Downlink Asynchronous Non-Orthogonal Multiple Access with Quantizer Optimization

Xun Zou, *Student Member, IEEE*, Mehdi Ganji, *Student Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

Abstract

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

Index Terms

Asynchronous non-orthogonal multiple access, limited feedback, quantizer optimization

I. Introduction

Non-orthogonal multiple access (NOMA) has been regarded as one of the key technologies to meet the challenges of the next generation wireless communications. The key feature of NOMA is that different users' signals can share the same time and frequency resources. The NOMA consists of power-domain NOMA, code-domain NOMA, and the joint design of both [1]. In power-domain NOMA, the superposition coding and the successive interference cancellation (SIC) are utilized for multi-user transmission and detection, respectively.

Recently, a novel scheme called asynchronous NOMA (ANOMA) has been proposed to further improve the performance of NOMA, for example, in the uplink system [2] and the cooperative network [3]. In ANOMA, the intentional introduction of timing mismatch at the transmitter and the oversampling technique at the receiver result in the sampling diversity [2–4]. It has been demonstrated that ANOMA outperforms NOMA in terms of the throughput performance, the power consumption, etc. Moreover, the time asynchrony has also been exploited in multi-user transmit beamforming [5].

The channel state information (CSI) plays a critical role in optimizing the system performance. At the transmitter side, the CSI is employed to conduct the adaptive power/rate allocation and generate the beamforming vectors in multiple-input multiple-output (MIMO) systems. In time-division duplexing (TDD) systems, the channel reciprocity is exploited by the base station (BS) to utilize the CSI estimated via the uplink training. In frequency-division duplexing (FDD) systems, a prevailing technique is using the limited feedback from users. The NOMA with limited feedback has been studied in the existing literature, for example, the one-bit feedback scheme in the massive MIMO NOMA systems [6] and the multi-user single antenna systems [7], and the scalar quantizer design in downlink power-domain NOMA [8]. To the best of our knowledge, the analysis of limited feedback schemes in ANOMA systems and the optimal quantizer design for NOMA/ANOMA are still absent. In fact, the limited feedback design in NOMA/ANOMA systems is more challenging compared with that in the orthogonal multiple access (OMA) scenario, e.g., in [9]. It is because the CSI of each user not only affects its own but also other users' performance due to the inter-user interference (IUI) ingrained in the non-orthogonal transmission. In more details, the CSI is used for both allocating powers and determining the SIC order, which further complicates the rate expressions and the system design.

In this letter, we consider a downlink ANOMA system with limited feedback. We employ the max-min criterion for the power allocation and the scalar quantizer for channel quantization, respectively. We derive the closed-form expressions for the upper and lower bounds of the max-min rate. It is manifested that ANOMA can achieve the same or even higher average max-min rate with a lower feedback rate compared with NOMA. Moreover, we propose a quantizer optimization method which applies to both NOMA and ANOMA systems. A gradient descent algorithm is designed to optimize the quantization levels. Simulation results show that a higher average max-min rate is achieved by using the optimized quantizer compared with the conventional uniform quantizer in [8], especially for the low-rate feedback scenario.

II. Preliminaries

A. NOMA and ANOMA

In this letter, we consider a two-user downlink system shown in Fig. 1 where the signals for Users 1 and 2 are superimposed and then transmitted by the BS. Both BS and users are equipped with a single antenna. As shown in Fig. 2, a timing mismatch
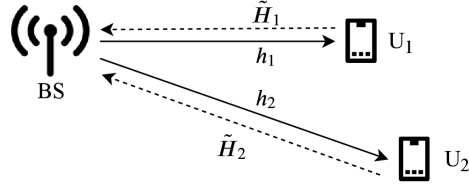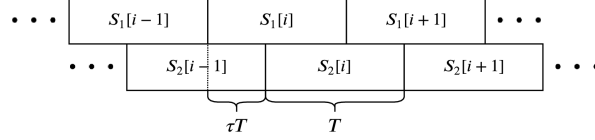
Fig. 1: Downlink scenario with limited feedback.

Fig. 2: Illustration of the superimposed signal in ANOMA.

of τT is intentionally added between the superimposed signals, where $0 < \tau < 1$ and T is the symbol interval. Let $p(t)$ represent the pulse shape. The transmitted signal is given by $S_1[i]p(t - iT) + S_2[i]p(t - iT - \tau T)$ where $S_u[i]$, $u = 1, 2$, is the ith symbol transmitted to User u. We assume that users know τ perfectly since it can be decided a priori, for example, to be the asymptotically optimal value of $\tau = 0.5$ [2, 3]. The block fading channel model is employed. At the receiver, the composite signal is sampled at iT and $(i + \tau)T$ after matched filtering. This sampling method is called "oversampling" and the details have already been presented in [2, 3]. For the sake of brevity, we omit it in this work. If $\tau = 0$, ANOMA degrades to NOMA and the received samples at iT and $(i + \tau)T$ are identical. The number of received samples is doubled in ANOMA ($\tau \neq 0$) compared with NOMA, which then results in sampling diversity [2–4].

We assume that Users 1 and 2 denote the strong and weak users, respectively, i.e., $|h_1|^2 > |h_2|^2$. User 1 employs the block-wise SIC [3], i.e., first detects the block of the signal for User 2, removes it, and then detects its own signal. As shown in [3], for a relatively large block length and the rectangular pulse shape, the rates of Users 1 and 2 with perfect CSI are given by

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

and (2) at the bottom of this page, $H_i = |h_i|^2$ is the channel gain of User i, $Q = 2\tau(1 - \tau)$, P is the total transmit power of BS, $\alpha \in (0, 1)$ is the power coefficient for the strong user (User 1 in this case), i.e., the powers allocated to Users 1 and 2 are αP and $(1 - \alpha)P$, respectively. Note that NOMA can be considered as a special case of ANOMA, simply by setting $\tau = 0$ in (2). It has been shown that R_{weak} in ANOMA is higher than that in NOMA while R_{strong} is the same for ANOMA and NOMA [3].

B. Limited Feedback

In this letter, we assume that CSI is perfectly estimated by users and fed back to the BS via an error and delay-free link. For example, in Fig. 1, User 1 knows h_1 and quantizes the channel gain H_1 as $\tilde{H}_1 = q(|h_1|^2)$ using a scalar quantizer q, and then sends it back to the BS. The BS determines the power coefficient α and the order of SIC according to the feedback. If $\tilde{H}_1 > \tilde{H}_2$ ($\tilde{H}_1 < \tilde{H}_2$), User 1 (User 2) is notified to conduct SIC. If $\tilde{H}_1 = \tilde{H}_2$, BS can randomly assign one user to utilize SIC. Without loss of generality, we assume that User 1 will be notified to conduct SIC if $\tilde{H}_1 = \tilde{H}_2$.

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

Fig. 3: Illustration of a scalar quantizer.

As shown in Fig. 3, we employ a scalar quantizer with $N = 2^b$ quantization levels, $q_0, \cdots, q_{N-1}$, where b is the number of bits used in a quantization codeword. b is also defined as the feedback rate per user. The quantized value for a given x is calculated by

$$q(x) = \begin{cases} q_i, & q_i \le x < q_{i+1}, \ i = 0, \cdots, N-2, \\ q_{N-1}, & x \ge q_{N-1}. \end{cases} \tag{3}$$

For example, in Fig. 3, x_0 and x_1 are quantized as q_0 and q_{N-1}, respectively. In this work, the quantizer is used to quantize the positive channel gain. Thus, we set $q_0 = 0$.

The BS transmits the signals to users based on the rates which are calculated according to the quantized channel gains. For example, if $\tilde{H}_1 > \tilde{H}_2$, the BS transmits to User 1 with the rate of $R_{\text{strong}}(\tilde{H}_1)$ while the actual channel capacity is $R_{\text{strong}}(H_1)$. Note that the proposed quantizer is designed to satisfy $q(x) \le q$ in order to avoid outage, i.e, to guarantee that the transmission rates do not exceed the channel capacities. It is trivial to derive that $R_{\text{strong}}(\tilde{H}_1) \le R_{\text{strong}}(H_1)$ if $\tilde{H}_1 \le H_1$ and $R_{\text{weak}}(\tilde{H}_2) \le R_{\text{weak}}(H_2)$ if $\tilde{H}_2 \le H_2$, which indicates that the quantizer in (3) avoids outage.

III. POWER ALLOCATION

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

$$\alpha^* = \arg\max_{\alpha} \min \{R_{\text{strong}}, R_{\text{weak}}\}. \tag{4}$$

According to (1) and (2), it is trivial to show that R_{strong} is an increasing function of α and R_{weak} is a decreasing function of α. Intuitively, if User 1 is the strong user, by increasing α (i.e., allocating more power to User 1 and less power to User 2), R_{strong} increases and R_{weak} decreases, and vice versa. As a result, the optimal power coefficient α^* can be obtained by solving $R_{\text{strong}} = R_{\text{weak}}$. According to [8], the optimal α in NOMA is given by

$$\alpha_{\text{N}}^* = \frac{2\tilde{H}_{\min}}{\sqrt{(\tilde{H}_1 + \tilde{H}_2)^2 + 4P\tilde{H}_1\tilde{H}_2\tilde{H}_{\min}} + \tilde{H}_1 + \tilde{H}_2}, \tag{5}$$

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

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

Theorem 1: The optimal power coefficients of NOMA and ANOMA, α_{N}^* and α_{A}^*, respectively, satisfy the following inequality

$$\alpha_{\text{N}}^* \le \alpha_{\text{A}}^*(0.5) \le \alpha_{\text{A}}^* \le \alpha_{\text{A}}^*(1), \tag{6}$$

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

$$\alpha_{\text{A}}^*(z) = \frac{2\tilde{H}_{\min}}{\sqrt{(\tilde{H}_1 + \tilde{H}_2 - zP\tilde{H}_{\min}^2 Q)^2 + 4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2 + zP\tilde{H}_{\min}^2 Q)} + \tilde{H}_1 + \tilde{H}_2 - zP\tilde{H}_{\min}^2 Q}. \tag{7}$$

$\qquad$ *Proof:* See Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\blacksquare$

Since the general expression for α_{A}^* is intractable for further analysis, $\alpha_{\text{A}}^*(0.5)$ and $\alpha_{\text{A}}^*(1)$ provide the simple lower and upper bounds of α_{A}^*, respectively. Compared with (5), (7) introduces an extra term $zP\tilde{H}_{\min}^2 Q$. By setting $\tau = 0$, $Q = 0$ and then (7) coincides with (5). The max-min rate of ANOMA is given by $R_{\text{A}}^*(\tilde{H}_1, \tilde{H}_2) = \log\left(1 + \alpha_{\text{A}}^* P\tilde{H}_{\max}\right) \in \left[\log\left(1 + \alpha_{\text{A}}^*(0.5)P\tilde{H}_{\max}\right), \log\left(1 + \alpha_{\text{A}}^*(1)P\tilde{H}_{\max}\right)\right]$.

The average max-min rate is expressed by

$$\mathbb{E}[R^*] = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \int_{q_{i,1}}^{q_{i+1,1}} \int_{q_{j,2}}^{q_{j+1,2}} R^*(q_{i,1}, q_{j,2}) \cdot f_1(H_1) f_2(H_2) \mathrm{d}H_1 \mathrm{d}H_2, \tag{8}$$

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

$$\mathbb{E}[R^*] < \overline{R}^* \triangleq \int_0^\infty \int_0^\infty R^*(H_1, H_2) f_1(H_1) f_2(H_2) \mathrm{d}H_1 \mathrm{d}H_2. \tag{9}$$

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

Corollary 1: ANOMA can achieve the average max-min rate of NOMA with a lower feedback rate.

Proof: Let us define q and q' as two quantizers which can be given by (3) but with different quantization levels. According to Theorem 1, ANOMA achieves a higher max-min rate compared with NOMA. Thus, $\overline{R}_{\mathrm{N}}^* < \overline{R}_{\mathrm{A}}^*$ and $\mathbb{E}\left[R_{\mathrm{N}}^*\right] < \mathbb{E}\left[R_{\mathrm{A}}^*\right]$ by using the quantizer q. The quantizer q' is designed such that the ANOMA using q' achieves the same average max-min rate as the NOMA using q, i.e., $\mathbb{E}\left[R_{\mathrm{A}}^*\right]' = \mathbb{E}\left[R_{\mathrm{N}}^*\right] < \mathbb{E}\left[R_{\mathrm{A}}^*\right]$. For ANOMA, using the quantizer q' results in a higher distortion compared with using q, i.e., $D\left[R_{\mathrm{A}}^*\right]' = \overline{R}_{\mathrm{A}}^* - \mathbb{E}\left[R_{\mathrm{A}}^*\right]' > D\left[R_{\mathrm{A}}^*\right] = \overline{R}_{\mathrm{A}}^* - \mathbb{E}\left[R_{\mathrm{A}}^*\right]$. According to the rate-distortion theory, there is a trade-off between the distortion and the quantization rate (equivalent to the feedback rate in this work). A lower distortion can be achieved by using a quantizer with a higher feedback rate and vice versa. As a result, the quantizer q' can have a lower feedback rate compared with q. ANOMA using q' can achieve a lower feedback rate while keeping the same average max-min rate as NOMA using q. The proof is complete. ∎

IV. SCALAR QUANTIZER OPTIMIZATION

The scalar quantizer shown in (3) is completely characterized by the quantization levels. Our goal is to optimize the quantization levels to maximize the average max-min rate $\mathbb{E}\left[R^*\right]$, i.e.,

$$
\begin{aligned}
\left[\mathbf{q}_1^*, \mathbf{q}_2^*\right] = \underset{\left[\mathbf{q_1}, \mathbf{q_2}\right]}{\arg\ \max}\, \mathbb{E}\left[R^*\right], \\
s.t.\ q_{0,i} < q_{1,i} < \cdots < q_{N_i-1,i},\ i = 1 \text{ or } 2.
\end{aligned}
\tag{10}
$$

Since $\overline{R}^*$ is not a function of quantization levels, (10) is equivalent to minimizing the distortion $D[R^*]$. In what follows, we propose a gradient descent algorithm to optimize the quantization levels. Let $\mathbb{E}\left[R^*\right]_{i,j}$ denote the (i,j)th term of $\mathbb{E}\left[R^*\right]$ in (8), i.e., $\mathbb{E}\left[R^*\right]_{i,j} = R^*(q_{i,1}, q_{j,2}) \int_{q_{i,1}}^{q_{i+1,1}} \int_{q_{j,2}}^{q_{j+1,2}} f_1(H_1) f_2(H_2) \mathrm{d}H_1 \mathrm{d}H_2$. $\mathbb{E}\left[R^*\right]_{i,j}$ is a function of $q_{i,1}, q_{i+1,1}, q_{j,2}$, and $q_{j+1,2}$. The gradients of $\mathbb{E}\left[R^*\right]_{i,j}$ in terms of $q_{i,1}$ and $q_{i+1,1}$ are calculated by

$$
\frac{\partial \mathbb{E}\left[R^*\right]_{i,j}}{\partial q_{i,1}} = \left[\frac{\partial R^*(q_{i,1}, q_{j,2})}{\partial q_{i,1}} \int_{q_{i,1}}^{q_{i+1,1}} f_1(H_1)\mathrm{d}H_1 - R^*(q_{i,1}, q_{j,2})f_1(q_{i,1}) \right] \int_{q_{j,2}}^{q_{j+1,2}} f_2(H_2)\mathrm{d}H_2,
\tag{11}
$$

$$
\frac{\partial \mathbb{E}\left[R^*\right]_{i,j}}{\partial q_{i+1,1}} = R^*(q_{i,1}, q_{j,2})f_1(q_{i+1,1}) \int_{q_{j,2}}^{q_{j+1,2}} f_2(H_2)\mathrm{d}H_2,
\tag{12}
$$

respectively. Similarly, we can derive $\frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j,2}}$ and $\frac{\partial \mathbb{E}[R^*]_{i,j}}{\partial q_{j+1,2}}$. Based on the gradients, we propose the quantizer optimization algorithm in Algorithm 1. At each iteration, $\mathbb{E}\left[R^*\right]$ does not decrease which is guaranteed by the gradient descent. Besides, $\mathbb{E}\left[R^*\right]$ is upper bounded by a constant as shown in (9). Hence, Algorithm 1 converges as the number of iterations increases.

Algorithm 1 Algorithm to optimize the scalar quantizer.

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

The computation complexity of Algorithm 1 is $O(N_1 N_2)$ where N_1 and N_2 are the number of quantization levels for Users 1 and 2, respectively. Furthermore, the maximum quantization level of the conventional uniform quantizer is set manually according to certain criterion. For example, in [8], the maximum quantization level is determined by considering the quantization loss. The advantage of optimizing the quantization levels is that the maximum quantization level can also be optimized using Algorithm 1 with no manual intervention, which will be shown in the next section.

Fig. 4: The average max-min rate vs. the number of feedback bits per channel for NOMA and ANOMA systems.

V. SIMULATION RESULTS

In this section, we present simulation results for NOMA/ANOMA systems with limited feedback. In our simulations, we employ the complex Gaussian channel model. Therefore, the channel gain follows the exponential distribution. We assume that $H_1 \sim \mathrm{Exp}(0.5)$ and $H_2 \sim \mathrm{Exp}(1)$. We set the total transmit power of the BS $P = 10$ and $\tau = 0.5$ since it has been proved in [2, 3] that $\tau = 0.5$ is the asymptotically optimal value to maximize the user throughput. For comparison, we employ the uniform quantizer proposed in [8] where the maximum quantization level L is derived by solving $L = \frac{1}{\lambda \Delta} \log\left(\frac{1}{\Delta}\right)$, λ is the parameter of the exponential distribution, and Δ is the quantization bin width.

Fig. 4 shows how the average max-min rate changes as a function of the number of feedback bits per user using the uniform quantizer. As the number of feedback bits increases, the average max-min rate converges to the full-CSI rate calculated by (9). Besides, using the same number of feedback bits, the upper bound of the max-min rate for ANOMA is always higher than its lower bound, which is then higher than that for NOMA. Equivalently, to achieve the same or even higher average max-min rate, ANOMA needs less feedback bits compared with NOMA, which verifies Corollary 1.

Fig. 5 shows how the average max-min rate in NOMA and ANOMA systems changes as the quantizer optimization algorithm runs. We employ the 3-bit quantizer for each channel and the lower bound of the max-min rate for ANOMA as an example. As the number of iterations increases, the average max-min rate converges. For the full-CSI case, the optimized quantizer, and the uniform quantizer, the average max-min rate for ANOMA is always higher than that for NOMA.
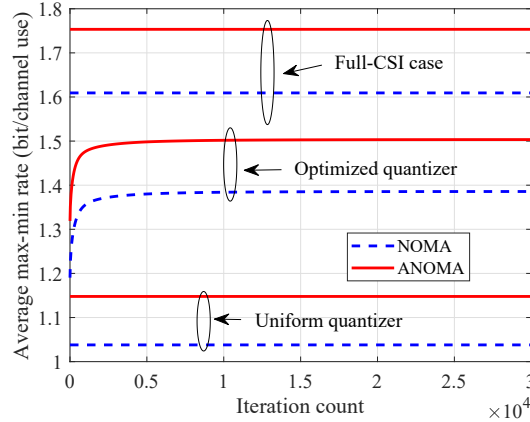

Fig. 5: The average max-min rate vs. the iteration count for NOMA and ANOMA systems.
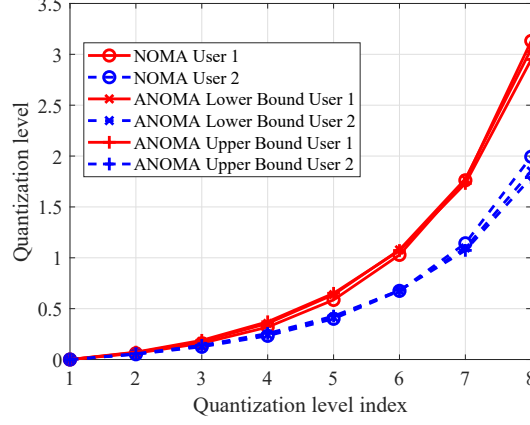
Fig. 6: The quantization levels for NOMA and ANOMA systems when 3 bits are used to quantize the channel gain.

Fig. 6 presents the optimal quantization levels when 3 bits (i.e., $2^3 = 8$ quantization levels) are used. First, it is shown that the optimal quantization levels are non-uniform. The range of the quantization levels for User 1 with a larger mean channel gain is wider than that of User 2. Furthermore, the quantization levels optimized using the upper and lower bounds of the max-min rate for ANOMA are very similar to each other and also close to those for NOMA.

VI. CONCLUSION

In this letter, we considered a two-user downlink ANOMA system with limited feedback. We reveal the advantage of ANOMA over NOMA in terms of the feedback rate. Furthermore, the quantizer optimization method is proposed for both ANOMA and NOMA to further improve the average max-min rate. Considering more than two users in the ANOMA systems is an interesting topic for the future work.

APPENDIX A
PROOF OF THEOREM 1

Proof: When $\tau = 0$, $Q = 0$ which then results in $\alpha_N^* = \alpha_A^*(x)$ for any finite x. For $\tau \neq 0$, we first prove Theorem 1 for the case of $\tilde{H}_1 \geq \tilde{H}_2$. The case of $\tilde{H}_1 < \tilde{H}_2$ SIC will be discussed later. By setting $R_{\text{strong}}(\tilde{H}_1) = R_{\text{weak}}(\tilde{H}_2)$, we obtain

$$2[1 + \alpha_A^* P(\tilde{H}_1 + \tilde{H}_2) + (\alpha_A^*)^2 P^2 \tilde{H}_1 \tilde{H}_2]$$
$$= \sqrt{[1 + P\tilde{H}_2 + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q]^2 - [\alpha_A^*(1 - \alpha_A^*)]^2 P^4 \tilde{H}_2^4 Q^2}$$
$$+ 1 + P\tilde{H}_2 + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q. \tag{13}$$

By cancelling out the square root, (13) becomes a quartic equation. The optimal power coefficient α_A^* is one root of the quartic equation which can be given by the general formula for quartic roots. However, it is intractable for further analysis. Therefore, we derive the upper and lower bounds to approximate the actual value of α_A^*.

For the upper bound, since $[1 + P\tilde{H}_2 + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q]^2 - (\alpha_A^*)^2(1 - \alpha_A^*)^2 P^4 \tilde{H}_2^4 Q^2 < [1 + P\tilde{H}_2 + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q]^2$, (13) becomes

$$1 + \alpha_A^* P(\tilde{H}_1 + \tilde{H}_2) + (\alpha_A^*)^2 P^2 \tilde{H}_1 \tilde{H}_2 < 1 + P\tilde{H}_2 + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q, \tag{14}$$

which results in

$$\alpha_A^* < \alpha_U^* \triangleq 2\tilde{H}_2 / \left[\tilde{H}_1 + \tilde{H}_2 - P\tilde{H}_2^2 Q + \sqrt{(\tilde{H}_1 + \tilde{H}_2 - P\tilde{H}_2^2 Q)^2 + 4\tilde{H}_2(P\tilde{H}_1 \tilde{H}_2 + P\tilde{H}_2^2 Q)} \right]. \tag{15}$$

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

$$2[1 + \alpha_A^* P(\tilde{H}_1 + \tilde{H}_2) + (\alpha_A^*)^2 P^2 \tilde{H}_1 \tilde{H}_2] > 2(1 + P\tilde{H}_2) + \alpha_A^*(1 - \alpha_A^*)P^2 \tilde{H}_2^2 Q. \tag{16}$$

Then, we obtain

$$\alpha_A^* > \alpha_L^* \triangleq 2\tilde{H}_2 / \left[\tilde{H}_1 + \tilde{H}_2 - \frac{P\tilde{H}_2^2 Q}{2} + \sqrt{\left(\tilde{H}_1 + \tilde{H}_2 - \frac{P\tilde{H}_2^2 Q}{2}\right)^2 + 4\tilde{H}_2\left(P\tilde{H}_1 \tilde{H}_2 + \frac{P\tilde{H}_2^2 Q}{2}\right)} \right]. \tag{17}$$

If $\tilde{H}_1 < \tilde{H}_2$, i.e., User 2 employs SIC, we can also derive the expressions for α_L^* and α_U^* by setting $R_{\text{strong}}(\tilde{H}_2) = R_{\text{weak}}(\tilde{H}_1)$. In fact, α_L^* and α_U^* for $\tilde{H}_1 < \tilde{H}_2$ are given by simply switching $\tilde{H}_1$ and $\tilde{H}_2$ in (15) and (17). Both α_L^* and α_U^* can be incorporated into a general expression in (7) by introducing a parameter z, i.e. $\alpha_L^* = \alpha_A^*(0.5)$ and $\alpha_U^* = \alpha_A^*(1)$.

To show the inequality in (6), let us define $g(x) = \sqrt{(\tilde{H}_1 + \tilde{H}_2 - xP\tilde{H}_{\min}^2 Q)^2 + 4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2 + xP\tilde{H}_{\min}^2 Q)} + \tilde{H}_1 + \tilde{H}_2 - xP\tilde{H}_{\min}^2 Q$ which is the denominator in (7). Then,

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

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

$$-\tilde{H}_1 - \tilde{H}_2 + \tilde{H}_{\min} < 0 < P\tilde{H}_1\tilde{H}_2$$
$$\Longrightarrow 4\tilde{H}_{\min}(-\tilde{H}_1 - \tilde{H}_2 + \tilde{H}_{\min} + xP\tilde{H}_{\min}^2 Q) < 4\tilde{H}_{\min}(P\tilde{H}_1\tilde{H}_2 + xP\tilde{H}_{\min}^2 Q)$$
$$\Longrightarrow |\tilde{H}_1 + \tilde{H}_2 - \tilde{H}_{\min}(2 + xP\tilde{H}_{\min}Q)| < \sqrt{(\tilde{H}_1 + \tilde{H}_2 - xP\tilde{H}_{\min}^2 Q)^2 + 4P\tilde{H}_{\min}(\tilde{H}_1\tilde{H}_2 + x\tilde{H}_{\min}^2 Q)},$$

which also results in $\frac{\partial g(x)}{\partial x} < 0$. Therefore, $\alpha_A^*(x)$ increases with x. Thus, $\alpha_N^* = \alpha_A^*(0) < \alpha_A^*(0.5) < \alpha_A^* < \alpha_A^*(1)$. The proof is complete. ∎

REFERENCES

[1] X. Zou, M. Ganji, and H. Jafarkhani, "Trellis-coded non-orthogonal multiple access," vol. 9, no. 4, pp. 538–542, Apr. 2020.

[2] X. Zou, B. He, and H. Jafarkhani, "An analysis of two-user uplink asynchronous non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1404–1418, Feb. 2019.

[3] X. Zou, M. Ganji, and H. Jafarkhani, "Cooperative asynchronous non-orthogonal multiple access with power minimization under QoS constraints," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1503–1518, Mar. 2020.

[4] M. Ganji and H. Jafarkhani, "Improving NOMA multi-carrier systems with intentional frequency offsets," vol. 8, no. 4, pp. 1060–1063, Aug. 2019.

[5] M. Ganji, X. Zou, and H. Jafarkhani, "Exploiting time asynchrony in multi-user transmit beamforming," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3156–3169, May 2020.

[6] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.

[7] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with 1-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[8] X. Liu and H. Jafarkhani, "Downlink non-orthogonal multiple access with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 6151–6164, Sep. 2017.

[9] E. Karamad, B. Khoshnevis, and R. S. Adve, "Quantization and bit allocation for channel state feedback in relay-assisted wireless networks," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 327–339, Jan. 2013.