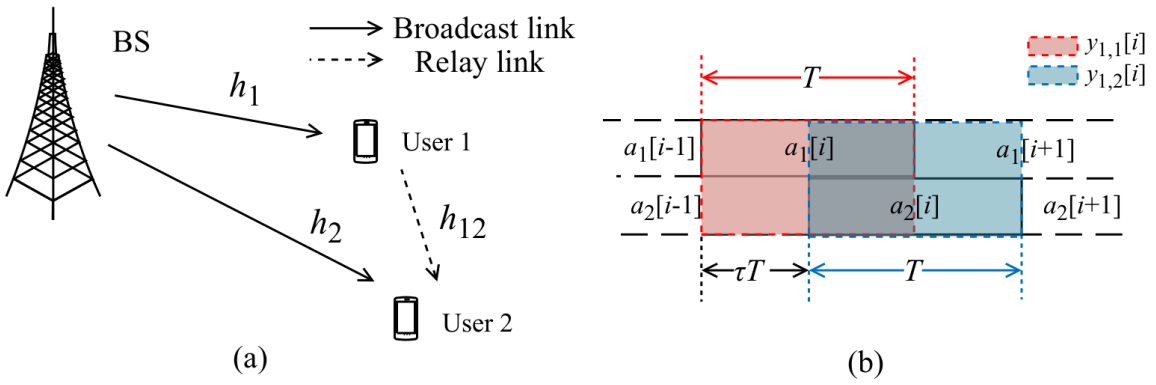

Figure 4.1: (a) Illustration of a three-node C-ANOMA/C-NOMA system with user relaying. (b) Illustration of the sampling for the broadcast phase in C-ANOMA systems.

4.2 System Model

In this chapter, as shown in Fig. 4.1 (a), we consider a downlink half-duplex C-ANOMA system which includes a single BS and two users equipped with a single antenna. User 1 (strong user) acts as a relay for User 2 (weak user) and adopts the decode-and-forward (DF) protocol, i.e., decodes and forwards the message to User 2 via the relay link. The downlink transmission is done in blocks, including two phases, i.e., the broadcast phase and the relay phase. In the broadcast phase, the BS broadcasts one block of superposed signal to two users simultaneously while User 1 is silent. In the relay phase, User 1 transmits the block of decoded signal to User 2 while the BS keeps silent. We assume that the channel is static within each block [57] and all the channel information is perfectly known at the BS, Users 1 and 2 [52, 47]. The channel coefficient between the BS and User i is denoted as h_i $(i = 1, 2)$ and the channel coefficient between Users 1 and 2 is denoted as h_{12}. In what follows, we present our analysis in the broadcast phase and the relay phase separately.

4.2.1 Broadcast Phase

4.2.1.1 C-ANOMA

In the C-ANOMA systems, a symbol mismatch is intentionally introduced in the downlink signal. As shown in Fig. 4.1 (b), the intended timing mismatch between the symbols for Users 1 and 2 is denoted by τT, where T is the symbol interval and τ, $0 \leq \tau < 1$, is the normalized timing mismatch. We assume that τ can be perfectly known at users. The timing mismatch information can be transmitted as part of the downlink control information through the downlink control channel, such as the physical downlink control channel (PDCCH) in the long term evolution (LTE) system. The downlink control channel is designed to be robust in order to ensure the successful reception of the control information, e.g., by applying a low-rate coding scheme. Note that the C-ANOMA system becomes a synchronous C-NOMA system when $\tau = 0$.

Let $a_1[i] = \sqrt{P_1}s_1[i]$ and $a_2[i] = \sqrt{P_2}s_2[i]$, where $s_j[i]$ denotes the ith symbol sent to User j, $j = 1, 2$, P_j stands for the power allocated to User j. The transmitted signal at the BS is given by

$$s(t) = \sum_{i=1}^{N} a_1[i]p(t - iT) + \sum_{i=1}^{N} a_2[i]p(t - iT - \tau T). \tag{4.1}$$

where N denotes the number of symbols in a block, i.e., the block length, $p(\cdot)$ denotes the pulse-shaping filter. Without loss of generality, the rectangular pulse shape is adopted, i.e., $p(t) = 1/\sqrt{T}$ when $t \in [0, T]$ and $p(t) = 0$ otherwise.

The received signal at User 1 is given by

$$y_1(t) = h_1 s(t) + n_1(t)$$

$$=h_1\left(\sum_{i=1}^{N} a_1[i]p(t-iT)+\sum_{i=1}^{N} a_2[i]p(t-iT-\tau T)\right)+n_1(t), \tag{4.2}$$

where $n_1(t) \sim \mathcal{CN}(0,1)$ denotes the normalized additive white Gaussian noise (AWGN).

The oversampling technique [6, 31, 7], depicted in Fig. 4.1 (b), is employed at the receiver to take advantage of sampling diversity in asynchronous systems. As shown in Fig. 4.1 (b), the receiver uses the matched filter, sampling at iT and $(i+\tau)T$, $i = 1, \cdots, N$, to obtain two sample vectors, denoted by $[y_{1,1}[1], \cdots, y_{1,1}[N]]^T$ and $[y_{1,2}[1], \cdots, y_{1,2}[N]]^T$. Specifically, the ith element in the first sample vector is given by

$$\begin{aligned} y_{1,1}[i] &= \int_0^\infty y_1(t)p(t-iT)dt \\ &= \int_0^\infty h_1 a_1[i]p(t-iT)p(t-iT)dt + \int_0^\infty \{h_1 a_2[i-1]p(t-(i+1+\tau)T) \\ &\quad + h_1 a_2[i]p(t-(i+\tau)T)\}\, p(t-iT)dt + n_{1,1}[i] \\ &= h_1 a_1[i] + \tau h_1 a_2[i-1] + (1-\tau)h_1 a_2[i] + n_{1,1}[i], \end{aligned} \tag{4.3}$$

where $n_{1,1}[i] = \int_0^\infty n_1(t)p(t-iT)dt$ denotes the additive noise. The ith element in the second sample vector is given by

$$\begin{aligned} y_{1,2}[i] &= \int_0^\infty y_1(t)p(t-iT-\tau T)dt \\ &= h_1 a_2[i] + \tau h_1 a_1[i+1] + (1-\tau)h_1 a_1[i] + n_{1,2}[i], \end{aligned} \tag{4.4}$$

where $n_{1,2}[i] = \int_0^\infty n_1(t)p(t-iT-\tau T)dt$ denotes the additive noise.

We can write the outputs of the two matched filters at User 1 in a matrix form as

$$\mathbf{Y}_1 = h_1\sqrt{P_1}\mathbf{R}\mathbf{G_1}\mathbf{S_1} + h_1\sqrt{P_2}\mathbf{R}\mathbf{G_2}\mathbf{S_2} + \mathbf{N}_1 \tag{4.5}$$

where $\mathbf{Y}_1 = [y_{1,1}[1]\ y_{1,2}[1]\ \cdots\ y_{1,1}[N]\ y_{1,2}[N]]^T$, $\mathbf{G}_1$ and $\mathbf{G}_2$ are $2N$-by-N matrices given

by $\mathbf{G_1} = \mathbf{I}_N \otimes [1\ 0]^T$ and $\mathbf{G_2} = \mathbf{I}_N \otimes [0\ 1]^T$, $\mathbf{S_i} = [s_i[1] \cdots s_i[N]]^T$ $(i = 1, 2)$, $\mathbf{N}_1 = [n_{1,1}[1]\ n_{1,2}[1] \cdots n_{1,1}[N]\ n_{1,2}[N]]^T$, and

$$\mathbf{R} = \begin{bmatrix} 1 & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1 & \tau & 0 & \cdots & 0 \\ 0 & \tau & 1 & 1-\tau & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \tau & 1 & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1 \end{bmatrix}. \tag{4.6}$$

Note that multiplying $\mathbf{R}$ by $\mathbf{G_i}$ outputs a $2N$-by-N matrix whose columns are equal to the odd (if $i = 1$) or even (if $i = 2$) columns of $\mathbf{R}$.

We assume that the transmitted symbols are normalized and independent to each other, such that $\mathbb{E}\left[\mathbf{S_i}{\mathbf{S_i}}^H\right] = \mathbf{I}$ $(i = 1, 2)$. Note that the noise terms in (4.3) and (4.4) are colored due to the oversampling, and we have

$$\mathbb{E}\left\{n_{1,1}[i]n_{1,2}^H[i]\right\} = \int_0^\infty \int_0^\infty \mathbb{E}\left\{n_1(t)n_1^H(s)\right\} p\left(t - iT\right) p\left(s - iT - \tau T\right) dtds = 1 - \tau. \tag{4.7}$$

Thus, the covariance matrix of $\mathbf{N}_1$ in (4.5) is given by

$$\mathbf{R}_{\mathbf{N}_1} = \mathbb{E}\left\{\mathbf{N}_1\mathbf{N}_1^H\right\} = \mathbf{R}. \tag{4.8}$$

Similarly, the received samples at User 2 in the broadcast phase can be written as

$$\mathbf{Y}_2 = h_2\sqrt{P_1}\mathbf{R}\mathbf{G}_1\mathbf{S_1} + h_2\sqrt{P_2}\mathbf{R}\mathbf{G}_2\mathbf{S_2} + \mathbf{N}_2, \tag{4.9}$$

where the covariance matrix $\mathbf{R}_{\mathbf{N}_2} = \mathbb{E}\left\{\mathbf{N}_2\mathbf{N}_2^H\right\} = \mathbf{R}$.

4.2.1.2 C-NOMA

By setting $\tau = 0$, the C-ANOMA system becomes the C-NOMA system. For the C-NOMA systems, users do not use the oversampling technique. The ith sample at Users 1 and 2 in the broadcast phase will be

$$y_1[i] = h_1\sqrt{P_1}s_1[i] + h_1\sqrt{P_2}s_2[i] + n_1[i], \tag{4.10}$$

$$y_2[i] = h_2\sqrt{P_1}s_1[i] + h_2\sqrt{P_2}s_2[i] + n_2[i], \tag{4.11}$$

where $n_j[i] = \int_0^\infty n_j(t)p(t - iT)dt$, $j = 1, 2$. Note that (4.10) and (4.11) can also be derived from (4.5) and (4.9), respectively, by letting $\tau = 0$.

4.2.2 Relay Phase

In the relay phase, User 2 receives another copy of the desired signal from User 1. The ith sample received at User 2 in the relay phase is given by

$$y_{12}[i] = h_{12}\sqrt{P_r}s_2[i] + n_{12}[i], \tag{4.12}$$

where P_r is the transmit power of User 1 and $n_{12}[i] = \int_0^\infty n_{12}(t)p(t - iT)dt$ is the additive noise. Note that the C-NOMA and C-ANOMA systems coincide in the relay phase.

For ease of the following analysis, we rewrite the received samples from the relay link in (4.12) into the matrix format, i.e.,

$$\mathbf{Y}_{12} = h_{12}\sqrt{P_r}\mathbf{S}_2 + \mathbf{N}_{12}, \tag{4.13}$$

where $\mathbf{Y}_{12} = [y_{12}[1], y_{12}[2], \cdots, y_{12}[N]]^T$, $\mathbf{N}_{12} = [n_{12}[1], n_{12}[2], \cdots, n_{12}[N]]^T$, and the covari-

ance matrix $\mathbf{R}_{\mathbf{N}_{12}} = \mathbb{E}\left\{\mathbf{N}_{12}\mathbf{N}_{12}^H\right\} = \mathbf{I}_N$.

Combining all the received samples of User 2 in C-ANOMA systems, i.e., $\mathbf{Y}_2$ in (4.9) and $\mathbf{Y}_{12}$ in (4.13), we have

$$\tilde{\mathbf{Y}}_2 = \left[\begin{smallmatrix}\mathbf{Y}_2\\ \mathbf{Y}_{12}\end{smallmatrix}\right] = \underbrace{\left[\begin{smallmatrix}h_2\sqrt{P_1}\mathbf{R}\mathbf{G}_1\\ \mathbf{0}_N\end{smallmatrix}\right]}_{\mathbf{W}_1}\mathbf{S}_1 + \underbrace{\left[\begin{smallmatrix}h_2\sqrt{P_2}\mathbf{R}\mathbf{G}_2\\ h_{12}\sqrt{P_r}\mathbf{I}_N\end{smallmatrix}\right]}_{\mathbf{W}_2}\mathbf{S}_2 + \underbrace{\left[\begin{smallmatrix}\mathbf{N}_2\\ \mathbf{N}_{12}\end{smallmatrix}\right]}_{\mathbf{N}}. \tag{4.14}$$

Applying $\mathbb{E}\left\{\mathbf{N}_2\mathbf{N}_2^H\right\} = \mathbf{R}$ and $\mathbb{E}\left\{\mathbf{N}_{12}\mathbf{N}_{12}^H\right\} = \mathbf{I}_N$, the covariance matrix of the concatenated noise vector $\mathbf{N}$ is given by

$$\mathbf{R}_{\mathbf{N}} = \mathbb{E}\left\{\mathbf{N}\mathbf{N}^H\right\} = \begin{bmatrix}\mathbb{E}\left\{\mathbf{N}_2\mathbf{N}_2^H\right\} & \mathbb{E}\left\{\mathbf{N}_2\mathbf{N}_{12}^H\right\}\\ \mathbb{E}\left\{\mathbf{N}_{12}\mathbf{N}_2^H\right\} & \mathbb{E}\left\{\mathbf{N}_{12}\mathbf{N}_{12}^H\right\}\end{bmatrix} = \left[\begin{smallmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{smallmatrix}\right]. \tag{4.15}$$

4.3 Performance Analysis of C-ANOMA Systems

In this section, we analyze the individual throughput of users in the C-ANOMA and C-NOMA systems, including the strong and weak users.

4.3.1 Strong User

4.3.1.1 C-ANOMA

In C-ANOMA systems, the block-wise SIC is adopted at User 1, i.e., it first decodes the block of symbols intended for User 2, subtracts it from the received signal, and then decodes its intended symbols. Note that the BS transmits one block of symbols via two time blocks in the half-duplex mode [45]. Besides, an extra τT time is utilized to create the sampling diversity in the symbol-asynchronous transmission. Hence, in the half-duplex C-ANOMA

systems, a block of N symbols are transmitted via $2N+\tau$ channel uses to Users 1 and 2. By considering (4.5) as a virtual multiple-input multiple-output (MIMO) system and treating the symbols for User 1 as noise, the throughput of User 1 to detect User 2's message is given by

$$\begin{aligned} R_{2\to1}^{\text{ANOMA}} =& \frac{1}{2N+\tau}\log\det\left[\mathbf{I}_{2N}+\left(\mathbf{R}_{\mathbf{N_1}}+P_1|h_1|^2\mathbf{R}\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}^H\right)^{-1}P_2|h_1|^2\mathbf{R}\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}^H\right] \\ \overset{(a)}{=}& \frac{1}{2N+\tau}\log\det\left[\mathbf{I}_{2N}+\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}P_2|h_1|^2\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right], \end{aligned} \tag{4.16}$$

where (a) is derived by applying $\mathbf{R}_{\mathbf{N_1}}=\mathbf{R}$ and $\mathbf{R}^H=\mathbf{R}$.

Under the assumption of perfect SIC, by subtracting User 2's message from the superposed signal in (4.5), the throughput of User 1 to detect its own message is calculated as

$$\begin{aligned} R_1^{\text{ANOMA}} =& \frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{R}_{\mathbf{N}_1}^{-1}\mathbf{R}\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}^H\right) \\ =& \frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right). \end{aligned} \tag{4.17}$$

After matrix calculations, we can rewrite the throughput expressions for User 1 in (4.16) and (4.17) as functions of the receive signal-to-noise ratios (SNRs), i.e., μ_1 and μ_2, the normalized timing mismatch, τ, and the block length, N, i.e.,

$$R_{2\to1}^{\text{ANOMA}} = \frac{1}{2N+\tau}\log\frac{\left(r_1^{N+1}-r_2^{N+1}\right)+\tau^2\left(r_1^N-r_2^N\right)}{r_1-r_2}+\frac{N}{2N+\tau}\log\left(\frac{\mu_1\mu_2}{1+\mu_1}\right), \tag{4.18}$$

$$R_1^{\text{ANOMA}} = \frac{N}{2N+\tau}\log\left(1+\mu_1\right), \tag{4.19}$$

where

$$\mu_1 = P_1|h_1|^2, \mu_2 = P_2|h_1|^2, Q = 2\tau(1-\tau), \tag{4.20}$$

$$r_1 = \frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q+\sqrt{\left(\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q\right)^2-Q^2}}{2}, \tag{4.21}$$

$$r_2 = \frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q - \sqrt{\left(\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q\right)^2 - Q^2}}{2}. \tag{4.22}$$

The detailed derivation of (4.18) and (4.19) is presented in Appendix B.1.

4.3.1.2 C-NOMA

In conventional (synchronous) NOMA systems, with perfect SIC, the throughputs of User 1 are given by [41, 45]

$$R_{2\to 1}^{\text{NOMA}} = \frac{1}{2}\log\left(1+\frac{\mu_2}{1+\mu_1}\right), \tag{4.23}$$

$$R_1^{\text{NOMA}} = \frac{1}{2}\log(1+\mu_1). \tag{4.24}$$

We note that by setting $\tau = 0$, we obtain $Q = 0$, $r_2 = 0$, and $r_1 = \mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1}$. Thus, $R_{2\to 1}^{\text{ANOMA}}|_{\tau=0} = R_{2\to 1}^{\text{NOMA}}$ and $R_1^{\text{ANOMA}}|_{\tau=0} = R_1^{\text{NOMA}}$.

4.3.1.3 Comparison between C-ANOMA and C-NOMA

To study the throughput performance in the systems with a relatively large block length, we consider the asymptotic case of $N \to \infty$. According to (4.19), the throughput of User 1 to decode its own message if $N \to \infty$ is given by

$$R_{1,\text{asymp}}^{\text{ANOMA}} \triangleq \lim_{N\to\infty} R_1^{\text{ANOMA}} = \frac{1}{2}\log\left(1+\mu_1\right) = R_1^{\text{NOMA}}. \tag{4.25}$$

We note from (4.25) that User 1 in C-ANOMA and C-NOMA systems can achieve the same throughput to detect its own message for a sufficiently large block length. It is because with perfect SIC, the throughput of User 1 to detect its own message is not affected by the

symbol asynchrony of the signal for User 2. Furthermore, we derive the following theorem to compare the throughputs of User 1 to detect User 2's message in the C-ANOMA and C-NOMA systems.

Theorem 4.1. *The throughputs of User 1 to detect User 2's message in the C-NOMA and C-ANOMA systems satisfy the following inequalities*

$$\begin{aligned} R_{2\to1}^{\mathrm{NOMA}} &\le R_{2\to1,L}^{\mathrm{ANOMA}} \triangleq \frac{1}{2}\log\left(1+\frac{\mu_2+\frac{1}{2}\mu_1\mu_2 Q}{1+\mu_1}\right) \\ &\le R_{2\to1,\mathrm{asymp}}^{\mathrm{ANOMA}} \triangleq \lim_{N\to\infty} R_{2\to1}^{\mathrm{ANOMA}} = \frac{1}{2}\log\left(\frac{\mu_1\mu_2 r_1}{1+\mu_1}\right) \\ &\le R_{2\to1,U}^{\mathrm{ANOMA}} \triangleq \frac{1}{2}\log\left(1+\frac{\mu_2+\mu_1\mu_2 Q}{1+\mu_1}\right), \end{aligned} \tag{4.26}$$

where $Q = 2\tau(1-\tau)$, all the equal signs are achieved if and only if $\tau = 0$.

Proof. See Appendix B.2. □

We note from Theorem 4.1 that for a relatively large block length, User 1 in C-ANOMA systems can achieve a higher throughput to decode User 2's message compared with that in C-NOMA systems. Besides, comparing the expressions for $R_{2\to1,L}^{\mathrm{ANOMA}}$ and $R_{2\to1}^{\mathrm{NOMA}}$, we find that the gain of C-ANOMA systems is related to the term $\mu_1\mu_2 Q$ which increases as the channel qualities improve.

In practice, the block length, N, is determined by several factors, such as the channel coherence time, the modulation, the sampling rate, etc., which are beyond the scope of this chapter. We assume that the block length N is a predetermined parameter in this chapter. We will show in Section 4.5 that the asymptotic throughput approximates the accurate one for not-so-large values of N, e.g, $N > 50$.

4.3.2 Weak User

In the half-duplex cooperative relaying scenario, the weak user, User 2, receives two blocks of symbols, one from the BS with the superposed signal through the broadcast link and the other one from User 1 with only the intended signal through the relay link.

4.3.2.1 C-ANOMA

Treating (4.14) as a virtual MIMO system and considering User 1's message as noise, the combining throughput of User 2 can be calculated as

$$R_2^{\text{ANOMA}} = \frac{1}{2N+\tau} \log\det\left[\mathbf{I}_{3N} + \left(\mathbf{R_N} + \mathbf{W}_1\mathbf{W}_1^H\right)^{-1}\mathbf{W}_2\mathbf{W}_2^H\right]. \tag{4.27}$$

The combining throughput of User 2 can be written as a function of the transmit powers, the channel gains, the normalized timing mismatch, and the block length in the following theorem.

Theorem 4.2. *In the half-duplex C-ANOMA systems, the combining throughput of User 2 is given by*

$$R_2^{\text{ANOMA}} = \frac{1}{2N+\tau} \log\frac{\left(z_1^{N+1} - z_2^{N+1}\right) + \tau^2\left(z_1^N - z_2^N\right)}{z_1 - z_2} + \frac{N}{2N+\tau}\log\left(\frac{P_1P_2|h_2|^4}{1+P_1|h_2|^2}\right), \tag{4.28}$$

where

$$\nu_1 = P_1|h_2|^2, \nu_2 = \frac{P_2|h_2|^2}{1+P_r|h_{12}|^2}, Q = 2\tau(1-\tau) \tag{4.29}$$

$$z_1 = \frac{\nu_1^{-1}+\nu_2^{-1}+\nu_1^{-1}\nu_2^{-1}+Q + \sqrt{\left[\nu_1^{-1}+\nu_2^{-1}+\nu_1^{-1}\nu_2^{-1}+Q\right]^2 - Q^2}}{2}, \tag{4.30}$$

$$z_2 = \frac{\nu_1^{-1}+\nu_2^{-1}+\nu_1^{-1}\nu_2^{-1}+Q - \sqrt{\left[\nu_1^{-1}+\nu_2^{-1}+\nu_1^{-1}\nu_2^{-1}+Q\right]^2 - Q^2}}{2}. \tag{4.31}$$

Proof. See Appendix B.3. □

4.3.2.2 C-NOMA

In C-NOMA systems, User 2 adopts the maximal ratio combining (MRC) to combine the signals from the direct and relay links [46, 45]. Then, the combining throughput of User 2 is given by

$$R_2^{\mathrm{NOMA}} = \frac{1}{2}\log\left(1 + P_r|h_{12}|^2 + \frac{P_2|h_2|^2}{P_1|h_2|^2+1}\right). \tag{4.32}$$

Note that by setting $\tau = 0$, we have $Q = 0$, $z_2 = 0$, and $z_1 = \nu_1^{-1}+\nu_2^{-1}+\nu_1^{-1}\nu_2^{-1}$. Thus, the expression for the combining throughput of User 2 in C-ANOMA systems coincides with that in C-NOMA systems, i.e., $R_2^{\mathrm{ANOMA}}|_{\tau=0} = R_2^{\mathrm{NOMA}}$.

4.3.2.3 Comparison between C-ANOMA and C-NOMA

We derive the following theorem which compares the throughputs of the C-ANOMA and C-NOMA systems for $N \to \infty$.

Theorem 4.3. *In C-ANOMA systems, the combining throughput of User 2 for the asymptotic case of $N \to \infty$ is given by*

$$\begin{aligned} R_{2,\mathrm{asymp}}^{\mathrm{ANOMA}} &\triangleq \lim_{N\to\infty} R_2^{\mathrm{ANOMA}} \\ &= \frac{1}{2}\log\left\{\frac{1+P_r|h_{12}|^2}{2} + \frac{P_2|h_2|^2 + P_1P_2|h_2|^4Q}{2(1+P_1|h_2|^2)}\right. \end{aligned}$$

$$+\frac{1}{2}\left[\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2-\left(\frac{P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2\right]^{\frac{1}{2}}\Bigg\}, \tag{4.33}$$

where $Q=2\tau(1-\tau)$. *The combining throughputs of User 2 for the C-NOMA and C-ANOMA systems satisfy the following inequalities*

$$\begin{aligned}
R_2^{\text{NOMA}} &\leq R_{2,L}^{\text{ANOMA}} \\
&\triangleq \frac{1}{2}\log\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+\frac{1}{2}P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right) \\
&\leq R_{2,\text{asymp}}^{\text{ANOMA}} \\
&\leq R_{2,U}^{\text{ANOMA}} \\
&\triangleq \frac{1}{2}\log\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right),
\end{aligned} \tag{4.34}$$

where the equal signs are achieved if and only if $\tau=0$.

Proof. See Appendix B.4. □

We note from (4.34) that the gain of C-ANOMA over C-NOMA depends on the term $P_1P_2|h_2|^4Q$, thus, a better direct channel between User 2 and the BS results in a greater performance improvement of C-ANOMA systems compared with C-NOMA systems. Moreover, according to (4.25), Theorems 4.1 and 4.3, it is shown that for $N\to\infty$, the throughputs of both users to detect the weak user's message in the C-ANOMA systems are larger than those in the C-NOMA systems while the throughput of the strong user to detect its own message is identical for the C-ANOMA and C-NOMA systems. In Section 4.5, we will show by numerical results that the C-ANOMA systems outperform the C-NOMA systems in terms of the throughput to decode the weak user's message with a relatively small value of N, e.g., $N>20$.

Furthermore, in both C-ANOMA and C-NOMA systems, the actual throughput of User 2 is affected by both the throughput of User 1 to detect User 2's message, $R_{2\to1}$, and the combining throughput of User 2, R_2. Since User 2's message needs to be detected by both Users 1 and 2, the minimum of $R_{2\to1}$ and R_2 is the bottleneck of the actual throughput of User 2, i.e., $R_{\text{act},2} = \min\{R_{2\to1}, R_2\}$ [52]. According to Theorems 4.1 and 4.3, it is trivial to derive that for a sufficiently large N and $\tau \neq 0$, $R_{\text{act},2}^{\text{ANOMA}} > R_{\text{act},2}^{\text{NOMA}}$.

4.4 C-ANOMA System Design

In this section, we study the optimal design of the C-ANOMA systems, including the optimal timing mismatch and the power control strategy.

4.4.1 Optimal Timing mismatch

We first investigate the optimal normalized timing mismatch, τ^*. Although the optimal normalized timing mismatch to maximize $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} is analytically intractable for a general finite block length N, we can numerically obtain τ^* for a given finite N by simply searching in the range of $0 \leq \tau < 1$ as done in Section 4.5. To derive the optimal τ for a large N, we study the asymptotic case of $N \to \infty$. According to (4.25), the throughput of User 1 to detect its own message is independent of τ. According to (4.26) and (4.33), it is easy to show that $R_{2\to1,\text{asymp}}^{\text{ANOMA}}$ and $R_{2,\text{asymp}}^{\text{ANOMA}}$ are increasing functions of Q which is given by $2\tau(1-\tau)$. Thus, maximizing $R_{2\to1,\text{asymp}}^{\text{ANOMA}}$ and $R_{2,\text{asymp}}^{\text{ANOMA}}$ is equivalent to maximizing the term $\tau(1-\tau)$. Therefore, the optimal τ to maximize the throughputs of both users to detect User 2's message converges to 0.5, i.e.,

$$\tau^* \triangleq \arg\max_{\tau} R_{2\to1,\text{asymp}}^{\text{ANOMA}} = \arg\max_{\tau} R_{2,\text{asymp}}^{\text{ANOMA}} = 0.5. \tag{4.35}$$

In practice, in order to reduce the resource consumption of the control information, the normalized timing mismatch can be fixed to the default value of 0.5 at both the BS and the users.

4.4.2 Power Minimization

In this chapter, we consider the delay-tolerant mode where the BS and the relay user can dynamically adjust their transmit powers according to the channel states in order to avoid outage and satisfy the minimum rate requirements [45]. Our objective is to minimize the weighted sum transmit power of the BS and the relay user under the minimum rate (i.e., QoS) requirements and the individual power constraints. Then, the power minimization problem can be formulated as

$$\min_{P_1,P_2,P_r} \quad \omega_s(P_1+P_2)+\omega_r P_r, \tag{4.36a}$$

$$\text{s.t.} \quad R_{2\to1}^{\text{ANOMA}} \geq R_2^*, R_1^{\text{ANOMA}} \geq R_1^*, R_2^{\text{ANOMA}} \geq R_2^*, \tag{4.36b}$$

$$P_1+P_2 < P_{s,\max}, P_r < P_{r,\max}, \tag{4.36c}$$

where ω_s and ω_r are the non-negative weights for the transmit powers of the BS and User 1, respectively, such that $\omega_s+\omega_r=1$. $P_{s,\max}$ and $P_{r,\max}$ stand for the maximum available powers of the BS and User 1, respectively. R_1^* and R_2^* are the target rates of Users 1 and 2's messages. Note that the choice of ω_s and ω_r provides a trade-off between the power consumption of the BS and that of the relay user. For instance, if one wants to further restrict the power consumption of the relay user due to its limited battery capacity, ω_r should be chosen greater than ω_s.

The exact expressions of $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} in (4.18) and (4.28) make the optimization problem (4.36) analytically intractable. To simplify the optimization problem, we replace

$R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} in (4.36b) with their *asymptotic lower bounds*, which can provide a suboptimal solution for the original optimization problem (4.36), i.e.,

$$\min_{P_1,P_2,P_r} \quad \omega_s(P_1+P_2)+\omega_r P_r, \tag{4.37a}$$

$$\text{s.t.} \quad R_{2\to1,L}^{\text{ANOMA}} \geq R_2^*, \tag{4.37b}$$

$$R_1^{\text{ANOMA}} \geq R_1^*, \tag{4.37c}$$

$$R_{2,L}^{\text{ANOMA}} \geq R_2^*, \tag{4.37d}$$

$$P_1+P_2 < P_{s,\max}, P_r < P_{r,\max}. \tag{4.37e}$$

For sufficiently large values of N, Eqs. (4.37b) and (4.37d) are stronger constraints for $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} compared with those in (4.36b), which means that the solution of (4.36) can do at least as good as that of (4.37). In what follows, we explain that (4.37) can also provide a suboptimal solution of (4.36) for a finite N. By definition, as N increases, the exact throughputs can be arbitrarily close to the asymptotic ones. We assume that $R_{2\to1}^{\text{ANOMA}} \geq R_{2\to1,L}^{\text{ANOMA}}$ for any $N \geq N_1$ and $R_2^{\text{ANOMA}} \geq R_{2,L}^{\text{ANOMA}}$ for any $N \geq N_2$. By choosing a proper N^*, for example, $N^* = \max\{N_1, N_2\}$, we can ensure that $R_{2\to1}^{\text{ANOMA}} \geq R_{2\to1,L}^{\text{ANOMA}}$ and $R_2^{\text{ANOMA}} \geq R_{2,L}^{\text{ANOMA}}$ for the given N^*. We will show that N^* can be a reasonable value (e.g., $N^* = 100$) in the numerical results section. In practice, the actual block length is usually greater than 100. For example, in global system for mobile communications (GSM), there are approximately 156 symbols in a normal burst (a physical channel carrying information on traffic and control channels) [37]. As a result, the optimization problem (4.37) can provide a suboptimal solution for the problem (4.36) with the block length used in practical communication systems.

By simplifying (4.37b), (4.37c), and (4.37d), we obtain

$$P_2 \geq \frac{\gamma_2}{|h_1|^2}\frac{1+P_1|h_1|^2}{1+\frac{1}{2}QP_1|h_1|^2}, \tag{4.38}$$

$$P_1 \geq \frac{\gamma_1 + \epsilon}{|h_1|^2}, \tag{4.39}$$

$$P_r \geq \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2}\frac{1+\frac{1}{2}QP_1|h_2|^2}{1+P_1|h_2|^2}. \tag{4.40}$$

where $\gamma_i = 2^{2R_i^*} - 1$, $i = 1, 2$, is the target signal-to-interference-plus-noise ratio (SINR) to detect User i's message and $\epsilon = 2^{2R_1^*}(2^{\frac{\tau}{N}R_1^*} - 1)$. The value of ϵ can be made arbitrary small with increasing N. For a sufficiently large N, i.e., $N > N^*$, we have $\epsilon < \epsilon^* \triangleq 2^{2R_1^*}(2^{\frac{\tau}{N^*}R_1^*} - 1)$, hence, we can substitute (4.39) with a stronger constraint, i.e.,

$$P_1 \geq \frac{\gamma_1 + \epsilon^*}{|h_1|^2}. \tag{4.41}$$

Then, by replacing the constraints with (4.38), (4.40), and (4.41), the optimization problem (4.37) becomes

$$\min_{P_1,P_2,P_r} \quad \omega_s(P_1 + P_2) + \omega_r P_r, \tag{4.42a}$$

$$\text{s.t.} \quad \frac{\gamma_1 + \epsilon^*}{|h_1|^2} \leq P_1 \leq P_{s,\max}, \tag{4.42b}$$

$$\frac{\gamma_2}{|h_1|^2}\frac{1+P_1|h_1|^2}{1+\frac{1}{2}QP_1|h_1|^2} \leq P_2 \leq P_{s,\max} - P_1, \tag{4.42c}$$

$$\zeta_r \leq P_r \leq P_{r,\max}, \tag{4.42d}$$

where $\zeta_r \triangleq \max\left\{0, \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2}\frac{1+\frac{1}{2}QP_1|h_2|^2}{1+P_1|h_2|^2}\right\}$. Note that (4.42d) indicates that the feasible domain of P_r depends on P_1 and P_2 while the constraints of P_1 and P_2 in (4.42b) and (4.42c) do not rely on P_r. For any given P_1 and P_2, the weighted sum power is minimized when P_r is equal to the least possible value, i.e., $P_r = \zeta_r$. Besides, we note that increasing P_1 improves R_1^{ANOMA} while worsens $R_{2,L}^{\text{ANOMA}}$ and $R_{2\to1,L}^{\text{ANOMA}}$ due to the increased interference from User 1's message. Then, the powers P_2 and P_r have to increase to counteract the interference of User 1's message. As a result, P_1 should also be chosen as the least possible value within the feasible domain (4.42b) to minimize the weighted sum power, i.e., $P_1 = \frac{\gamma_1+\epsilon^*}{|h_1|^2}$. By

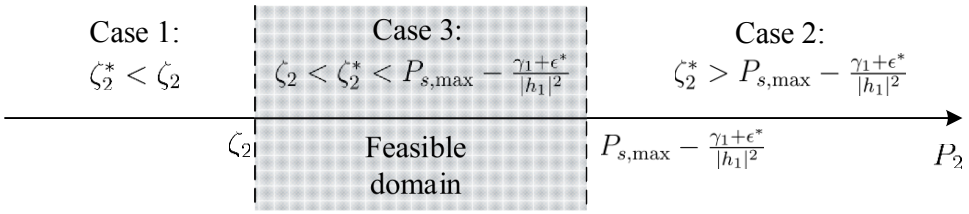

Figure 4.2: Illustration of the relationship among ζ_2^*, ζ_2, and $P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$.

substituting the values of P_r and P_1, the optimization problem (4.42) becomes

$$\min_{P_2} \quad \omega_s P_2 + \omega_r \max\left\{0, \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2}\frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}\right\}, \tag{4.43a}$$

$$\text{s.t.} \quad \zeta_2 \le P_2 \le P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}, \tag{4.43b}$$

where

$$\zeta_2 = \max\left\{\frac{\gamma_2}{|h_1|^2}\frac{1+\gamma_1+\epsilon^*}{1+\frac{1}{2}Q(\gamma_1+\epsilon^*)}, \left(\frac{\gamma_2}{|h_{12}|^2} - P_{r,\max}\right)\frac{|h_{12}|^2}{|h_2|^2}\frac{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}\right\}, \tag{4.44}$$

and the rightmost term in (4.44) is derived by setting $\zeta_r < P_{r,\max}$.

We note from (4.42b) and (4.43b) that if $\frac{\gamma_1+\epsilon^*}{|h_1|^2} > P_{s,\max}$ or $\zeta_2 > P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, there is no valid solution for the power minimization problem, i.e., QoS requirements cannot be satisfied with the limited transmit powers of the BS and the relay user. The following analysis is under the assumption that there are valid solutions for the power minimization problem.

By setting $\frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2}\frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2} = 0$, we obtain

$$P_2 = \zeta_2^* \triangleq \frac{\gamma_2}{|h_2|^2}\frac{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}. \tag{4.45}$$

Then, the objective function (4.43a) becomes $\omega_s P_2 + \omega_r\left(\frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2}\frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}\right)$ if $P_2 < \zeta_2^*$ and $\omega_s P_2$ otherwise. As shown in Fig. 4.2, we separate the following analysis into

three cases according to the relationship among ζ_2^*, ζ_2, and $P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$.

4.4.2.1 Case 1

If $\zeta_2^* < \zeta_2$, the optimization problem (4.43) becomes

$$\min_{P_2} P_2, \quad \text{s.t. } \zeta_2 \leq P_2 \leq P_{s,\max} - \frac{\gamma_1 + \epsilon^*}{|h_1|^2}, \tag{4.46}$$

In this case, it is easy to obtain that the optimal transmit powers are $P_1^* = \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, $P_2^* = \zeta_2$, and $P_r^* = 0$. Intuitively, this case indicates that the channel between the BS and User 2 is strong enough such that no relay transmission is needed to satisfy the QoS requirements at User 2.

4.4.2.2 Case 2

If $\zeta_2^* > P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, the optimization problem (4.43) becomes

$$\min_{P_2} \quad \omega_s P_2 + \omega_r \left(\frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2 + \frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2 + (\gamma_1+\epsilon^*)|h_2|^2} \right), \tag{4.47a}$$

$$\text{s.t.} \quad \zeta_2 \leq P_2 \leq P_{s,\max} - \frac{\gamma_1 + \epsilon^*}{|h_1|^2}. \tag{4.47b}$$

By omitting the constant terms, the objective function (4.47a) becomes

$$\min_{P_2} \left(\omega_s - \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2 + \frac{1}{2}Q(\gamma_1 + \epsilon^*)|h_2|^2}{|h_1|^2 + (\gamma_1 + \epsilon^*)|h_2|^2} \right) P_2. \tag{4.48}$$

Note that the solution of (4.47) depends on the values of ω_s and ω_r. The solutions can be

given as follows: If $\omega_s = \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, the optimal transmit powers are

$$P_1^* = \frac{\gamma_1+\epsilon^*}{|h_1|^2}, P_2^* \in \left[\zeta_2, P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}\right], P_r^* = \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2^*|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}, \tag{4.49}$$

where P_2^* can be any value in the given range. We provide the intuitive explanation for the solution as follows: One can observe from (4.49) that P_r^* decreases with P_2^*. Under certain conditions on ω_r and ω_s, i.e., $\omega_s = \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, the weighted sum power will be constant as the decrease in P_r^* is equal to the increase in P_2^*.

If $\omega_s > \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, P_2^* should be chosen the least possible value. The optimal transmit powers are given by

$$P_1^* = \frac{\gamma_1+\epsilon^*}{|h_1|^2}, P_2^* = \zeta_2, P_r^* = \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2^*|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}. \tag{4.50}$$

If $\omega_s < \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, P_2^* should choose the largest possible value. The optimal transmit powers are given by

$$P_1^* = \frac{\gamma_1+\epsilon^*}{|h_1|^2}, P_2^* = P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}, P_r^* = \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2^*|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}. \tag{4.51}$$

4.4.2.3 Case 3

If $\zeta_2 < \zeta_2^* < P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, the optimization problem (4.43) becomes two sub-problems, i.e.,

$$\min_{P_2} P_2, \quad \text{s.t. } \zeta_2^* \le P_2 \le P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}, \tag{4.52}$$

and

$$\min_{P_2} P_2 \left(\omega_s - \omega_r \frac{P_2|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2 + \frac{1}{2}Q(\gamma_1 + \epsilon^*)|h_2|^2}{|h_1|^2 + (\gamma_1 + \epsilon^*)|h_2|^2} \right),$$
$$\text{s.t. } \zeta_2 \leq P_2 \leq \zeta_2^*. \tag{4.53}$$

By following the derivation of Case 1, one can solve the problem (4.52). Similarly, by following the steps of Case 2, one can solve the problem (4.53). We assume that the optimal transmit powers for (4.52) and (4.53) are $[\tilde{P}_1,\ \tilde{P}_2,\ \tilde{P}_r]$ and $[\bar{P}_1,\ \bar{P}_2,\ \bar{P}_r]$, respectively. Then, the solution for Case 3 is given by

$$[P_1^*, P_2^*, P_r^*] = \underset{[P_1,P_2,P_r] \in \left\{ [\tilde{P}_1,\tilde{P}_2,\tilde{P}_r], [\bar{P}_1,\bar{P}_2,\bar{P}_r] \right\}}{\arg\min} \omega_s(P_1 + P_2) + \omega_r P_r. \tag{4.54}$$

To summarize the solutions, we provide Algorithm 1 to solve the problem (4.42). Note that Algorithm 1 is implemented by using only the conditional statements. All the expressions used in Algorithm 1 are closed-form. Hence, Algorithm 1 can run in constant time.

4.4.3 Comparison with C-NOMA

According to (4.25), (4.26), Theorems 4.1 and 4.3, the expressions for the throughputs in C-ANOMA systems, $R_{2\to1,L}^{\text{ANOMA}}$, R_1^{ANOMA}, and $R_{2,L}^{\text{ANOMA}}$, become those in C-NOMA systems, $R_{2\to1}^{\text{NOMA}}$, R_1^{NOMA}, and R_2^{NOMA}, by setting $\tau = 0$. Therefore, the solutions derived in the previous subsection can be applied to the C-NOMA systems simply by setting $\tau = 0$ which then results in $\epsilon^* = 0$ and $Q = 0$. For the C-NOMA systems, the power minimization problem (4.42) becomes

$$\min_{P_1,P_2,P_r} \quad \omega_s(P_1 + P_2) + \omega_r P_r, \tag{4.55a}$$

Algorithm 1 Algorithm to find the optimal powers under QoS constraints

function SOLVE_CASE_1(L, U)
  return $P_2^* = L$, $P_r^* = 0$.
end function
function SOLVE_CASE_2(L, U)
  if $\omega_s = \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, **then** $P_2^* = \text{random}\left(\left[\zeta_2, P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}\right]\right)$.
  else if $\omega_s > \omega_r \frac{|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$, **then** $P_2^* = L$.
  else $P_2^* = U$.
  return P_2^*, $P_r^* = \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2^*|h_2|^2}{|h_{12}|^2} \frac{|h_1|^2+\frac{1}{2}Q(\gamma_1+\epsilon^*)|h_2|^2}{|h_1|^2+(\gamma_1+\epsilon^*)|h_2|^2}$.
end function
if $\frac{\gamma_1+\epsilon^*}{|h_1|^2} > P_{s,\max}$ or $\zeta_2 > P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, **then** there is no solution, break.
$P_1^* = \frac{\gamma_1+\epsilon^*}{|h_1|^2}$
if $\zeta_2^* < \zeta_2$, **then** P_2^*, P_r^* = SOLVE_CASE_1(ζ_2, $P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$).
else if $\zeta_2^* > P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$, **then** P_2^*, P_r^* = SOLVE_CASE_2(ζ_2, $P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2}$).
else
  $\tilde{P}_2^*, \tilde{P}_r^* = \text{SOLVE\_CASE\_1}(\zeta_2^*, P_{s,\max} - \frac{\gamma_1+\epsilon^*}{|h_1|^2})$.
  $\hat{P}_2^*, \hat{P}_r^* = \text{SOLVE\_CASE\_2}(\zeta_2, \zeta_2^*)$.
  $P_2^*, P_r^* = \arg\min_{[P_2,P_r]\in\{[\tilde{P}_2^*,\tilde{P}_r^*],[\hat{P}_2^*,\hat{P}_r^*]\}} \omega_s P_2 + \omega_r P_r$.
return P_1^*, P_2^*, P_r^*.

$$\text{s.t.} \quad \frac{\gamma_1}{|h_1|^2} \le P_1 \le P_{s,\max}, \tag{4.55b}$$

$$\frac{\gamma_2\left(1+P_1|h_1|^2\right)}{|h_1|^2} \le P_2 \le P_{s,\max} - P_1, \tag{4.55c}$$

$$\max\left\{0, \frac{\gamma_2}{|h_{12}|^2} - \frac{P_2|h_2|^2}{|h_{12}|^2\left(1+P_1|h_2|^2\right)}\right\} \le P_r \le P_{r,\max}. \tag{4.55d}$$

Note that the feasible domains of P_2 and P_r in (4.55c) and (4.55d) are the subsets of those in (4.42c) and (4.42d), respectively. As a result, for a sufficiently large N, the minimization problem (4.42) for the C-ANOMA systems is a relaxation of the minimization problem (4.55) for the C-NOMA systems [58]. That is, the problem (4.42) provides a solution to minimize the weighted sum power within a wider feasible domain compared with (4.55). In other words, if $[P_{1,\text{ANOMA}}^*, P_{2,\text{ANOMA}}^*, P_{r,\text{ANOMA}}^*]$ and $[P_{1,\text{NOMA}}^*, P_{2,\text{NOMA}}^*, P_{r,\text{NOMA}}^*]$ are the optimal

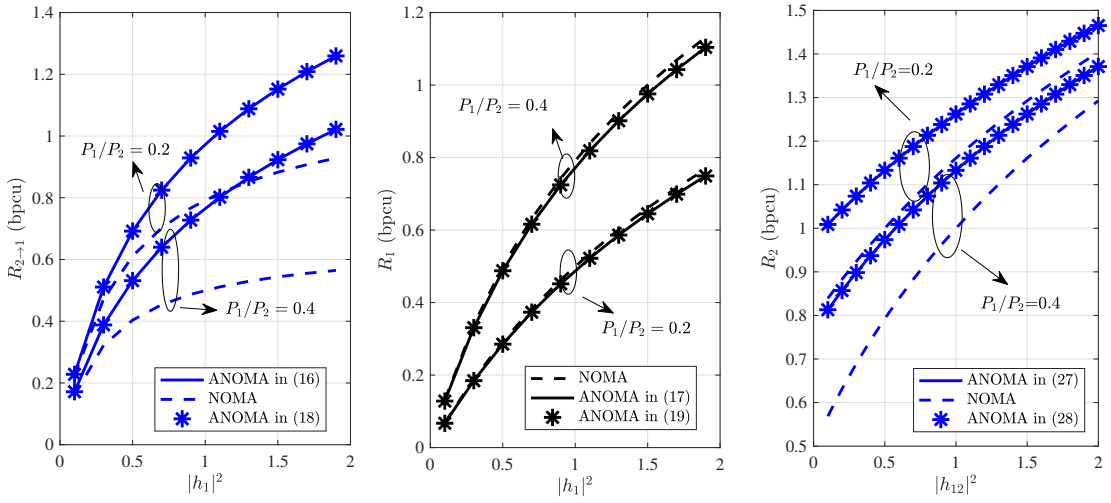

Figure 4.3: The throughputs $R_{2\to1}$, R_1, and R_2 as functions of the channel gain $|h_1|^2$ or $|h_{12}|^2$ for C-ANOMA and C-NOMA systems when $N = 10$, $\tau = 0.5$, $P_1 + P_2 = 5$, $P_r = 2$, $|h_2|^2 = 1$.

solutions for (4.42) and (4.55), respectively, we have

$$\omega_s(P^*_{1,\text{ANOMA}} + P^*_{2,\text{ANOMA}}) + \omega_r P^*_{r,\text{ANOMA}} \leq \omega_s(P^*_{1,\text{NOMA}} + P^*_{2,\text{NOMA}}) + \omega_r P^*_{r,\text{NOMA}}. \tag{4.56}$$

We note from (4.56) that for a sufficiently large block length, the C-ANOMA systems can consume less power compared with the C-NOMA systems in order to guarantee the same QoS. We will illustrate this phenomenon with numerical results in Section 4.5.

4.5 Numerical Results

In this section, we present numerical results to compare the throughputs and power consumptions of C-NOMA and C-ANOMA systems.

First, we compare the throughputs of Users 1 and 2 in the C-NOMA and C-ANOMA systems with different ratios of P_1 to P_2 in Fig. 4.3. The curves of "ANOMA in (4.16)/(4.18)/(4.17)

/(4.19)/(4.27)/(4.28)" are derived directly from the expressions in (4.16)/(4.18)/(4.17)/(4.19) /(4.27)/(4.28). First, in Fig. 4.3, it is shown that the throughputs calculated by (4.18), (4.19), and (4.28) completely align with the results of (4.16), (4.17), and (4.27), respectively, which verifies the correctness of (4.18), (4.19), and (4.28). Second, it is demonstrated that the throughputs $R_{2\to1}$ and R_2 in the C-ANOMA systems are higher than those in the C-NOMA systems. And R_1^{ANOMA} is less than but very close to R_1^{NOMA} even for a relatively small block length $N = 10$, especially when $|h_1|^2$ is small. Third, Fig. 4.3 shows that the throughputs in both C-ANOMA and C-NOMA systems increase with the channel gain. More specifically, the gaps of $R_{2\to1}$ between the C-ANOMA and C-NOMA systems grow wider as $|h_1|^2$ increases. In contrast, the gaps of R_2 between the C-ANOMA and C-NOMA systems shrink as $|h_{12}|^2$ increases. Note that R_2 depends on both the broadcast link from the BS and the relay link from User 1. The sampling diversity can only be obtained through the asynchronous transmission from the broadcast link. As $|h_{12}|^2$ increases, the quality of the relay link becomes more and more dominant in calculating R_2. Accordingly, the throughput gain from the sampling diversity becomes less and less noticeable as $|h_{12}|^2$ increases while $|h_2|^2$ is constant. Finally, it is evident that the actual throughput of User 2 for C-ANOMA is better than that of C-NOMA since $R_2^{\text{ANOMA}} > R_2^{\text{NOMA}}$ and $R_{2\to1}^{\text{ANOMA}} > R_{2\to1}^{\text{NOMA}}$.

We compare the throughput performances of C-ANOMA and C-NOMA systems in the full-duplex or half-duplex mode in Fig. 4.4. In our simulation, we calculate the users' throughputs in the full-duplex C-NOMA systems based on the SINR expressions derived in the existing literature, for example [45]. Using the notation in [45], $|h_{\text{LI}}|^2$ stands for the level of the residual loop self-interference at the relay user caused by the full-duplex operation. For the full-duplex C-ANOMA and C-NOMA, the throughput of the relay user is calculated by treating the self-interference as noise. In Fig. 4.4, it is shown that in both half-duplex and full-duplex modes, $R_{2\to1}$ and R_2 in C-ANOMA outperform those in C-NOMA while R_1 in C-ANOMA is very close to that in C-NOMA. These results align with the results in Fig. 4.3. Besides, in some cases, the performance of full-duplex systems can be worse than that of

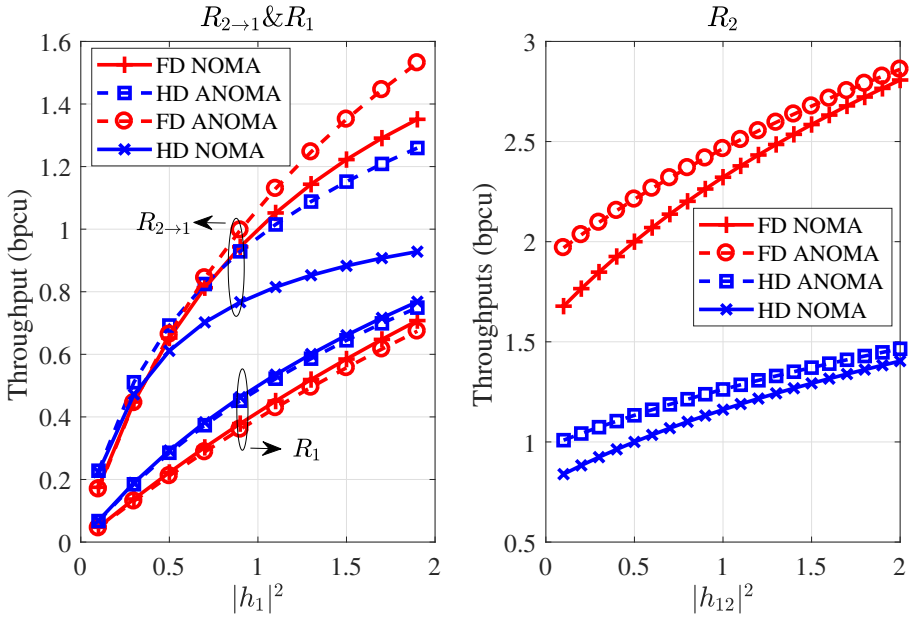

Figure 4.4: The throughputs as functions of the channel gain for the full-duplex or half-duplex C-NOMA and C-ANOMA systems when $N = 10$, $\tau = 0.5$, $P_1 = 1$, $P_2 = 4$, $P_r = 2$, $|h_2|^2 = 1$, $|h_{\text{LI}}|^2 = 1$.

the half-duplex systems due to the residual loop self-interference. For example, R_1 in the full-duplex C-ANOMA or C-NOMA system is worse than that in half-duplex C-ANOMA or C-NOMA system. As studied in the existing literature, the self-interference cancellation plays a crucial role in the full-duplex systems [59]. Moreover, it is also shown that R_2 in the full-duplex C-ANOMA or C-NOMA system is better than that in the half-duplex C-ANOMA or C-NOMA system because the relay user in the full-duplex mode can relay the signal to the weak user without consuming an additional time block. Note that the full-duplex operation and the realization of the self-interference require higher hardware complexity and power consumption at the relay user compared with the half-duplex operation.

We compare the performances of C-ANOMA, C-NOMA, and C-OMA in Fig. 4.5. Following the comparison between C-NOMA and C-OMA in the existing literature [46, 52, 45, 44], we adopt the time division multiple access (TDMA) C-OMA as an example of C-OMA systems. Specifically, the transmission in C-OMA occupies three time slots. In the first time slot, the BS transmits User 1's messages to User 1. In the second time slot, the BS broadcasts

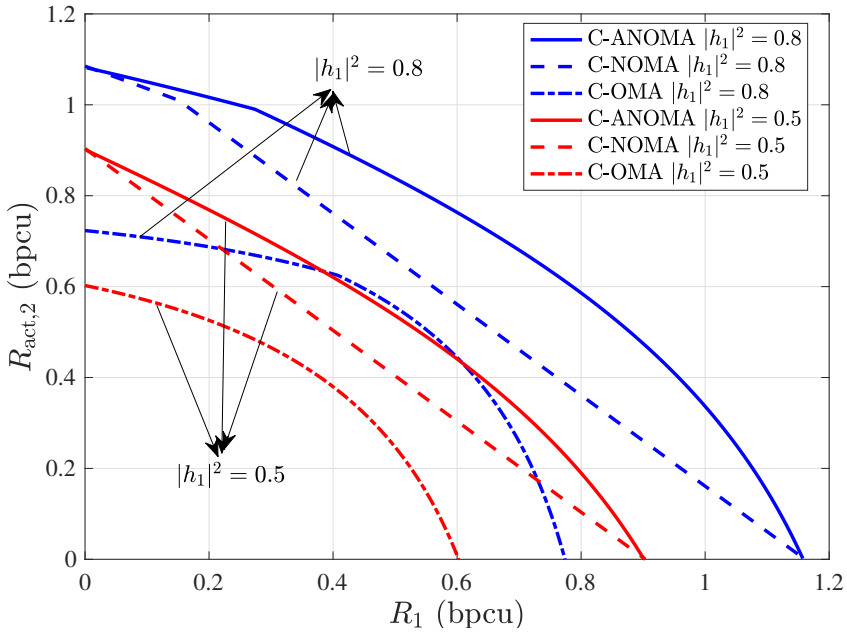

Figure 4.5: The throughputs R_1 vs. $R_{\text{act},2}$ for C-ANOMA, C-NOMA, and C-OMA systems when $|h_{12}|^2 = 1$, $|h_2|^2 = 0.5$, $N = 100$, $\tau = 0.5$, $P_1 + P_2 = 5$, $P_r = 1$.

User 2's messages to Users 1 and 2. In the last time slot, User 1 relays User 2's messages to User 2. In Fig. 4.5, it is shown that C-NOMA achieves a better performance compared with C-OMA. Besides, it is also demonstrated that C-ANOMA outperforms C-NOMA because the oversampling technique provides extra sampling diversity [6, 31, 7].

In Fig. 4.6, we show how the throughputs in C-ANOMA systems change with the block length N. Since the expression for R_1^{ANOMA} in (4.19) is simple, the curves of R_1^{ANOMA} are omitted in Fig. 4.6. It is shown that as the block length increases, the accurate throughputs $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} converge to the asymptotic ones calculated by (4.26) and (4.33), respectively. We note that the asymptotic throughputs, $R_{2\to1,\text{asymp}}^{\text{ANOMA}}$ and $R_{2,\text{asymp}}^{\text{ANOMA}}$, perfectly approximate the accurate throughputs, $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA}, when $N > 50$. And for both $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA}, the accurate throughputs exceed their asymptotic lower bounds when $N > 20$. As a result, for $N > 20$, it is reasonable to use the lower bounds of the asymptotic throughputs as the constraints (4.37b) and (4.37d) in order to simplify the optimization problem. Besides, Fig. 4.6 verifies Theorems 4.1 and 4.3 in addition to showing that the

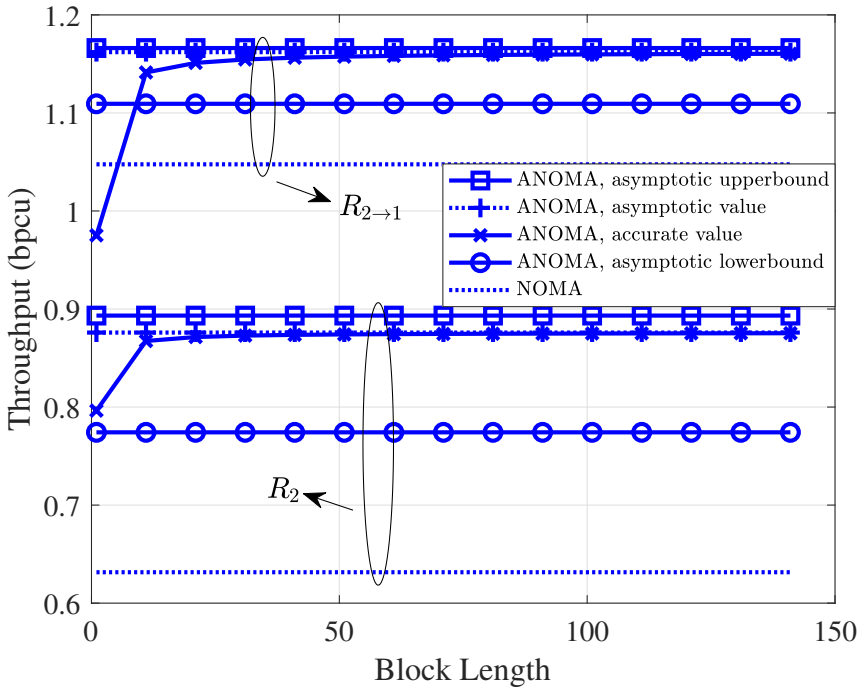

Figure 4.6: The throughputs R_2 and $R_{2\to1}$ as functions of the block length N for C-ANOMA and C-NOMA systems when $\tau = 0.5$, $P_1 = 1.5$, $P_2 = 3.5$, $P_r = 2$, $|h_1|^2 = 1$, $|h_2|^2 = 0.8$, $|h_{12}|^2 = 1$.

C-ANOMA systems outperform the C-NOMA systems for relatively small values of N.

Based on the results in Fig. 4.6, we discuss the time delay of the message delivery. In the non-cooperative NOMA or full-duplex C-NOMA systems, one message block is delivered to two users simultaneously via one time block. In contrast, in the half-duplex C-NOMA or C-ANOMA systems, one additional time block is needed to transmit the same message block. In order to reduce the delay, the block length is expected to be small. Besides, as shown in Fig. 4.6, the accurate throughputs in C-ANOMA systems increase with the block length. It is worth mentioning that when the block length is large, e.g., $N > 50$, a greater block length only results in a very subtle throughput improvement. Therefore, considering the delay and the throughput, a modest block length is desired.

We also study the optimal design of C-ANOMA systems. Fig. 4.7 (a) shows the optimal normalized timing mismatch τ^* to maximize R_2^{ANOMA} or $R_{2\to1}^{\text{ANOMA}}$ as a function of the block

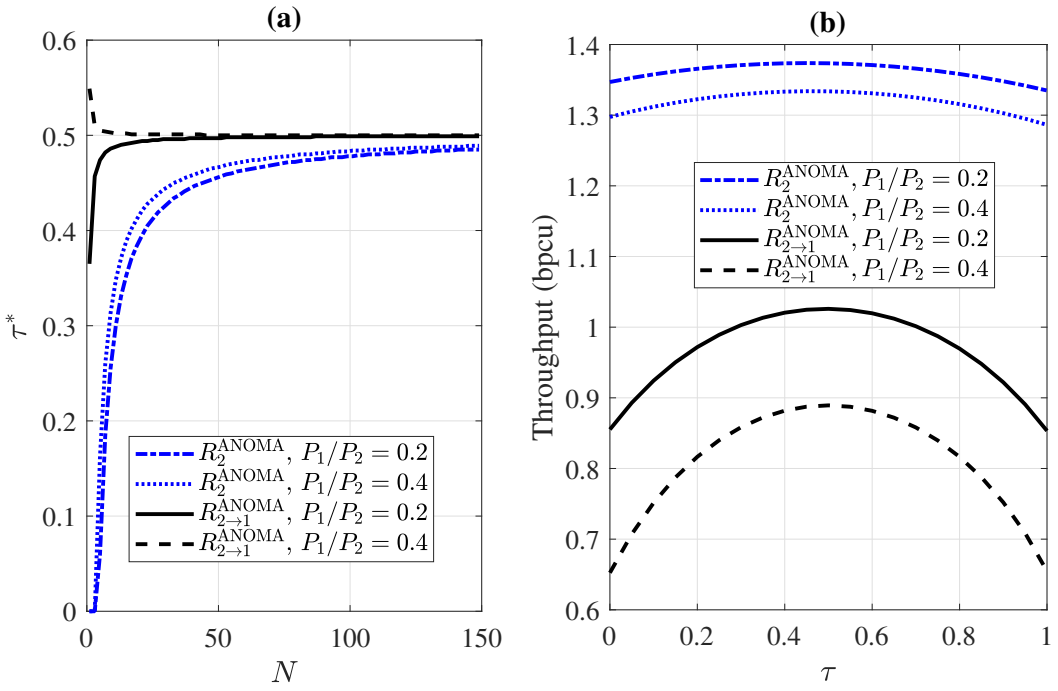

Figure 4.7: (a) The optimal normalized timing mismatch τ^* to maximize the throughputs R_2^{ANOMA} and $R_{2\to1}^{\text{ANOMA}}$ as a function of the block length N when $P_1 + P_2 = 5$, $P_r = 2$, $|h_1|^2 = 1$, $|h_2|^2 = 0.5$, $|h_{12}|^2 = 2$. (b) The throughputs R_2 and $R_{2\to1}$ as functions of the normalized timing mismatch τ when $N = 50$, $P_1 + P_2 = 5$, $P_r = 2$, $|h_1|^2 = 1$, $|h_2|^2 = 0.5$, $|h_{12}|^2 = 2$.

length N. In our simulation, τ^* is found by exhaustive search. Although τ^* varies a lot when N is relatively small, τ^* converges to 0.5 steadily as N increases for both R_2^{ANOMA} and $R_{2\to1}^{\text{ANOMA}}$ with different ratios of P_1 to P_2, as predicted by our analytical results. This is because the timing mismatch only exists in the asynchronous transmission in the broadcast phase and will affect $R_{2\to1}^{\text{ANOMA}}$ and R_2^{ANOMA} in the same way. Moreover, Fig. 4.7 (b) presents how the throughputs change with the normalized timing mismatch. It is demonstrated that for both R_2^{ANOMA} and $R_{2\to1}^{\text{ANOMA}}$, the throughputs are maximized when $\tau \approx 0.5$, which verifies the results shown in Fig. 4.7 (a). Compared with $R_{2\to1}^{\text{ANOMA}}$, the choice of τ has a greater impact on R_1^{ANOMA}. It is because the relay link dominates the performance of User 2 if the channel of the relay link is good. Besides, it is shown that for $\tau \in [0.4, 0.6]$, the choice of τ only has a subtle effect on $R_{2\to1}^{\text{ANOMA}}$ and R_1^{ANOMA}.

Moreover, we show the minimized weighted sum power under the QoS constraints as a

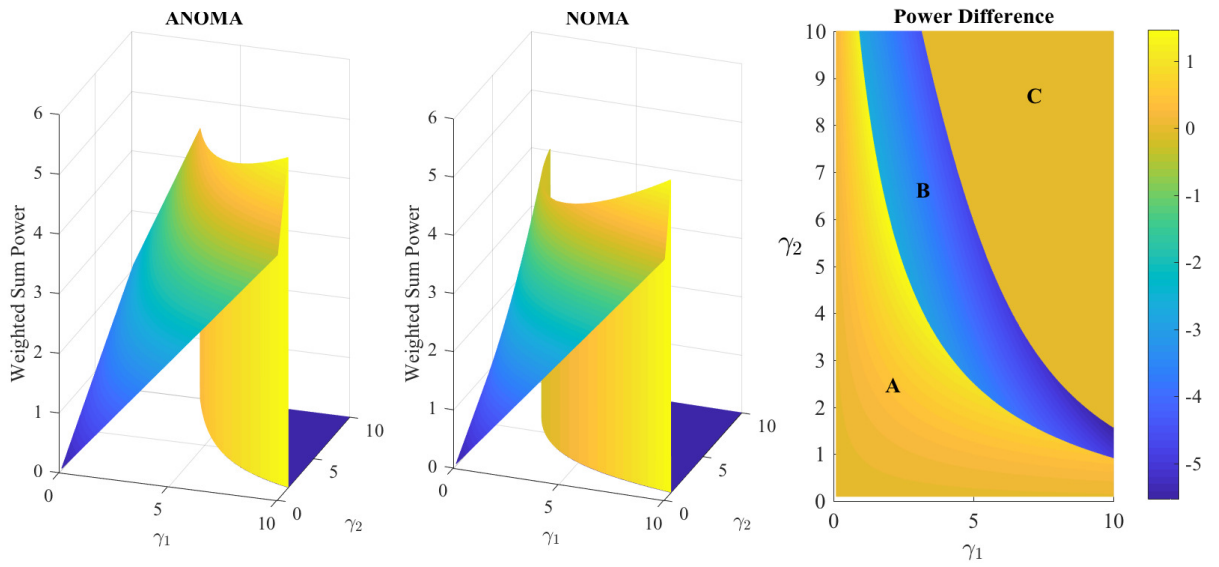

Figure 4.8: The minimized weighted sum power under the QoS constraints as a function of the target SINRs, γ_1 and γ_2, for the C-NOMA and C-ANOMA systems when $\tau = 0.5$, $P_{s,\max} = 20$, $P_{r,\max} = 5$, $\omega_s = 0.2$, $\omega_r = 0.8$, $|h_1|^2 = 1$, $|h_2|^2 = 0.5$, $|h_{12}|^2 = 2$, $N = 100$. The right-most figure illustrates the difference between the minimized weighted sum power in C-NOMA systems and that in C-ANOMA systems. In that figure, A stands for the area where QoS constraints can be satisfied for both ANOMA and NOMA and $P_{\text{sum}}^{\text{ANOMA}} < P_{\text{sum}}^{\text{NOMA}}$. B stands for the area where QoS constraints can be satisfied for ANOMA, but not for NOMA. C stands for the area where QoS constraints cannot be satisfied for either ANOMA or NOMA.

function of target SINRs, γ_1 and γ_2, for C-NOMA and C-ANOMA systems in Fig. 4.8. We set ω_s and ω_r as 0.2 and 0.8, respectively, because the power consumption of the relay user with limited battery capacity has a higher priority in the power minimization problem. In Fig. 4.8, the weighted sum power is calculated by solving the power optimization problem (4.42) for the C-NOMA (setting $\tau = 0$) and C-ANOMA (setting $\tau = 0.5$) systems. In our simulation, we assume that the BS and the relay user will stop transmission (i.e., $P_1 = P_2 = P_r = 0$) if the QoS constraints cannot be satisfied. For both C-NOMA and C-ANOMA systems, it is shown in Fig. 4.8 that the weighted sum power increases with the target SINRs until the BS and the relay user reach their power limits and stop transmission. To further compare the power consumptions, we calculate the difference of the weighted sum powers between the C-NOMA and C-ANOMA systems and provide the results in Fig. 4.8. As shown in Fig. 4.8, C-ANOMA systems can consume less power compared with C-NOMA systems to guarantee

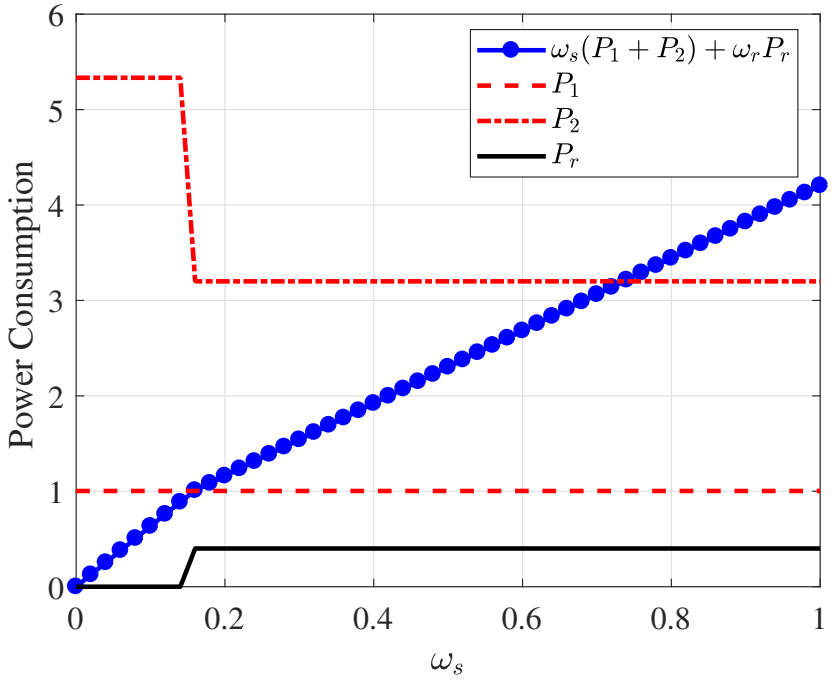

Figure 4.9: The power consumptions as functions of the weight allocated to the transmit power of the BS, i.e., ω_s, for the C-ANOMA systems when $\tau = 0.5$, $P_{s,\max} = 20$, $P_{r,\max} = 5$, $|h_1|^2 = 1$, $|h_2|^2 = 0.5$, $|h_{12}|^2 = 2$, $N = 100$.

the same QoS in the area A. In the area B, it is shown that C-ANOMA systems can still satisfy the QoS with limited transmit powers while C-NOMA systems cannot. When both γ_1 and γ_2 are large, i.e., in the area C, neither C-NOMA nor ANOMA systems can satisfy the QoS with the limited transmit powers.

Finally, we show how the power consumptions change with the weight ω_s in the power minimization problem in Fig. 4.9. In our simulation, we set $\omega_r = 1 - \omega_s$. In Fig. 4.9, the power allocated to User 1 does not change with ω_s as long as the BS has enough transmit power to support the QoS of User 1. If ω_s is large, the BS can save a large amount of power (decreases by about 2) under the help of the relay user (transmit power increases by about 0.5) because the channel of the relay link is better than that of the broadcast link between the BS and User 2. When ω_s is small (ω_r is large), the relay user keeps silent to reduce energy consumption. When ω_s is large (ω_r is small), the BS communicates with User 2

under the help of the relay user, which takes advantage of the relay link to complement the large path loss between the BS and User 2. Hence, Fig. 4.9 shows that one can make a trade-off between the power consumption of the BS and that of the relay user by adjusting the weight ω_s.

4.6 Conclusion

In this chapter, we study the half-duplex C-ANOMA systems with user relaying. We analytically prove that for a sufficiently large block length, the strong user in C-ANOMA systems can achieve the same throughput as that in C-NOMA systems while the weak user in C-ANOMA systems benefits from the symbol-asynchronous transmission. Moreover, we analyze the optimal design of the C-ANOMA systems. As the block length increases, the optimal timing mismatch converges to half of the symbol interval. Besides, we solve a weighted sum power minimization problem under QoS constraints. Numerical results demonstrate that C-ANOMA systems can consume less power to satisfy the same QoS requirements compared with C-NOMA systems.

There are several directions worth studying in the future. For example, since the analysis of this chapter is under the assumption of perfect channel estimation, the impact of imperfect channel estimation on the system performance is an interesting future work.

Chapter 5

Downlink Asynchronous Non-Orthogonal Multiple Access with Imperfect Channel State Information

5.1 Introduction

NOMA has been extensively studied in academia. Zhu et al. [60] investigated resource allocation in downlink NOMA systems abd analytically characterized the optimal power allocation with given channel assignment over multiple channels under different performance criteria. Ding et al. [61] studied the performance of NOMA in a cellular downlink scenario with randomly deployed users and showed the superiority of NOMA in terms of ergodic sum rates. In addition, the NOMA systems have also been studied with practical consideration, such as the imperfect channel state information (CSI). For example, Yang et al. investigated the performance of NOMA with partial CSI, including the case with imperfect CSI and that with second order statistics. Fang et al. [62] investigated energy efficiency improvement

for a downlink NOMA single-cell network by considering imperfect CSI. The system optimization problem has also been studied in the NOMA system with imperfect CSI, e.g., the optimal resource allocation for power-efficient multicarrier NOMA [63] and the optimization of scheduling[64]. Besides, Liu and Jafarkhani analyzed downlink NOMA networks with limited channel feedback information in [2].

Recently, a scheme called asynchronous NOMA (ANOMA) has been proposed and studied [48, 49, 65]. Compared with conventional (synchronous) NOMA, a timing mismatch is intentionally introduced in the superimposed signal and the oversampling technique is utilized at receiver to achieve the sampling diversity. In [48], the uplink ANOMA system was thoroughly studied and the impact of timing error on ANOMA systems was analyzed. It was demonstrated that the ANOMA can achieve higher sum throughput than the NOMA. In [49], the downlink ANOMA system was also investigated and the superiority of ANOMA over the conventional NOMA was demonstrated. However, to the best of our knowledge, the ANOMA systems with imperfect CSI have never been studied and little is known about the performance comparison between ANOMA and NOMA with imperfect CSI.

In this chapter, we consider a downlink ANOMA system with imperfect CSI. With channel estimation error, we prove that with a relatively large frame length, the ANOMA systems can achieve lower outage probability compared with the NOMA systems. To this end, we derive the expressions for the individual throughput of each user and greatly simplify them for the asymptotic case of infinite frame length. Besides, we analytically prove that the optimal timing mismatch converges to half of a single symbol length as the frame length goes up.

The remainder of the chapter is organized as follows. The downlink ANOMA system model is presented in Section 5.2. The throughput performance of the ANOMA system is analyzed in Section 5.3. We discuss the outage probability in Section 5.4. Numerical and simulation results are presented in Section 5.5. Finally, we draw the conclusions in Section 5.6.

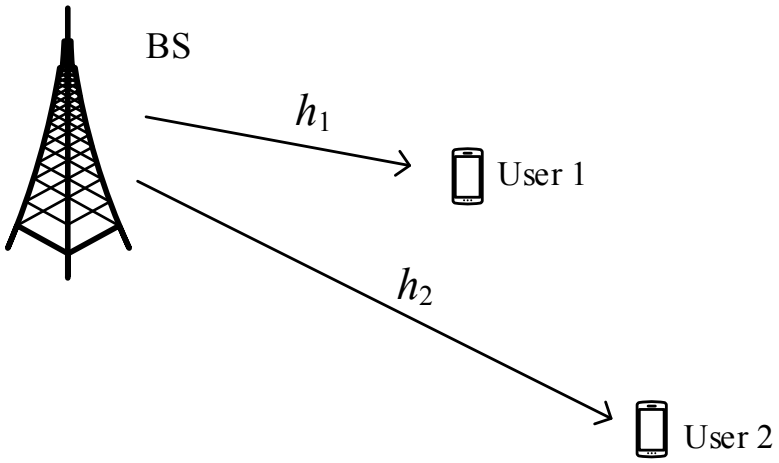

Figure 5.1: Illustration of a two-user downlink system.

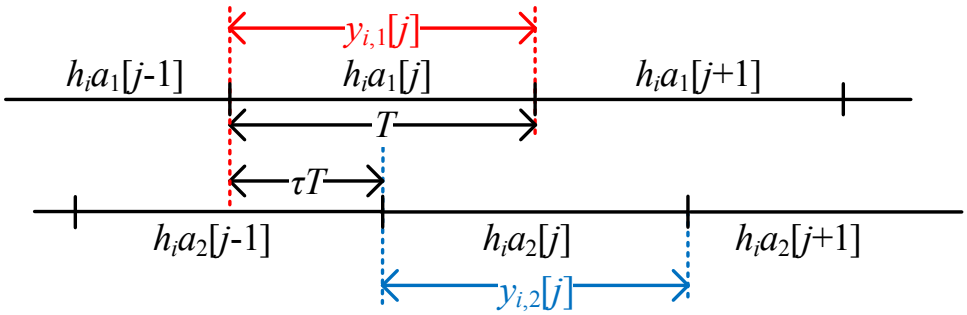

Figure 5.2: Illustration of oversampling in ANOMA systems.

Notations: $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^T$ denotes the transpose, $(\cdot)^{-1}$ denotes the inverse operation, $\otimes$ denotes the Kronecker product, $|x|$ denotes the absolute value of x, $\bar{x}$ denotes the complex conjugate of x, $\mathbb{E}[\cdot]$ denotes the expectation operation, $\mathcal{CN}(0,1)$ denotes the complex normal distribution with zero mean and unit variance. diag($\mathbf{x}$) stands for a diagonal matrix whose k-th diagonal element is equal to the k-th entry of vector $\mathbf{x}$.

5.2 System Model

In this chapter, as shown in Fig. 5.1, we consider a downlink ANOMA system, including a single base station (BS) and two users which are equipped with a single antenna. User 1 has the strong channel while User 2 has the weak channel. We assume that the channel remains

static within each frame and varies independently from one frame to another. The channel coefficient between BS and User i is denoted as h_i $(i = 1, 2)$ and $|h_1|^2 > |h_2|^2$.

In ANOMA systems, a timing mismatch is intentionally introduced in the superimposed signal. As shown in Fig. 5.2, the intended timing mismatch between the symbols for Users 1 and 2 is denoted by τT, where T is the symbol length and τ, $0 \leq \tau < 1$, is the normalized timing mismatch. We assume that τ can be perfectly known at users as the downlink control information. Note that the ANOMA system becomes a synchronous NOMA system when $\tau = 0$.

Let $a_1[j] = \sqrt{P_1}s_1[j]$ and $a_2[j] = \sqrt{P_2}s_2[j]$, where $s_i[j]$ denotes the jth symbol sent to User i, $i = 1, 2$, P_i stands for the power allocated to User i. The transmitted signal at BS is given by

$$s(t) = \sum_{j=1}^{N} a_1[j]p(t - jT) + \sum_{j=1}^{N} a_2[j]p(t - jT - \tau T). \tag{5.1}$$

where N denotes the number of symbols in a frame, i.e., the frame length, $p(\cdot)$ denotes the pulse-shaping filter. Without loss of generality, the rectangular pulse shape is adopted, i.e., $p(t) = 1/\sqrt{T}$ when $t \in [0, T]$ and $p(t) = 0$ otherwise. Then, the received signal of User i is given by

$$y_i(t) = h_i s(t) + n_i(t). \tag{5.2}$$

The oversampling technique [6, 31, 7], depicted in Fig. 3.2, is utilized at the receiver to create extra samples. As shown in Fig. 3.2, the receiver uses the matched filter, sampling at jT and $(j + \tau)T$, $j = 1, \cdots, N$, to obtain two sample vectors, denoted by $[y_{i,1}[1], \cdots, y_{i,1}[N]]^T$ and $[y_{i,2}[1], \cdots, y_{i,2}[N]]^T$. Specifically, the jth element in the first sample vector of User i is

given by

$$y_{i,1}[j] = \int_0^\infty y_i(t)p(t-jT)dt = h_i a_1[j] + \tau h_i a_2[j-1] + (1-\tau)h_i a_2[j] + n_{i,1}[j], \quad (5.3)$$

where $n_{i,1}[j] = \int_0^\infty n_i(t)p(t-jT)dt$ denotes the additive noise. The jth element in the second sample vector is given by

$$y_{i,2}[j] = \int_0^\infty y_i(t)p(t-jT-\tau T)dt = h_i a_2[j] + \tau h_i a_1[j+1] + (1-\tau)h_i a_1[j] + n_{i,2}[j], \quad (5.4)$$

where $n_{i,2}[j] = \int_0^\infty n_i(t)p(t-jT-\tau T)dt$ denotes the additive noise.

Combining (5.3) with (5.4), the received signal at User i is given by

$$\mathbf{Y}_i = h_i\sqrt{P_1}\mathbf{R}\mathbf{G}_1\mathbf{S}_1 + h_i\sqrt{P_2}\mathbf{R}\mathbf{G}_2\mathbf{S}_2 + \mathbf{N}_i \quad (5.5)$$

where $\mathbf{Y}_i = [y_{i,1}[1]\ y_{i,2}[1]\ \cdots\ y_{i,1}[N]\ y_{i,2}[N]]^T$, $\mathbf{G}_1$ and $\mathbf{G}_2$ are $2N$-by-N matrices given by $\mathbf{G}_1 = \mathbf{I}_N \otimes [1\ 0]^T$ and $\mathbf{G}_2 = \mathbf{I}_N \otimes [0\ 1]^T$, $\mathbf{S}_i = [s_i[1]\ \cdots\ s_i[N]]^T$ $(i = 1, 2)$, $\mathbf{N}_i = [n_{i,1}[1]\ n_{i,2}[1]\ \cdots\ n_{i,1}[N]\ n_{i,2}[N]]^T$, and

$$\mathbf{R} = \begin{bmatrix} 1 & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1 & \tau & 0 & \cdots & 0 \\ 0 & \tau & 1 & 1-\tau & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \tau & 1 & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1 \end{bmatrix}. \quad (5.6)$$

Note that multiplying $\mathbf{R}$ by $\mathbf{G}_i$ outputs a $2N$-by-N matrix whose columns are equal to the odd (if $i = 1$) or even (if $i = 2$) columns of $\mathbf{R}$.

We assume that the transmitted symbols are normalized and independent to each other, such that $\mathbb{E}\left[\mathbf{S}_i\mathbf{S}_i^H\right] = \mathbf{I}$. Note that the noise terms in (3.2) and (3.3) are colored due to the

oversampling, and we have

$$\mathbb{E}\left\{n_{i,1}[j]n_{i,2}^{H}[j]\right\} = \int_{0}^{\infty}\int_{0}^{\infty}\mathbb{E}\left\{n_i(t)n_i^{H}(s)\right\}p\left(t-jT\right)p\left(s-jT-\tau T\right)dtds = 1-\tau. \tag{5.7}$$

Thus, the covariance matrix of $\mathbf{N}_i$ in (5.5) is given by

$$\mathbf{R}_{\mathbf{N}_i} = \mathbb{E}\left\{\mathbf{N}_i\mathbf{N}_i^{H}\right\} = \mathbf{R}. \tag{5.8}$$

5.3 Performance of ANOMA Systems with Imperfect CSI

In this chapter, we assume that BS can perfectly estimate CSI via uplink channel training. Then, BS feeds CSI back to users and the users fail to decode the feedback message perfectly. We model the channel coefficient between BS and User i as

$$h_i = \hat{h}_i + \epsilon_i, \tag{5.9}$$

where $\hat{h}_i$ is the estimated channel coefficient, $\epsilon_i \sim \mathcal{CN}(0, \sigma_{\epsilon_i}^2)$ denotes the channel estimation error at User i. We assume that $\hat{h}_i$ and ϵ_i are uncorrelated to each other.

Substituting (5.9), (5.5) can be rewritten as

$$\mathbf{Y}_i = \hat{h}_i\mathbf{RPG_1S_1} + \hat{h}_i\mathbf{RPG_2S_2} + \epsilon_i\mathbf{RPS} + \mathbf{N}_i, \tag{5.10}$$

where $\mathbf{S} = \mathbf{G_1S_1} + \mathbf{G_2S_2}$.

With stronger channel, User 1 adopts the SIC technique, i.e., first detects User 2's message,

removes it, and then detects its own message, which implies that User 2's message will be detected at both users. By treating (5.10) as a virtual multiple-input multiple-output (MIMO) system, the throughput of User i to decode User 2's message is calculated as

$$\begin{aligned}
R_{2\to i}^{\mathrm{A}} =& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\Big(\mathbf{R}_{\mathbf{N}i}+|\hat{h}_i|^2\mathbf{R}\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H\mathbf{R} \\
&+\sigma_{\epsilon_i}^2\mathbf{R}\mathbf{P}\mathbf{P}^H\mathbf{R}\Big)^{-1}|\hat{h}_i|^2\mathbf{R}\mathbf{P}\mathbf{G}_2\mathbf{G}_2^H\mathbf{P}^H\mathbf{R}\Big] \\
=& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\Big(\mathbf{I}+|\hat{h}_i|^2\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H\mathbf{R} \\
&+\sigma_{\epsilon_i}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\Big)^{-1}|\hat{h}_i|^2\mathbf{P}\mathbf{G}_2\mathbf{G}_2^H\mathbf{P}^H\mathbf{R}\Big] \\
=& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\mathbf{P}\left(|\hat{h}_i|^2\mathbf{G}_1\mathbf{G}_1^H+|\hat{h}_i|^2\mathbf{G}_2\mathbf{G}_2^H+\sigma_{\epsilon_i}^2\mathbf{I}\right)\mathbf{P}^H\mathbf{R}\Big] \\
&-\frac{1}{N+\tau}\log\det\Big[\mathbf{I}+|\hat{h}_i|^2\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H\mathbf{R}+\sigma_{\epsilon_i}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\Big] \\
\overset{(a)}{=}& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\left(|\hat{h}_i|^2+\sigma_{\epsilon_i}^2\right)\mathbf{P}\mathbf{P}^H\mathbf{R}\Big] \\
&-\frac{1}{N+\tau}\log\det\Big[\mathbf{I}+|\hat{h}_i|^2\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H\mathbf{R}+\sigma_{\epsilon_i}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\Big], \qquad (5.11)
\end{aligned}$$

where (a) is derived by applying $\mathbf{G}_1\mathbf{G}_1^H+\mathbf{G}_2\mathbf{G}_2^H=\mathbf{I}$.

Under the assumption of perfect SIC, the throughput of User 1 to decode its own message is computed as

$$\begin{aligned}
R_1^{\mathrm{A}} =& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\left(\mathbf{R}_{\mathbf{N}}+\sigma_{\epsilon_1}^2\mathbf{R}\mathbf{P}\mathbf{P}^H\mathbf{R}\right)^{-1}|\hat{h}_1|^2\mathbf{R}\mathbf{P}\mathbf{G}_{\mathbf{1}}\mathbf{G}_{\mathbf{1}}{}^H\mathbf{P}^H\mathbf{R}\Big] \\
=& \frac{1}{N+\tau}\log\det\Big[\mathbf{I}+\sigma_{\epsilon_1}^2\mathbf{P}\mathbf{P}^H\mathbf{R}+|\hat{h}_1|^2\mathbf{P}\mathbf{G}_{\mathbf{1}}\mathbf{G}_{\mathbf{1}}{}^H\mathbf{P}^H\mathbf{R}\Big] \\
&-\frac{1}{N+\tau}\log\det\left[\mathbf{I}+\sigma_{\epsilon_1}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\right] \qquad (5.12)
\end{aligned}$$

According to [65], for $\omega_i\neq 0$ $(i=1,2)$, we have

$$\frac{1}{N+\tau}\log\det\left[\mathbf{I}+\mathbf{\Omega}\mathbf{R}\right]$$

$$=\frac{N}{N+\tau}\log\left(\omega_1\omega_2\right)+\underbrace{\frac{1}{N+\tau}\log\frac{\left[r_1^{N+1}-r_2^{N+1}\right]+\tau^2\left[r_1^N-r_2^N\right]}{r_1-r_2}}_{\triangleq\phi(\omega_1,\omega_2,\tau)}, \tag{5.13}$$

where

$$\mathbf{\Omega}=\text{diag}\left(\omega_1,\omega_2,\cdots,\omega_1,\omega_2\right), Q=2\tau(1-\tau), \tag{5.14}$$

$$r_1=\frac{\omega_1^{-1}+\omega_2^{-1}+\omega_1^{-1}\omega_2^{-1}+Q}{2}+\frac{\sqrt{\left[\omega_1^{-1}+\omega_2^{-1}+\omega_1^{-1}\omega_2^{-1}+Q\right]^2-Q^2}}{2}, \tag{5.15}$$

$$r_2=\frac{\omega_1^{-1}+\omega_2^{-1}+\omega_1^{-1}\omega_2^{-1}+Q}{2}-\frac{\sqrt{\left[\omega_1^{-1}+\omega_2^{-1}+\omega_1^{-1}\omega_2^{-1}+Q\right]^2-Q^2}}{2}, \tag{5.16}$$

and

$$\lim_{N\to\infty}\phi(\omega_1,\omega_2,\tau)=\log r_1. \tag{5.17}$$

According to (5.13), we obtain

$$\frac{1}{N+\tau}\log\det\left[\mathbf{I}+\left(|\hat{h}_i|^2+\sigma_{\epsilon_i}^2\right)\mathbf{P}\mathbf{P}^H\mathbf{R}\right]=\frac{N}{N+\tau}\log\left(\mu_{1i}\mu_{2i}\right)+\phi(\mu_{1i},\mu_{2i},\tau), \tag{5.18}$$

$$\begin{aligned}&\frac{1}{N+\tau}\log\det\left[\mathbf{I}+|\hat{h}_i|^2\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H\mathbf{R}+\sigma_{\epsilon_i}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\right]\\&\overset{(a)}{=}\frac{1}{N+\tau}\log\det\left[\mathbf{I}+|\hat{h}_i|^2P_1\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}+\sigma_{\epsilon_i}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\right]\\&=\frac{N}{N+\tau}\log\left(\nu_{1i}\nu_{2i}\right)+\phi(\nu_{1i},\nu_{2i},\tau),\end{aligned} \tag{5.19}$$

$$\frac{1}{N+\tau}\log\det\left[\mathbf{I}+\sigma_{\epsilon_1}^2\mathbf{P}\mathbf{P}^H\mathbf{R}\right]=\frac{N}{N+\tau}\log(\rho_1\rho_2)+\phi(\rho_1,\rho_2,\tau), \tag{5.20}$$

where (a) is derived by applying $\mathbf{P}\mathbf{G}_1\mathbf{G}_1^H\mathbf{P}^H=P_1\mathbf{G}_1\mathbf{G}_1^H$,

$$\mu_{1i}=P_1\left(|\hat{h}_i|^2+\sigma_{\epsilon_i}^2\right),\mu_{2i}=P_2\left(|\hat{h}_i|^2+\sigma_{\epsilon_i}^2\right), \tag{5.21}$$

$$\nu_{1i}=P_1\left(|\hat{h}_i|^2+\sigma_{\epsilon_i}^2\right),\nu_{2i}=P_2\sigma_{\epsilon_i}^2. \tag{5.22}$$

$$\rho_1 = P_1\sigma_{\epsilon_1}^2, \rho_2 = P_2\sigma_{\epsilon_1}^2. \tag{5.23}$$

Note that $\mu_{1i} = \nu_{1i}$ and $\nu_{21} = \rho_2$. Applying (5.18), (5.19), and (5.20), we can easily simplify the matrix expressions for $R_{2\to i}^{\mathrm{A}}$ and R_1^{A} in (5.11) and (5.12), i.e.,

$$R_{2\to i}^{\mathrm{A}} = \frac{N}{N+\tau}\log\left(\frac{\mu_{1i}\mu_{2i}}{\nu_{1i}\nu_{2i}}\right) + \phi(\mu_{1i},\mu_{2i},\tau) - \phi(\nu_{1i},\nu_{2i},\tau), \tag{5.24}$$

$$R_1^{\mathrm{A}} = \frac{N}{N+\tau}\log\left(\frac{\nu_{11}\nu_{21}}{\rho_1\rho_2}\right) + \phi(\nu_{11},\nu_{21},\tau) - \phi(\rho_1,\rho_2,\tau). \tag{5.25}$$

Note that the throughput expressions for the ANOMA systems with imperfect CSI, i.e., (5.24) and (5.25), are consistent with those for the NOMA systems with imperfect CSI. For example, by setting $\tau = 0$, we obtain the throughput expressions for the NOMA systems with imperfect CSI, i.e., [62, 66]

$$R_{2\to i}^{\mathrm{N}} = \log\frac{1+\mu_{1i}+\mu_{2i}}{1+\nu_{1i}+\nu_{2i}} = \log\left(1+\frac{P_2|\hat{h}_i|^2}{1+(P_1+P_2)\sigma_{\epsilon_i}^2+P_1|\hat{h}_i|^2}\right), \tag{5.26}$$

$$R_1^{\mathrm{N}} = \log\frac{1+\nu_{11}+\nu_{21}}{1+\rho_1+\rho_2} = \log\left(1+\frac{P_1|\hat{h}_1|^2}{1+(P_1+P_2)\sigma_{\epsilon_1}^2}\right). \tag{5.27}$$

Note that the general expressions for R_1^{A} and $R_{2\to i}^{\mathrm{A}}$ for a finite N, (5.24) and (5.25), are too complicated to analyze. In addition, the frame length in practical systems are relatively large, e.g., $N \approx 156$ in global system for mobile communications (GSM). As a result, we analyze throughputs in ANOMA systems for the asymptotic case of $N \to \infty$. Applying (5.17), we have

$$R_{2\to i,\infty}^{\mathrm{A}} \triangleq \lim_{N\to\infty} R_{2\to i}^{\mathrm{A}} = \log\frac{1+Q\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1}+\sqrt{1+2Q\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1}}}{1+Q\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1}+\sqrt{1+2Q\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1}}} + \log\frac{1+\mu_{1i}+\mu_{2i}}{1+\nu_{1i}+\nu_{2i}}, \tag{5.28}$$

$$R_{1,\infty}^{\mathrm{A}} \triangleq \lim_{N\to\infty} R_1^{\mathrm{A}} = \log \frac{1+Q\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1}+\sqrt{1+2Q\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1}}}{1+Q\frac{\rho_1\rho_2}{\rho_1+\rho_2+1}+\sqrt{1+2Q\frac{\rho_1\rho_2}{\rho_1+\rho_2+1}}}+\log\frac{1+\nu_{11}+\nu_{21}}{1+\rho_1+\rho_2}. \tag{5.29}$$

To compare the throughputs of ANOMA and NOMA systems, we derive the following theorem.

Theorem 5.1. *For the asymptotic case of infinite frame length, the users in ANOMA systems can achieve higher throughputs compared with those in NOMA systems, i.e.,*

$$R_{2\to i,\infty}^{\mathrm{A}} > R_{2\to i}^{\mathrm{N}}, \tag{5.30}$$

$$R_{1,\infty}^{\mathrm{A}} > R_1^{\mathrm{N}}. \tag{5.31}$$

Proof. See Appendix C.1. □

Based on (5.28) and (5.29), the optimal timing mismatch can be obtained to maximize the individual throughput, which is shown in the following theorem.

Theorem 5.2. *As the frame length goes to infinity, the optimal timing mismatch to maximize the individual throughputs in ANOMA systems converges to half of the frame length, i.e.,*

$$\arg_\tau \max R_{2\to i,\infty}^{\mathrm{A}} = 0.5, \tag{5.32}$$

$$\arg_\tau \max R_{1,\infty}^{\mathrm{A}} = 0.5. \tag{5.33}$$

Proof. See Appendix C.2. □

Note that Theorem 5.2 is consistent with the results provided in [48] (the uplink ANOMA system with perfect CSI) and [65] (the cooperative ANOMA system with perfect CSI). In other words, the availability of perfect CSI does not affect the optimal timing mismatch.

5.4 Outage Performance of ANOMA Systems

Due to the fact that only the imperfect CSI is available at the user side, the outage event may happen. Since SIC is employed at User 1, the outage event happens at User 1 if it cannot detect either User 2's message or its own message. Thus, the outage probability of User 1 is computed as

$$P_{\text{out},1}^{\text{N}} \triangleq 1 - P\left\{R_{2\to1}^{\text{N}} > C_2, R_1^{\text{N}} > C_1\right\} \overset{(a)}{\geq} 1 - P\left\{R_{2\to1,\infty}^{\text{A}} > C_2, R_{1,\infty}^{\text{A}} > C_1\right\} \triangleq P_{\text{out},1,\infty}^{\text{A}}. \tag{5.34}$$

where (a) is derived by applying Theorem 5.1 and C_1 and C_2 are the target throughput for User 1 and 2, respectively. Similarly, the outage event happens if User 2 cannot detect its own message, i.e.,

$$P_{\text{out},2}^{\text{N}} \triangleq P\left\{R_{2\to2}^{\text{N}} < C_2\right\} \geq P\left\{R_{2\to2,\infty}^{\text{A}} < C_2\right\} \triangleq P_{\text{out},2,\infty}^{\text{A}}. \tag{5.35}$$

We note from (5.34) and (5.35) that with a sufficiently large frame length, the ANOMA systems can achieve lower outage probability compared with the NOMA systems. We will show in the simulation results section that the condition of "sufficiently large frame length" can be satisfied with a reasonable finite N. Besides, maximizing $R_{2\to2,\infty}^{\text{A}}$ and $R_{1,\infty}^{\text{A}}$ are equivalent to minimizing the outage probabilities. Therefore, Theorem 5.2 still applies when minimizing $P_{\text{out},1,\infty}^{\text{A}}$ and $P_{\text{out},2,\infty}^{\text{A}}$, i.e., the optimal timing mismatch to minimize the outage probability is half of the frame length for the infinite frame length.

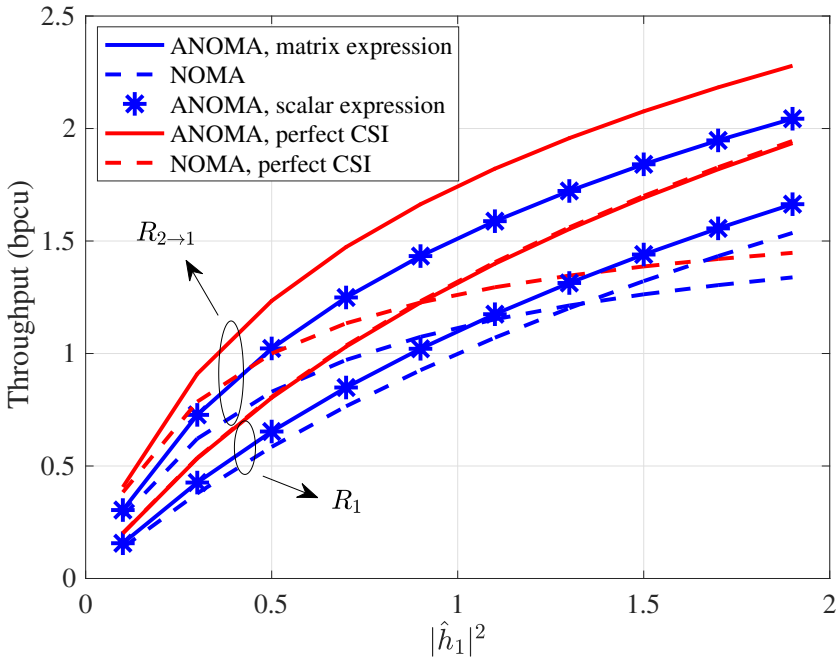

Figure 5.3: The throughput as a function of the estimated channel gain $|\hat{h}_1|^2$ for ANOMA and NOMA systems when $N = 100$, $\tau = 0.5$, $P_1 = 1.5$, $P_2 = 3.5$, $\sigma^2_{\epsilon_1} = 0.1$.

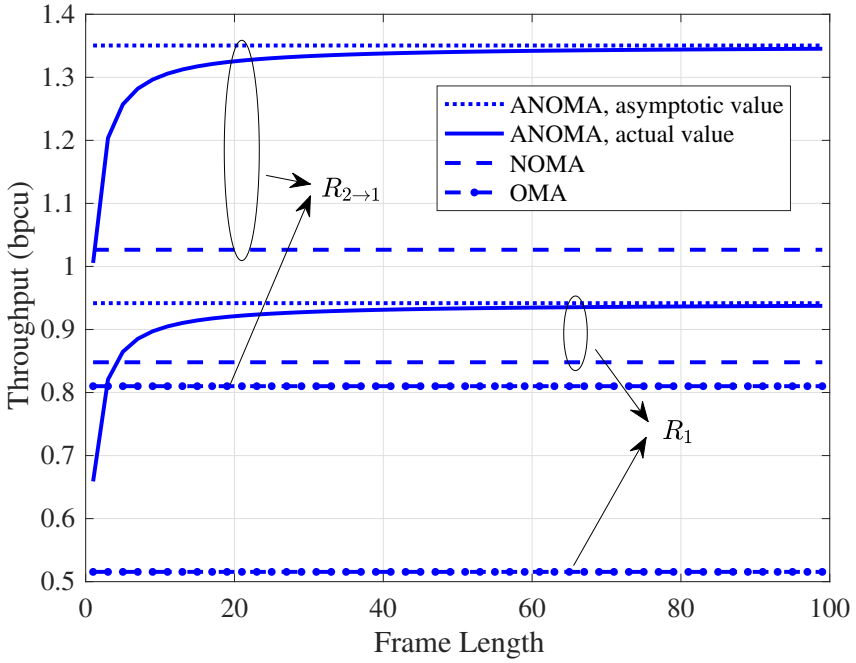

Figure 5.4: The throughput as a function of the frame length N for ANOMA and NOMA systems when $|\hat{h}_1|^2 = 0.8$, $\tau = 0.5$, $P_1 = 1.5$, $P_2 = 3.5$, $\sigma^2_{\epsilon_1} = 0.1$.

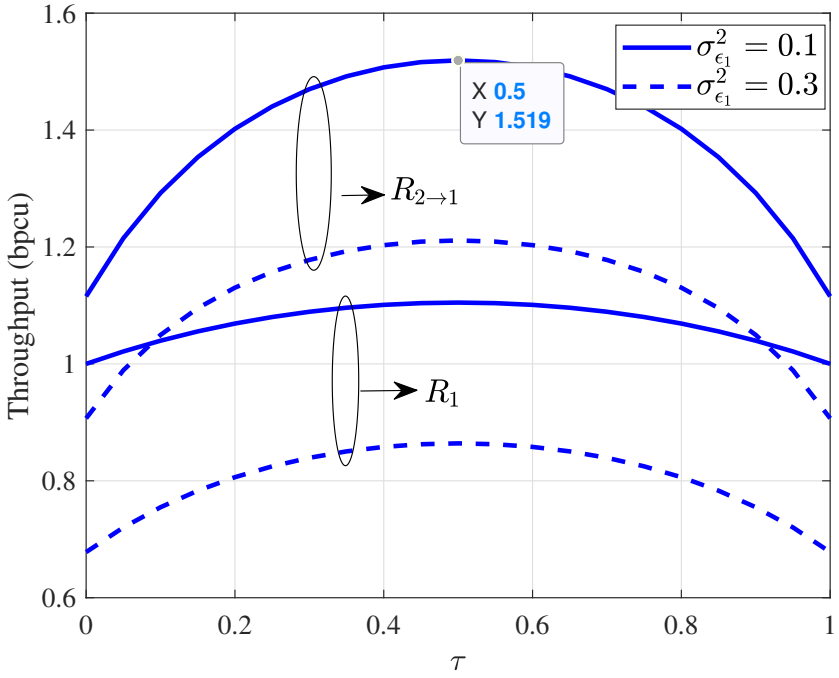

Figure 5.5: The throughput as a function of the normalized timing mismatch τ for ANOMA systems when $|\hat{h}_1|^2 = 1$, $P_1 = 1.5$, $P_2 = 3.5$ for $\sigma^2_{\epsilon_1} = 0.1$ or 0.3.

5.5 Numerical and Simulation Results

In this section, we provide numerical and simulation results to compare the performances of ANOMA and NOMA systems. Note that $R^{\mathrm{A}}_{2\to1}$ and $R^{\mathrm{A}}_{2\to2}$ given by (5.11) only differs in the channel coefficients. Hence, for the sake of brevity, we only present the results of $R^{\mathrm{A}}_{2\to1}$ and compare it with $R^{\mathrm{N}}_{2\to1}$ in Figs. 5.3, 5.4, and 5.5. The conclusions drawn from $R^{\mathrm{A}}_{2\to1}$ and $R^{\mathrm{N}}_{2\to1}$ also apply to $R^{\mathrm{A}}_{2\to2}$ and $R^{\mathrm{N}}_{2\to2}$.

At first, we compare the throughput performances of ANOMA and NOMA systems as the function of the estimated channel gain in Fig. 5.3. The curves "ANOMA, matrix expression" are calculated according to (5.11) and (5.12) while the curves "ANOMA, scalar expression" are computed by applying (5.24) and (5.25). In Fig. 5.3, it is shown that the throughputs calculated by the matrix expression completely align with those by the scalar expression, which verifies the correctness of (5.24) and (5.25). Besides, for both $R_{2\to1}$ and R_1, the

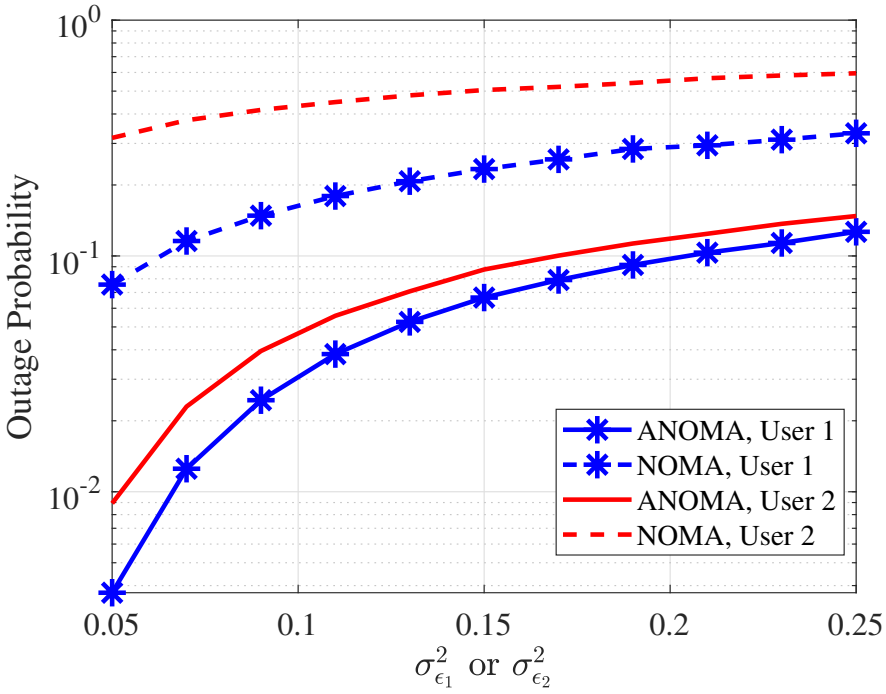

Figure 5.6: The outage probability as a function of the variance of channel estimation error $\sigma^2_{\epsilon_1}$ for ANOMA and NOMA systems when $N = 100$, $\tau = 0.5$, $P_1 = 1.5$, $P_2 = 3.5$, $h_1 = \sqrt{5}$, $h_2 = \sqrt{3}$, $C_1 = C_2 = 1.5$.

throughputs in the ANOMA systems are larger than those in the NOMA systems.

Second, we show how the throughouts in the ANOMA systems change with the frame length in Fig. 5.4. The curves "ANOMA, asympototic value" are calculated by (5.28) and (5.29). In Fig. 5.4, it is demonstrated that the actual throughputs steadily converge to the asymptotic ones in ANOMA systems. In addition, for a relatively large frame length, e.g., $N > 20$, the ANOMA systems outperform the NOMA systems in terms of the throughput, which verifies Theorem 5.1.

Furthermore, we study the throughputs in the ANOMA systems as a function of the normalized timing mismatch τ for the asymptotic case of $N \to \infty$ in Fig. 5.5. It is shown that for both $R_{2\to1}$ and R_1, the optimal normalized timing mismatch to maximize the throughput is 0.5, which verifies Theorem 5.2.

At last, we simulate and compare the outage probabilities of ANOMA and NOMA systems with imperfect CSI in Fig. 5.6. It is shown that the outage probability increases with the variance of channel estimation error. Besides, for both Users 1 and 2, the ANOMA systems can achieve a lower outage probability compared with the NOMA systems.

5.6 Conclusion

In this chapter, the downlink ANOMA systems with imperfect CSI are studied. We conclude that with the channel estimation error, the ANOMA systems outperform the NOMA systems in terms of outage probability. In addition, we also discover that with imperfect CSI, the optimal timing mismatch converges to half of a single symbol length, which is consistent with the ANOMA systems with perfect CSI.

Chapter 6

Downlink Asynchronous Non-Orthogonal Multiple Access with Limited Feedback

6.1 Introduction

The channel state information (CSI) plays a critical role in optimizing the system performance. At the receiver side, CSI is utilized to perform the coherent detection which improves the signal estimation performance. More importantly, the CSI at the transmitter side is employed to conduct the adaptive power/rate allocation and generate the beamforming vectors in multiple-input multiple-output (MIMO) systems. In time-division duplexing (TDD) systems, the channel reciprocity is exploited by the base station (BS) to utilize the CSI estimated via the uplink training. In frequency-division duplexing (FDD) systems, a prevailing technique is using the limited feedback from users. The NOMA with limited feedback has been studied in the existing literature, for example, the one-bit feedback scheme

in the massive MIMO NOMA systems [67] and the multi-user single antenna systems [68], and the scalar quantizer design in downlink power-domain NOMA [2]. To the best of our knowledge, the analysis of limited feedback schemes in ANOMA systems and the optimal quantizer design for NOMA/ANOMA are still absent. In fact, the limited feedback design in NOMA/ANOMA systems is more challenging compared with that in the orthogonal multiple access (OMA) scenario, e.g., in [69]. It is because the CSI of each user not only affects its own but also other users' performance due to the inter-user interference (IUI) ingrained in the non-orthogonal transmission. In more details, the CSI is used for both allocating powers and determining the SIC order, which further complicates the rate expressions and the system design.

In this chapter, we consider a downlink ANOMA system with limited feedback. We employ the max-min criterion for the power allocation and the scalar quantizer for channel quantization, respectively. We derive the closed-form expressions for the upper and lower bounds of the max-min rate. It is manifested that ANOMA can achieve the same or even higher average max-min rate with a lower feedback rate compared with NOMA. Moreover, we propose a quantizer optimization method which applies to both NOMA and ANOMA systems. A gradient descent algorithm is designed to optimize the quantization levels. Simulation results show that a higher average max-min rate is achieved by using the optimized quantizer compared with the conventional uniform quantizer in [2], especially for the low-rate feedback scenario.

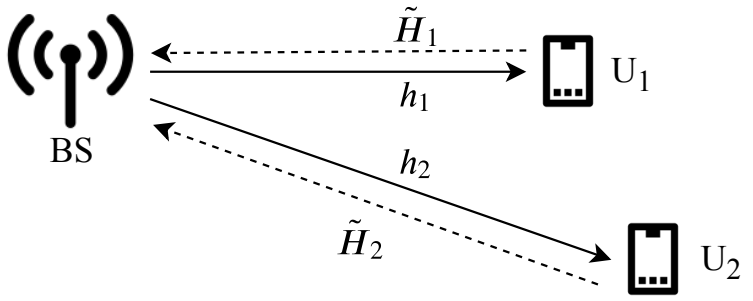

Figure 6.1: Downlink scenario with limited feedback.

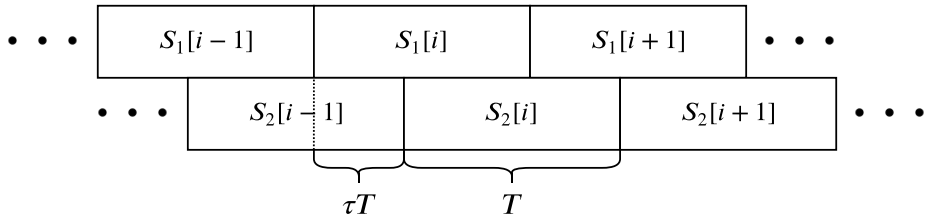

Figure 6.2: Illustration of the superimposed signal in ANOMA.

6.2 Preliminaries

6.2.1 NOMA and ANOMA

In this chapter, we consider a two-user downlink system shown in Fig. 6.1 where the signals for Users 1 and 2 are superimposed and then transmitted by the BS. As shown in Fig. 6.2, a timing mismatch of τT is intentionally added between the superimposed signals, where $0 < \tau < 1$ and T is the symbol interval. Let $p(t)$ represent the pulse shape. The transmitted signal is given by $S_1[i]p(t - iT) + S_2[i]p(t - iT - \tau T)$ where $S_u[i]$, $u = 1, 2$, is the ith symbol transmitted to User u. At the receiver, the composite signal is sampled at iT and $(i + \tau)T$ after matched filtering. This sampling method is called "oversampling" and the details have already been presented in [48, 65]. For the sake of brevity, we omit it in this work. If $\tau = 0$, ANOMA degrades to NOMA and the received samples at iT and $(i+\tau)T$ are identical. The number of received samples is doubled in ANOMA ($\tau \neq 0$) compared with NOMA, which then results in sampling diversity [48, 65, 50].

We assume that Users 1 and 2 denote the strong and weak users, respectively, i.e., $|h_1|^2 > |h_2|^2$. User 1 employs SIC, i.e., first detects the messages for User 2, removes it, and then detects its own message. As shown in [65], for a relatively large block length and the rectangular pulse shape, the rates of Users 1 and 2 with perfect CSI are given by

$$R_{\text{strong}}(H_1) = \log_2\left(1+\alpha P H_1\right), \tag{6.1}$$

$$R_{\text{weak}}(H_2) = \log_2\left(\frac{1+PH_2+\alpha(1-\alpha)P^2H_2^2Q}{2(1+\alpha PH_2)} + \frac{\sqrt{[1+PH_2+\alpha(1-\alpha)P^2H_2^2Q]^2-\alpha^2(1-\alpha)^2P^4H_2^4Q^2}}{2(1+\alpha PH_2)}\right). \tag{6.2}$$

where $H_i = |h_i|^2$ is the channel gain of User i, $Q = 2\tau(1-\tau)$, P is the total transmit power of BS, $\alpha \in (0,1)$ is the power coefficient for the strong user (User 1 in this case), i.e., the powers allocated to Users 1 and 2 are αP and $(1-\alpha)P$, respectively. Note that NOMA can be considered as a special case of ANOMA, simply by setting $\tau = 0$ in (6.1) and (6.2). It has been shown that R_{weak} in ANOMA is higher than that in NOMA while R_{strong} is the same for ANOMA and NOMA [65].

6.2.2 Limited Feedback

Figure 6.3: Illustration of a scalar quantizer.

In this chapter, we assume that CSI is perfectly estimated by users and fed back to the BS via an error and delay-free link. For example, in Fig. 6.1, User 1 knows h_1 and quantizes the channel gain H_1 as $\tilde{H}_1 = q(|h_1|^2)$ using a scalar quantizer q, and then sends it back to the BS. The BS determines the power coefficient α and the order of SIC according to the

feedback. If $\tilde{H}_1 > \tilde{H}_2$ ($\tilde{H}_1 < \tilde{H}_2$), User 1 (User 2) is notified to conduct SIC. If $\tilde{H}_1 = \tilde{H}_2$, BS can randomly assign one user to utilize SIC. Without loss of generality, we assume that User 1 will be notified to conduct SIC if $\tilde{H}_1 = \tilde{H}_2$.

As shown in Fig. 6.3, we employ a scalar quantizer with $N = 2^b$ quantization levels, $q_0, \cdots, q_{N-1}$, where b is the number of bits used in a quantization codeword. b is also defined as the feedback rate per user. The quantized value for a given x is calculated by

$$q(x) = \begin{cases} q_i, & q_i \leq x < q_{i+1},\ i = 0, \cdots, N-2, \\ q_{N-1}, & x \geq q_{N-1}. \end{cases} \tag{6.3}$$

For example, in Fig. 6.3, x_0 and x_1 are quantized as q_0 and q_{N-1}, respectively. In this work, the quantizer is used to quantize the positive channel gain. Thus, we set $q_0 = 0$.

The BS transmits the signals to users based on the rates which are calculated according to the quantized channel gains. For example, if $\tilde{H}_1 > \tilde{H}_2$, the BS transmits to User 1 with the rate of $R_{\text{strong}}(\tilde{H}_1)$ while the actual channel capacity is $R_{\text{strong}}(H_1)$. Note that the proposed quantizer is designed to satisfy $q(x) \leq q$ in order to avoid outage, i.e, to guarantee that the transmission rates do not exceed the channel capacities. It is trivial to derive that $R_{\text{strong}}(\tilde{H}_1) \leq R_{\text{strong}}(H_1)$ if $\tilde{H}_1 \leq H_1$ and $R_{\text{weak}}(\tilde{H}_2) \leq R_{\text{weak}}(H_2)$ if $\tilde{H}_2 \leq H_2$, which indicates that the quantizer in (6.3) avoids outage.

6.3 Power Allocation

To attain fairness among users, we employ the max-min criterion for the power allocation, i.e., the power is allocated to users such that the minimum rate is maximized. The optimal

power coefficient α^* is given by

$$\alpha^* = \arg\max_{\alpha} \min \left\{ R_{\text{strong}}, R_{\text{weak}} \right\}. \tag{6.4}$$

According to (6.1) and (6.2), it is trivial to show that R_{strong} is an increasing function of α and R_{weak} is a decreasing function of α. Intuitively, if User 1 is the strong user, by increasing α (i.e., allocating more power to User 1 and less power to User 2), R_{strong} increases and R_{weak} decreases, and vice versa. As a result, the optimal power coefficient α^* can be obtained by solving $R_{\text{strong}} = R_{\text{weak}}$. According to [2], the optimal α in NOMA is given by

$$\alpha_{\text{N}}^* = \frac{2\tilde{H}_{\min}}{\sqrt{(\tilde{H}_1+\tilde{H}_2)^2+4P\tilde{H}_1\tilde{H}_2\tilde{H}_{\min}}+\tilde{H}_1+\tilde{H}_2}, \tag{6.5}$$

where $\tilde{H}_{\min} = \min\left\{\tilde{H}_1, \tilde{H}_2\right\}$ is defined to incorporate the rate expressions for both $\tilde{H}_1 \geq \tilde{H}_2$ and $\tilde{H}_1 < \tilde{H}_2$. If User 1 employs SIC, the max-min rate of NOMA is given by $\log\left(1+\alpha_{\text{N}}^* P\tilde{H}_1\right)$ according to (6.1). If User 2 employs SIC, the max-min rate is given by $\log\left(1+\alpha_{\text{N}}^* P\tilde{H}_2\right)$. To summarize, the max-min rate of NOMA is expressed by $R_{\text{N}}^*(\tilde{H}_1, \tilde{H}_2) = \log\left(1+\alpha_{\text{N}}^* P\tilde{H}_{\max}\right)$ where $\tilde{H}_{\max} = \max\left\{\tilde{H}_1, \tilde{H}_2\right\}$.

The optimal power coefficient for ANOMA, α_{A}^*, is presented in the following theorem.

Theorem 6.1. *The optimal power coefficients of NOMA and ANOMA, α_{N}^* and α_{A}^*, respectively, satisfy the following inequality*

$$\alpha_{\text{N}}^* \leq \alpha_{\text{A}}^*(0.5) \leq \alpha_{\text{A}}^* \leq \alpha_{\text{A}}^*(1), \tag{6.6}$$

where the equal sign is achieved when $\tau = 0$. The expression for $\alpha_{\mathrm{A}}^*(x)$ is given by.

$$\alpha_{\mathrm{A}}^*(x) = \frac{2\tilde{H}_{\min}}{\sqrt{(\tilde{H}_1+\tilde{H}_2 - xP\tilde{H}_{\min}^2Q)^2+4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2+xP\tilde{H}_{\min}^2Q)}+\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q}. \tag{6.7}$$

Proof. See Appendix D.1. □

Compared with (6.5), (6.7) introduces an extra term $xP\tilde{H}_{\min}^2Q$. By setting $\tau = 0$, $Q = 0$ and then (6.7) coincides with (6.5). The max-min rate of ANOMA is given by $R_{\mathrm{A}}^*(\tilde{H}_1, \tilde{H}_2) = \log\left(1 + \alpha_{\mathrm{A}}^* P\tilde{H}_{\max}\right) \in \left[\log\left(1 + \alpha_{\mathrm{A}}^*(0.5)P\tilde{H}_{\max}\right), \log\left(1 + \alpha_{\mathrm{A}}^*(1)P\tilde{H}_{\max}\right)\right]$.

The average max-min rate is expressed by

$$\mathbb{E}\left[R^*\right] = \sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1}\int_{q_{i,1}}^{q_{i+1,1}}\int_{q_{j,2}}^{q_{j+1,2}} R^*(q_{i,1}, q_{j,2}) f_1(H_1) f_2(H_2)\mathrm{d}H_1\mathrm{d}H_2, \tag{6.8}$$

where $f_i(H_i)$ is the distribution function of User i's channel gain, $q_{j,i}$ represents the jth quantization level of the quantizer used by User i, and R^* can be R_{N}^* or R_{A}^*. Note that the average max-min rate for the limited feedback is upper bounded by that for the full-CSI case, i.e.,

$$\mathbb{E}\left[R^*\right] < \overline{R}^* \triangleq \int_0^\infty\int_0^\infty R^*(H_1,H_2) f_1(H_1) f_2(H_2)\mathrm{d}H_1\mathrm{d}H_2. \tag{6.9}$$

Let us define the quantization distortion as $D\left[R^*\right] = \overline{R}^* - \mathbb{E}\left[R^*\right]$.

Corollary 6.1. *ANOMA can achieve the average max-min rate of NOMA with a lower feedback rate.*

Proof. Let us define q and q' as two quantizers which can be given by (6.3) but with different quantization levels. According to Theorem 6.1, ANOMA achieves a higher max-min rate

compared with NOMA. Thus, $\overline{R}^*_{\mathrm{N}} < \overline{R}^*_{\mathrm{A}}$ and $\mathbb{E}\left[R^*_{\mathrm{N}}\right] < \mathbb{E}\left[R^*_{\mathrm{A}}\right]$ by using the quantizer q. The quantizer q' is designed such that the ANOMA using q' achieves the same average max-min rate as the NOMA using q, i.e., $\mathbb{E}\left[R^*_{\mathrm{A}}\right]' = \mathbb{E}\left[R^*_{\mathrm{N}}\right] < \mathbb{E}\left[R^*_{\mathrm{A}}\right]$. For ANOMA, using the quantizer q' results in a higher distortion compared with using q, i.e., $D\left[R^*_{\mathrm{A}}\right]' = \overline{R}^*_{\mathrm{A}} - \mathbb{E}\left[R^*_{\mathrm{A}}\right]' > D\left[R^*_{\mathrm{A}}\right] = \overline{R}^*_{\mathrm{A}} - \mathbb{E}\left[R^*_{\mathrm{A}}\right]$. According to the rate-distortion theory, there is a trade-off between the distortion and the quantization rate (equivalent to the feedback rate in this work). A lower distortion can be achieved by using a quantizer with a higher feedback rate and vice versa. As a result, the quantizer q' can have a lower feedback rate compared with q. ANOMA using q' can achieve a lower feedback rate while keeping the same average max-min rate as NOMA using q. The proof is complete. □

6.4 Scalar Quantizer Optimization

The scalar quantizer shown in (6.3) is completely characterized by the quantization levels. Our goal is to optimize the quantization levels to maximize the average max-min rate $\mathbb{E}\left[R^*\right]$, i.e.,

$$\begin{aligned}\left[\mathbf{q}_1^*, \mathbf{q}_2^*\right] = & \underset{[\mathbf{q_1}, \mathbf{q_2}]}{\arg\max}\, \mathbb{E}\left[R^*\right], \\ & s.t.\ q_{0,i} < q_{2,i} < \cdots < q_{N_i-1,i},\ i = 1 \text{ or } 2. \end{aligned} \tag{6.10}$$

Since $\overline{R}^*$ is not a function of quantization levels, (6.10) is equivalent to minimizing the distortion $D[R^*]$. In what follows, we propose a gradient descent algorithm to optimize the quantization levels. Let $\mathbb{E}\left[R^*\right]_{i,j}$ denote the (i,j)th term of $\mathbb{E}\left[R^*\right]$ in (6.8), i.e., $\mathbb{E}\left[R^*\right]_{i,j} = R^*(q_{i,1}, q_{j,2}) \int_{q_{i,1}}^{q_{i+1,1}} \int_{q_{j,2}}^{q_{j+1,2}} f_1(H_1) f_2(H_2) \mathrm{d}H_1 \mathrm{d}H_2$. $\mathbb{E}\left[R^*\right]_{i,j}$ is a function of $q_{i,1}$, $q_{i+1,1}$, $q_{j,2}$, and

$q_{j+1,2}$. The gradients of $\mathbb{E}\left[R^*\right]_{i,j}$ in terms of $q_{i,1}$ and $q_{i+1,1}$ are calculated by

$$\frac{\partial \mathbb{E}\left[R^*\right]_{i,j}}{\partial q_{i,1}} = \left[\frac{\partial R^*(q_{i,1}, q_{j,2})}{\partial q_{i,1}} \int_{q_{i,1}}^{q_{i+1,1}} f_1(H_1)\mathrm{d}H_1 - R^*(q_{i,1}, q_{j,2}) f_1(q_{i,1})\right] \int_{q_{j,2}}^{q_{j+1,2}} f_2(H_2)\mathrm{d}H_2, \tag{6.11}$$

$$\frac{\partial \mathbb{E}\left[R^*\right]_{i,j}}{\partial q_{i+1,1}} = R^*(q_{i,1}, q_{j,2}) f_1(q_{i+1,1}) \int_{q_{j,2}}^{q_{j+1,2}} f_2(H_2)\mathrm{d}H_2, \tag{6.12}$$

respectively. Similarly, we can derive $\frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j,2}}$ and $\frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j+1,2}}$. Based on the gradients, we propose the quantizer optimization algorithm in Algorithm 2. At each iteration, $\mathbb{E}\left[R^*\right]$ does not decrease which is guaranteed by the gradient descent. Besides, $\mathbb{E}\left[R^*\right]$ is upper bounded by a constant as shown in (6.9). Hence, Algorithm 2 converges as the number of iterations increases.

Algorithm 2 Algorithm to optimize the scalar quantizer.

Initialize the step size Δ, the maximum number of iterations $I_{\max}$, the number of quantization levels for User 1 N_1 and that for User 2 N_2, the iteration count $I = 0$.
Initialize the quantization levels $\mathbf{q}_1 = [0, q_{1,1}, \cdots, q_{N_1-1,1}, \infty]$, $\mathbf{q}_2 = [0, q_{1,2}, \cdots, q_{N_2-1,2}, \infty]$
while $I < I_{\max}$ **do**
    Initialize $E_R = 0$, $\mathbf{dq}_1 = [0, \cdots, 0]_{1\times(N_1+1)}$, $\mathbf{dq}_2 = [0, \cdots, 0]_{1\times(N_2+1)}$.
    for $i = 2, \cdots, N_1$ **do**
        for $j = 2, \cdots, N_2$ **do**
            $E_R = E_R + R^*(q_1[i], q_2[j]) \int_{q_1[i]}^{q_1[i+1]} \int_{q_2[j]}^{q_2[j+1]} \cdot f_1(H_1) f_2(H_2)\mathrm{d}H_1\mathrm{d}H_2$.
            $\mathbf{dq}_1[i] = \mathbf{dq}_1[i] + \frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{i,1}}$.
            $\mathbf{dq}_2[j] = \mathbf{dq}_2[j] + \frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j,2}}$.
            if $i \neq N_1$ **then**
                $\mathbf{dq}_1[i+1] = \mathbf{dq}_1[i+1] + \frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{i+1,1}}$.
            end if
            if $j \neq N_2$ **then**
                $\mathbf{dq}_2[j+1] = \mathbf{dq}_2[j+1] + \frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j+1,2}}$.
            end if
        end for
    end for
    $\mathbf{q}_1 = \mathbf{q}_1 + \Delta * \mathbf{dq}_1$, $\mathbf{q}_2 = \mathbf{q}_2 + \Delta * \mathbf{dq}_2$, $I = I + 1$.
end while

The computation complexity of Algorithm 2 is $O(N_1 N_2)$ where N_1 and N_2 are the number of quantization levels for Users 1 and 2, respectively. Furthermore, the maximum quantization level of the conventional uniform quantizer is set manually according to certain criterion. For example, in [2], the maximum quantization level is determined by considering the quantization loss. The advantage of optimizing the quantization levels is that the maximum quantization level can also be optimized using Algorithm 2 with no manual intervention, which will be shown in the next section.

6.5 Simulation Results

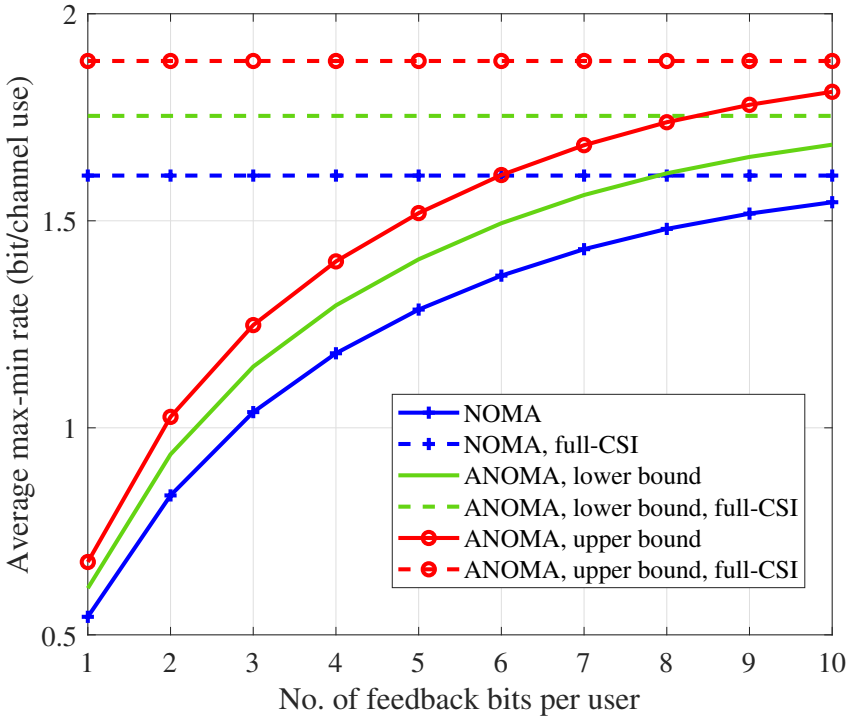

Figure 6.4: The average max-min rate vs. the number of feedback bits per channel for NOMA and ANOMA systems.

In this section, we present simulation results for NOMA/ANOMA systems with limited feedback. In our simulations, we employ the complex Gaussian channel model. Therefore, the channel gain follows the exponential distribution. We assume that $H_1 \sim \text{Exp}(0.5)$ and

$H_2 \sim \text{Exp}(1)$. We set the total transmit power of the BS $P = 10$. For comparison, we employ the uniform quantizer proposed in [2] where the maximum quantization level T is derived by solving $T = \frac{1}{\lambda\Delta}\log\left(\frac{1}{\Delta}\right)$, λ is the parameter of the exponential distribution, and Δ is the quantization bin width.

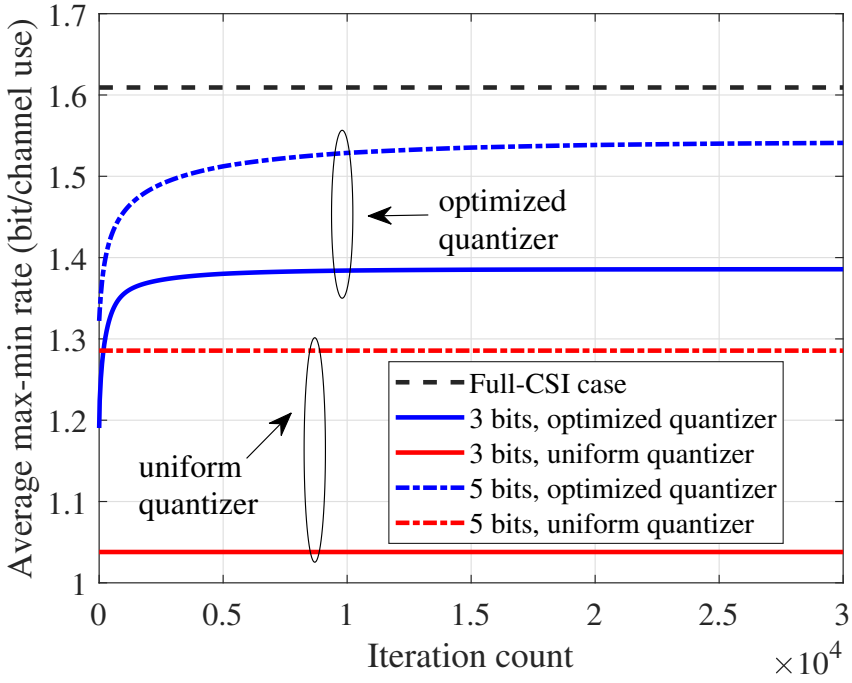

Figure 6.5: The average max-min rate vs. the iteration count for NOMA systems.

Fig. 6.4 shows how the average max-min rate changes as a function of the number of feedback bits per user using the uniform quantizer. As the number of feedback bits increases, the average max-min rate converges to the full-CSI rate calculated by (6.9). Besides, using the same number of feedback bits, the upper bound of the max-min rate for ANOMA is always higher than its lower bound, which is then higher than that for NOMA. Equivalently, to achieve the same or even higher average max-min rate, ANOMA needs less feedback bits compared with NOMA, which verifies Corollary 6.1.

Fig. 6.5 shows how the average max-min rate in NOMA systems changes as the quantizer optimization algorithm runs. As the number of iterations increases, the average max-min

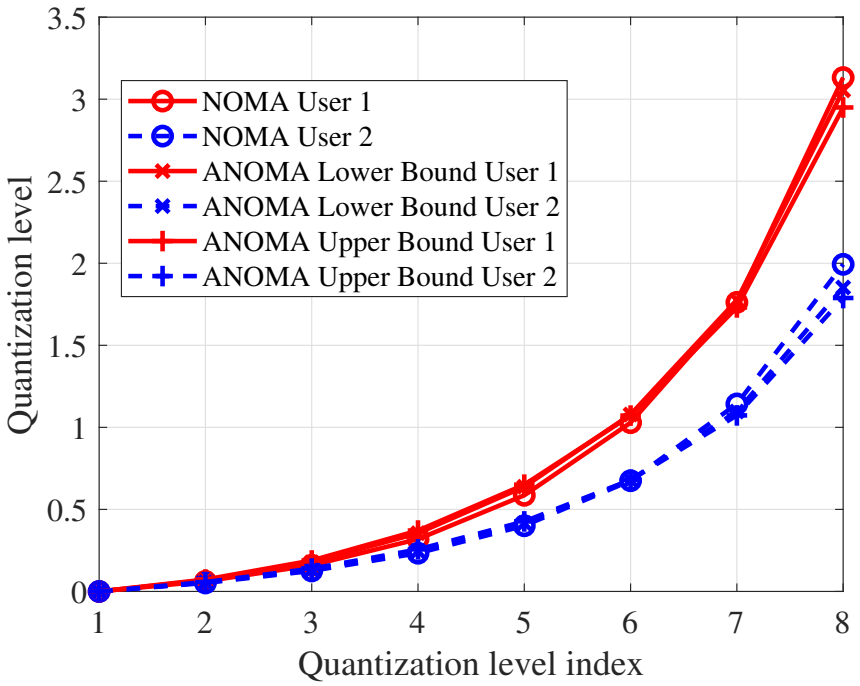

Figure 6.6: The quantization levels for NOMA and ANOMA systems when 3 bits are used to quantize the channel gain.

rate converges. The quantizer with a higher feedback rate achieves a higher average max-min rate. Moreover, the rate gap between the optimized quantizer and the uniform quantizer shrinks if a higher feedback rate is used. Similar results hold for ANOMA although for the sake of brevity, we skip the corresponding figure.

Fig. 6.6 presents the optimal quantization levels when 3 bits (i.e., $2^3 = 8$ quantization levels) are used. First, it is shown that the optimal quantization levels are non-uniform. The range of the quantization levels for User 1 with a larger mean channel gain is wider than that of User 2. Furthermore, the quantization levels optimized using the upper and lower bounds of the max-min rate for ANOMA are very similar to each other and also close to those for NOMA.

6.6 Conclusion

In this chapter, we considered a downlink ANOMA system with limited feedback. We reveal the advantage of ANOMA over NOMA in terms of the feedback rate. Furthermore, the quantizer optimization method is proposed for both ANOMA and NOMA to further improve the average max-min rate.

Chapter 7

Conclusions and Future Work

In this dissertation, we thoroughly studied the ANOMA systems and demonstrated the benefits of applying ANOMA compared with the conventional NOMA systems. It is manifested that ANOMA outperforms NOMA in terms of the throughput performance, the outage probability, the power consumption, and the feedback rate. Besides, the optimal designs of ANOMA systems are provided, including the optimal timing mismatch in the two-user system, the optimal power allocation, and the optimal scalar quantizer in the limited feedback scheme.

While this dissertation mainly focuses on two-user scenarios, the ANOMA scheme can be extended to the multi-user (more than 2 users) scenario. In a K-user scenario, the BS should use K samples per symbol length, each aligned with one of the users to obtain sampling diversity. The study of ANOMA in a more than two-user scenario is an interesting future work. It is also an interesting problem to study the ANOMA in a multicell scenario. Besides, it is worthwhile extending the considered ANOMA into the OFDM systems, since current wireless communication systems are often based on the OFDM technique. For OFDM systems, one can introduce the frequency-domain asynchrony instead of the time-domain

asynchrony and apply similar ideas to design ANOMA OFDM schemes. Moreover, a further analysis on the bit error rate performance of the ANOMA systems can be a good future work.

Bibliography

[1] 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 J. Sel. Areas Commun.*, 35(10):2181–2195, October 2017.

[2] Xiaoyi Liu and Hamid Jafarkhani. Downlink non-orthogonal multiple access with limited feedback. *IEEE Trans. Wireless Commun.*, 16(9):6151–6164, September 2017.

[3] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[4] Hamid Jafarkhani. *Space-time coding: theory and practice*. Cambridge university press, 2005.

[5] Sergio Verdu. The capacity region of the symbol-asynchronous Gaussian multiple-access channel. *IEEE Trans. Inf. Theory*, 35(4):733–751, July 1989.

[6] Xun Zou and Hamid Jafarkhani. Asynchronous channel training in massive MIMO systems. In *Proc. IEEE GLOBECOM*, pages 1–6, Washington, DC, USA, December 2016.

[7] Mehdi Ganji and Hamid Jafarkhani. Interference mitigation using asynchronous transmission and sampling diversity. In *Proc. IEEE GLOBECOM*, pages 1–6, Washington, DC, USA, December 2016.

[8] Huseyin Haci, Huiling Zhu, and Jiangzhou Wang. Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique. *IEEE Trans. Commun.*, 65(3):1319–1335, March 2017.

[9] Aniruddha Das and Bhaskar D Rao. MIMO systems with intentional timing offset. *EURASIP J. Adv. Signal Process.*, 2011(1):1–14, December 2011.

[10] Xuehua Zhang, Mehdi Ganji, and Hamid Jafarkhani. Exploiting asynchronous signaling for multiuser cooperative networks with analog network coding. In *Proc. IEEE WCNC*, pages 1–6, San Francisco, CA, USA, March 2017.

[11] Xuehua Zhang and Hamid Jafarkhani. Asynchronous network coding for multiuser cooperative communications. *IEEE Trans. Commun.*, 16(12):8250–8260, December 2017.

[12] Sina Poorkasmaei and Hamid Jafarkhani. Asynchronous orthogonal differential decoding for multiple access channels. *IEEE Trans. Wireless Commun.*, 14(1):481–493, January 2015.

[13] MR Avendi and Hamid Jafarkhani. Differential distributed space-time coding with imperfect synchronization in frequency-selective channels. *IEEE Trans. Wireless Commun.*, 14(4):1811–1822, April 2015.

[14] Qi Wang, Rong Zhang, Lie-Liang Yang, and Lajos Hanzo. Non-orthogonal multiple access: A unified perspective. *IEEE Wireless Commun. Mag.*, 25(2):10–16, April 2018.

[15] Jinho Choi. Power allocation for max-sum rate and max-min rate proportional fairness in NOMA. *IEEE Commun. Lett.*, 20(10):2055–2058, October 2016.

[16] Sundaram Vanka, Sunil Srinivasa, Zhenhua Gong, Peter Vizi, Kostas Stamatiou, and Martin Haenggi. Superposition coding strategies: Design and experimental evaluation. *IEEE Trans. Wireless Commun.*, 11(7):2628–2639, July 2012.

[17] Rong Zhang and Lajos Hanzo. A unified treatment of superposition coding aided communications: Theory and practice. *IEEE Commun. Surveys Tuts.*, 13(3):503–520, Third Quarter 2011.

[18] Hosein Nikopour and Hadi Baligh. Sparse code multiple access. In *Proc. IEEE PIMRC*, pages 332–336, London, UK, September 2013.

[19] Tor Aulin and Ramon Espineira. Trellis coded multiple access (TCMA). In *Proc. IEEE ICC*, pages 1177–1181, Vancouver, BC, Canada, June 1999.

[20] Boya Di, Lingyang Song, Yonghui Li, and Geoffrey Ye Li. TCM-NOMA: Joint multi-user codeword design and detection in trellis coded modulation based NOMA for beyond 5G. *IEEE J. Sel. Topics Signal Process.*, 13(3):766–780, June 2019.

[21] Hamid Jafarkhani and V Tarokh. Design of successively refinable trellis-coded quantizers. *IEEE Trans. Inf. Theory*, 45(5):1490–1497, July 1999.

[22] Hamid Jafarkhani and V Tarokh. Multiple description trellis-coded quantization. *IEEE Trans. Commun.*, 47(6):799 – 803, June 1999.

[23] Andrew Viterbi. Convolutional codes and their performance in communication systems. *IEEE Trans. Comm. Tech.*, 19(5):751–772, October 1971.

[24] Fredrik Brannstrom, Tor M Aulin, and Lars K Rasmussen. Iterative detectors for trellis-code multiple-access. *IEEE Trans. Commun.*, 50(9):1478–1485, September 2002.

[25] Gottfried Ungerboeck. Trellis-coded modulation with redundant signal sets part I: Introduction. *IEEE Commun. Mag.*, 25(2):5–11, February 1987.

[26] Sergio Benedetto and Ezio Biglieri. *Principles of digital transmission: with wireless applications*. Springer Science & Business Media, 1999.

[27] Jianxiong Cui, Guangliang Dong, Shengli Zhang, Haitao Li, and Guinian Feng. Asynchronous NOMA for downlink transmissions. *IEEE Commun. Lett.*, 21(2):402–405, October 2017.

[28] Stefan Parkvall, Erik Strom, and Björn Ottersten. The impact of timing errors on the performance of linear DS-CDMA receivers. *IEEE J. Sel. Areas Commun.*, 14(8):1660–1668, October 1996.

[29] Heidi Steendam and Marc Moeneclaey. The effect of synchronisation errors on MC-CDMA performance. In *Proc. of IEEE ICC*, pages 1510–1514, Vancouver, BC, Canada, June 1999.

[30] Ramesh Chembil Palat, Annamalai Annamalai, and Jeffrey H Reed. Accurate bit-error-rate analysis of bandlimited cooperative OSTBC networks under timing synchronization errors. *IEEE Trans. Veh. Technol.*, 58(5):2191–2200, June 2009.

[31] Xun Zou, Biao He, and Hamid Jafarkhani. On uplink asynchronous non-orthogonal multiple access systems with timing error. In *Proc. IEEE ICC*, pages 1–6, Kansas City, MO, USA, May 2018.

[32] Stefania Sesia, Matthew Baker, and Issam Toufik. *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[33] Andrew C Marcum, James V Krogmeier, David J Love, and Alex Sprintson. Analysis and implementation of asynchronous physical layer network coding. *IEEE Trans. Wireless Commun.*, 14(12):6595–6607, December 2015.

[34] Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity limits of MIMO channels. [Online], Available: https://web.stanford.edu/class/archive/ee/ee359/ee359.1062/cup_mimo.pdf.

[35] *Release 10: Physical Layer Procedures for Evolved Universal Terrestrial Radio Access (E-UTRA)*. document TS 36.213, v10.1.0, April 2011.

[36] Jan R Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. Wiley series in probability and mathematical statistics, NYC, NY, 1988.

[37] Jerry D Gibson. *Mobile communications handbook*. CRC press, 2012.

[38] J Nicholas Laneman, David NC Tse, and Gregory W Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inf. Theory*, 50(12):3062–3080, December 2004.

[39] Erdem Koyuncu and Hamid Jafarkhani. Distributed beamforming in wireless multiuser relay-interference networks with quantized feedback. *IEEE Trans. Inf. Theory*, 58(7):4538–4576, July 2012.

[40] Yindi Jing and Hamid Jafarkhani. Single and multiple relay selection schemes and their achievable diversity orders. *IEEE Trans. Wireless Commun.*, 8(3):1414–1423, March 2009.

[41] Jung-Bin Kim and In-Ho Lee. Capacity analysis of cooperative relaying systems using non-orthogonal multiple access. *IEEE Commun. Lett.*, 19(11):1949–1952, November 2015.

[42] Jinjin Men and Jianhua Ge. Non-orthogonal multiple access for multiple-antenna relaying networks. *IEEE Commun. Lett.*, 19(10):1686–1689, October 2015.

[43] Xuesong Liang, Yongpeng Wu, Derrick Wing Kwan Ng, Yiping Zuo, Shi Jin, and Hongbo Zhu. Outage performance for cooperative NOMA transmission with an AF relay. *IEEE Commun. Lett.*, 21(11):2428–2431, November 2017.

[44] Zhiguo Ding, Mugen Peng, and H Vincent Poor. Cooperative non-orthogonal multiple access in 5G systems. *IEEE Commun. Lett.*, 19(8):1462–1465, August 2015.

[45] Xinwei Yue, Yuanwei Liu, Shaoli Kang, Arumugam Nallanathan, and Zhiguo Ding. Exploiting full/half-duplex user relaying in NOMA systems. *IEEE Trans. Commun.*, 66(2):560–575, February 2018.

[46] Zhengquan Zhang, Zheng Ma, Ming Xiao, Zhiguo Ding, and Pingzhi Fan. Full-duplex device-to-device-aided cooperative nonorthogonal multiple access. *IEEE Trans. Veh. Technol.*, 66(5):4467–4471, May 2017.

[47] Zhongxiang Wei, Xu Zhu, Sumei Sun, Jingjing Wang, and Lajos Hanzo. Energy efficient full-duplex cooperative non-orthogonal multiple access. *IEEE Trans. Veh. Technol.*, 67(10):10123–10128, October 2018.

[48] Xun Zou, Biao He, and Hamid Jafarkhani. An analysis of two-user uplink asynchronous non-orthogonal multiple access systems. *IEEE Trans. Wireless Commun.*, 18(2):1404–1418, February 2019.

[49] Mehdi Ganji and Hamid Jafarkhani. Time asynchronous NOMA for downlink transmission. In *Proc. IEEE WCNC*, pages 1–6, Marrakech, Morocco, April 2019.

[50] Mehdi Ganji and Hamid Jafarkhani. Improving NOMA multi-carrier systems with intentional frequency offsets. 8(4):1060–1063, August 2019.

[51] Shabnam Sodagari and Hamid Jafarkhani. Enhanced spectrum sharing and cognitive radio using asynchronous primary and secondary users. *IEEE Commun. Lett.*, 22(4):832–835, April 2018.

[52] Gang Liu, Xianhao Chen, Zhiguo Ding, Zheng Ma, and F Richard Yu. Hybrid half-duplex/full-duplex cooperative non-orthogonal multiple access with transmit power adaptation. *IEEE Trans. Wireless Commun.*, 17(1):506–519, January 2018.

[53] Lei Lei, Di Yuan, and Peter Värbrand. On power minimization for non-orthogonal multiple access (NOMA). *IEEE Commun. Lett.*, 20(12):2458–2461, December 2016.

[54] Yaru Fu, Yi Chen, and Chi Wan Sung. Distributed power control for the downlink of multi-cell NOMA systems. *IEEE Trans. Wireless Commun.*, 16(9):6207–6220, September 2017.

[55] Yindi Jing and Hamid Jafarkhani. Network beamforming using relays with perfect channel information. *IEEE Trans. Inf. Theory*, 55(6):2499–2517, June 2009.

[56] Xianhao Chen, Gang Liu, Zheng Ma, F. Richard Yu, and Zhiguo Ding. Power allocation for cooperative non-orthogonal multiple access systems. In *Proc. IEEE GLOBECOM*, pages 1–6, Singapore, Singapore, December 2017.

[57] Yuanwei Liu, Zhiguo Ding, Maged Elkashlan, and H Vincent Poor. Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer. *IEEE J. Sel. Areas Commun.*, 34(4):938–953, April 2016.

[58] Arthur M Geoffrion. Duality in nonlinear programming: a simplified applications-oriented development. *SIAM review*, 13(1):1–37, January 1971.

[59] Ahmed Masmoudi and Tho Le-Ngoc. Self-interference cancellation limits in full-duplex communication systems. In *Proc. IEEE GLOBECOM*, pages 1–6, Washington, DC, USA, December 2016.

[60] Jianyue Zhu, Jiaheng Wang, Yongming Huang, Shiwen He, Xiaohu You, and Luxi Yang. On optimal power allocation for downlink non-orthogonal multiple access systems. *IEEE J. Sel. Areas Commun.*, 35(12):2744–2757, December 2017.

[61] Zhiguo Ding, Zheng Yang, Pingzhi Fan, and H Vincent Poor. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Lett.*, 21(12):1501–1505, December 2014.

[62] Fang Fang, Haijun Zhang, Julian Cheng, Sébastien Roy, and Victor CM Leung. Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI. *IEEE J. Sel. Areas Commun.*, 35(12):2874–2885, December 2017.

[63] Zhiqiang Wei, Derrick Wing Kwan Ng, Jinhong Yuan, and Hui-Ming Wang. Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information. *IEEE Trans. Commun.*, 65(9):3944–3961, September 2017.

[64] Jianhua He, Zuoyin Tang, Zuowen Tang, Hsiao-Hwa Chen, and Cong Ling. Design and optimization of scheduling and non-orthogonal multiple access algorithms with imperfect channel state information. *IEEE Trans. Veh. Technol.*, 67(11):10800–10814, November 2018.

[65] Xun Zou, Mehdi Ganji, and Hamid Jafarkhani. Cooperative asynchronous non-orthogonal multiple access with power minimization under QoS constraints. *IEEE Trans. Wireless Commun.*, 19(3):1503–1518, March 2020.

[66] Zheng Yang, Zhiguo Ding, Pingzhi Fan, and George K Karagiannidis. On the performance of non-orthogonal multiple access systems with partial channel information. *IEEE Trans. Commun.*, 64(2):654–667, February 2016.

[67] Zhiguo Ding and H Vincent Poor. Design of massive-MIMO-NOMA with limited feedback. *IEEE Signal Process. Lett.*, 23(5):629–633, May 2016.

[68] Peng Xu, Yi Yuan, Zhiguo Ding, Xuchu Dai, and Robert Schober. On the outage performance of non-orthogonal multiple access with 1-bit feedback. *IEEE Trans. Wireless Commun.*, 15(10):6716–6730, October 2016.

[69] Ehsan Karamad, Behrouz Khoshnevis, and Raviraj S Adve. Quantization and bit allocation for channel state feedback in relay-assisted wireless networks. *IEEE Trans. Signal Process.*, 61(2):327–339, January 2013.

[70] David Poole. *Linear algebra: A modern introduction*. Cengage Learning, 2014.

Appendix A

Supplementary Proofs for Chapter 3

A.1 Proof of Theorem 3.1

Proof. According to (3.14), we can rewrite $\det\left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right)$ as

$$\begin{aligned}\det\left(\mathbf{I}_{2N}+\mathbf{H}\mathbf{H}^H\mathbf{R}\right) &= \det\left(\mathbf{H}\mathbf{H}^H\right)\det\left((\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}\right)\\ &= \left(P_1|h_1|^2\right)^N\left(P_2|h_2|^2\right)^N\det\left((\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}\right). \qquad \text{(A.1)}\end{aligned}$$

According to (3.8) and (3.9), $(\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}$ is a $2N\times 2N$ matrix calculated by (A.2) at the bottom of this page.

For simplicity of presentation, we define $\mu_1 = P_1|h_1|^2$, $\mu_2 = P_2|h_2|^2$, and d_m in (A.3) at the bottom of this page.

$$(\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}=\begin{bmatrix} 1+\left(P_1|h_1|^2\right)^{-1} & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1+\left(P_2|h_2|^2\right)^{-1} & \tau & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1-\tau & 1+\left(P_2|h_2|^2\right)^{-1} & \tau & 0 \\ 0 & \cdots & 0 & \tau & 1+\left(P_1|h_1|^2\right)^{-1} & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1+\left(P_2|h_2|^2\right)^{-1} \end{bmatrix}. \tag{A.2}$$

$$d_m=\begin{cases} \det\left(\begin{bmatrix} 1+\mu_1^{-1} & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1+\mu_2^{-1} & \tau & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1-\tau & 1+\mu_2^{-1} & \tau & 0 \\ 0 & \cdots & 0 & \tau & 1+\mu_1^{-1} & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1+\mu_2^{-1} \end{bmatrix}_{m\times m}\right), & \text{if } m \text{ is even}, \\ \det\left(\begin{bmatrix} 1+\mu_1^{-1} & 0 & 1-\tau & \cdots & \cdots & 0 \\ 1-\tau & 1+\mu_2^{-1} & \tau & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \tau & 1+\mu_1^{-1} & 1-\tau & 0 \\ 0 & \cdots & 0 & 1-\tau & 1+\mu_2^{-1} & \tau \\ 0 & \cdots & \cdots & 0 & \tau & 1+\mu_1^{-1} \end{bmatrix}_{m\times m}\right), & \text{if } m \text{ is odd}. \end{cases} \tag{A.3}$$

Thus, $\det\left((\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}\right)=d_{2N}$.

By the method of cofactor expansion [70], the determinant of $\det\left((\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}\right)$ can be expressed as a weighted sum of the determinants of its minors. The minor $M_{i,j}$ is defined as the determinant of the matrix that results from $(\mathbf{H}\mathbf{H}^H)^{-1}+\mathbf{R}$ by removing the ith row and the jth column. Then, we have

$$\begin{aligned} d_{2N}&=\sum_{j=1}^{2N}(-1)^{2N+j}a_{2N,j}M_{2N,j} \\ &=(-1)^{2N+2N}\left(1+\mu_2^{-1}\right)\underbrace{\det\left(\begin{bmatrix} 1+\mu_1^{-1} & 1-\tau & \cdots & \cdots & 0 \\ 1-\tau & 1+\mu_2^{-1} & \tau & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1-\tau & 1+\mu_2^{-1} & \tau \\ 0 & \cdots & 0 & \tau & 1+\mu_1^{-1} \end{bmatrix}_{(2N-1)\times(2N-1)}\right)}_{d_{2N-1}} \end{aligned}$$

$$
\begin{aligned}
&+ (-1)^{2N+2N-1}(1-\tau)\det\left(\begin{bmatrix} 1+\mu_1^{-1} & 1-\tau & \cdots & \cdots & 0 \\ 1-\tau & 1+\mu_2^{-1} & \tau & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1-\tau & 1+\mu_2^{-1} & 0 \\ 0 & \cdots & 0 & \tau & 1-\tau \end{bmatrix}_{(2N-1)\times(2N-1)}\right) \\
&= \left(1+\mu_2^{-1}\right) d_{2N-1} - (1-\tau)^2(-1)^{4N-2} \\
&\cdot \underbrace{\det\left(\begin{bmatrix} 1+\mu_1^{-1} & 1-\tau & 0 & \cdots & \cdots & 0 \\ 1-\tau & 1+\mu_2^{-1} & \tau & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1-\tau & 1+\mu_2^{-1} & \tau & 0 \\ 0 & \cdots & 0 & \tau & 1+\mu_1^{-1} & 1-\tau \\ 0 & \cdots & \cdots & 0 & 1-\tau & 1+\mu_2^{-1} \end{bmatrix}_{(2N-2)\times(2N-2)}\right)}_{d_{2N-2}} \\
&= \left(1+\mu_2^{-1}\right) d_{2N-1} - (1-\tau)^2 d_{2N-2}, \qquad \text{(A.4)}
\end{aligned}
$$

where $N \geq 2$ and $a_{i,j}$ denotes the element of the matrix $(\mathbf{H}\mathbf{H}^H)^{-1} + \mathbf{R}$ at the ith row and the jth column. Similarly, we can also write the recursive formula for d_{2N-1} as

$$d_{2N-1} = (1+\mu_1^{-1})d_{2N-2} - \tau^2 d_{2N-3}. \qquad \text{(A.5)}$$

From (A.4) and (A.5), we obtain

$$d_{2N} = \left[\mu_1^{-1} + \mu_2^{-1} + \mu_1^{-1}\mu_2^{-1} + 2\tau(1-\tau)\right] d_{2N-2} - \tau^2(1-\tau)^2 d_{2N-4}. \qquad \text{(A.6)}$$

To formalize (A.6) as the recursion formula of a geometric progression, (A.6) can be rewritten as

$$d_{2N} - r_1 d_{2N-2} = r_2(d_{2N-2} - r_1 d_{2N-4}), \qquad \text{(A.7)}$$

$$d_{2N} - r_2 d_{2N-2} = r_1(d_{2N-2} - r_2 d_{2N-4}), \qquad \text{(A.8)}$$

where

$$r_1 = \frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+2\tau(1-\tau)}{2} + \frac{\sqrt{\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+2\tau(1-\tau)\right]^2-4\tau^2(1-\tau)^2}}{2}, \tag{A.9}$$

$$r_2 = \frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+2\tau(1-\tau)}{2} - \frac{\sqrt{\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+2\tau(1-\tau)\right]^2-4\tau^2(1-\tau)^2}}{2}. \tag{A.10}$$

Since $\mu_1 > 0$, $\mu_2 > 0$, and $\tau \in [0, 1)$, we note that the part under the square root symbol in (A.9) and (A.10) is always positive, such that

$$\begin{aligned}&\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+2\tau(1-\tau)\right]^2 - 4\tau^2(1-\tau)^2 \\ &= \left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+4\tau(1-\tau)\right]\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}\right] > 0.\end{aligned} \tag{A.11}$$

From (A.7) and (A.8), we obtain

$$d_{2N} - r_1 d_{2N-2} = r_2^{N-1}(d_2 - r_1 d_0), \tag{A.12}$$

$$d_{2N} - r_2 d_{2N-2} = r_1^{N-1}(d_2 - r_2 d_0). \tag{A.13}$$

Solving d_{2N} from the equation group constituted by (A.12) and (A.13), we derive

$$d_{2N} = \frac{r_1^N(d_2 - r_2 d_0) - r_2^N(d_2 - r_1 d_0)}{r_1 - r_2}. \tag{A.14}$$

Substituting $d_0 = 1$ and

$$d_2 = \begin{bmatrix} 1+\mu_1^{-1} & 1-\tau \\ 1-\tau & 1+\mu_2^{-1} \end{bmatrix} = \left(1+\mu_1^{-1}\right)\left(1+\mu_2^{-1}\right) - (1-\tau)^2 = r_1 + r_2 + \tau^2 \tag{A.15}$$

into (A.14), we have

$$d_{2N} = \frac{(r_1^{N+1} - r_2^{N+1}) + \tau^2(r_1^N - r_2^N)}{r_1 - r_2}. \tag{A.16}$$

Finally, based on (A.1) and (A.16), we obtain the throughput as

$$R^{\text{ANOMA}} = \frac{1}{N+\tau} \log \frac{(r_1^{N+1} - r_2^{N+1}) + \tau^2(r_1^N - r_2^N)}{r_1 - r_2} + \frac{N}{N+\tau} \log\left(\mu_1\mu_2\right). \tag{A.17}$$

This completes the proof. □

A.2 Proof of Corollary 3.1

Proof. Note that μ_1, μ_2, r_1, r_2, and τ are all independent of N. We then have

$$\begin{aligned}
\lim_{N\to\infty} R^{\text{ANOMA}} &= \lim_{N\to\infty} \frac{N}{N+\tau} \log(\mu_1\mu_2) \\
&\quad + \lim_{N\to\infty} \frac{\log\left[(r_1^{N+1} - r_2^{N+1}) + \tau^2(r_1^N - r_2^N)\right] - \log(r_1 - r_2)}{N+\tau} \\
&\overset{(a)}{=} \log(\mu_1\mu_2) + \lim_{N\to\infty} \frac{(r_1^{N+1}\log r_1 - r_2^{N+1}\log r_2) + \tau^2(r_1^N \log r_1 - r_2^N \log r_2)}{(r_1^{N+1} - r_2^{N+1}) + \tau^2(r_1^N - r_2^N)} \\
&\overset{(b)}{=} \log(\mu_1\mu_2) + \lim_{N\to\infty} \frac{(r_1\alpha^N \log r_1 - r_2 \log r_2) + \tau^2(\alpha^N \log r_1 - \log r_2)}{(r_1\alpha^N - r_2) + \tau^2(\alpha^N - 1)} \\
&= \log(\mu_1\mu_2) + \lim_{N\to\infty} \frac{\alpha^N(r_1+\tau^2)\log r_1 - (r_2+\tau^2)\log r_2}{\alpha^N(r_1+\tau^2) - (r_2+\tau^2)} \\
&\overset{(c)}{=} \log\left(\mu_1\mu_2 r_1\right),
\end{aligned} \tag{A.18}$$

where $\alpha = r_1/r_2$, (a) is derived by applying L'Hospital's rule, (b) is derived by dividing both the numerator and the denominator by r_2^N, and (c) is obtained from the facts that $r_1 > r_2 > 0$ and $\alpha > 1$ according to (3.17) and (3.18). This completes the proof. □

A.3 Proof of Theorem 3.2

Proof. The expressions for $\lim_{N\to\infty} R^{\text{ANOMA}}$ and R^{NOMA} are given by

$$\begin{aligned}\lim_{N\to\infty} R^{\text{ANOMA}} &= \log\left(\mu_1\mu_2 r_1\right) \\ &= \log\left\{\frac{1+\mu_1+\mu_2+\mu_1\mu_2(2\tau-2\tau^2)}{2}\right. \\ &\quad \left.+\frac{\sqrt{\left(1+\mu_1+\mu_2\right)^2+2\left(1+\mu_1+\mu_2\right)\mu_1\mu_2(2\tau-2\tau^2)}}{2}\right\} \end{aligned} \tag{A.19}$$

and $R^{\text{NOMA}} = \log(1+\mu_1+\mu_2)$, respectively.

If $\tau = 0$, it is easy to find that $\lim_{N\to\infty} R^{\text{ANOMA}} = \log(1+\mu_1+\mu_2) = R^{\text{NOMA}}$.

If $\tau \neq 0$, i.e., $\tau \in (0,1)$, we have $2\tau - 2\tau^2 > 0$. According to (A.19), since $\mu_1 > 0$ and $\mu_2 > 0$, we obtain

$$\begin{aligned}\lim_{N\to\infty} R^{\text{ANOMA}} &= \log\left\{\frac{1+\mu_1+\mu_2+\mu_1\mu_2(2\tau-2\tau^2)}{2}\right. \\ &\quad \left.+\frac{\sqrt{\left(1+\mu_1+\mu_2\right)^2+2\left(1+\mu_1+\mu_2\right)\mu_1\mu_2(2\tau-2\tau^2)}}{2}\right\} \\ &> \log\left\{\frac{1+\mu_1+\mu_2}{2}+\frac{\sqrt{\left(1+\mu_1+\mu_2\right)^2}}{2}\right\} = R^{\text{NOMA}}. \end{aligned} \tag{A.20}$$

Until now, we have proved $\lim_{N\to\infty} R^{\text{ANOMA}} = R^{\text{NOMA}}$ if $\tau = 0$ and $\lim_{N\to\infty} R^{\text{ANOMA}} >$

R^{NOMA} if $\tau \neq 0$. Next, we need to prove $\tau = 0$ if $\lim_{N\to\infty} R^{\text{ANOMA}} = R^{\text{NOMA}}$.

If $\lim_{N\to\infty} R^{\text{ANOMA}} = R^{\text{NOMA}}$, we have $\lim_{N\to\infty} R^{\text{ANOMA}} = \log(\mu_1\mu_2 r_1) = \log(1+\mu_1+\mu_2) = R^{\text{NOMA}}$.

After simplifications, we have

$$\sqrt{(1+\mu_1+\mu_2)^2+2\,(1+\mu_1+\mu_2)\,\mu_1\mu_2(2\tau-2\tau^2)} = 1+\mu_1+\mu_2-\mu_1\mu_2(2\tau-2\tau^2). \quad \text{(A.21)}$$

Note that (A.21) holds only if the right side of (A.21) is non-negative, i.e.,

$$1+\mu_1+\mu_2-\mu_1\mu_2(2\tau-2\tau^2) \geq 0. \quad \text{(A.22)}$$

Squaring both sides of the equal sign in (A.21), we obtain

$$4(1+\mu_1+\mu_2)(2\tau-2\tau^2) = \mu_1\mu_2(2\tau-2\tau^2)^2. \quad \text{(A.23)}$$

Then, (A.23) holds if $2\tau-2\tau^2 = 0$ or $4(1+\mu_1+\mu_2) = \mu_1\mu_2(2\tau-2\tau^2)$. It is easy to prove that $4(1+\mu_1+\mu_2) = \mu_1\mu_2(2\tau-2\tau^2)$ contradicts (A.22). As a result, (A.23) holds only if $2\tau-2\tau^2 = 0$ which then leads to $\tau = 0$.

Therefore, $\lim_{N\to\infty} R^{\text{ANOMA}} \geq R^{\text{NOMA}}$ is always true and the equal sign is achieved if and only if $\tau = 0$. This completes the proof. □

A.4 Proof of Theorem 3.3

Proof. From Theorem 3.1, we have

$$\begin{aligned} R^{\text{ANOMA}} &= \frac{N}{N+\tau}\log(\mu_1\mu_2) + \frac{1}{N+\tau}\log\frac{(r_1^{N+1}-r_2^{N+1})+\tau^2(r_1^N-r_2^N)}{r_1-r_2} \\ &\overset{(a)}{=} \frac{N}{N+\tau}\log(\mu_1\mu_2) + \frac{1}{N+\tau}\log\left[\sum_{i=0}^{N} r_1^i r_2^{N-i} + \tau^2\sum_{i=0}^{N-1} r_1^i r_2^{N-1-i}\right] \\ &= \frac{1}{N+\tau}\log\left[\sum_{i=0}^{N} \mu_1^N\mu_2^N r_1^i r_2^{N-i} + \tau^2\sum_{i=0}^{N-1} \mu_1^N\mu_2^N r_1^i r_2^{N-1-i}\right] \\ &= \frac{1}{N+\tau}\log\left[\sum_{i=0}^{N} (\mu_1\mu_2)^{N-i}(\mu_1\mu_2 r_1)^i r_2^{N-i} + \tau^2\sum_{i=0}^{N-1} (\mu_1\mu_2)^{N-i}(\mu_1\mu_2 r_1)^i r_2^{N-1-i}\right], \end{aligned} \tag{A.24}$$

where (a) is derived by applying $a^N - b^N = (a-b)(\sum_{i=0}^{N-1} a^i b^{N-1-i})$. In what follows, we prove that r_2 is a non-decreasing function of μ_1 and μ_2, and $\mu_1\mu_2 r_1$ increases as μ_1 and μ_2 increase, so that R^{ANOMA} increases as μ_1 and μ_2 increase.

From (3.17), we can find that $\frac{\partial r_1}{\partial \mu_1} < 0$ and $\frac{\partial r_1}{\partial \mu_2} < 0$. Since $r_2 = \tau^2(1-\tau)^2/r_1$, we further find that

$$\frac{\partial r_2}{\partial \mu_1} = -\frac{\tau^2(1-\tau)^2}{r_1^2}\frac{\partial r_1}{\partial \mu_1} \geq 0 \text{ and } \frac{\partial r_2}{\partial \mu_2} = -\frac{\tau^2(1-\tau)^2}{r_1^2}\frac{\partial r_1}{\partial \mu_2} \geq 0. \tag{A.25}$$

With (3.17), we have

$$\begin{aligned} \mu_1\mu_2 r_1 = &\frac{1+\mu_1+\mu_2+\mu_1\mu_2(2\tau-2\tau^2)}{2} \\ &+ \frac{\sqrt{(1+\mu_1+\mu_2)^2 + 2(1+\mu_1+\mu_2)\mu_1\mu_2(2\tau-2\tau^2)}}{2}. \end{aligned} \tag{A.26}$$

From (A.26), we can derive that

$$\frac{\partial(\mu_1\mu_2 r_1)}{\partial \mu_1} > 0 \text{ and } \frac{\partial(\mu_1\mu_2 r_1)}{\partial \mu_2} > 0. \tag{A.27}$$

Based on (A.25) and (A.27), we note that r_2 is a non-decreasing function of μ_1 and μ_2, and $\mu_1\mu_2 r_1$ increases as μ_1 and μ_2 increase. In addition, since μ_1, μ_2, r_2, and $\mu_1\mu_2 r_1$ are positive, the term $(\mu_1\mu_2)^j(\mu_1\mu_2 r_1)^i r_2^{M-i}$ $(i = 0, \cdots, M-1,\, j \geq 0)$ is an increasing function of μ_1 and μ_2 for any positive M. Then, R^{ANOMA} is an increasing function of μ_1 and μ_2 because it is constituted by a sum of $(\mu_1\mu_2)^j(\mu_1\mu_2 r_1)^i r_2^{M-i}$ $(i = 0, \cdots, M-1,\, j \geq 0,\, M > 0)$. Hence, maximizing the throughput is equivalent to maximizing μ_1 and μ_2 simultaneously, which means that the two users should transmit at full power. This completes the proof. □

A.5 Proof of Theorem 3.4

Proof.

$$\begin{aligned}
\tau^* &= \arg\max_{\tau} \lim_{N\to\infty} R^{\text{ANOMA}} = \arg\max_{\tau} \log\left(\mu_1\mu_2 r_1\right) \\
&= \arg\max_{\tau} \log \Bigg\{ \frac{1+\mu_1+\mu_2+\mu_1\mu_2(2\tau-2\tau^2)}{2} \\
&\quad + \frac{\sqrt{\left(1+\mu_1+\mu_2\right)^2 + 2\left(1+\mu_1+\mu_2\right)\mu_1\mu_2(2\tau-2\tau^2)}}{2} \Bigg\} \\
&\overset{(a)}{=} \arg\max_{\tau} \left[2\tau - 2\tau^2\right] = 0.5,
\end{aligned} \tag{A.28}$$

where (a) is derived due to the fact that μ_1 and μ_2 are positive and independent of τ. This completes the proof. □

Appendix B

Supplementary Proofs for Chapter 4

B.1 Derivation of (4.18) and (4.19)

Substituting $\mathbf{G}_1$ and $\mathbf{R}$ by their expressions, the matrix determinant term in (4.17) becomes (B.1), where (a) and (b) are derived by applying the cofactor expansion [70], (c) is derived by applying the cofactor expansion iteratively. Thus, Eq. (4.19) is obtained.

According to (4.16), we have

$$\begin{aligned}
R_{2\to1}^{\text{ANOMA}} =& \frac{1}{2N+\tau}\log\det\left[\mathbf{I}_{2N}+\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}P_2|h_1|^2\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right]\\
=& \frac{1}{2N+\tau}\log\det\Big[\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\\
&\cdot\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}+P_2|h_1|^2\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right)\Big]\\
=& \frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}+P_2|h_1|^2\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right)\\
&-\frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)\\
\overset{(a)}{=}& \frac{1}{2N+\tau}\left[\log\det\left(\mathbf{I}_{2N}+\mathbf{HR}\right)-\log\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)\right], \qquad \text{(B.2)}
\end{aligned}$$

$$
\begin{aligned}
&\det\left(\mathbf{I}_{2N}+P_1|h_1|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)\\
=&\det\begin{bmatrix}
1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & 0 & \cdots & \cdots & 0\\
0 & 1 & 0 & 0 & \cdots & 0\\
0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & \cdots & 0\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \cdots & 0 & 1 & 0 & 0\\
0 & \cdots & 0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau)\\
0 & \cdots & \cdots & 0 & 0 & 1
\end{bmatrix}_{2N\times 2N}\\
\overset{(a)}{=}&\det\begin{bmatrix}
1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & 0 & \cdots & \cdots & 0\\
0 & 1 & 0 & 0 & \cdots & 0\\
0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & \cdots & 0\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \cdots & 0 & 0 & 1 & 0\\
0 & \cdots & 0 & 0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2
\end{bmatrix}_{(2N-1)\times(2N-1)}\\
\overset{(b)}{=}&\left(1{+}P_1|h_1|^2\right)\\
&\cdot\det\begin{bmatrix}
1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & 0 & \cdots & \cdots & 0\\
0 & 1 & 0 & 0 & \cdots & 0\\
0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau) & \cdots & 0\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
0 & \cdots & 0 & P_1|h_1|^2\tau & 1{+}P_1|h_1|^2 & P_1|h_1|^2(1{-}\tau)\\
0 & \cdots & \cdots & 0 & 0 & 1
\end{bmatrix}_{(2N-2)\times(2N-2)}\\
=&\cdots\overset{(c)}{=}(1+P_1|h_1|^2)^N. \qquad \text{(B.1)}
\end{aligned}
$$

where (a) is derived because $\mathbf{G}_i\mathbf{G}_i^H$ is a $2N$-by-$2N$ matrix whose odd (if $i=1$) or even (if $i=2$) diagonal elements are 1 and all the others are 0, and $\mathbf{H}=|h_1|^2\cdot\mathrm{diag}\left([P_1,P_2,\cdots,P_1,P_2]\right)$.

According to Theorem 1 in [48], the term $\log\det\left(\mathbf{I}_{2N}+\mathbf{HR}\right)$ in (B.2) can be written as

$$
\log\det\left(\mathbf{I}_{2N}+\mathbf{HR}\right)=N\log\left(\mu_1\mu_2\right)+\log\frac{\left(r_1^{N+1}-r_2^{N+1}\right)+\tau^2\left(r_1^N-r_2^N\right)}{r_1-r_2}, \qquad \text{(B.3)}
$$

where $\mu_1=P_1|h_1|^2,\mu_2=P_2|h_1|^2,Q=2\tau(1-\tau)$,

$$
r_1=\frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q+\sqrt{\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q\right]^2-Q^2}}{2}, \qquad \text{(B.4)}
$$

$$
r_2=\frac{\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q-\sqrt{\left[\mu_1^{-1}+\mu_2^{-1}+\mu_1^{-1}\mu_2^{-1}+Q\right]^2-Q^2}}{2}. \qquad \text{(B.5)}
$$

Thus, Eq. (4.18) can be easily derived according to (B.1) and (B.3).

B.2 Proof of Theorem 4.1

Proof. According to Corollary 1 in [48], we have

$$\lim_{N\to\infty} \frac{1}{N+\tau} \log \frac{\left(r_1^{N+1} - r_2^{N+1}\right) + \tau^2 \left(r_1^N - r_2^N\right)}{r_1 - r_2} = \log r_1. \tag{B.6}$$

As a result, the combining throughput of User 2 for $N \to \infty$ is calculated as

$$\begin{aligned} R_{2,\text{asymp}}^{\text{ANOMA}} =& \frac{1}{2} \log \left(\frac{\mu_1 \mu_2 r_1}{1+\mu_1} \right) \\ =& \frac{1}{2} \log \left(\frac{1+\mu_1+\mu_2+\mu_1\mu_2 Q + \sqrt{(1+\mu_1+\mu_2)^2 + 2\left(1+\mu_1+\mu_2\right)\mu_1\mu_2 Q}}{2(1+\mu_1)} \right), \end{aligned}$$

where $\mu_1 > 0$, $\mu_2 > 0$, $\tau \in [0,1)$, and $Q = 2\tau(1-\tau) > 0$. One can easily derive

$$\begin{aligned} 1+\mu_1+\mu_2 &\leq \sqrt{(1+\mu_1+\mu_2)^2 + 2\left(1+\mu_1+\mu_2\right)\mu_1\mu_2 Q} \\ &= \sqrt{(1+\mu_1+\mu_2+\mu_1\mu_2 Q)^2 - (\mu_1\mu_2 Q)^2} \leq 1+\mu_1+\mu_2+\mu_1\mu_2 Q, \end{aligned}$$

and the equal sign is achieved if and only if $\tau = 0$. As a result,

$$\frac{1}{2} \log \left(\frac{1+\mu_1+\mu_2+0.5\mu_1\mu_2 Q}{1+\mu_1} \right) \leq R_{2,\text{asymp}}^{\text{ANOMA}} \leq \frac{1}{2} \log \left(\frac{1+\mu_1+\mu_2+\mu_1\mu_2 Q}{1+\mu_1} \right).$$

Note that,

$$\frac{1}{2} \log \left(\frac{1+\mu_1+\mu_2+0.5\mu_1\mu_2 Q}{1+\mu_1} \right) \geq \frac{1}{2} \log \left(\frac{1+\mu_1+\mu_2}{1+\mu_1} \right) = R_2^{\text{NOMA}},$$

where the equal sign is achieved if and only if $\tau = 0$. The proof is complete. □

B.3 Proof of Theorem 4.2

Proof. According to (4.27), the combining throughput of User 2 is given by (B.7),

$$
\begin{aligned}
&R_2^{\text{ANOMA}} \\
=&\frac{1}{2N+\tau}\log\det\left[\mathbf{I}_{3N}+\left(\mathbf{R_N}+\mathbf{W}_1\mathbf{W}_1^H\right)^{-1}\mathbf{W}_2\mathbf{W}_2^H\right] \\
=&\frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{3N}+\left[\begin{smallmatrix}\mathbf{R}+P_1|h_2|^2\mathbf{R}\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}^H & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_N\end{smallmatrix}\right]^{-1}\left[\begin{smallmatrix}P_2|h_2|^2\mathbf{R}\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}^H & \sqrt{P_2P_r}h_2\bar{h}_{12}\mathbf{R}\mathbf{G}_2\\ \sqrt{P_2P_r}h_{12}\bar{h}_2\mathbf{G}_2^H\mathbf{R}^H & P_r|h_{12}|^2\mathbf{I}_N\end{smallmatrix}\right]\right) \\
=&\frac{1}{2N+\tau} \\
&\cdot\log\det\left[\begin{smallmatrix}\mathbf{I}_{2N}+P_2|h_2|^2\left(\mathbf{R}+P_1|h_2|^2\mathbf{R}\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\mathbf{R}\mathbf{G}_2\mathbf{G}_2^H\mathbf{R} & \sqrt{P_2P_r}h_2\bar{h}_{12}\left(\mathbf{R}+P_1|h_2|^2\mathbf{R}\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\mathbf{R}\mathbf{G}_2\\ \sqrt{P_2P_r}h_{12}\bar{h}_2\mathbf{G}_2^H\mathbf{R} & \mathbf{I}_N+P_r|h_{12}|^2\mathbf{I}_N\end{smallmatrix}\right] \\
\overset{(a)}{=}&\frac{1}{2N+\tau}\log\left\{\det\left[(1+P_r|h_{12}|^2)\mathbf{I}_N\right]\det\left[\mathbf{I}_{2N}+P_2|h_2|^2\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\right.\right. \\
&\left.\left.\cdot\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}-\frac{P_2P_r|h_2|^2|h_{12}|^2}{1+P_r|h_{12}|^2}\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\right]\right\} \\
=&\frac{N}{2N+\tau}\log\left(1+P_r|h_{12}|^2\right)+\frac{1}{2N+\tau}\log\det\left\{\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1}\right. \\
&\left.\cdot\left[\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}+\left(P_2|h_2|^2-\frac{P_2P_r|h_2|^2|h_{12}|^2}{1+P_r|h_{12}|^2}\right)\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right)\right]\right\} \\
=&\frac{N}{2N+\tau}\log\left(1+P_r|h_{12}|^2\right)+\frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)^{-1} \\
&+\frac{1}{2N+\tau}\log\det\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}+\frac{P_2|h_2|^2}{1+P_r|h_{12}|^2}\mathbf{G}_2\mathbf{G}_2^H\mathbf{R}\right) \\
=&\frac{N}{2N+\tau}\log\left(1+P_r|h_{12}|^2\right)-\frac{\log\det\left(\mathbf{I}_{2N}+P_1|h_2|^2\mathbf{G}_1\mathbf{G}_1^H\mathbf{R}\right)}{2N+\tau}+\frac{\log\det(\mathbf{I}_{2N}+\tilde{\mathbf{H}}\mathbf{R})}{2N+\tau}.
\end{aligned}
\tag{B.7}
$$

where (a) is derived by applying the determinant of the block matrix, i.e., if $\mathbf{D}$ is invertible,

$$
\det\left(\begin{smallmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{smallmatrix}\right)=\det(\mathbf{D})\det(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}) \tag{B.8}
$$

and $\tilde{\mathbf{H}}=\text{diag}\left(\left[P_1|h_2|^2,\frac{P_2|h_2|^2}{1+P_r|h_{12}|^2},\cdots P_1|h_2|^2,\frac{P_2|h_2|^2}{1+P_r|h_{12}|^2}\right]\right)$.

Applying (B.1) and (B.3), Eq. (B.7) can be rewritten as (4.28). The proof is complete. □

B.4 Proof of Theorem 4.3

Proof. Applying (B.6), the combining throughput of User 2 for $N \to \infty$ is computed as

$$\begin{aligned} R_{2,\text{asymp}}^{\text{ANOMA}} =& \frac{1}{2}\log\left(\frac{P_1P_2|h_2|^4}{1+P_1|h_2|^2}z_1\right) \\ \overset{(a)}{=}& \frac{1}{2}\log\Bigg[\frac{1+P_r|h_{12}|^2}{2}+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{2(1+P_1|h_2|^2)} \\ &+\frac{1}{2}\sqrt{\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2-\left(\frac{P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2}\Bigg], \end{aligned}$$

where (a) is derived by replacing z_1 with its expression in (4.30). Since $Q \geq 0$,

$$\begin{aligned} &1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{1+P_1|h_2|^2} \\ \geq&\sqrt{\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2+P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2-\left(\frac{P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\right)^2} \\ =&\left[\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2}{1+P_1|h_2|^2}\right)^2+\frac{2P_1P_2|h_2|^4Q}{1+P_1|h_2|^2}\left(1+P_r|h_{12}|^2+\frac{P_2|h_2|^2}{1+P_1|h_2|^2}\right)\right]^{\frac{1}{2}} \\ \geq&1+P_r|h_{12}|^2+\frac{P_2|h_2|^2}{1+P_1|h_2|^2}, \end{aligned}$$

where the equal signs are achieved if and only if $Q=0$ which results in $\tau=0$. The proof is complete. □

Appendix C

Supplementary Proofs for Chapter 5

C.1 Proof of Theorem 5.1

Proof. According to (5.26), (5.27), (5.28), and (5.29), we have

$$R^{\mathrm{A}}_{2\to i,\infty} - R^{\mathrm{N}}_{2\to i} = \log \frac{1+Q\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1} + \sqrt{1+2Q\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1}}}{1+Q\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1} + \sqrt{1+2Q\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1}}},$$

$$R^{\mathrm{A}}_{1,\infty} - R^{\mathrm{N}}_{1} = \log \frac{1+Q\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1} + \sqrt{1+2Q\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1}}}{1+Q\frac{\rho_{1}\rho_{2}}{\rho_{1}+\rho_{2}+1} + \sqrt{1+2Q\frac{\rho_{1}\rho_{2}}{\rho_{1}+\rho_{2}+1}}}.$$

For the ease of proof, we define the function $f(x) = 1+Qx+\sqrt{1+2Qx}$. Then, we can rewrite $R^{\mathrm{A}}_{2\to i,\infty}-R^{\mathrm{N}}_{2\to i} = \log\left[f\left(\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1}\right)/f\left(\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1}\right)\right]$ and $R^{\mathrm{A}}_{1,\infty}-R^{\mathrm{N}}_{1} = \log\left[f\left(\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1}\right)/f\left(\frac{\rho_{1}\rho_{2}}{\rho_{1}+\rho_{2}+1}\right)\right]$. It is obvious that the function $f(x)$ increases with x. To compare $R^{\mathrm{A}}_{2\to i,\infty}$ and $R^{\mathrm{N}}_{2\to i}$, we study the relationship between $\frac{\mu_{1i}\mu_{2i}}{\mu_{1i}+\mu_{2i}+1}$ and $\frac{\nu_{1i}\nu_{2i}}{\nu_{1i}+\nu_{2i}+1}$.

Since $\mu_{2i} > \nu_{2i}$ and $\mu_{1i} = \nu_{1i}$,

$$\begin{aligned}
&\frac{1+\mu_{1i}}{\mu_{2i}} < \frac{1+\nu_{1i}}{\nu_{2i}} \\
\Longrightarrow &\frac{1+\mu_{1i}+\mu_{2i}}{\mu_{2i}} < \frac{1+\nu_{1i}+\nu_{2i}}{\nu_{2i}} \\
\Longrightarrow &\frac{1+\mu_{1i}+\mu_{2i}}{\mu_{1i}\mu_{2i}} < \frac{1+\nu_{1i}+\nu_{2i}}{\nu_{1i}\nu_{2i}} \\
\Longrightarrow &\frac{\mu_{1i}\mu_{2i}}{1+\mu_{1i}+\mu_{2i}} > \frac{\nu_{1i}\nu_{2i}}{1+\nu_{1i}+\nu_{2i}}.
\end{aligned}$$

As a result, $f\left(\frac{\mu_{1i}\mu_{2i}}{1+\mu_{1i}+\mu_{2i}}\right) > f\left(\frac{\nu_{1i}\nu_{2i}}{1+\nu_{1i}+\nu_{2i}}\right)$ which results in $R^{\mathrm{A}}_{2\to i,\infty} - R^{\mathrm{N}}_{2\to i} > 0$. Similarly, by deriving $\frac{\nu_{11}\nu_{21}}{\nu_{11}+\nu_{21}+1} > \frac{\rho_1\rho_2}{\rho_1+\rho_2+1}$, we obtain $R^{\mathrm{A}}_{1,\infty} > R^{\mathrm{N}}_1$. □

C.2 Proof of Theorem 5.2

Proof. By setting $g(x,y,z) = 1 + x + y + xyz + \sqrt{(1+x+y)^2 + 2xyz(1+x+y)}$, $R^{\mathrm{A}}_{2\to i,\infty} = \log \frac{g(\mu_{1i},\mu_{2i},Q)}{g(\nu_{1i},\nu_{2i},Q)}$. Hence,

$$\begin{aligned}
\max R^{\mathrm{A}}_{2\to i,\infty} &\Longleftrightarrow \max \frac{g(\mu_{1i},\mu_{2i},Q)}{g(\nu_{1i},\nu_{2i},Q)} \\
&\Longleftrightarrow \max \left[g(\mu_{1i},\mu_{2i},Q) - g(\nu_{1i},\nu_{2i},Q) \right].
\end{aligned}$$

The derivative of $g(x,y,Q)$ with respect to Q is calculated as

$$\begin{aligned}
\frac{\partial g(x,y,Q)}{\partial Q} &= xy + \frac{Q(1+x+y)xy}{\sqrt{(1+x+y)^2 + 2xyQ(1+x+y)}} \\
&= xy\left[1 + Q\sqrt{\frac{1+x+y}{1+x+y+2Qxy}}\right] \\
&= xy + Q\sqrt{xy}\sqrt{\frac{(1+x+y)xy}{1+x+y+2Qxy}}.
\end{aligned}$$

It is trivial to show that xy and $\frac{(1+x+y)xy}{1+x+y+2Qxy}$ are increasing functions of y which is equivalent to $\frac{\partial^2 g(x,y,Q)}{\partial Q \partial y} > 0$. Since $\mu_{1i} = \nu_{1i}$ and $\mu_{2i} > \nu_{2i}$, we obtain

$$\frac{\partial[g(\mu_{1i}, \mu_{2i}, Q) - g(\nu_{1i}, \nu_{2i}, Q)]}{\partial Q} > 0.$$

As a result,

$$\begin{aligned} &\arg\max_{\tau} \left[g(\mu_{1i}, \mu_{2i}, Q) - g(\nu_{1i}, \nu_{2i}, Q)\right] \\ &\iff \arg\max_{\tau} Q = 0.5. \end{aligned}$$

Hence, $\arg\max_\tau R^{\mathrm{A}}_{2\to i,\infty} = 0.5$. Similarly, since $\nu_{11} > \rho_1$ and $\nu_{21} = \rho_2$, one can prove $\arg\max_\tau R^{\mathrm{A}}_{1,\infty} = 0.5$ by following the same steps above. □

Appendix D

Supplementary Proofs for Chapter 6

D.1 Proof of Theorem 6.1

Proof. When $\tau = 0$, $Q = 0$ which then results in $\alpha_{\mathrm{N}}^* = \alpha_{\mathrm{A}}^*(x)$ for any finite x. For $\tau \neq 0$, we first prove Theorem 6.1 for the case of $\tilde{H}_1 \geq \tilde{H}_2$. The case of $\tilde{H}_1 < \tilde{H}_2$ SIC will be discussed later. By setting $R_{\mathrm{strong}}(\tilde{H}_1) = R_{\mathrm{weak}}(\tilde{H}_2)$, we obtain

$$\begin{aligned} &2[1 + \alpha_{\mathrm{A}}^* P(\tilde{H}_1 + \tilde{H}_2) + (\alpha_{\mathrm{A}}^*)^2 P^2 \tilde{H}_1 \tilde{H}_2] \\ =&\sqrt{[1+P\tilde{H}_2+\alpha_{\mathrm{A}}^*(1-\alpha_{\mathrm{A}}^*)P^2\tilde{H}_2^2Q]^2-[\alpha_{\mathrm{A}}^*(1-\alpha_{\mathrm{A}}^*)]^2P^4\tilde{H}_2^4Q^2} \\ &+ 1 + P\tilde{H}_2 + \alpha_{\mathrm{A}}^*(1 - \alpha_{\mathrm{A}}^*)P^2\tilde{H}_2^2Q. \end{aligned} \tag{D.1}$$

By cancelling out the square root, (D.1) becomes a quartic equation. The optimal power coefficient α_{A}^* is one root of the quartic equation which can be given by the general formula for quartic roots. However, it is intractable for further analysis. Therefore, we derive the upper and lower bounds to approximate the actual value of α_{A}^*.

For the upper bound, since $[1+P\tilde{H}_2+\alpha_{\mathrm{A}}^*(1-\alpha_{\mathrm{A}}^*)P^2\tilde{H}_2^2Q]^2-(\alpha_{\mathrm{A}}^*)^2(1-\alpha_{\mathrm{A}}^*)^2P^4\tilde{H}_2^4Q^2 < [1+$

$P\tilde{H}_2+\alpha_{\rm A}^*(1-\alpha_{\rm A}^*)P^2\tilde{H}_2^2Q]^2$, (D.1) becomes

$$1+\alpha_{\rm A}^*P(\tilde{H}_1+\tilde{H}_2)+(\alpha_{\rm A}^*)^2P^2\tilde{H}_1\tilde{H}_2 < 1+P\tilde{H}_2+\alpha_{\rm A}^*(1-\alpha_{\rm A}^*)P^2\tilde{H}_2^2Q, \tag{D.2}$$

which results in

$$\alpha_{\rm A}^* < \alpha_U^* \triangleq 2\tilde{H}_2 / \left[\tilde{H}_1+\tilde{H}_2 - P\tilde{H}_2^2Q+\sqrt{(\tilde{H}_1+\tilde{H}_2-P\tilde{H}_2^2Q)^2+4\tilde{H}_2(P\tilde{H}_1\tilde{H}_2+P\tilde{H}_2^2Q)}\right]. \tag{D.3}$$

For the lower bound, as $[1+P\tilde{H}_2+\alpha_{\rm A}^*(1-\alpha_{\rm A}^*)P^2\tilde{H}_2^2Q]^2-(\alpha_{\rm A}^*)^2(1-\alpha_{\rm A}^*)^2P^4\tilde{H}_2^4Q^2 > [1+P\tilde{H}_2]^2$,

$$2[1+\alpha_{\rm A}^*P(\tilde{H}_1+\tilde{H}_2)+(\alpha_{\rm A}^*)^2P^2\tilde{H}_1\tilde{H}_2] > 2(1+P\tilde{H}_2)+\alpha_{\rm A}^*(1-\alpha_{\rm A}^*)P^2\tilde{H}_2^2Q. \tag{D.4}$$

Then, we obtain

$$\alpha_{\rm A}^* > \alpha_L^* \triangleq \frac{2\tilde{H}_2}{\tilde{H}_1+\tilde{H}_2-\frac{P\tilde{H}_2^2Q}{2}+\sqrt{\left(\tilde{H}_1+\tilde{H}_2-\frac{P\tilde{H}_2^2Q}{2}\right)^2+4\tilde{H}_2\left(P\tilde{H}_1\tilde{H}_2+\frac{P\tilde{H}_2^2Q}{2}\right)}}. \tag{D.5}$$

If $\tilde{H}_1 < \tilde{H}_2$, i.e., User 2 employs SIC, we can also derive the expressions for α_L^* and α_U^* by setting $R_{\rm strong}(\tilde{H}_2) = R_{\rm weak}(\tilde{H}_1)$. In fact, α_L^* and α_U^* for $\tilde{H}_1 < \tilde{H}_2$ are given by simply switching $\tilde{H}_1$ and $\tilde{H}_2$ in (D.3) and (D.5). Both α_L^* and α_U^* can be incorporated into a general expression in (6.7) by introducing a parameter x, i.e. $\alpha_L^* = \alpha_{\rm A}^*(0.5)$ and $\alpha_U^* = \alpha_{\rm A}^*(1)$.

To show the inequality in (6.6), let us define

$$g(x) = \sqrt{(\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q)^2+4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2+xP\tilde{H}_{\min}^2Q)}+\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q$$

which is the denominator in (6.7). Then,

$$\frac{\partial g(x)}{\partial x} = -P\tilde{H}_{\min}^2 Q \left(1 + \frac{\tilde{H}_1 + \tilde{H}_2 - \tilde{H}_{\min}(2 + xP\tilde{H}_{\min}Q)}{\sqrt{(\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q)^2+4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2+xP\tilde{H}_{\min}^2Q)}}\right).$$

If $\tilde{H}_1 + \tilde{H}_2 - \tilde{H}_{\min}(2 + P\tilde{H}_{\min}Qx) > 0$, it is obvious that $\frac{\partial g(x)}{\partial x} < 0$. Otherwise,

$$\begin{aligned}
&-\tilde{H}_1 - \tilde{H}_2 + \tilde{H}_{\min} < 0 < P\tilde{H}_1\tilde{H}_2 \\
\Longrightarrow &4\tilde{H}_{\min}(-\tilde{H}_1 - \tilde{H}_2 + \tilde{H}_{\min} + xP\tilde{H}_{\min}^2Q) < 4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2 + xP\tilde{H}_{\min}^2Q) \\
\Longrightarrow &\Big[\tilde{H}_1+\tilde{H}_2-\tilde{H}_{\min}(2+xP\tilde{H}_{\min}Q)\Big]^2-(\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q)^2 \\
&< 4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2 + xP\tilde{H}_{\min}^2Q) \\
\Longrightarrow &|\tilde{H}_1 + \tilde{H}_2 - \tilde{H}_{\min}(2 + xP\tilde{H}_{\min}Q)| \\
&<\sqrt{(\tilde{H}_1+\tilde{H}_2-xP\tilde{H}_{\min}^2Q)^2+4P\tilde{H}_{\min}(\tilde{H}_1\tilde{H}_2+x\tilde{H}_{\min}^2Q)},
\end{aligned}$$

which also results in $\frac{\partial g(x)}{\partial x} < 0$. Therefore, $\alpha_{\mathrm{A}}^*(x)$ increases with x. Thus, $\alpha_{\mathrm{N}}^* = \alpha_{\mathrm{A}}^*(0) < \alpha_{\mathrm{A}}^*(0.5) < \alpha_{\mathrm{A}}^* < \alpha_{\mathrm{A}}^*(1)$. The proof is complete. □