



저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

Master's Thesis

Optimizing Communication Beamforming for New
Multiple Access under Low-Resolution Quantization:
A Spectral and Energy Efficiency Perspective

Seokjun Park

Department of Electrical Engineering

Ulsan National Institute of Science and Technology

2023

Optimizing Communication Beamforming for New
Multiple Access under Low-Resolution Quantization:
A Spectral and Energy Efficiency Perspective

Seokjun Park

Department of Electrical Engineering

Ulsan National Institute of Science and Technology

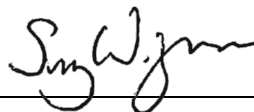
Optimizing Communication Beamforming for New
Multiple Access under Low-Resolution Quantization:
A Spectral and Energy Efficiency Perspective

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Master of Science

Seokjun Park

05.23.2023 of submission

Approved by



Advisor Sung Whan Yoon

Optimizing Communication Beamforming for New
Multiple Access under Low-Resolution Quantization:
A Spectral and Energy Efficiency Perspective

Seokjun Park

This certifies that the thesis/dissertation of Seokjun Park is approved.

05.23.2023 of submission

Signature



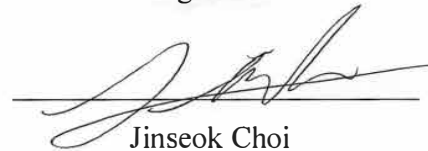
Advisor: Sung Whan Yoon

Signature



Hyoil Kim

Signature



Jinseok Choi

Abstract

Currently, there is growing interest in 6G wireless communication beyond the era of 5G. In addition, the hardware devices require high-speed wireless communication and low-power communications. For example, there are applications such as the internet-of-things (IoT), where devices are limited by battery capacity and have low computing capabilities but require high spectral efficiency. In order to address the issue of power consumption in wireless communication, low-power hardware such as low-resolution analog-to-digital converter (ADC) and digital-to-analog converter (DAC) systems are having attention as a promising transceiver architecture. This is because the power consumption of quantizers decreases exponentially as the number of quantization bits decreases. In this dissertation, low-resolution quantizer system is considered to achieve the trade-off between high spectral efficiency and energy efficiency. Another challenge that needs to be addressed in the development of 6G wireless communications is the severe inter-user interference resulting from the exponential increase in the number of smart devices. For example, in IoT communications, the large number of IoT devices and high channel correlation among them can lead to a significant amount of inter-user interference, which in turn can cause considerable degradation in spectral performance. In this regard, new multiple access approaches are introduced such as rate-splitting multiple access (RSMA), non-orthogonal multiple access (NOMA), spatial-division multiple access (SDMA), and orthogonal multiple access (OMA) to control the inter-user interference. Specifically, I consider rate-splitting multiple access to boost the spectral efficiency because rate-splitting multiple access provides extra achievable antenna degree-of freedom by dividing the messages into common and private messages. It is difficult to optimize rate-splitting multiple access precoders due to the minimum rate constraint involved in determining the common rate. Furthermore, the designing quantized precoders is more highly challenging to solve the optimization problem. In this dissertation, I develop a promising RSMA precoder algorithm coupled with quantization errors to maximize the spectral efficiency. To make the optimization problem in smooth function, I first approximate the spectral efficiency of common stream utilizing the Log-Sum Exp technique. Then, I derive the first-order optimality condition in terms of the nonlinear eigenvalue problem (NEP). I suggest computationally efficient method to find a sub-optimal solution for obtaining the principal eigen-vector of the nonlinear eigenvalue problem. In addition, I propose the weighted minimum mean square error-based RSMA precoding algorithm to the considered quantization system. Simulation results demonstrate the performance of the proposed algorithm in terms of the spectral efficiency, and more importantly, rate-splitting multiple access can achieve key benefit than spatial-division multiple access by balancing between the channel gain and quantization error utilizing the common stream in multiuser MIMO systems.

Contents

I	Introduction	1
1.1	Background	1
1.2	Wireless Communication Systems	1
1.3	Low-resolution Communication Systems	2
1.4	Rate-Splitting Multiple Access Communication Systems	3
1.5	Motivation	4
II	Contribution	5
III	System Model	7
IV	Performance Metrics and Problem Formulation	11
V	Precoder Optimization Using Generalized Power Iteration	13
5.1	Problem Reformulation	13
5.2	First-Order Optimality Condition	15
5.3	Quantized Generalized Power Iteration for Rate-Splitting	16
5.4	Extension of Existing Approach: WMMSE Approach	17
VI	Numerical Results	20
6.1	Homogeneous Quantization Resolution	21
6.2	Heterogeneous Quantization Resolution	22

6.3	Effect of DAC and ADC Quantization to RSMA	26
6.4	Convergence Analysis	28
VII	Proof of Lemma 1	29
VIII	Conclusion	30
8.1	Future work	30
IX	Vita	32
	References	33

List of Figures

1	A low-resolution quantization regime rate-splitting multiple access downlink system with multiuser MIMO.	7
2	The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{DAC},n} = 4$ DAC bits, $\forall n$, and $b_{\text{ADC},k} = 6$ ADC bits, $\forall k$	21
3	The sum spectral efficiency versus angular difference of users for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{DAC},n} = 4$ DAC bits, $\forall n$, $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$, and SNR= 50 dB.	22
4	The sum spectral efficiency versus SNR for $N = 6$ AP antennas, $K = 4$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. The numbers of DACs are set uniformly from 2 to 8 bits.	23
5	The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.	24
6	The spectral efficiency of the common and private streams versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.	24
7	The average antenna power ratios of Q-GPI-RS and Q-GPI-SEM versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.	25
8	The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 4$ users, $b_{\text{DAC},n} = 10$ DAC bits, $\forall n$, and $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$ and for $N = 8$ AP antennas, $K = 4$ users, and $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$. We consider the half of the DACs are equipped with 3 bits and the other half are equipped with 10 bits.	26
9	The sum spectral efficiency versus the number of DAC bits versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and SNR = 40 dB with $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$	27

- 10 The sum spectral efficiency versus the number of ADC bits versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and SNR = 40 dB with $b_{\text{DAC},n} = 10$ DAC bits, $\forall n$ 27
- 11 The power ratio versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and SNR = 40 dB. 28
- 12 Convergence results in terms of the sum spectral efficiency for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$, and SNR $\in \{-10, 0, 10, 20\}$ dB transmit power with mixed-resolution DACs. For the mixed-resolution DAC case, we consider that a single DAC is an 8-bit DAC and the rest are 3-bit DACs. 28

List of Acronyms

I Introduction

In this chapter of this dissertation, I briefly overview the background and motivation for this dissertation. First I represent the background of research for low-resolution quantizer beamforming architectures. Then, I introduce about the prior work of rate-splitting multiple access. In addition, the motivation and contribution of the proposed algorithm are represented in the subsection.

Currently, 6G wireless communication has attracted considerable attention, surpassing the 5G era [1]. Consequently, wireless communication systems that are low-power yet high-speed have become more critical, especially for applications like the Internet-of-Things (IoT), where devices are typically constrained by limited battery life and low computing capability but require high spectral efficiency [2]. To address the power consumption issue, low-power hardware, such as low-resolution analog-to-digital converters (ADCs) and digital-to-analog converters (DACs), can be utilized since the power consumption of quantizers reduces exponentially with the decrease in quantization bits [3]. The potential for power-saving by using low-resolution quantizers has led to extensive research into low-resolution quantization systems or mixed-resolution quantization systems, where high-resolution and low-resolution quantizers coexist [4–6].

One of the major challenges in realizing 6G wireless communications is the significant inter-user interference caused by the substantial increase in the number of smart devices. For instance, in IoT communications, the growing number of IoT devices and the high channel correlation among them [2] can result in significant inter-user interference, leading to a considerable decrease in spectral efficiency. To address this issue, rate-splitting multiple access (RSMA) was introduced in [7] as a unified multiple access scheme for downlink multi-antenna wireless networks, which can effectively overcome the limitations of the spectral efficiency gain in multiuser multiple-input multiple-output (MU-MIMO) systems by reducing the inter-user interference [8–11]. In this dissertation, we investigate the use of RSMA for downlink MU-MIMO systems with low-resolution quantizers and propose a novel and computationally efficient precoding method to maximize spectral efficiency.

1.1 Background

1.2 Wireless Communication Systems

Cellular networks consist of a large number of users who use cellular devices such as mobile phones and tablets, and a large number of base stations (BSs) that are fixed and arranged to provide coverage to the

This chapter is based on the work accepted to IEEE Transactions on Wireless Communications in the journal paper: S. Park, J. Choi, J. Park, W. Shin, and B. Clerckx, “Rate-Splitting Multiple Access for Quantized Multiuser MIMO Communications,” IEEE Trans. Wireless Commun., 2022. The author of this work acknowledges the valuable feedback and contributions provided by Prof. Jinseok Choi, Prof. Jeonghun Park, Prof. Wonjae Shin, and Prof. Bruno.Clerckx who served as supervisor and collaborator, respectively, and helped to improve the quality of this paper.

users. The physical area covered by a BS is referred to as a cell. Each mobile user in a cell is connected to an associated BS.

Wireless communications are subject to two key impairments, fading and interference, which make the problem more challenging than wireline communications. Fading is the time variation of channel strengths that is caused by the small-scale effect of multi-path fading and the large-scale effects known as path loss and shadowing. Interference is generated by other signals and can cause inter-symbol interference when different delays on multiple paths from the transmitter to the receiver cause interference at the receiver for subsequent transmissions. Interference between users communicating with the BS on the same time and frequency resource is called inter-user interference.

In a multi-cell environment, incoming signals from other cells can interfere with the co-channel signals of the associated cell, known as inter-cell interference. Addressing such interference is crucial to the design of wireless communications. When the channel is in deep fade, i.e., the channel strength is very low, it is almost impossible to achieve reliable communications. To overcome this issue, diversity techniques have been developed.

Diversity can be obtained over time via coding and interleaving, over frequency when the channel is frequency selective, and over space when multiple antennas are spaced sufficiently at the transmitter and/or receivers. Numerous interference mitigation techniques have been created in addition to diversity techniques to manage various types of interference. Linear equalizers, such as maximum ratio combining, zero-forcing combining, and minimum mean squared error combining, and nonlinear equalizers are widely utilized and can be implemented over time, frequency, and space. Multiple access techniques like code-division multiple access and orthogonal frequency division multiple access have also been developed to cater to multiple users without causing interference between them. Cell sectorization is another technique utilized to minimize interference among co-channel cells. This method spatially divides each cell by using directional antennas at the BS, which considerably decreases interference without requiring the acquisition of new BS sites.

Using multiple antennas not only provides diversity gain but also power gain when the receiver has multiple antennas or the transmitter knows the channel state information. A multiple-input multiple-output (MIMO) system, which has both multiple transmit and receive antennas, offers a new way to use multiple antennas. MIMO systems add an extra spatial dimension and provide a degree-of-freedom gain, which can be exploited by spatially multiplexing multiple data streams onto the MIMO channel [12]. This results in an increase in channel capacity proportional to the degree-of-freedom. MIMO techniques have become the primary tool in wireless communication for increasing both capacity and reliability.

1.3 Low-resolution Communication Systems

The number of bits used in DACs and ADCs affects the quality of the received signal, and their performance has been a topic of extensive research. The information-theoretic performance limits when using low-resolution DACs and ADCs have been widely studied in the literature. For instance, the capacity of uniformly distributed quadrature phase shift keying (QPSK) was achieved in a quantized

single-input single-output (SISO) additive white Gaussian noise (AWGN) channel [13]. In addition, the capacity of multiple-input single-output (MISO) fading channel with one-bit ADCs was derived in a closed form [14]. To better understand the analytical performance of low-resolution quantization systems, linear approximations of the quantization process have been employed, including the Bussgang decomposition [15] and additive quantization noise model (AQNM) [16]. In [15], the lower bound on the achievable rate of the quantized MIMO channel was derived based on the Bussgang decomposition. Additionally, optimal bandwidth and resolution of ADC for the SISO channel were analyzed employing the AQNM [16]. These studies provide insights into the performance limits of low-resolution DACs and ADCs in wireless communication systems, and can be used to optimize their design for specific applications.

Existing conventional precoding methods for perfect quantization systems have limited spectral efficiency due to non-negligible quantization errors that are not adequately accounted for [17, 18]. Accordingly, several approaches have been proposed to overcome the challenges posed by low-resolution DACs in massive MU-MIMO systems. For example, in [17], precoders based on conventional methods such as minimum mean square error (MMSE) and zero-forcing (ZF) were proposed for 3 to 4-bit DACs, which achieved performance comparable to that of high-resolution DAC systems. The inter-user interference minimization problem was addressed in [19] using the alternating direction method of multipliers (ADMM). A generalized power iteration-based algorithm was proposed in [18] for designing the precoder that maximizes energy efficiency in downlink MU-MIMO systems with heterogeneous-resolution DACs and ADCs. Furthermore, mixed DACs and ADCs architectures, which are special cases of heterogeneous-resolution DACs and ADCs, were investigated in [4, 20], revealing the potential for increased spectral efficiency compared to homogeneous-resolution DACs and ADCs, where the bits of DACs (and ADCs) are equally distributed.

The feasibility of one-bit quantization systems in developing advanced channel estimation and detection techniques has been investigated due to their practicality and analytical tractability [21–24]. A maximum likelihood detector [22] and Bussgang decomposition [22] have been employed for developing channel estimation and detection techniques for one-bit quantization systems. In addition, learning-based detection methods that do not require explicit channel estimation have been proposed [23, 24]. Although prior precoding methods for low-resolution quantization systems have resulted in notable improvement in spectral efficiency, their application is limited to conventional signaling methods, which may not be optimal for systems with significant inter-user interference. To address this issue, I propose an advanced signaling method for low-resolution quantization systems that can enhance spectral efficiency by managing inter-user interference and reducing power consumption at transceivers.

1.4 Rate-Splitting Multiple Access Communication Systems

In [25], RSMA was introduced as a theoretically optimal method for achieving channel degree-of-freedom with imperfect channel state information (CSI) by reducing interference. RSMA is a generalized form of the Han-Kobayashi scheme [26], which ensures an achievable rate of one bit per second per

hertz of the capacity in a Gaussian interference channel model [27]. The primary concept behind RSMA involves dividing each user's stream into a common stream and a private stream. The common stream's codebook is shared with all users, enabling them to eliminate the common stream through successive interference cancellation (SIC) when decoding their private streams. Consequently, the private streams experience a reduced amount of interference, leading to an improvement in spectral efficiency [7, 25].

According to [28], an analysis of the achievable sum spectral efficiency in the MISO channel for two receivers was conducted using a randomly generated precoder for the common stream and ZF precoder for the private streams [29]. To maximize the spectral efficiency in RSMA, a linear precoding method was proposed for the downlink MISO system based on weighted MMSE (WMMSE) [30] in [25]. In [25], the problem of maximizing spectral efficiency was transformed into a quadratically constrained quadratic program (QCQP) by minimizing mean square error (MSE). Additionally, the RSMA precoder design was represented through the convex-convex procedure (CCCP) [10] by approximating the optimization problem in convex form. Moreover, to maximize the sum rate, rate allocation and power control optimization for RSMA were studied in [31]. A hierarchical RSMA architecture was proposed for the downlink massive MIMO system in which the decodability of more than one common stream is dependent on the hierarchy [9]. In [7], a generalized RSMA framework was introduced to connect, generalize, and outperform existing multiple access techniques like orthogonal multiple access (OMA), non-orthogonal multiple access (NOMA), spatial-division multiple access (SDMA), and multiuser MIMO. The optimal rate allocation and power control algorithm for the common and private streams were proposed in [32]. Additionally, in [33], an RSMA precoder design algorithm based on the generalized power iteration (GPI) algorithm was presented to maximize the sum spectral efficiency in the downlink MU-MIMO system, considering channel estimation errors.

RSMA has gained attention as a promising technology in various applications [34, 35]. Its performance in low-resolution quantization systems with regularized zero-forcing (RZF) was analyzed in [36]. In [37], a precoder design algorithm based on the ADMM was proposed for a joint radar-communication (JRC) system with multiple antennas in RSMA, considering low-resolution DACs. Additionally, an ADMM-based algorithm for energy-efficient dual-functional radar communication in RSMA was proposed in [38], considering low-resolution DACs and RF chain selection.

1.5 Motivation

It is worth noting that while previous studies have demonstrated the advantages of RSMA and proposed advanced transmission techniques, a comprehensive investigation on the precoding design problem for downlink RSMA, taking into account the number of DAC and ADC bits, is still lacking. Although other state-of-the-art methods in [37, 38] have analyzed the spectral and energy efficiencies of RSMA with low-resolution DACs, their results are limited to systems with homogeneous DACs at the transmitter and perfect ADCs at the receiver. Moreover, the proposed methods in [37, 38] are optimized for joint radar and communication systems, which may not be optimal for communication-only scenarios. Furthermore, computing the spectral efficiency in RSMA requires considering the minimum rate of

the common stream, which is a non-smooth function, and existing optimization methods based on the CVX toolbox have high computational complexity. Therefore, it is necessary to develop a computationally efficient precoding method with improved performance to maximize the sum spectral efficiency in downlink RSMA systems with heterogeneous DACs and ADCs.

II Contribution

In this dissertation, a novel precoding optimization framework is proposed, and the impact of RSMA on coarse quantization systems is investigated. The contributions of this work can be summarized as follows:

- In my dissertation, I consider a downlink RSMA system where a multi-antenna access point (AP) transmits a common stream and private streams to single-antenna users. The AP is equipped with low-resolution DACs, which means that the digital-to-analog conversion process is subject to quantization error. We allow for arbitrary resolutions of the DACs, which means that the number of bits used for each DAC may be different. This is in contrast to previous work, such as [25, 37, 38], where either no quantization error or only DAC quantization error was considered, and the DACs were assumed to have the same resolution. In addition to considering heterogeneous DACs, we also consider low-resolution ADCs for the RSMA users. Again, we allow for arbitrary resolutions of the ADCs, which means that the number of bits used for each ADC may be different. This makes the problem more complicated and challenging than previous work that only considered perfect or high-resolution ADCs. To maximize the sum spectral efficiency of the system, we formulate an optimization problem that takes into account the common stream that needs to be decodable by all users. Specifically, the common rate is the minimum rate that all users can decode the stream, which makes the problem a non-smooth optimization problem. I aim to find a linear precoding scheme that maximizes the sum spectral efficiency while satisfying power constraints and accounting for the quantization errors of the DACs and ADCs. Overall, my dissertation work aims to provide a more realistic and comprehensive analysis of RSMA in the presence of low-resolution DACs and ADCs with arbitrary resolutions. By considering heterogeneous quantization at both the transmitter and receiver, we hope to provide insights that can inform the design of practical RSMA systems that operate in real-world scenarios.
- In this dissertation, I propose a novel and computationally efficient precoding method, which is called the Quantized Generalized Power Iteration for Rate-Splitting (Q-GPI-RS) algorithm. This method is designed to solve the sum spectral efficiency maximization problem for downlink RSMA systems with heterogeneous DACs and ADCs. To address the non-smoothness of the minimum rate function, I first utilize the LogSumExp approximation technique, which converts the function into a tractable form. This technique helps to reduce the algorithm's complexity and avoids introducing inequality constraints for each user's common stream, which is often done in previous studies [25, 37, 38]. Furthermore, we establish the first-order optimality condition for the

sum spectral efficiency maximization problem, which is used to identify stationary points. We interpret this condition as a nonlinear eigenvalue problem (NEP) [39]. The eigenvalue and eigenvector of the NEP correspond to the sum spectral efficiency and the precoding vector, respectively. Hence, finding the leading eigenvector is equivalent to finding the best local optimal point. Based on this insight, we propose the Q-GPI-RS algorithm, which leverages the efficiency of power iteration to seek the principal eigenvector. We also extend the Weighted Minimum Mean Square Error (WMMSE) based RSMA precoding method [25] to the considered system for comprehensiveness. This extension is called Quantization-aware WMMSE-based Alternating Optimization (Q-WMMSE-AO). Overall, the proposed Q-GPI-RS algorithm and Q-WMMSE-AO extension offer a computationally efficient and effective approach to solving the sum spectral efficiency maximization problem for downlink RSMA systems with heterogeneous DACs and ADCs.

- The simulation results indicate that Q-GPI-RS outperforms the conventional linear precoding method, such as RZF, in terms of spectral efficiency. Specifically, in various scenarios, Q-GPI-RS achieves a significant gain in spectral efficiency compared to other baselines. Furthermore, Q-WMMSE-AO improves the spectral efficiency and exhibits higher robustness compared to the conventional WMMSE-AO approach, as demonstrated in the simulation results [25]. Moreover, Q-GPI-RS outperforms Q-WMMSE-AO in most environments, providing higher spectral efficiency gains. Moreover, the simulation results show that RSMA outperforms SDMA in the medium to high signal-to-noise ratio (SNR) regime where the performance is limited by inter-user interference and quantization error. This indicates that RSMA offers a noticeable spectral efficiency gain over SDMA. The proposed methods' effectiveness and the benefit of RSMA over SDMA are thus validated by the simulation results.
- The numerical results of our study reveal several key findings. Firstly, we confirm that RSMA can significantly enhance the spectral efficiency, particularly in channels with high correlation, even in the presence of quantization errors from both DACs and ADCs. Secondly, we find that the spectral efficiency gain of RSMA increases as the resolution of DACs and ADCs increases, because the quantization errors associated with the common stream cannot be canceled out by the users. Notably, we also observe that the effect of ADC is more dominant than that of DAC in determining the rate of the common stream, demonstrating the need to consider both types of quantization errors in designing transmission strategies. Finally, in mixed-resolution DAC systems, antennas with low-resolution DACs tend to be deactivated in high SNR scenarios to minimize the quantization error for RSMA, a phenomenon not observed in SDMA. This phenomenon causes overload due to insufficient active antennas, emphasizing the importance of leveraging RSMA to maximize sum spectral efficiency thanks to the common stream.

Notation: a is a scalar, \mathbf{a} is a vector and \mathbf{A} is a matrix. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote matrix transpose, Hermitian, and inversion, respectively. $\mathbb{E}[\cdot]$ and $\text{tr}(\cdot)$ represent expectation operation and trace of a matrix, respectively. \mathbf{I}_K is the identity matrix of size $K \times K$. $\mathbf{0}_N$ is the zero matrix of size

$N \times N$ and $\mathbf{0}_{N \times 1}$ is the zero vector of size $N \times 1$. $\mathbf{A} = \text{blkdiag}(\mathbf{A}_1, \dots, \mathbf{A}_n, \dots, \mathbf{A}_N)$ is a block diagonal matrix with block diagonal entries of $\mathbf{A}_1, \dots, \mathbf{A}_N$.

III System Model

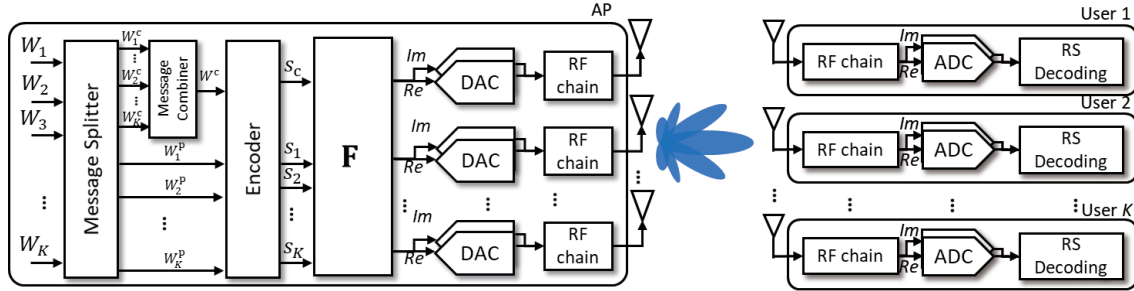


Figure 1: A low-resolution quantization regime rate-splitting multiple access downlink system with multiuser MIMO.

We consider a multiuser multiple-input multiple-output (MU-MIMO) downlink system, where K single-antenna users are served by an access point (AP) equipped with N antennas, as illustrated in Fig. 1. The DACs at the AP are characterized by their respective resolutions, denoted by $b_{\text{DAC},n}$ for the n -th antenna. Similarly, each user is assumed to be equipped with an ADC pair, whose quantization resolution is denoted by $b_{\text{ADC},k}$ for the k -th user.

In our system, we adopt RSMA transmission, utilizing a 1-layer rate-splitting (RS) architecture as depicted in Fig. 1. The RSMA approach splits an individual message W_k into two parts: a common part W_k^c and a private part W_k^p . The common parts are then combined and jointly encoded to generate a common message $W^c = (W_1^c, \dots, W_K^c)$. This common message is then encoded into a common stream s_c using a public codebook, which is intended to be decoded by all users in the network. The private message W_k^p is encoded to be a private stream s_k via a private codebook, allowing it to be decoded only by the corresponding user. By receiving a signal, a user first decodes the common stream s_c , eliminates it through successive interference cancellation (SIC), and then decodes the corresponding private stream s_k . It is important to note that once the common stream s_c is decoded, user k can recover its common message part W_k^c from the common message W^c based on the approach proposed in [11, 25] for RSMA. In addition, the received signal is assumed by Gaussian signaling $s_c, s_k \sim \mathcal{CN}(0, 1)$, i.e., a complex Gaussian distribution with zero mean and unit variance.

The common stream and private streams are linearly precoded at the AP. The digital baseband signal $\mathbf{x} \in \mathbb{C}^N$ is

$$\mathbf{x} = \sqrt{P} \mathbf{f}_0 s_c + \sqrt{P} \sum_{k=1}^K \mathbf{f}_k s_k, \quad (1)$$

where $\mathbf{f}_0 \in \mathbb{C}^N$ and $\mathbf{f}_k \in \mathbb{C}^N$ are precoding vectors for the common and private streams, respectively, and P is the maximum transmit power. Since \mathbf{x} is quantized at the DACs, we adopt the AQNM method [40]

to convert the quantization process approximately in the linear form. Then, applying the AQNM, the quantized signal is represented as

$$Q(\mathbf{x}) \approx \mathbf{x}_q = \sqrt{P}\Phi_{\alpha_{\text{DAC}}}\mathbf{f}_0s_c + \sqrt{P}\Phi_{\alpha_{\text{DAC}}}\sum_{k=1}^K \mathbf{f}_k s_k + \mathbf{q}_{\text{DAC}}, \quad (2)$$

where $Q(\cdot)$ is a scalar quantizer which adopts for each real and imaginary part, $\Phi_{\alpha_{\text{DAC}}} = \text{diag}(\alpha_{\text{DAC},1}, \dots, \alpha_{\text{DAC},N}) \in \mathbb{C}^{N \times N}$ denotes a diagonal matrix of quantization loss, $\Phi_{\beta_{\text{DAC}}} = \text{diag}(\beta_{\text{DAC},1}, \dots, \beta_{\text{DAC},N}) \in \mathbb{C}^{N \times N}$, and $\mathbf{q}_{\text{DAC}} \in \mathbb{C}^N$ is a DAC quantization noise vector. The quantization loss of the n -th DAC $\alpha_{\text{DAC},n} \in (0, 1)$ is defined as $\alpha_{\text{DAC},n} = 1 - \beta_{\text{DAC},n}$, where $\beta_{\text{DAC},n}$ is a normalized mean squared quantization error $\beta_{\text{DAC},n} = \frac{\mathbb{E}[|x - Q_n(x)|^2]}{\mathbb{E}[|x|^2]}$ [5,40]. The values of $\beta_{\text{DAC},n}$ vary according to the number of quantization bits $b_{\text{DAC},n}$. In particular, $\beta_{\text{DAC},n}$ is represented in Table 1 in [41] when the value of $b_{\text{DAC},n}$ is less than 5 and if the value of $b_{\text{DAC},n}$ is larger than 5, $\beta_{\text{DAC},n}$ can be approximated as $\frac{\pi\sqrt{3}}{2}2^{-2b_{\text{DAC},n}}$. The quantization noise is uncorrelated with digital baseband signal \mathbf{x} and follows as $\mathbf{q}_{\text{DAC}} \sim \mathcal{CN}(\mathbf{0}_{N \times 1}, \mathbf{R}_{\mathbf{q}_{\text{DAC}}\mathbf{q}_{\text{DAC}}})$ which is the worst case in terms of the spectral efficiency. Let $\mathbf{F} = [\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_K] \in \mathbb{C}^{N \times (K+1)}$. Then, the covariance matrix of \mathbf{q}_{DAC} is [40]

$$\mathbf{R}_{\mathbf{q}_{\text{DAC}}\mathbf{q}_{\text{DAC}}} = \Phi_{\alpha_{\text{DAC}}}\Phi_{\beta_{\text{DAC}}}\text{diag}\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]\right). \quad (3)$$

The received analog baseband signal vector at K users is

$$\mathbf{y} = \mathbf{H}^H \mathbf{x}_q + \mathbf{n}, \quad (4)$$

where $\mathbf{H}^H \in \mathbb{C}^{K \times N}$ is a downlink channel matrix between the AP and K users, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{K \times 1}, \sigma^2 \mathbf{I}_K)$ is an AWGN with zero mean and variance of σ^2 . We assume that the perfect channel state information (CSI) is obtained at the AP and users. The ADCs with $b_{\text{ADC},k}$ bits quantize the received analog signal y_k at each user. Subsequently, the received digital baseband signals are expressed as [40]:

$$\begin{aligned} Q(\mathbf{y}) &\approx \mathbf{y}_q \\ &= \Phi_{\alpha_{\text{ADC}}}\mathbf{y} + \mathbf{q}_{\text{ADC}} \\ &= \sqrt{P}\Phi_{\alpha_{\text{ADC}}}\mathbf{H}^H\Phi_{\alpha_{\text{DAC}}}\mathbf{f}_0s_c + \sqrt{P}\Phi_{\alpha_{\text{ADC}}}\mathbf{H}^H\Phi_{\alpha_{\text{DAC}}}\sum_{k=1}^K \mathbf{f}_k s_k + \Phi_{\alpha_{\text{ADC}}}\mathbf{H}^H\mathbf{q}_{\text{DAC}} + \Phi_{\alpha_{\text{ADC}}}\mathbf{n} + \mathbf{q}_{\text{ADC}}, \end{aligned} \quad (5)$$

where $\Phi_{\alpha_{\text{ADC}}} = \text{diag}(\alpha_{\text{ADC},1}, \dots, \alpha_{\text{ADC},K}) \in \mathbb{C}^{K \times K}$ consists of a diagonal matrix of quantization loss, and $\mathbf{q}_{\text{ADC}} \in \mathbb{C}^K$ denotes quantization noise vector of ADC. The ADC quantization loss of the k -th ADC $\alpha_{\text{ADC},k}$ and $\beta_{\text{ADC},k}$ is formulated similarly with DAC quantization loss. The ADC quantization noise \mathbf{q}_{ADC} is assumed to be uncorrelated with the analog baseband signal \mathbf{y} . It is considered to follow a complex Gaussian distribution with zero mean and a covariance matrix $\mathbf{q}_{\text{ADC}} \sim \mathcal{CN}(\mathbf{0}_{K \times 1}, \mathbf{R}_{\mathbf{q}_{\text{ADC}}\mathbf{q}_{\text{ADC}}})$. And then, this worst-case scenario is assumed for evaluating the spectral efficiency. Accordingly, we evaluate the covariance of ADC as

$$\mathbf{R}_{\mathbf{q}_{\text{ADC}}\mathbf{q}_{\text{ADC}}} = \Phi_{\alpha_{\text{ADC}}}\Phi_{\beta_{\text{ADC}}}\text{diag}\left(\mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]\right). \quad (6)$$

The digital baseband signal at user k is formulated as

$$y_{q,k} = \sqrt{P}\alpha_{\text{ADC},k}\mathbf{h}_k^H\Phi_{\alpha_{\text{DAC}}}\mathbf{f}_0s_c + \sqrt{P}\alpha_{\text{ADC},k}\mathbf{h}_k^H\Phi_{\alpha_{\text{DAC}}}\mathbf{f}_ks_k + \sqrt{P}\alpha_{\text{ADC},k}\sum_{i=1,i\neq k}^K\mathbf{h}_k^H\Phi_{\alpha_{\text{DAC}}}\mathbf{f}_is_i + \alpha_{\text{ADC},k}\mathbf{h}_k^H\mathbf{q}_{\text{DAC}} + \alpha_{\text{ADC},k}n_k + q_{\text{ADC},k}, \quad (7)$$

where \mathbf{h}_k refers to the k -th column of the channel matrix \mathbf{H} . We note that the quantization errors, namely $\alpha_{\text{ADC},k}\mathbf{h}_k^H\mathbf{q}_{\text{DAC}}$ and $q_{\text{ADC},k}$, appearing in (7), are associated with both the common stream and the private streams along with their respective precoders. This implies that achieving the potential of RSMA in low-resolution quantization systems requires proper design of the precoder for the common stream. In other words, it is more challenging to achieve the desired performance as compared to the systems with low-resolution or perfect quantization of DACs.

[AQNM Computation] Before determining the quantization loss, let us first explain the derivation of the additive quantization noise model (AQNM). Under the assumptions of scalar MMSE quantization and a zero-mean quantization input, i.e., $\mathbb{E}[y|\mathcal{Q}(y)] = \mathcal{Q}(y) = y_q$ and $\mathbb{E}[y] = 0$, Lemma 1 states the relationship between a *quantization input* y and *quantization output* y_q regardless of how the *quantization input* y is composed of as

$$y_q = \alpha y + \sigma_q \omega_q$$

where $\alpha = 1 - \beta$ is the quantization loss, $\sigma_q^2 = \alpha(1 - \alpha)\mathbb{E}[|y|^2]$, $\omega_q \sim \mathcal{CN}(0, 1)$, and $\beta = \mathbb{E}[|y - y_q|^2]/\mathbb{E}[|y|^2]$.

In addition, the proof of Lemma 1 with the quantization input y and output y_q can be shown by following the same proof as follows: assuming $\mathbb{E}[y|\mathcal{Q}(y)] = \mathcal{Q}(y) = y_q$, we have $\mathbb{E}[yy_q^*|y_q] = y_q^*\mathbb{E}[y|y_q] = |y_q|^2$, and thus, we also have

$$\mathbb{E}[yy_q^*] = \mathbb{E}[|y_q|^2]. \quad (8)$$

Using $\beta = \mathbb{E}[|y - y_q|^2]/\mathbb{E}[|y|^2]$ and (8), we obtain

$$\beta\mathbb{E}[|y|^2] = \mathbb{E}[|y - y_q|^2] = \mathbb{E}[|y|^2] - 2\mathbb{E}[yy_q^*] + \mathbb{E}[|y_q|^2] = \mathbb{E}[|y|^2] - \mathbb{E}[|y_q|^2].$$

Equivalently, we have

$$\mathbb{E}[|y_q|^2] = (1 - \beta)\mathbb{E}[|y|^2]. \quad (9)$$

Now, let q be the additive quantization noise $q = y_q - (1 - \beta)y$ so that $y_q = (1 - \beta)y + q$. Since y and y_q are zero mean, q is also zero mean. Using (8) and (9), we obtain

$$\mathbb{E}[yq^*] = \mathbb{E}[y(y_q - (1 - \beta)y)^*] = (1 - \beta)\mathbb{E}[|y|^2] - (1 - \beta)\mathbb{E}[|y|^2] = 0$$

which implies that y and q are uncorrelated. Finally, the quantization noise variance becomes

$$\begin{aligned} \mathbb{E}[|q|^2] &= \mathbb{E}[|y_q - (1 - \beta)y|^2] \\ &= \mathbb{E}[|y_q|^2] - 2(1 - \beta)\mathbb{E}[yy_q^*] + (1 - \beta)^2\mathbb{E}[|y|^2] \\ &= ((1 - \beta) - 2(1 - \beta)^2 + (1 - \beta)^2)\mathbb{E}[|y|^2] \\ &= \beta(1 - \beta)\mathbb{E}[|y|^2]. \end{aligned}$$

Equivalently, with $\alpha = 1 - \beta$, we have

$$\mathbb{E}[|q|^2] = \alpha(1 - \alpha)\mathbb{E}[|y|^2]. \quad (10)$$

[Weight of AQNM] For the optimal scalar quantizer, the optimal quantization distortion $W(b)$ is approximated as

$$W(b) = h\sigma^2 2^{-2b} \quad (11)$$

where $\sigma = \mathbb{E}[|y|^2]$ and the constant h for a zero mean Gaussian quantization input is given by

$$\begin{aligned} h &= \frac{1}{12} \left\{ \int_{-\infty}^{\infty} \left(\frac{e^{-x^2/2}}{\sqrt{2\pi}} \right)^{1/3} dx \right\}^3 \\ &= \frac{\pi\sqrt{3}}{2}. \end{aligned} \quad (12)$$

Since the AQNM assumes the scalar MMSE quantizer which provides optimal centroid condition, using (11) and (12) we have the normalized mean squared quantization error as

$$\beta = \frac{W(b)}{\mathbb{E}[|y|^2]} \approx \frac{\pi\sqrt{3}}{2} 2^{-2b} \quad (13)$$

which is a reasonable approximation in the low to medium resolution cases and an accurate approximation in the high resolution case [R3.2]. Since the values of β for $b \leq 5$ are measured and presented in Table 1 (refer to $N = 2, 4, 8, 16, 32$ in Table 1), the approximation in (13) is used for $b > 5$ and the measured values in the table are used for the value of β for $b \leq 5$.

In our work, we assume heterogeneous-resolution DACs and ADCs and thus, each antenna and each user have different quantization loss. Accordingly, we define the diagonal matrix of quantization loss as $\Phi_{\alpha_{\text{DAC}}} = \text{diag}(\alpha_{\text{DAC},1}, \dots, \alpha_{\text{DAC},N}) \in \mathbb{C}^{N \times N}$. Then, the quantization loss of the n -th DAC $\alpha_{\text{DAC},n} \in (0, 1)$ is determined as $\alpha_{\text{DAC},n} = 1 - \beta_{\text{DAC},n}$, where $\beta_{\text{DAC},n}$ is a normalized mean squared quantization error $\beta_{\text{DAC},n} = \frac{\mathbb{E}[|x - Q_n(x)|^2]}{\mathbb{E}[|x|^2]}$. Based on the approximation in (13), $\beta_{\text{DAC},n}$ can be approximated as $\frac{\pi\sqrt{3}}{2} 2^{-2b_{\text{DAC},n}}$ for $b_{\text{DAC},n} > 5$. In addition, for $b_{\text{DAC},n} \leq 5$ we use the values in the following table. Same applies for ADCs.

Table 1: Values of $\beta_{\text{DAC},n}$ with respect to $b_{\text{DAC},n}$

$b_{\text{DAC},n}$	1	2	3	4	5
$\beta_{\text{DAC},n}$	0.3634	0.1175	0.003454	0.009497	0.002499

IV Performance Metrics and Problem Formulation

The decoding principle of RSMA, as discussed in [25], involves each user decoding the common stream s_c by treating their private streams as noise. Once the common stream is successfully decoded, each user can cancel it from the received signal \mathbf{y} using successive interference cancellation (SIC). By combining message splitting and SIC, the RSMA technique can partially decode interference and treat the remaining interference as noise, leading to an increase in the sum spectral efficiency [7]. However, in order to perform SIC successfully, the common stream must be designed carefully to ensure that it can be decoded by all RSMA users. The rate of the common stream s_c is determined by the minimum of the spectral efficiencies of the common stream for all users. The spectral efficiency of s_c can then be defined as:

$$R_c = \min_{k \in \mathcal{K}} \left\{ \log_2 \left(1 + \frac{P\alpha_{\text{ADC},k}^2 |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0|^2}{\text{IUI}_c + \text{QE}_k + \alpha_{\text{ADC},k}^2 \sigma^2} \right) \right\} = \min_{k \in \mathcal{K}} \{R_{c,k}\}, \quad (14)$$

where

$$\begin{aligned} \text{IUI}_c &= P\alpha_{\text{ADC},k}^2 \sum_{i=1}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2, \\ \text{QE}_k &= \alpha_{\text{ADC},k}^2 \mathbf{h}_k^H \mathbf{R}_{\text{qDACqDAC}} \mathbf{h}_k + r_{\text{qADC},k} \alpha_{\text{ADC},k}. \end{aligned} \quad (15)$$

Once the common stream s_c is decoded and eliminated using SIC, the achievable spectral efficiency of the private stream s_k for user k is formulated as follows:

$$R_k = \log_2 \left(1 + \frac{P\alpha_{\text{ADC},k}^2 |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k|^2}{\text{IUI}_k + \text{QE}_k + \alpha_{\text{ADC},k}^2 \sigma^2} \right). \quad (16)$$

where $\text{IUI}_k = P\alpha_{\text{ADC},k}^2 \sum_{i=1, i \neq k}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2$. We remark that the individual rate of the common stream $R_{c,k}$ does not have an impact on the overall sum rate of RSMA. However, different power allocation between the common and private streams can result in different overall sum rate performance. Therefore, our main objective is to maximize the sum rate of RSMA by jointly optimizing the precoding vectors for the rates of the common stream and private streams. This implies that the proposed precoding scheme can determine the optimal rate division between the common stream and private streams to maximize the sum spectral efficiency.

Remark 1 (SIC and error propagation) It is important to note that during SIC, users can only eliminate the quantized common stream $\sqrt{P}\alpha_{\text{ADC},k} \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0 s_c$ and not the quantization errors that are present in the common stream. To account for this, we consider the worst-case scenario for the additive quantization error by assuming it follows a Gaussian distribution under the AQNM. Consequently, the formulated spectral efficiency of the common stream is conservative. To avoid error propagation during SIC, we assume perfect decoding of the common stream if the derived rate of the common stream is less than or equal to the minimum of the common rates.

We note that the user signals undergo quantization at both the DACs and ADCs, leading to quantization noise error terms that affect the rates of both the common and private streams. In order to achieve maximum sum spectral efficiency, we have formulated the problem as follows

$$\underset{\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_K}{\text{maximize}} \quad R_c + \sum_{k=1}^K R_k = R_\Sigma \tag{17}$$

$$\text{subject to} \quad \text{tr} \left(\mathbb{E} \left[\mathbf{x}_q \mathbf{x}_q^H \right] \right) \leq P, \tag{18}$$

where (18) represents that transmit power constraint. Due to the interdependence between the weight matrix of the quantization losses, precoder, and channel matrix, and the introduction of the common rate with a quantization loss matrix in RSMA transmission, it is important to carefully consider the optimization problem to incorporate the quantization error. Therefore, in the next section, I introduce a promising and computationally efficient with respect to complexity order precoding method to solve (17).

V Precoder Optimization Using Generalized Power Iteration

Due to the non-convexity and non-smoothness of the problem stated in (17), finding a direct solution is highly challenging. Therefore, we propose several techniques to convert the problem into a tractable form that can be solved efficiently.

5.1 Problem Reformulation

We first simplify the constraint in (18). By simplifying the constraint in (18), the DAC covariance matrix of \mathbf{q}_{DAC} in (3) is derived as

$$\mathbf{R}_{\mathbf{q}_{\text{DAC}}\mathbf{q}_{\text{DAC}}} = \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(P \mathbf{f}_0 \mathbf{f}_0^H + P \sum_{i=1}^K \mathbf{f}_i \mathbf{f}_i^H \right) \quad (19)$$

$$= \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(P \mathbf{F} \mathbf{F}^H \right). \quad (20)$$

In addition, the power constraint in (18) is reconstructed as

$$\text{tr} \left(\mathbb{E} \left[\mathbf{x}_q \mathbf{x}_q^H \right] \right) = \text{tr} \left(P \Phi_{\alpha_{\text{DAC}}} \sum_{i=0}^K \mathbf{f}_i \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}}^H + \mathbf{R}_{\mathbf{q}_{\text{DAC}}\mathbf{q}_{\text{DAC}}} \right) \quad (21)$$

$$\stackrel{(a)}{=} \text{tr} \left(P \Phi_{\alpha_{\text{DAC}}} \mathbf{F} \mathbf{F}^H \right) \quad (22)$$

$$\leq P, \quad (23)$$

where (a) comes from (20) and $\Phi_{\beta_{\text{DAC}}} = \mathbf{I}_N - \Phi_{\alpha_{\text{DAC}}}$. Consequently, the transmit power constraint in (18) is reduced to

$$\text{tr} \left(\Phi_{\alpha_{\text{DAC}}} \mathbf{F} \mathbf{F}^H \right) \leq 1. \quad (24)$$

To overcome the non-smoothness of the minimum operation and non-convexity of the spectral efficiency, we propose an approach to approximate the minimum rate of $R_{c,k}$ and reformulate the problem into a tractable form. Accordingly, we utilize a positive constant $\tau > 0$ and utilize the LogSumExp technique to approximate the minimum function, proposed in [42]. Then, the approximated minimum function is

$$\min_{i=1, \dots, N} \{x_i\} \approx -\tau \ln \left(\sum_{i=1}^N \exp \left(-\frac{1}{\tau} x_i \right) \right), \quad (25)$$

where (25) becomes tight as $\tau \rightarrow 0$. Adopting (25) to the rate of common stream in (14), we have

$$\min_{k \in \mathcal{K}} \{R_{c,k}\} \approx -\tau \ln \left(\sum_{k=1}^K \exp \left(-\frac{1}{\tau} R_{c,k} \right) \right). \quad (26)$$

We note that after obtaining precoders using the approximation in (26), we need to compute the actual common rate by using the minimum operation without approximation to determine the achievable spectral efficiency. Even though the approximated function in (26) becomes smooth objective function in (17), the optimization problem still remains non-convexity, coupled with the quantization errors, which

are functions of the precoding vectors, further complicate the problem. To address this issue, we reformulate the expression for QE_k in (15) by rearranging the term related to the covariance of the DAC quantization error as

$$\mathbf{h}_k^H \mathbf{R}_{\mathbf{q}_{\text{DAC}} \mathbf{q}_{\text{DAC}}} \mathbf{h}_k = \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(P \sum_{i=0}^K \mathbf{f}_i \mathbf{f}_i^H \right) \mathbf{h}_k \quad (27)$$

$$= P \sum_{i=0}^K \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(\mathbf{h}_k \mathbf{h}_k^H \right) \mathbf{f}_i, \quad (28)$$

and the quantization error covariance-related term of ADC as

$$\frac{r_{\mathbf{q}_{\text{ADC},k} \mathbf{q}_{\text{ADC},k}}}{\alpha_{\text{ADC},k} \beta_{\text{ADC},k}} \quad (29)$$

$$= \mathbf{h}_k^H \mathbb{E}[\mathbf{x}_q \mathbf{x}_q^H] \mathbf{h}_k + \sigma^2$$

$$= \mathbf{h}_k^H \left(P \Phi_{\alpha_{\text{DAC}}} \sum_{i=0}^K \mathbf{f}_i \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}}^H + \mathbf{R}_{\mathbf{q}_{\text{DAC}} \mathbf{q}_{\text{DAC}}} \right) \mathbf{h}_k + \sigma^2 \quad (30)$$

$$\stackrel{(a)}{=} P \sum_{i=0}^K \mathbf{f}_i^H \left(\Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} + \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(\mathbf{h}_k \mathbf{h}_k^H \right) \right) \mathbf{f}_i + \sigma^2 \quad (31)$$

where $r_{\mathbf{q}_{\text{ADC},k} \mathbf{q}_{\text{ADC},k}}$ presents that the k th diagonal entry of (6) and (a) is from (28). Based on the reformulated term (28) and (31), the SINRs of the common stream of user k is reformulated as

$$\gamma_{c,k} = \frac{\alpha_{\text{ADC},k} |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0|^2}{\sum_{i=0}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2 - \alpha_{\text{ADC},k} |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0|^2 + \sum_{i=0}^K \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(\mathbf{h}_k \mathbf{h}_k^H \right) \mathbf{f}_i + \frac{\sigma^2}{P}}. \quad (32)$$

Similarly, the SINR of the private stream of user k is reorganized as

$$\gamma_k = \frac{\alpha_{\text{ADC},k} |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k|^2}{\sum_{i=0}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2 - \alpha_{\text{ADC},k} (|\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k|^2 + |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0|^2) + \sum_{i=0}^K \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag} \left(\mathbf{h}_k \mathbf{h}_k^H \right) \mathbf{f}_i + \frac{\sigma^2}{P}}. \quad (33)$$

Let us define the k -th user's weighted precoding vector as

$$\mathbf{w}_k = \alpha_{\text{ADC},k}^{1/2} \mathbf{f}_k. \quad (34)$$

Then, we define $\mathbf{W} = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_K]$, and then the vectorized weighted precoding matrix \mathbf{W} as $\bar{\mathbf{w}} = \text{vec}(\mathbf{W})$. Here, we assume $\text{tr}(\mathbf{W}\mathbf{W}^H) = 1$ which indicates that the AP uses the maximum transmit power P , which is optimal in terms of maximizing the spectral efficiency. As we assume that $\text{tr}(\mathbf{W}\mathbf{W}^H) = 1$ where this assumption implies that the access point (AP) uses the maximum transmit power P , which is considered optimal for maximizing the spectral efficiency. We assume $\mathbf{G}_k = (\Phi_{\alpha_{\text{DAC}}}^{1/2})^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}}^{1/2} + \Phi_{\beta_{\text{DAC}}} \text{diag} \left(\mathbf{h}_k \mathbf{h}_k^H \right)$. In addition, based on the SINRs in (32) and (33) with the vectorized weighted precoder $\bar{\mathbf{w}}$, we reorganize $R_{c,k}$ in a Rayleigh quotient form as

$$R_{c,k} = \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right), \quad (35)$$

where

$$\mathbf{A}_{c,k} = \text{blkdiag}(\mathbf{G}_k, \dots, \mathbf{G}_k) + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}, \quad (36)$$

$$\mathbf{B}_{c,k} = \mathbf{A}_{c,k} - \text{blkdiag}\left(\alpha_{\text{ADC},k}(\Phi_{\alpha_{\text{DAC}}}^{1/2})^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}}^{1/2}, \mathbf{0}_N, \dots, \mathbf{0}_N\right), \quad (37)$$

where $\mathbf{A}_{c,k}$ and $\mathbf{B}_{c,k}$ are the diagonal matrices of size $N(K+1) \times N(K+1)$. In a similar manner, we can express R_k in the Rayleigh quotient form as follows

$$R_k = \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right), \quad (38)$$

where

$$\begin{aligned} \mathbf{A}_k &= \text{blkdiag}\left(\mathbf{G}_k - \alpha_{\text{ADC},k}(\Phi_{\alpha_{\text{DAC}}}^{1/2})^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}}^{1/2}, \mathbf{G}_k, \dots, \mathbf{G}_k\right) + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}, \\ \mathbf{B}_k &= \mathbf{A}_k - \text{blkdiag}\left(\mathbf{0}_N, \dots, \underbrace{\alpha_{\text{ADC},k}(\Phi_{\alpha_{\text{DAC}}}^{1/2})^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}}^{1/2}}_{\text{the } (k+1)\text{th block}}, \mathbf{0}_N, \dots, \mathbf{0}_N\right). \end{aligned} \quad (39)$$

According to (26), (35), and (38), the reformulated optimization problem in (17) is

$$\underset{\bar{\mathbf{w}}}{\text{maximize}} \quad \ln \left(\sum_{k=1}^K \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)^{-\frac{1}{\tau \ln 2}} \right)^{-\tau} + \frac{1}{\ln 2} \sum_{k=1}^K \ln \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right) \quad (40)$$

$$\text{subject to } \|\bar{\mathbf{w}}\| = 1. \quad (41)$$

The equality constraint in (41) is derived from utilizing the maximum transmit power. However, the problem in (40) remains unchanged up to a scaling factor of $\bar{\mathbf{w}}$. Therefore, we can disregard the constraint in (41). However, the problem remains non-convex with respect to the precoding vector $\bar{\mathbf{w}}$, which makes it impossible to find the global optimal solution. It is worth noting that the problem in (40) can be seen as a variant of the Rayleigh quotient problem. Hence, we can identify the first-order KKT conditions of the problem, which can be interpreted as an eigenvalue problem for the precoding vector. By finding the principal eigenvector of the problem, we can derive a local optimal solution. However, it is important to keep in mind that the solution obtained is only locally optimal and does not guarantee a global optimal solution.

5.2 First-Order Optimality Condition

In this subsection, the first-order optimality condition of (40) is derived in terms of $\bar{\mathbf{w}}$ as

Lemma 1 *The first-order optimality condition of the optimization problem (40) is satisfied if the following holds:*

$$\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}) \bar{\mathbf{w}} = \lambda(\bar{\mathbf{w}}) \bar{\mathbf{w}}, \quad (42)$$

where

$$\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}) = \lambda_{\text{num}}(\bar{\mathbf{w}}). \quad (43)$$

$$\sum_{k=1}^K \left[\frac{\exp\left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)\right)}{\sum_{\ell=1}^K \exp\left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,\ell} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,\ell} \bar{\mathbf{w}}} \right)\right)} \frac{\mathbf{A}_{c,k}}{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}} + \frac{\mathbf{A}_k}{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}} \right]$$

$$\mathbf{B}_{\text{KKT}}(\bar{\mathbf{w}}) = \lambda_{\text{den}}(\bar{\mathbf{w}}). \quad (44)$$

$$\sum_{k=1}^K \left[\frac{\exp\left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)\right)}{\sum_{\ell=1}^K \exp\left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,\ell} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,\ell} \bar{\mathbf{w}}} \right)\right)} \frac{\mathbf{B}_{c,k}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} + \frac{\mathbf{B}_k}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right],$$

with

$$\lambda(\bar{\mathbf{w}}) = \left\{ \frac{1}{K \ln 2} \sum_{k=1}^K \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)^{-\frac{1}{\tau}} \right\}^{-\frac{\tau}{\ln 2}} \prod_{k=1}^K \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right), \quad (45)$$

$$\lambda_{\text{num}} = \prod_{k=1}^K \left(\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}} \right), \quad (46)$$

$$\lambda_{\text{den}} = \left\{ \frac{1}{K \ln 2} \sum_{k=1}^K \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)^{-\frac{1}{\tau}} \right\}^{\frac{\tau}{\ln 2}} \prod_{k=1}^K \left(\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}} \right).$$

Proof 1 See Appendix VII.

Here, Lemma 1 states that the obtained first-order optimality condition can be considered as a nonlinear eigenvalue problem for $\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}})$. Furthermore, the objective function of the problem (40) is represented as $\ln \lambda(\bar{\mathbf{w}})$. As a result, we can obtain the best local optimal solution by finding the principal eigenvector of (42). This is because any eigenvector of (42) corresponds to one of the stationary points of the problem in (40), which can be summarized as follows:

Proposition 1 Denoting the leading eigenvector for the problem (42) to be $\bar{\mathbf{w}}^*$ and its corresponding eigenvalue to be λ^* , i.e., $\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}^*) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}^*) \bar{\mathbf{w}}^* = \lambda^* \bar{\mathbf{w}}^*$, the eigenvector $\bar{\mathbf{w}}^*$ is the stationary point that achieves the best local optimal solution of the problem (40).

Using the above insights, we have developed a computationally efficient RSMA precoding algorithm that aims to maximize the sum spectral efficiency by finding the best local optimal point.

5.3 Quantized Generalized Power Iteration for Rate-Splitting

By utilizing the generalized power iteration-based method [43], we can find the leading eigenvector of (42). Let $\bar{\mathbf{w}}_t$ be the vectorized weighted precoder at the t -th iteration. At the t -th iteration, the algorithm builds $\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t)$ and $\mathbf{B}_{\text{KKT}}(\bar{\mathbf{w}}_t)$ according to (43) and (44), and then $\bar{\mathbf{w}}_{t+1}$ is updated as

$$\bar{\mathbf{w}}_{t+1} = \frac{\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}_t) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t) \bar{\mathbf{w}}_t}{\|\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}_t) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t) \bar{\mathbf{w}}_t\|}. \quad (47)$$

Algorithm 1 Quantized Generalized Power Iteration for Rate-Splitting (Q-GPI-RS)

- 1: **initialize:** $\bar{\mathbf{w}}_0$
 - 2: Set the iteration count $t = 0$.
 - 3: **While** $\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{w}}_t\| > \varepsilon$ & $t \leq t_{\max}$ **do**
 Build matrix $\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t)$ in (43)
 Build matrix $\mathbf{B}_{\text{KKT}}(\bar{\mathbf{w}}_t)$ in (44)
 Compute $\bar{\mathbf{w}}_{t+1} = \mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}_t)\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t)\bar{\mathbf{w}}_t$.
 Normalize $\bar{\mathbf{w}}_{t+1} \leftarrow \frac{\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}_t)\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t)\bar{\mathbf{w}}_t}{\|\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}}_t)\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}_t)\bar{\mathbf{w}}_t\|}$.
 $t \leftarrow t + 1$.
 - 4: **return** $\bar{\mathbf{w}}_t$.
-

We repeat the iteration until the convergence criterion is satisfied, i.e., $\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{w}}_t\| < \varepsilon$ in which $\varepsilon > 0$ is the threshold or a maximum iteration count t_{\max} is reached according to a system requirement. Algorithm 1 summarizes the steps. Algorithm 1 is our proposed algorithm. Before proceeding with the numerical evaluation of our proposed method, we introduce an optimization approach based on extending the direction of an existing state-of-the-art precoding approach. This is done to enable a performance comparison between the two methods in the next section.

Remark 2 (Algorithm Complexity) The complexity of Q-GPI-RS is dominated by the calculation of $\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{w}})$. Since $\mathbf{B}_{\text{KKT}}(\bar{\mathbf{w}})$ is a block diagonal and symmetric matrix, we perform the computing inversion with computational complexity order of $\mathcal{O}(T_{\text{GPI}}(K+1)N^3)$ where T_{GPI} is the number of outer iterations of Q-GPI-RS. Similarly, QCQP [25, 44] and CCCP [10] convex relaxation methods have the complexity orders of $\mathcal{O}(T_{\text{QCQP}}K^{3.5}N^{3.5})$ and $\mathcal{O}(T_{\text{CCCP}}N^6K^{0.5}2^{3.5K})$ where T_{QCQP} and T_{CCCP} are the number of outer iterations of QCQP and CCCP, respectively. Accordingly, the proposed Q-GPI-RS is more computationally efficient than the other state-of-the-art precoding methods.

5.4 Extension of Existing Approach: WMMSE Approach

In this section, we present an extension of the WMMSE-based alternating optimization (WMMSE-AO) approach [25] for precoding optimization in the considered system. The WMMSE-AO method was developed by [25] to solve the non-convex problem of maximizing the sum spectral efficiency with respect to the precoder, without considering the quantization error. To solve our optimization problem, we adopt the similar principle of the WMMSE-based algorithms in [25] and [37]. Firstly, we define $\hat{s}_{c,k}$ and \hat{s}_k as estimated values for $s_{c,k}$ and s_k , respectively. Next, we compute the mean squared errors (MSEs) of the common stream $\varepsilon_{c,k}$ and private stream ε_k received at user k by defining the scalar equalizers of the common and private streams for user k as $g_{c,k}$ and g_k , respectively.

$$\begin{aligned}
 \varepsilon_{c,k} &= \mathbb{E}[|\hat{s}_{c,k} - s_{c,k}|^2] = \mathbb{E}[|g_{c,k}y_{q,k} - s_{c,k}|^2] \\
 &= |g_{c,k}|^2 T_{c,k} - 2\text{Re} \left\{ \sqrt{P} g_{c,k} \alpha_{\text{ADC},k} \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0 \right\} + 1,
 \end{aligned} \tag{48}$$

$$\begin{aligned}\varepsilon_k &= \mathbb{E}[|\hat{s}_k - s_k|^2] = \mathbb{E}[|g_k y_{q,k} - s_k|^2] \\ &= |g_k|^2 T_k - 2\text{Re} \left\{ \sqrt{P} g_k \alpha_{\text{ADC},k} \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k \right\} + 1.\end{aligned}\quad (49)$$

Furthermore, $T_{c,k}$ and T_k are formulated as

$$T_{c,k} = P \alpha_{\text{ADC},k}^2 \sum_{i=0}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2 + \alpha_{\text{ADC},k}^2 \mathbf{h}_k^H \mathbf{R}_{\text{qDAC}} \mathbf{h}_k + r_{\text{qADC},k} \alpha_{\text{ADC},k} + \alpha_{\text{ADC},k}^2 \sigma^2, \quad (50)$$

$$T_k = P \alpha_{\text{ADC},k}^2 \sum_{i=1}^K |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_i|^2 + \alpha_{\text{ADC},k}^2 \mathbf{h}_k^H \mathbf{R}_{\text{qDAC}} \mathbf{h}_k + r_{\text{qADC},k} \alpha_{\text{ADC},k} + \alpha_{\text{ADC},k}^2 \sigma^2. \quad (51)$$

The minimum MSEs is achieved when $g_{c,k}^{\text{MMSE}} = \sqrt{P} \alpha_{\text{ADC},k} \mathbf{f}_0^H \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k T_{c,k}^{-1}$ and $g_k^{\text{MMSE}} = \sqrt{P} \alpha_{\text{ADC},k} \mathbf{f}_k^H \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k T_k^{-1}$. According to (50), the MMSE of the common stream is defined as

$$\varepsilon_{c,k}^{\text{MMSE}} = T_{c,k}^{-1} (T_{c,k} - P \alpha_{\text{ADC},k}^2 |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0|^2) \quad (52)$$

and the MMSE of the private stream for user k is

$$\varepsilon_k^{\text{MMSE}} = T_k^{-1} (T_k - P \alpha_{\text{ADC},k}^2 |\mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k|^2). \quad (53)$$

Based on (48), the augmented WMSE of the common stream is defined by

$$\begin{aligned}\xi_{c,k} &= u_{c,k} \varepsilon_{c,k} - \log_2(u_{c,k}) \\ &= P \sum_{i=0}^K \mathbf{f}_i^H \left(\alpha_{\text{ADC},k}^2 u_{c,k} |g_{c,k}|^2 \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \right) \mathbf{f}_i - 2\text{Re} \left\{ \sqrt{P} u_{c,k} g_{c,k} \alpha_{\text{ADC},k} \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_0 \right\} \\ &\quad + \alpha_{\text{ADC},k}^2 u_{c,k} |g_{c,k}|^2 \mathbf{h}_k^H \mathbf{R}_{\text{qDAC}} \mathbf{h}_k + u_{c,k} |g_{c,k}|^2 r_{\text{qADC},k} \alpha_{\text{ADC},k} + \alpha_{\text{ADC},k}^2 u_{c,k} |g_{c,k}|^2 \sigma^2 \\ &\quad + u_{c,k} - \log_2(u_{c,k}).\end{aligned}\quad (54)$$

In the same manner, using (49), the augmented WMSE of the private stream for user k is defined as

$$\begin{aligned}\xi_k &= u_k \varepsilon_k - \log_2(u_k) \\ &= P \sum_{i=1}^K \mathbf{f}_i^H \left(\alpha_{\text{ADC},k}^2 u_k |g_k|^2 \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \right) \mathbf{f}_i \\ &\quad - 2\text{Re} \left\{ \sqrt{P} u_k g_k \alpha_{\text{ADC},k} \mathbf{h}_k^H \Phi_{\alpha_{\text{DAC}}} \mathbf{f}_k \right\} + \alpha_{\text{ADC},k}^2 u_k |g_k|^2 \mathbf{h}_k^H \mathbf{R}_{\text{qDAC}} \mathbf{h}_k + u_k |g_k|^2 r_{\text{qADC},k} \alpha_{\text{ADC},k} \\ &\quad + \alpha_{\text{ADC},k}^2 u_k |g_k|^2 \sigma^2 + u_k - \log_2(u_k).\end{aligned}\quad (55)$$

The optimal weights is computed to obtain the minimum of $\xi_{c,k}$ and ξ_k as $u_{c,k} = 1/\varepsilon_{c,k}^{\text{MMSE}}$ and $u_k = 1/\varepsilon_k^{\text{MMSE}}$. Consequently, for given equalizers $\xi_{c,k}$, ξ_k , and weights $u_{c,k}$, u_k , the sum spectral efficiency maximization optimization problem is formulated by the following WMSE minimization problem:

$$\underset{\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_K, \xi_c}{\text{minimize}} \quad \xi_c + \sum_{k=1}^K \xi_k \quad (56)$$

$$\text{subject to} \quad \text{tr} \left(\Phi_{\alpha_{\text{DAC}}} \mathbf{F} \mathbf{F}^H \right) \leq 1, \quad (57)$$

$$\xi_{c,k} \leq \xi_c, \forall k \in \mathcal{K}. \quad (58)$$

By remarking that the problem (56) is the QCQP, the optimization problem can be solved by CVX.

Therefore, we compute the equalizers, weights, and precoders in the alternating manner as follows:

1. *Update of equalizers and weights:* we update the equalizers and weights by computing $g_{c,k}^{\text{MMSE}}$, g_k^{MMSE} , $u_{c,k} = 1/\varepsilon_{c,k}^{\text{MMSE}}$ and $u_k = 1/\varepsilon_k^{\text{MMSE}}$ for given precoding vectors.
2. *Update of precoders and ξ_c :* then, the precoders and ξ_c can be derived by solving (56) via CVX for given equalizers and weights.
3. *Repeat steps 1 and 2:* the steps 1 and 2 are repeated until convergence.

VI Numerical Results

In this section, we conduct a comparison of the sum spectral efficiency achieved by the proposed Q-GPI-RS method and several existing baseline methods. The channel vector \mathbf{h}_k is derived from its spatial covariance matrix $\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H]$, which is computed as the expectation of the outer product of \mathbf{h}_k and its Hermitian transpose. To generate \mathbf{R}_k , we adopt the one-ring channel model as described in [45]. Specifically, the channel covariance matrix \mathbf{R}_k at the n -th antenna and m -th antenna is defined as $[\mathbf{R}_k]_{n,m} = \frac{1}{2\Delta_k} \int_{\theta_k - \Delta_k}^{\theta_k + \Delta_k} e^{-j\frac{2\pi}{\psi} \Psi(x)(\mathbf{r}_n - \mathbf{r}_m)} dx$, where ψ is a signal wavelength, Δ_k is the angular spread of user k , θ_k is angle-of-departure (AoD) of user k , $\Psi(x) = [\cos(x), \sin(x)]$, and \mathbf{r}_n is the position vector of n -th antenna. The channel vector \mathbf{h}_k is expressed as a decomposition using the Karhunen-Loeve model, where $\mathbf{h}_k = \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{g}_k$. The matrix \mathbf{U}_k is the eigenvectors of \mathbf{R}_k , a spatial covariance matrix, $\mathbf{\Lambda}_k$ is a diagonal matrix of eigenvalues of \mathbf{R}_k , and \mathbf{g}_k is a random vector with each entry following a complex normal distribution with mean zero and variance one. The rank of \mathbf{R}_k is denoted by r_k , and \mathbf{g}_k has dimensions of \mathbb{C}^{r_k} . The channel is assumed to be constant within one transmission block, which follows a block fading channel model. The parameter θ_k is dependent on the user's location and varies accordingly. When the users are randomly located, the angle of departure (AoD) θ_k follows an independent and identically distributed (IID) uniform distribution between 0 and π for all users. In the case of correlated users, the differences of θ_k between all users are randomly distributed within $\pi/6$. The simulation is conducted under the settings $\Delta_k = \pi/6$, $\varepsilon = 0.01$, and $\sigma^2 = 1$. The initialization process involves setting the precoding vector $\mathbf{f}_k = \mathbf{h}_k$ for each user, which corresponds to maximum ratio transmission (MRT). For the common stream, the average of channel vectors is adopted, i.e., $\mathbf{f}_0 = \frac{1}{K} \mathbf{H} \cdot \mathbf{1}_{K \times 1}$. The stacked precoding vector $\bar{\mathbf{w}}_0$ is then initialized based on \mathbf{F} . We define an effective channel vector as $\mathbf{h}_k^{\text{eff}} = \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k \alpha_{\text{ADC},k}^H$. Based on this definition, we compare the proposed Q-GPI-RS method with the following baseline methods:

- **GPI-RS (RSMA)**: The GPI algorithm based RSMA [33].
- **Q-WMMSE-AO (RSMA)**: The quantized WMMSE alternating optimization (Q-WMMSE-AO) algorithm [46].
- **WMMSE-AO (RSMA)**: The WMMSE-AO RSMA precoding [25].
- **WMMSE-CCCP (RSMA)**: The WMMSE concave-convex procedure (CCCP) RSMA algorithm [10].
- **Q-GPI-SEM (SDMA)**: The quantization-aware GPI-based spectral efficiency maximization method [18].
- **WMMSE (SDMA)**: The WMMSE-based precoding method [30].
- **Q-RZF (SDMA)**: The conventional linear precoders based on the effective channel $\mathbf{h}_k^{\text{eff}} = \Phi_{\alpha_{\text{DAC}}}^H \mathbf{h}_k$ such as quantization-aware RZF.

We note that for SDMA, WMMSE and Q-GPI-SEM algorithms are considered as state-of-the-art precoding methods, except for Q-RZF. Furthermore, to perform a thorough analysis of the proposed methods, we also evaluate the state-of-the-art benchmarks for RSMA, such as GPI-RS, Q-WMMSE-AO, WMMSE-AO, and WMMSE-CCCP. We use $\text{SNR} = P/\sigma^2$ to define the signal-to-noise ratio. To determine the value of the approximation parameter τ , we conduct experiments on the system configuration to obtain an appropriate value. Once we obtain the value of τ , we fix it for each evaluation point without any further online tuning. For computing the spectral efficiency in simulation results, we use equations (14) and (16) without approximating the minimum function.

6.1 Homogeneous Quantization Resolution

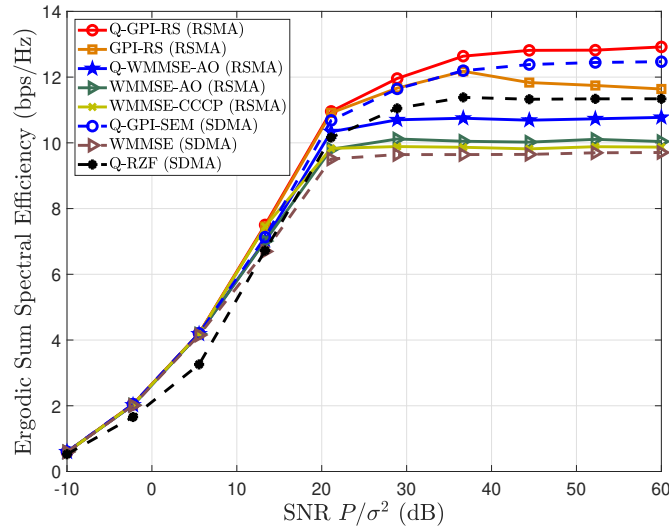


Figure 2: The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{DAC},n} = 4$ DAC bits, $\forall n$, and $b_{\text{ADC},k} = 6$ ADC bits, $\forall k$.

In Fig. 2, we consider randomly distributed users for $N = 4$ and $K = 2$ with $b_{\text{DAC},n} = 4$, $\forall n$ and $b_{\text{ADC},k} = 6$, $\forall k$. According to the results presented in Figure 2, it can be observed that Q-GPI-RS has achieved the highest spectral efficiency. It is important to note that Q-WMMSE-AO, even though it takes quantization effects into consideration, does not guarantee the best local optimal point and thus, reveals a performance limitation compared to Q-GPI-RS. Furthermore, GPI-RS has shown similar performance to Q-GPI-RS in the low SNR regime. However, in the high SNR regime, GPI-RS suffers from quantization errors, resulting in performance degradation compared to Q-GPI-SEM. Similarly, WMMSE-AO and WMMSE-CCCP approaches have also shown performance degradation compared to Q-WMMSE-AO due to their design not considering quantization errors. As a result, Q-RZF has outperformed the WMMSE-based methods as SNR increases. The proposed Q-GPI-RS method, with properly designed RSMA precoding, has shown RSMA gain by mitigating interference compared to Q-GPI-SEM, which is based on classical SDMA. This gain has also been observed by comparing WMMSE-AO with WMMSE. In conclusion, the simulation results show that the proposed algorithm Q-GPI-RS achieves the highest

spectral efficiency compared to the benchmarks. Therefore, it is crucial to consider the quantization error for realizing the potential of RSMA.

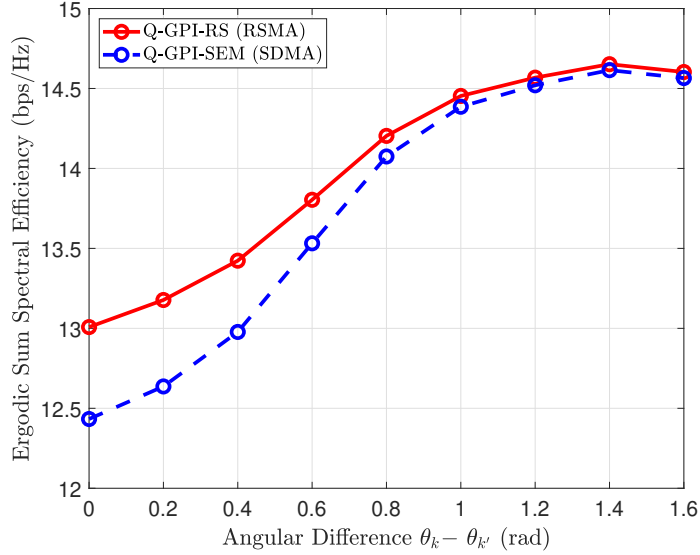


Figure 3: The sum spectral efficiency versus angular difference of users for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{DAC},n} = 4$ DAC bits, $\forall n$, $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$, and $\text{SNR} = 50$ dB.

In Fig. 3, we evaluate the spectral efficiency according to the user’s angular difference ($\theta_k - \theta_{k'}$) for $N = 4$, $K = 2$, $b_{\text{DAC},n} = 4$, $\forall n$, and $b_{\text{ADC},k} = 8$, $\forall k$. Based on the results presented in Figure 3, it can be observed that the gap between Q-GPI-RS and Q-GPI-SEM increases as the angular difference between users decreases. This finding aligns with the analytical results obtained for perfect quantization systems in [47], where it was shown that the gain of RSMA increases as user channel correlation increases. This phenomenon occurs because RSMA can exploit the common stream to reduce inter-user interference, which becomes more severe when channel correlation is high. On the other hand, SDMA suffers from high inter-user interference. In indoor communication environments, where low-power applications such as IoT communications are prevalent, user channels are typically correlated, and there are small angular differences among users [48]. Therefore, for the rest of the simulations, it is assumed that the user channels are correlated, with an angular difference between users less than $\pi/6$. This assumption is based on the fact that in such environments, the gain of RSMA over SDMA is expected to be more significant due to the higher level of channel correlation. Overall, the simulation results demonstrate that RSMA is a promising technique for enhancing the performance of wireless communication systems, especially in scenarios with highly correlated user channels. The results also highlight the importance of considering the impact of channel quantization when designing RSMA-based systems.

6.2 Heterogeneous Quantization Resolution

The spectral efficiency of a system with heterogeneous quantization bits at the AP’s DACs is analyzed in this dissertation paper. In the scenario where $N = 6$, $K = 4$, and $b_{\text{ADC},k} = 8$, $\forall k$, the number of DAC

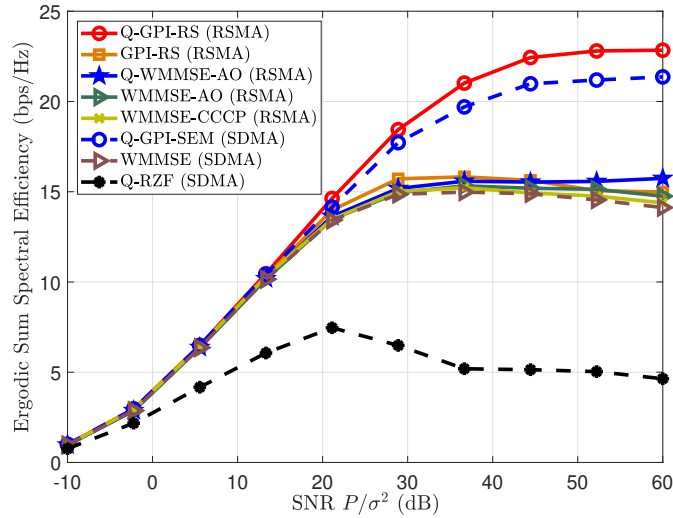


Figure 4: The sum spectral efficiency versus SNR for $N = 6$ AP antennas, $K = 4$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. The numbers of DACs are set uniformly from 2 to 8 bits.

bits is set randomly between 2 to 8 bits. The results are presented in Fig. 4, where Q-GPI-RS achieves the highest spectral efficiency compared to other baseline methods, with a more significant improvement over the previous random channel case. Moreover, Q-WMMSE-AO, which considers the quantization error, achieves higher spectral efficiency than WMMSE-AO. The performances of GPI-RS, WMMSE-AO, and WMMSE-CCCP decrease in the high SNR regime because they were not designed to consider the quantization error. In comparison to SDMA, Q-GPI-RS demonstrates the advantage of RSMA over Q-GPI-SEM since RSMA can reduce inter-user interference by utilizing the common stream as SNR increases. Additionally, the other RSMA algorithms, such as Q-WMMSE-AO and WMMSE-AO, provide higher spectral efficiency than WMMSE and Q-RZF due to higher channel correlation, which was not observed in Fig. 2. Therefore, it is essential to consider the quantization error in heterogeneous DAC systems to achieve RSMA gain with channel correlation.

In this dissertation paper, we investigate the mixed-resolution DAC scenario in which only one of the DACs has a higher resolution than the others. Specifically, we consider the case where $N = 4$ antennas are equipped with DACs, and $K = 2$ users are served by the system. The resolution of the ADCs is fixed at 8 bits for all antennas. The distribution of the DAC bits is $(3, 3, 3, 8)$. We evaluate the performance of three different precoding schemes, Q-GPI-RS, Q-GPI-SEM, and ZF, in terms of spectral efficiency. As shown in Fig.5, we observe that Q-GPI-RS achieves a significant improvement in spectral efficiency compared to Q-GPI-SEM in medium-to-high SNR regime. This gain is achieved by assigning a higher rate to the common stream as shown in Fig.6. Additionally, Fig. 7 demonstrates that Q-GPI-RS concentrates the transmit power on the antenna with the higher-resolution DAC to prevent the quantization error from significantly increasing as the SNR increases. We can analyze this trend by examining the SINR expressions, specifically the quantization error term from DACs in (32) and (33). This term, denoted by $\text{QE}_{\text{DAC}} = \sum_{i=0}^K \mathbf{f}_i^H \Phi_{\alpha_{\text{DAC}}} \Phi_{\beta_{\text{DAC}}} \text{diag}(\mathbf{h}_k \mathbf{h}_k^H) \mathbf{f}_i$, can be fol-

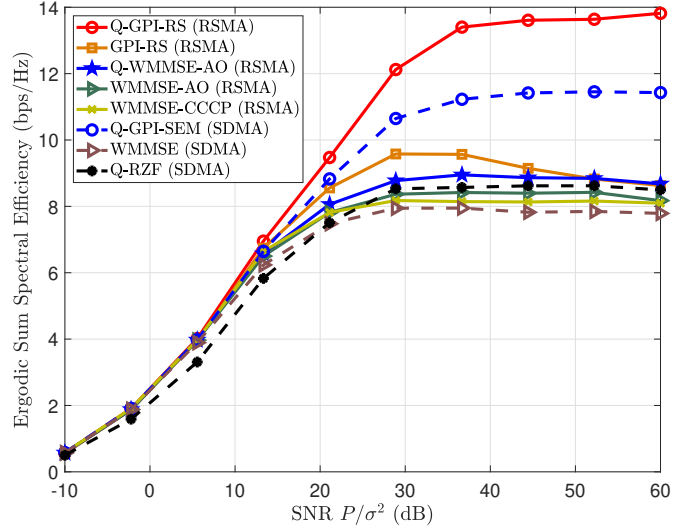


Figure 5: The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.

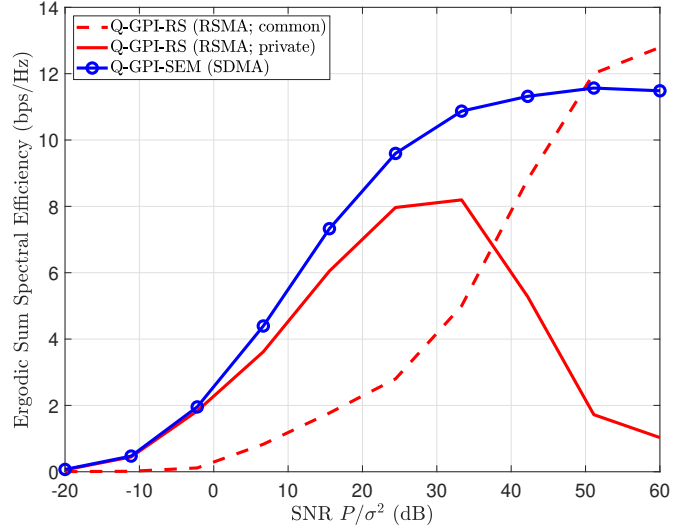


Figure 6: The spectral efficiency of the common and private streams versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.

lowed as the sum of quantization errors at each antenna $\text{QE}_{\text{DAC},n}$, i.e., $\text{QE}_{\text{DAC}} = \sum_{n=1}^N \text{QE}_{\text{DAC},n}$, where $\text{QE}_{\text{DAC},n} = |f_{i,n}|^2 \alpha_{\text{DAC},n} (1 - \alpha_{\text{DAC},n}) [\text{diag}(\mathbf{h}_k \mathbf{h}_k^H)]_{n,n}$. Here, $[\cdot]_{n,n}$ indicates the n th diagonal element of $[\cdot]$. We can observe that the quantization error term from the DACs decreases as the DAC resolution increases. This is because the value of $\alpha_{\text{DAC},n}$ becomes approximately 1 for high-resolution DACs intuitively. As a result, we can manage the quantization error by allocating transmit power mostly on the antennas with high-resolution DACs in the heterogeneous DAC systems. In an overloaded system, the

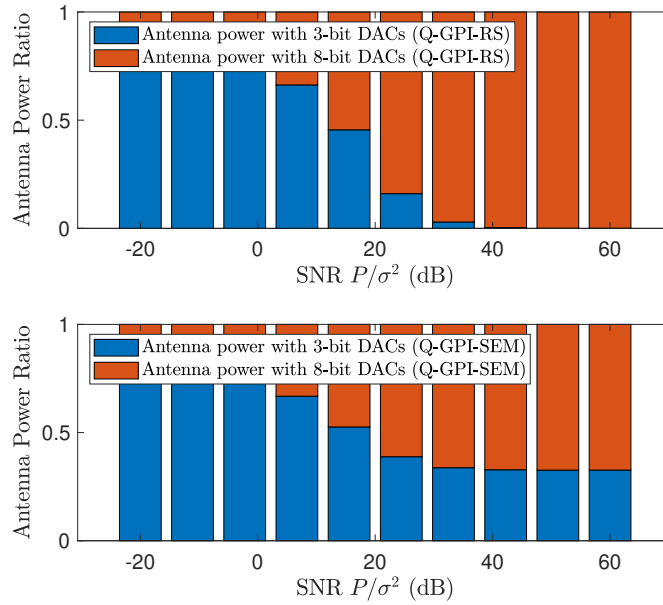


Figure 7: The average antenna power ratios of Q-GPI-RS and Q-GPI-SEM versus SNR for $N = 4$ AP antennas, $K = 2$ users, and $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$. We consider that one of the DACs is equipped with 8 bits and the rest are equipped with 3 bits.

number of active antennas needs to be reduced. RSMA transmission can deal with this situation by leveraging the common stream. Q-GPI-SEM, on the other hand, allocates the transmit power to all antennas, leading to a limitation in performance compared to Q-GPI-RS. While Q-GPI-SEM achieves higher spectral efficiency than the other baselines, it still cannot avoid the significant increase of quantization errors with the SNR.

The utilization of mixed ADC/DAC architecture, which is a type of heterogeneous ADC/DAC systems, has shown significant potential in improving spectral efficiency while reducing power consumption in MIMO systems. In this dissertation, we aim to evaluate the spectral efficiency versus signal-to-noise ratio (SNR) for mixed DACs. We consider a scenario where half of the DACs are equipped with 3 bits and the other half with 10 bits, while all ADCs are equipped with 10 bits. We set $N=8$ and $K=4$ for the number of antennas and user equipment, respectively. We compare the performance of the proposed Q-GPI-RS with that of high-resolution DAC case where all DACs have 10 bits. Our simulation results, shown in Fig. 8, demonstrate that the proposed Q-GPI-RS achieves the highest sum spectral efficiency in this setup. Moreover, the gain of RSMA becomes more significant as the SNR increases for both $N=8$ and $N=4$ cases. Interestingly, we observe that the spectral efficiency gap between the $N=4$ and $N=8$ cases reduces as the SNR increases. This phenomenon is due to the fact that Q-GPI-RS with mixed-resolution DACs reduces quantization error at high SNR by allocating more power to antennas with high-resolution DACs. As a result, the effective number of antennas becomes the number of antennas with high-resolution DACs at high SNR, making $N=8$ and $N=4$ cases similar while still providing a gap from the SDMA-based approaches. On the other hand, in the low-to-medium SNR, the gain of RSMA

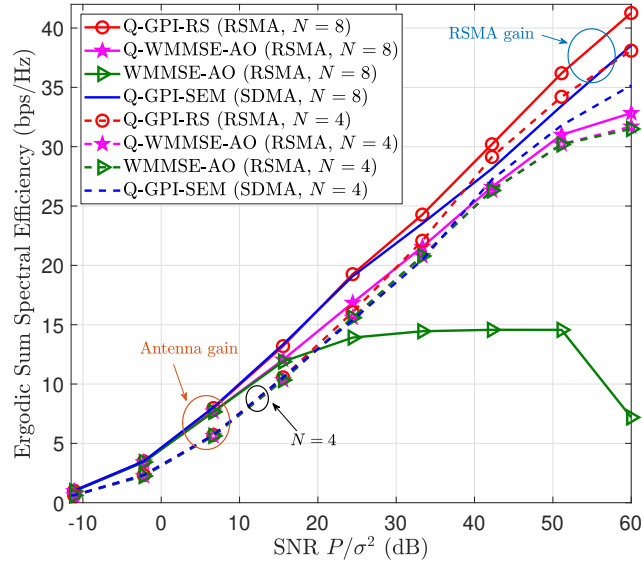


Figure 8: The sum spectral efficiency versus SNR for $N = 4$ AP antennas, $K = 4$ users, $b_{\text{DAC},n} = 10$ DAC bits, $\forall n$, and $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$ and for $N = 8$ AP antennas, $K = 4$ users, and $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$. We consider the half of the DACs are equipped with 3 bits and the other half are equipped with 10 bits.

is relatively small, but there is still an antenna gain for $N=8$ compared to $N=4$, thanks to additional antennas with low-resolution DACs. Therefore, deploying the proposed RSMA precoding method with mixed-resolution DACs can provide an antenna gain in the low-to-medium SNR with a small cost of deploying low-power hardware, while still providing the RSMA gain in high SNR. In summary, our research demonstrates that mixed-resolution DACs can significantly improve the spectral efficiency and power consumption in MIMO systems. The proposed Q-GPI-RS with mixed-resolution DACs provides the highest spectral efficiency and gains antenna benefits in both high and low-to-medium SNR, making it a promising candidate for practical MIMO systems.

6.3 Effect of DAC and ADC Quantization to RSMA

In this section, we observe the effect of the number of DAC and ADC bits for the proposed algorithm. In Fig. 9, we evaluate the spectral efficiency versus the number of DAC bits with $b_{\text{ADC},k} = 10$, $\forall k$ and the number of ADC bits with $b_{\text{DAC},n} = 10$, $\forall n$ in Fig. 10 for $N = 8$, $K = 4$, and SNR = 40 dB. We also analyze the allocated power ratio between the common and private streams according to the number of quantization bits in Fig. 11. The proposed algorithm obtains the highest spectral efficiency than the other baselines over the considered number of bit as shown in Fig. 9. Since RSMA concentrates the rate to the common stream in high resolution quantizer, the gain of spectral efficiency of RSMA becomes larger by increasing the resolutions of DAC and ADC, corresponding to the observation in [36]. Based on the Figure 11, it can be seen that as the quantizer resolution increases, there is an increase in the amount of

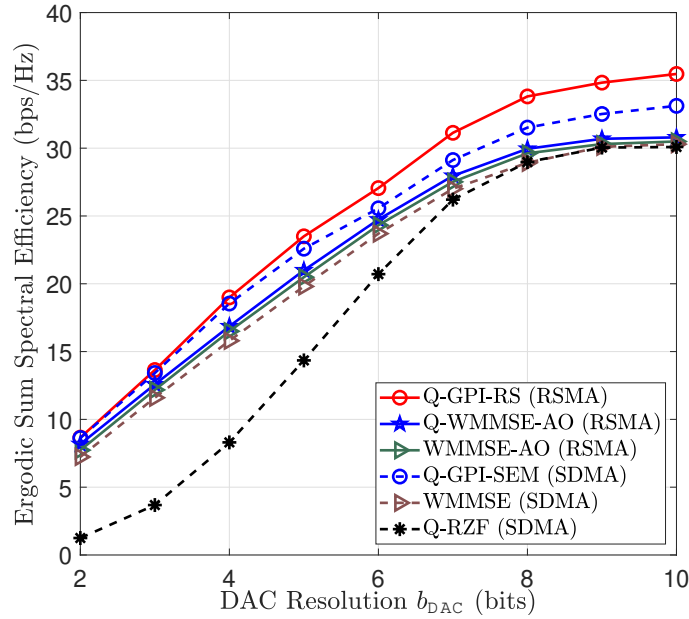


Figure 9: The sum spectral efficiency versus the number of DAC bits versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and $\text{SNR} = 40$ dB with $b_{\text{ADC},k} = 10$ ADC bits, $\forall k$.

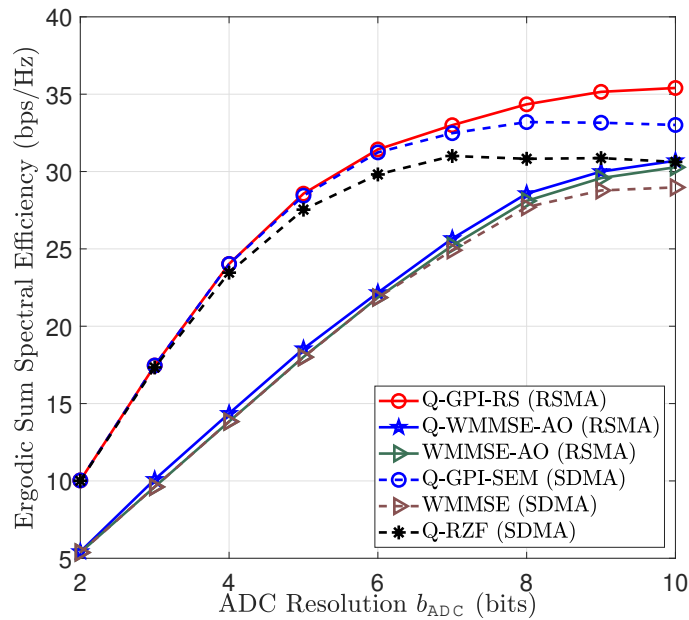


Figure 10: The sum spectral efficiency versus the number of ADC bits versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and $\text{SNR} = 40$ dB with $b_{\text{DAC},n} = 10$ DAC bits, $\forall n$.

power allocated to the common stream. The allocation of power to the common stream is more limited by the number of bits in the ADC compared to the number of bits in the DAC. As a result, the RSMA

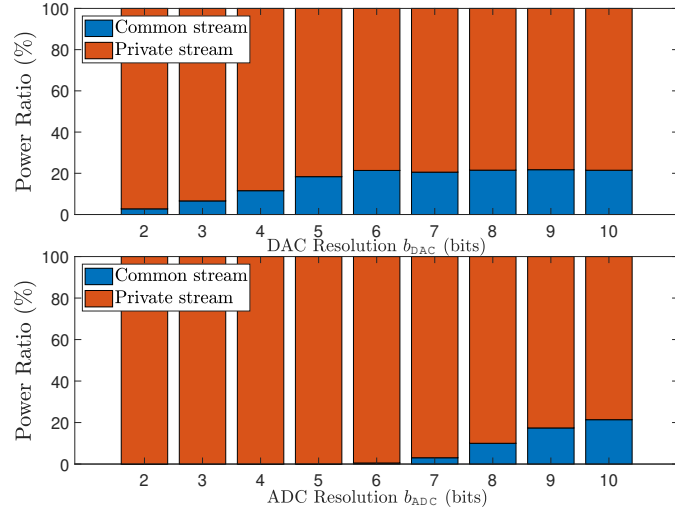


Figure 11: The power ratio versus DAC and ADC bit between common and private streams for $N = 8$ AP antennas, $K = 4$ users, and $\text{SNR} = 40$ dB.

gain is more affected by changes in the ADC resolution rather than changes in the DAC resolution.

6.4 Convergence Analysis

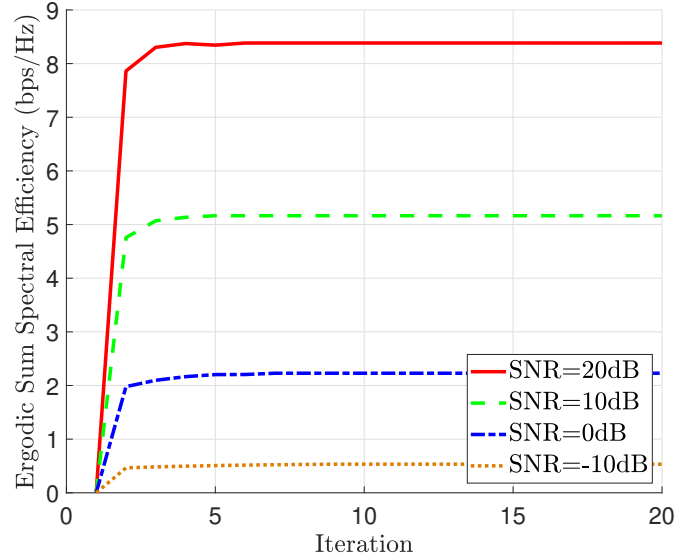


Figure 12: Convergence results in terms of the sum spectral efficiency for $N = 4$ AP antennas, $K = 2$ users, $b_{\text{ADC},k} = 8$ ADC bits, $\forall k$, and $\text{SNR} \in \{-10, 0, 10, 20\}$ dB transmit power with mixed-resolution DACs. For the mixed-resolution DAC case, we consider that a single DAC is an 8-bit DAC and the rest are 3-bit DACs.

In Fig. 12, we evaluate the proposed algorithm convergence in terms of the iteration count for $N = 4$, $K = 2$, $b_{\text{ADC},k} = 8$, $\forall k$, and $\text{SNR} \in \{-10, 0, 10, 20\}$ dB where a single DAC is 8-bit and the other DACs

are 3-bit. As shown in Fig. 12, In Fig. 12, we present the convergence result in terms of the iteration count for $N = 4$, $K = 2$, $b_{\text{ADC},k} = 8$, $\forall k$, and $\text{SNR} \in \{-10, 0, 10, 20\}$ dB considering that a single DAC is an 8-bit DAC and the rest are 3-bit DACs. The Q-GPI-RS algorithm achieves convergence within a small number of iterations, specifically $T_{\text{GPI}} = 5$, for all considered SNR values. This indicates that the algorithm converges quickly and efficiently, making it a promising choice for practical implementation. Therefore, the Q-GPI-RS algorithm's potential for practical implementation is further enhanced due to its low complexity, which reduces the computational resources required for its implementation as discussed in Remark 2.

VII Proof of Lemma 1

From the reformulated optimization problem, the Lagrangian function is defined as

$$L(\bar{\mathbf{w}}) = \ln \left(\frac{1}{K} \sum_{k=1}^K \exp \left(\log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right)^{-\frac{1}{\tau}} \right) \right)^{-\tau} + \sum_{k=1}^K \frac{1}{\ln 2} \ln \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right). \quad (59)$$

Furthermore, we calculate the partial derivatives of the Lagrangian function (59), and then determine the stationarity condition by setting (59) to zero. To simplify the notation, we denote the first and second terms in (59) as $L_1(\bar{\mathbf{w}})$ and $L_2(\bar{\mathbf{w}})$, respectively. The partial derivative of $L_1(\bar{\mathbf{w}})$ is represented as

$$\frac{\partial L_1(\bar{\mathbf{w}})}{\partial \bar{\mathbf{w}}^H} = \frac{1}{\ln 2} \sum_{k=1}^K \left[\frac{\exp \left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right) \right)}{\sum_{\ell=1}^K \exp \left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,\ell} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,\ell} \bar{\mathbf{w}}} \right) \right)} \left\{ \frac{\mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}} - \frac{\mathbf{B}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right\} \right]. \quad (60)$$

The derivative of the Lagrangian function $L_2(\bar{\mathbf{w}})$ is derived as

$$\frac{\partial L_2(\bar{\mathbf{w}})}{\partial \bar{\mathbf{w}}^H} = \frac{1}{\ln 2} \sum_{k=1}^K \left[\frac{\mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}} - \frac{\mathbf{B}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right]. \quad (61)$$

Using (60) and (61), the first-order optimality condition is given as

$$\frac{\partial L_1(\bar{\mathbf{w}})}{\partial \bar{\mathbf{w}}^H} + \frac{\partial L_2(\bar{\mathbf{w}})}{\partial \bar{\mathbf{w}}^H} \quad (62)$$

$$= \sum_{k=1}^K \left[\frac{\exp \left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right) \right)}{\sum_{\ell=1}^K \exp \left(-\frac{1}{\tau} \log_2 \left(\frac{\bar{\mathbf{w}}^H \mathbf{A}_{c,\ell} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,\ell} \bar{\mathbf{w}}} \right) \right)} \left\{ \frac{\mathbf{A}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{A}_{c,k} \bar{\mathbf{w}}} - \frac{\mathbf{B}_{c,k} \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_{c,k} \bar{\mathbf{w}}} \right\} \right] + \sum_{k=1}^K \left[\frac{\mathbf{A}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{A}_k \bar{\mathbf{w}}} - \frac{\mathbf{B}_k \bar{\mathbf{w}}}{\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}} \right] \quad (63)$$

$$= 0. \quad (64)$$

Rearranging (63), we derive $\mathbf{A}_{\text{KKT}}(\bar{\mathbf{w}}) \bar{\mathbf{w}} = \lambda(\bar{\mathbf{w}}) \mathbf{B}_{\text{KKT}}(\bar{\mathbf{w}}) \bar{\mathbf{w}}$. Since $\mathbf{B}_{c,k}$ and \mathbf{B}_k are Hermitian block diagonal matrices, \mathbf{B}_{KKT} is invertible. Accordingly, we finally obtain the condition in (42). This completes the proof. \blacksquare

VIII Conclusion

In my dissertation, we propose a promising precoding algorithm for downlink RSMA systems with low-resolution quantizers, with the aim of maximizing the sum spectral efficiency. The non-smooth problem is addressed by employing the LogSumExp technique to convert into tractable form. Furthermore, the non-convex problem is reformulated to the product of the Rayleigh quotients, due to making it more tractable. By deriving the first-order optimality condition, the stationary point is determined using the generalized eigenvalue problem. The resulting computationally efficient algorithm can be used to maximize the spectral efficiency. Simulation results demonstrate that the proposed method outperforms the baseline methods, achieving the highest sum spectral efficiency. It is observed that RSMA offers gains in most DAC and ADC resolutions, with the gain increasing as the resolutions of DACs and ADCs increase. This is because a higher resolution allows for a higher rate of the common stream. Additionally, the proposed Q-GPI-RS algorithm yields a significant gain of RSMA in heterogeneous DACs by decreasing the allocated transmit power for the antennas with low-resolution DACs. Accordingly, since the proposed algorithm can suppress the quantization error, Q-GPI-RS achieves the high performance gain taking advantage of using RSMA in heterogeneous resolution DACs. Based on the simulation results, it is confirmed that RSMA is beneficial for coarse quantization systems, and the proposed RSMA precoding algorithm significantly improves the spectral efficiency, offering high adaptation to the heterogeneous quantization bits thanks to RSMA. Consequently, the proposed algorithm can provide benefits in both increasing communication efficiency and designing low-power transceiver architectures. For future work, it is promising to develop an optimal RSMA precoding algorithm that explicitly considers both channel estimation error and quantization error.

8.1 Future work

In this dissertation, I addressed some of the main critical issues to adopt rate-splitting multiple access with low-resolution quantizers in MU-MIMO systems. There are still issues left that need to be resolved to successfully realize quantized RSMA communication systems. Therefore, I present promising future research directions related to the topics in this dissertation.

- **Partial channel state information of transmitter in quantized RSMA system:** Quantized rate-splitting multiple access system is a promising technology for improving the spectral efficiency and capacity of wireless communication networks. In this system, partial channel state information (CSI) at the transmitter is a critical parameter that can significantly impact the performance of the system. Partial CSI refers to the partial knowledge of the channel conditions available to the transmitter. Specifically, it involves information about the path loss, channel gains, and channel phases of the communication links. However, the CSI may not be known precisely due to channel estimation errors, feedback delay, and quantization noise. In quantized RSMA system, partial CSI is used to determine the power allocation and rate splitting coefficients that optimize the transmission strategy. The power allocation and rate splitting coefficients are calculated based on the

available partial CSI, and their values directly affect the performance of the system in terms of capacity and bit error rate (BER). For example, if the transmitter has partial CSI that underestimates the channel gains, it may allocate less power to the channel, resulting in a lower transmission rate and reduced capacity. Conversely, if the partial CSI overestimates the channel gains, it may allocate more power to the channel, leading to a higher transmission rate but a higher BER due to more significant noise. Therefore, accurate and timely estimation of partial CSI is crucial for optimizing the transmission strategy in the quantized RSMA system. In future work, research can focus on developing novel partial CSI estimation algorithms that can minimize the impact of channel estimation errors, feedback delay, and quantization noise. Additionally, the design of efficient transmission schemes that can adapt to the changes in the partial CSI can be explored to enhance the performance for future wireless communication systems.

- **Energy efficiency maximization in quantized RSMA system:** The energy efficiency maximization problem in quantized rate-splitting multiple access systems is challenging because it is in a fractional form. The objective function in this problem is to maximize the energy efficiency, which is the ratio of the sum rate to the total power consumption. The problem can be formulated as a non-convex optimization problem with fractional programming, which makes it difficult to solve using traditional optimization techniques. One approach to solving this problem is by using the Dinkelbach method, which converts the problem into a sequence of linear programming problems that can be solved using gradient descent. The Dinkelbach method involves introducing a new variable into the objective function to convert the fractional form into a linear form. Specifically, the original objective function is expressed as the ratio of two non-negative functions, where the denominator is the power constraint. The Dinkelbach method replaces the denominator with a variable, λ , and introduces a constraint that forces the value of λ to be equal to the denominator. Using the Dinkelbach method, the energy efficiency maximization problem can be transformed into a linear problem that are easier to solve. In each iteration, the problem is formulated as an linear form problem by fixing the value of λ and solving for the optimal power allocation. The solution of each linear form problem is used to update the value of λ , which is then used in the next iteration. The gradient descent algorithm is used to optimize the value of λ , which converges to the optimal solution of the energy efficiency maximization problem. The algorithm starts with an initial value of λ and iteratively updates it until the solution converges. In summary, the Dinkelbach method is a useful tool for solving the energy efficiency maximization problem in the quantized rate-splitting multiple access systems systems. By transforming the fractional form into linear form problem, the method allows for the use of gradient descent to find the optimal power allocation that maximizes the energy efficiency. Future research can explore further improvements to this approach, such as incorporating constraints related to fairness and quality of service into the optimization problem.
- **Coverage analysis of LEO Satellite Network:** As the many satellites are located in the non-terrestrial region, the analysis of coverage of satellite network needs to be studied. Accordingly,

for my future research, I plan to develop new models that use stochastic geometry to analyze the coverage of satellite networks. Prior satellite network models involve complex system-level simulations, limiting analytical understanding of the network. By developing the prior coverage analysis of satellite communication, I will model satellite and user locations using Poisson point processes on concentric spheres, and then derive analytical derivations for the coverage probability of a typical downlink user. In addition, to consider more practical environment, I consider path-loss and shadowing effect in satellite communication networks. Then, comparing with other benchmark for coverage probability, I validate the proposed method by simulation results. This work can have potential to study the satellite networks.

IX Vita

Seokjun Park received his B.S. in the Department of Electrical Engineering at Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea, in 2022. He is currently pursuing the M.S. degree with the Department of Electrical Engineering, UNIST under supervision of professor Jinseok Choi. His main research interest is to develop future wireless communication systems and to develop algorithms for intelligent devices by applying information theory, optimization, and machine learning.

References

- [1] P. Yang, Y. Xiao, M. Xiao, and S. Li, “6G Wireless Communications: Vision and Potential Techniques,” *Proc. of Int. Symp. on Model. and Opt. in Mobile, Ad Hoc, and Wireless Networks*, vol. 33, no. 4, pp. 70–75, Jul. 2019.
- [2] S. Pattar, R. Buyya, K. R. Venugopal, S. Iyengar, and L. Patnaik, “Searching for the IoT Resources: Fundamentals, Requirements, Comprehensive Review, and Future Directions,” *IEEE Commun. Mag.*, vol. 20, no. 3, pp. 2101–2132, Apr. 2018.
- [3] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, “On Low-Resolution ADCs in Practical 5G Millimeter-Wave Massive MIMO Systems,” *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 205–211, Apr. 2018.
- [4] N. Liang and W. Zhang, “Mixed-ADC Massive MIMO,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Mar. 2016.
- [5] J. Zhang, L. Dai, Z. He, B. Ai, and O. A. Dobre, “Mixed-ADC/DAC Multipair Massive MIMO Relaying Systems: Performance Analysis and Power Optimization,” *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 140–153, Sep. 2018.
- [6] J. Choi, G. Lee, A. Alkhateeb, A. Gatherer, N. Al-Dhahir, and B. L. Evans, “Advanced Receiver Architectures for Millimeter-Wave Communications with Low-Resolution ADCs,” *IEEE Commun. Mag.*, vol. 58, no. 8, pp. 42–48, Aug. 2020.
- [7] Y. Mao, B. Clerckx, and V. O. Li, “Rate-Splitting Multiple Access for Downlink Communication Systems: Bridging, Generalizing, and Outperforming SDMA and NOMA,” *EURASIP Jour. Wireless Comm. and Networking*, vol. 2018, no. 1, pp. 1–54, May 2018.
- [8] H. Joudeh and B. Clerckx, “Robust Transmission in Downlink Multiuser MISO Systems: A Rate-Splitting Approach,” *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.
- [9] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, “A Rate Splitting Strategy for Massive MIMO with Imperfect CSIT,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, Jul. 2016.
- [10] Z. Li, C. Ye, Y. Cui, S. Yang, and S. Shamai, “Rate Splitting for Multi-Antenna Downlink: Precoder Design and Practical Implementation,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1910–1924, Aug. 2020.

- [11] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate Splitting for MIMO Wireless Networks: A Promising PHY-Layer Strategy for LTE Evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.
- [12] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [13] J. Singh, O. Dabeer, and U. Madhow, "On the Limits of Communication with Low-Precision Analog-to-Digital Conversion at the Receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [14] J. Mo and R. W. Heath, "Capacity Analysis of One-Bit Quantized MIMO Systems with Transmitter Channel State Information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Jul. 2015.
- [15] A. Mezghani and J. A. Nossek, "Capacity Lower Bound of MIMO Channels with Output Quantization and Correlated Noise," in *Proc. IEEE Int. Symp. Info. Th.*, Jan. 2012, pp. 1–5.
- [16] O. Orhan, E. Erkip, and S. Rangan, "Low Power Analog-to-Digital Conversion in Millimeter Wave Systems: Impact of Resolution and Bandwidth on Performance," in *Proc. Info. Th. and Appl. Workshop*, Feb. 2015, pp. 191–198.
- [17] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized Precoding for Massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Jul. 2017.
- [18] J. Choi, J. Park, and N. Lee, "Energy Efficiency Maximization Precoding for Quantized Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, Feb. 2021.
- [19] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Finite-Alphabet Precoding for Massive MU-MIMO with Low-Resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4706–4720, Jul. 2018.
- [20] H. Pirzadeh and A. L. Swindlehurst, "Spectral Efficiency of Mixed-ADC Massive MIMO," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3599–3613, May 2018.
- [21] J. Choi, J. Mo, and R. W. Heath, "Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems with One-Bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
- [22] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel Estimation and Performance Analysis of One-Bit Massive MIMO Systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.
- [23] Y. Jeon, N. Hong, and N. Lee, "Supervised-Learning-Aided Communication Framework for MIMO Systems with Low-Resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7299–7313, May 2018.

- [24] J. Choi, Y. Cho, B. L. Evans, and A. Gatherer, “Robust Learning-Based ML Detection for Massive MIMO Systems with One-Bit Quantized Signals,” in *Proc. IEEE Glob. Comm. Conf.* IEEE, Dec. 2019, pp. 1–6.
- [25] H. Joudeh and B. Clerckx, “Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems with Partial CSIT: A Rate-Splitting Approach,” *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Aug. 2016.
- [26] Te Han and K. Kobayashi, “A New Achievable Rate Region for the Interference Channel,” *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [27] R. H. Etkin, D. N. C. Tse, and H. Wang, “Gaussian Interference Channel Capacity to within One Bit,” *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Nov. 2008.
- [28] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of Freedom of Time Correlated MISO Broadcast Channel with Delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
- [29] C. Hao, Y. Wu, and B. Clerckx, “Rate Analysis of Two-Receiver MISO Broadcast Channel with Finite Rate Feedback: A Rate-Splitting Approach,” *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, Sep. 2015.
- [30] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, “Weighted Sum-Rate Maximization Using Weighted MMSE for MIMO-BC Beamforming Design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.
- [31] Z. Yang, M. Chen, W. Saad, and M. Shikh-Bahaei, “Optimization of Rate Allocation and Power Control for Rate Splitting Multiple Access (RSMA),” *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5988–6002, Sep. 2021.
- [32] ———, “Downlink Sum-Rate Maximization for Rate Splitting Multiple Access (RSMA),” in *Proc. IEEE Int. Conf. Comm.*, Jul. 2020, pp. 1–6.
- [33] J. Park, J. Choi, N. Lee, W. Shin, and H. V. Poor, “Rate-Splitting Multiple Access for Downlink MIMO: A Generalized Power Iteration Approach,” *IEEE Trans. Wireless Commun.*, Sep. 2022.
- [34] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, “Rate-Splitting Multiple Access: Fundamentals, Survey, and Future Research Trends,” *IEEE Comm. Surveys Tuts*, Jul. 2022.
- [35] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, “A Primer on Rate-Splitting Multiple Access: Tutorial, Myths, and Frequently Asked Questions,” <https://arxiv.org/abs/2209.00491>, Sep. 2022.
- [36] R. K. Ahiadormey and K. Choi, “Performance Analysis of Rate Splitting in Massive MIMO Systems with Low Resolution ADCs/DACs,” *Applied Sciences*, vol. 11, no. 20, p. 9409, Oct. 2021.

- [37] O. Dizdar, A. Kaushik, B. Clerckx, and C. Masouros, “Rate-Splitting Multiple Access for Joint Radar-Communications with Low-Resolution DACs,” in *Proc. IEEE Int. Conf. Comm.*, Jul. 2021, pp. 1–6.
- [38] —, “Energy Efficient Dual-Functional Radar-Communication: Rate-Splitting Multiple Access, Low-Resolution DACs, and RF Chain Selection,” *IEEE Open J. Commun. Soc.*, vol. 3, pp. 986–1006, 2022.
- [39] Y. Cai, L.-H. Zhang, Z. Bai, and R.-C. Li, “On an Eigenvector-Dependent Nonlinear Eigenvalue Problem,” *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 3, pp. 1360–1382, Sep. 2018.
- [40] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, “Robust Predictive Quantization: Analysis and Design via Convex Optimization,” *IEEE Trans. Signal Process.*, vol. 1, no. 4, pp. 618–632, Dec. 2007.
- [41] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, “Uplink Achievable Rate for Massive MIMO Systems with Low-Resolution ADC,” *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Oct. 2015.
- [42] C. Shen and H. Li, “On the Dual Formulation of Boosting Algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2216–2231, Mar. 2010.
- [43] J. Choi, N. Lee, S. Hong, and G. Caire, “Joint User Selection, Power Allocation, and Precoding Design with Imperfect CSIT for Multi-Cell MU-MIMO Downlink Systems,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 162–176, Sep. 2020.
- [44] H. Joudeh and B. Clerckx, “Rate-Splitting for Max-Min Fair Multigroup Multicast Beamforming in Overloaded Systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7276–7289, Nov. 2017.
- [45] A. Adhikary, J. Nam, J. Ahn, and G. Caire, “Joint Spatial Division and Multiplexing - The Large-Scale Array Regime,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [46] S. Park, J. Choi, J. Park, W. Shin, and B. Clerckx, “Rate-Splitting Multiple Access for Quantized Multiuser MIMO Communications,” *IEEE Trans. Wireless Commun.*, 2022.
- [47] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, “Rate-Splitting Unifying SDMA, OMA, NOMA, and Multicasting in MISO Broadcast Channel: A Simple Two-User Rate Analysis,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, Nov. 2019.
- [48] F. Zafari, A. Gkelias, and K. K. Leung, “A Survey of Indoor Localization Systems and Technologies,” *IEEE Comm. Surveys Tuts*, vol. 21, no. 3, pp. 2568–2599, Apr. 2019.

