Closed-form Symbol Error Rate Expressions for Non-orthogonal Multiple Access Systems

Qinwei He, *Student Member, IEEE,* Yulin Hu, *Senior Member, IEEE,* Anke Schmeink, *Senior Member, IEEE,*

Abstract

In this work, the reliability performance of a non-orthogonal multiple access (NOMA) system is addressed. We characterize the reliability performance of the users in a downlink power domain NOMA system with arbitrary ordered pulse-amplitude modulation (PAM) and quadrature amplitude modulation (QAM). In particular, the exact symbol error rate (SER) expressions are derived. Furthermore, by taking the power allocation between users into account, a reliability optimization problem is formulated, which minimizes the SER of the user with a good channel quality while guaranteeing the reliability requirement for the user with a relatively poor channel. Based on the characterized SER model, a suboptimal power allocation approach for this problem is provided with a performance extremely close to the optimal solution. Through simulations, we validate the provided SER characterization and confirm the effectiveness of the proposed suboptimal power allocation policy.

Index Terms

Non-orthogonal multiple access (NOMA), power allocation, reliability, symbol error rate (SER).

I. Introduction

Future wireless communications are expected to provide significant high throughput and support connections for massive devices. To meet these requirements, numerous transmission and multiple access techniques have been proposed and investigated, such as massive multiple-input multiple-output (MIMO) [1], millimeter-wave (mmWave) [2], non-orthogonal multiple access (NOMA) [3]–[5], and sparse code multiple access (SCMA) [6]. Among these techniques, NOMA has been considered a key promising radio access scheme as it effectively improves the network performances with respect to spectrum efficiency [7], [8], energy efficiency [9] and security [10]. More importantly, NOMA can be combined with many existing and emerging techniques, such as MIMO [11], cognitive radio (CR) [12] and simultaneous wireless information and power transfer (SWIPT) [9], [13], to further enhance the system performance under different communication scenarios.

In comparison to the conventional orthogonal multiple access (OMA), NOMA exploits the channel gain differences between users by employing the successive interference cancellation (SIC) and thus, schedules multiple users non-orthogonally on the same time-spectrum resource. To take full advantage of a NOMA system, resource allocation

Q. He, Y. Hu, and A. Schmeink are with ISEK Research Group, RWTH Aachen University, Kopernikusstr. 16, 52074 Aachen, Germany (email:$he|hu|schmeink$@ti.rwth-aachen.de). Y. Hu is the corresponding author.

policies, including power allocation and channel assignment, have been widely considered in the system design [14]. Various resource allocation policies have been proposed for NOMA systems to improve/optimize different objectives, such as fairness among all users [15], [16], quality-of-service (QoS) [17], [18], and energy efficiency [19]–[21]. However, the above studies regarding the power allocation design generally consider the system throughput or energy efficiency (in the design of guaranteeing fairness or QoS), while the power allocation design which can optimize the system reliability performance, i.e., minimize the error rate for each user according to the service requirement, has not been addressed so far.

Clearly, having closed-form error probability expressions facilitates the reliability-optimal power allocation design of such NOMA systems. However, this has not been well studied so far. On the one hand, only a few works address this issue. On the other hand, these existing error probability expressions of NOMA users are conducted under scenarios with specific assumptions. For example, closed-form bit error rate (BER) expressions of the quadrature phase-shift keying (QPSK) based NOMA system are derived in [22]. In addition, the authors in [23] study the analytical symbol error rate (SER) of NOMA users in a visible light communications (VLC) network where the line of sight (LOS) link is always feasible. However, the reliability characterisation of NOMA users with more general assumptions is missing, i.e., for networks which are not of the specific VLC transmission, with general pathloss and channel assumptions, and with general modulation schemes, such as pulse-amplitude modulation (PAM) and quadrature amplitude modulation (QAM). Hence, characterizing the reliability and providing the closed-form error probability expressions for general NOMA systems with arbitrary modulation orders is of significant interest. More importantly, this enables us to analytically investigate the optimal power allocation design for the NOMA system when considering the reliability requirements of the users.

In this work, we aim at deriving the analytical SER expressions for general NOMA systems and providing the closed-form solution to the power allocation optimization problem while guaranteeing the reliability requirements of the NOMA users. First of all, by analyzing the constellation of the superposed symbol in the PAM based NOMA system, the exact SER expression for each user with arbitrary PAM order is obtained. In addition, considering the relationship between the PAM and QAM modulations, the exact SER expressions for the NOMA system with arbitrary QAM size are also derived. Secondly, we study a reliability optimal design for the considered NOMA system. After investigating the impact of power allocation parameters on users' SERs, we formulate the optimization problem, which maximize the reliability of the user with a good channel quality while guaranteeing the reliability requirement of the user with a relatively poor channel. A closed-form suboptimal power allocation solution for this problem is provided, which is shown to be extremely close to the optimal one. By simulations, we validate our analytical SER expressions and demonstrate the excellent performance of the proposed suboptimal power allocation policy.

The remainder of the paper is organized as follows. In Section II, we introduce the model of the considered NOMA system. The general SER analysis of the NOMA systems with arbitrary PAM and QAM schemes are provided in Section III and IV, respectively. In Section V, we study the influence of power allocation on the SER and propose our power allocation design which satisfies the reliability requirements of the users close to the optimum. In Section VI, We validate the exact SER expressions and evaluate the proposed power allocation approach via simulations.

Finally, we conclude the paper with a summary of our key contributions in Section VII.

II. System model

We consider a NOMA system, including a source and two users. The channels between the source and users are assumed to experience quasi-static fading, i.e., channels are constant within each transmission and vary from one transmission to the next. Without loss of generality, in each transmission, we always call the user with the relatively poor channel quality by User 1 and the one with the relatively better channel by User 2. During a transmission, two different information symbols s_1 and s_2 are required to be sent from the source to User 1 and User 2, respectively. The system operates under a NOMA scheme where these two symbols are transmitted simultaneously, i.e., the source transmits the superposition of the scaled symbols to both users at the same time after allocating transmit powers to these two symbols. Denote by p_1 and p_2 the transmit powers allocated for these two messages, then the superposed symbol z at the source side is given by

$$z = \sqrt{p_1}s_1 + \sqrt{p_2}s_2, \tag{1}$$

where s_1 and s_2 are encoded with the same modulation scheme, i.e., PAM or QAM. In addition, s_1 and s_2 are assumed with unit (average) power and the corresponding alphabet of these two constellations are denoted as $\mathcal{A}_1$ and $\mathcal{A}_2$. Thus, the energy of z is $E_s = p_1 + p_2$. The source allocates p_1 and p_2 for the two symbols according to a power allocation parameter γ. In particular, γ indicates the ratio of the power allocated to User 1 to the total power, i.e., $\gamma = p_1/E_s$.

Generally, the constellation of the superposed symbol z can be obtained by superimposing the constellation of User 2 on top of the constellation of User 1, which is influenced by the modulation methods of the two users. Fig. 1 and Fig. 4 offer two examples of the formation process for PAM and QAM based superposed symbols, respectively. Furthermore, the corresponding constellation points of these two kinds of superposed symbols will be analyzed in details in Section III-A and Section IV-A accordingly.

In the downlink NOMA system, the superposed symbol z is transmitted via two independent fading channels to User 1 and User 2. The received symbols at the two users are expressed as

$$\begin{aligned} y_1 &= h_1\left(\sqrt{p_1}s_1 + \sqrt{p_2}s_2\right) + w_1, \\ y_2 &= h_2\left(\sqrt{p_1}s_1 + \sqrt{p_2}s_2\right) + w_2, \end{aligned} \tag{2}$$

where w_1 and w_2 are either real or complex white Gaussian random noises depending on the modulation schemes. For the PAM based NOMA system, we assume that $w_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $w_2 \sim \mathcal{N}(0, \sigma_2^2)$. In addition, for the QAM based system, the real and imaginary parts of the noise for the same user are independent and identically distributed (i.i.d.) Gaussian random variables, thus, we have $\Re\{w_1\}$ and $\Im\{w_1\} \sim \mathcal{N}(0, \sigma_1^2)$, as well as $\Re\{w_2\}$ and $\Im\{w_2\} \sim \mathcal{N}(0, \sigma_2^2)$. Further, h_1 and h_2 are the corresponding channel coefficients from the source to the two users. Recall that we denote the user with the relatively low channel gain by User 1. Hence, in this work $|h_1| \leq |h_2|$

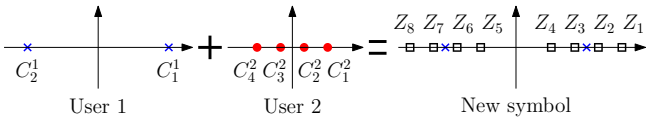

Fig. 1. The construction of the superposed symbol

always holds. The equivalent model of (2) can be achieved by dividing both sides by h_1 and h_2, respectively, which gives

$$\begin{aligned} r_1 &= \frac{y_1}{h_1} = (\sqrt{p_1}s_1 + \sqrt{p_2}s_2) + \hat{w}_1, \\ r_2 &= \frac{y_2}{h_2} = (\sqrt{p_1}s_1 + \sqrt{p_2}s_2) + \hat{w}_2, \end{aligned} \tag{3}$$

where $\hat{w}_1 = \frac{w_1}{h_1}$ and $\hat{w}_2 = \frac{w_2}{h_2}$ are scaled real or complex white Gaussian random noises. Obviously, the mean values of $\hat{w}_1$ and $\hat{w}_2$ remain 0 and their variances for the PAM based system or on one dimension for the QAM based NOMA are $\frac{N_1}{2} = \frac{\sigma_1^2}{\|h_1\|_2^2}$ and $\frac{N_2}{2} = \frac{\sigma_2^2}{\|h_2\|_2^2}$, respectively. It is worth mentioning that the key idea of the NOMA system is to allocate more power to the user with poorer channel conditions [5], thus, $p_1 > p_2$ always holds in our system.

In the NOMA system, the two users decode the received symbols in different manners. On the one hand, User 1 directly decodes s_1 by treating s_2 as interference. Thus, the estimation of s_1 at User 1 is obtained by the minimum Euclidean distance criterion and given by

$$\hat{s}_1 = \arg \min_{s_1 \in \mathcal{A}_1} \|r_1 - \sqrt{p_1}s_1\|_2^2, \tag{4}$$

where $\|\cdot\|_2$ represents the Euclidean norm. The simplest way to achieve $\hat{s}_1$ from r_1 is to demodulate the power normalized r_1 directly, i.e., $\hat{s}_1 = \mathcal{Q}_1(\frac{r_1}{\sqrt{p_1}})$, where $\mathcal{Q}_1(\cdot)$ denotes the corresponding PAM or QAM demodulation function for User 1. On the other hand, SIC is employed when User 2 retrieves its symbol. In particular, the decoding process for User 2 has the following steps: First, the estimation of s_1 at User 2 is obtained similar to (4), i.e., $\hat{s}_1' = \arg\min_{s_1 \in \mathcal{A}_1} \left\|r_2 - \sqrt{p_1}s_1\right\|_2^2$. Afterwards, the interference term from User 1 can be subtracted from the received symbol, which gives us the updated received symbol $\tilde{r}_2 = r_2 - \sqrt{p_1}\hat{s}_1'$. Finally, by utilizing the minimum distance criterion, the estimation of s_2 for User 2 can be achieved from $\tilde{r}_2$ by

$$\hat{s}_2 = \arg \min_{s_2 \in \mathcal{A}_2} \|\tilde{r}_2 - \sqrt{p_2}s_2\|_2^2. \tag{5}$$

III. SERs of NOMA users with PAM symbols

In this section, we study SER expressions for a PAM based NOMA system. In Section III-A, the superposed symbol z given in Equation (1) is characterized first. Based on that, we derive the exact closed-form SERs of User 1 and User 2 in Section III-B and Section III-C, respectively. Furthermore, to make our analysis and derivations easier to understand, we start each of these sections with a simple example and then extend the analysis and derivation process to a general case.

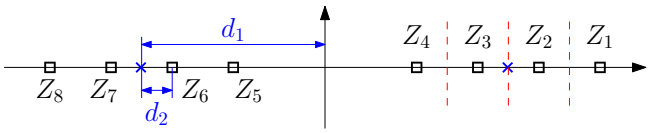

Fig. 2. The superposed symbol of 4-PAM on 2-PAM

A. Analysis of the superposed PAM symbol

We consider a general scenario that the constellation orders of PAM schemes can be different for encoding the messages for the two users. We assume that the messages for User 1 and User 2 are encoded via N-PAM and M-PAM schemes, respectively, i.e., N and M are not necessarily the same. In addition, without loss of generality, we assume that for each user the constellation points have the same occurrence probability. Denote the i^{th} constellation points of User 1 and User 2 by C_i^1 and C_i^2, respectively. Then, we have $P(C_i^1) = \frac{1}{N}$, $i = 1, ..., N$ for User 1 and $P(C_i^2) = \frac{1}{M}$, $i = 1, ..., M$ for User 2. Under the above assumption, in the following we study the superposed symbol of the PAM based NOMA system.

According to (1), the transmitted symbol, which is a superposition of an N-PAM symbol and an M-PAM symbol, can be treated as a PAM-liked symbol with a higher constellation order. For example, if 2-PAM and 4-PAM are utilized by User 1 and User 2, respectively, then the corresponding transmitted symbols is an 8-PAM-liked symbol and the construction of it is illustrated on the left of Fig. 1. The new constellations of the superposed symbol are illustrated by the black squares on the right side of Fig. 1, denoted as $Z_1, Z_2, \ldots, Z_8$.

The detailed constellation of the superposed symbol in Fig. 1 is provided in Fig. 2. As the constellation is symmetric with respect to the y-axis, we only show the decision region of each point on the right side, which is divided by the red dashed line. In the figure, d_1 and d_2 are the distant parameters of the two users, which are influenced by the transmit powers (p_1 and p_2) and constellation orders (N and M) of the users. In this example (N=2, M=4), we have $d_1 = \sqrt{p_1}$ and $d_2 = \sqrt{p_2/5}$.

Based on the above example, we further discuss the superposed symbol with general modulation orders, i.e., N-PAM for User 1 and M-PAM for User 2. In fact, the difference between the general case and the above special case (N=2, M=4) is on the distant parameters. In particular, we have $d_1 = \sqrt{\frac{3p_1}{(N-1)(N+1)}}$ and $d_2 = \sqrt{\frac{3p_2}{(M-1)(M+1)}}$ for the general case, where the fraction part in the square root is obtained from the power normalization factor for the PAM modulation. Moreover, the decision regions of the constellation points of the superposed symbol should be prevented from overlapping, e.g., the decision regions of Z_4 and Z_5 in Fig. 2 will be overlapped if the distant constraint is not fulfilled. Hence, d_1 and d_2 have to satisfy the constraint $d_1 > (M-1)d_2$. Finally, all the possible constellation points (Z_i, $i = 1, ..., MN$) of the superposed symbol have the same occurrence probability, i.e., $P(Z_i) = P(C_i^1) \cdot P(C_i^2) = \frac{1}{MN}$.

B. SER of User 1

We start also with discussing the example given in Fig. 1 before deriving the general SER expression. Assuming the symbol Z_4 is transmitted, then the corresponding information symbol of User 1 conveyed by it is C_1^1. By

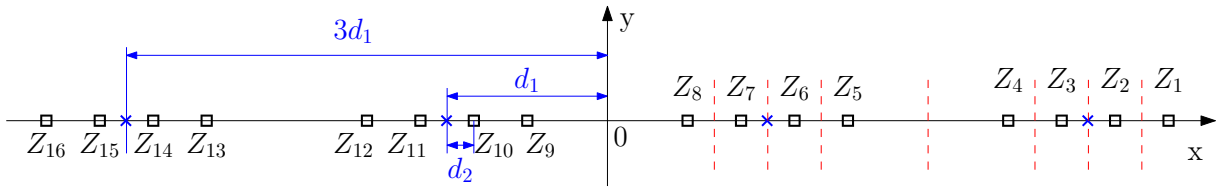

Fig. 3. The superposed symbol of 4-PAM on 4-PAM

examining Fig. 2, one can find that when Z_4 is sent, even if the received symbol is decoded as Z_1, Z_2 or Z_3, User 1 can still obtain the right information. Thus, the probability of a correct decoding for User 1 with Z_4 can be expressed as

$$\begin{aligned} &P(U_1=1|Z_4) \\ &\quad = P(Z_1|Z_4)+P(Z_2|Z_4)+P(Z_3|Z_4)+P(Z_4|Z_4), \end{aligned} \tag{6}$$

where $U_1=1$ means User 1 achieves the transmitted symbol successfully and the conditional probability $P(Z_1|Z_4)$ represents the probability of cataloguing the received PAM symbol to Z_1 when transmitting Z_4.

As given in Equation (3), the equalized received symbol at User 1 over channel h_1 is r_1. Thus, the conditional probability distribution function (PDF) of r_1 when Z_4 has been transmitted is:

$$P(r_1|Z_4)=\frac{1}{\sqrt{\pi N_1}}e^{\frac{-(r_1-(d_1-3d_2))^2}{N_1}}. \tag{7}$$

From Fig. 2 we can see that the correct decision region of User 1 given Z_4 is actually the whole right side of the y-axis, thus, Equation (6) can be written as $P(U_1=1|Z_4)=P(r_1>0|Z_4)$. By utilizing Equation (7), we have

$$\begin{aligned} P(r_1>0|Z_4) &= \int_0^{\infty} P(r_1|Z_4)dr_1 \\ &= \frac{1}{\sqrt{\pi N_1}}\int_{-d_1+3d_2}^{\infty} e^{\frac{-r^2}{N_1}}dr_1 = 1-Q\left(\frac{d_1-3d_2}{\sqrt{N_1/2}}\right), \end{aligned} \tag{8}$$

where $Q(\cdot)$ represents the Gaussian Q-function which is defined as $Q(x)=\frac{1}{\sqrt{2\pi}}\int_0^{\infty}e^{-\frac{x^2}{2}}dx$. Similar as the discussion from (6) to (8), the correct decoding probability of User 1 with other transmitted symbols can be obtained. Then, by employing the y-axis symmetry of the constellation, the overall correct decoding probability of User 1 is obtained as

$$\begin{aligned} P(U_1=1) &= 2\sum_{i=1}^{4}P(U_1=1|Z_i)P(Z_i)=\frac{1}{4}\sum_{i=1}^{4}P(U_1=1|Z_i) \\ &= 1-\frac{1}{4}\sum_{k=1}^{2}\left[Q\left(\frac{d_1-(2k-1)d_2}{\sqrt{N_1/2}}\right)+Q\left(\frac{d_1+(2k-1)d_2}{\sqrt{N_1/2}}\right)\right]. \end{aligned} \tag{9}$$

Apparently, the SER of User 1 in this case can be achieved from (9) easily, which is $\text{SER}_{\text{U1}}=1-P(U_1=1)$.

Now, we consider the general case with an M-PAM overlaid on an N-PAM constellation. The basic procedure to obtain the SER expression of it is very similar to the special case ($N=2$ and $M=4$) discussed above. The key difference between them lies in the correct decision region of each constellation point for User 1. In other words, the total number and the length of the different correct decision regions under the general case are different from the special one. In particular, there are only two kinds of correct decision regions in the above special case, i.e.,

$(0,+\infty)$ for points Z_1 to Z_4 and $(-\infty,0)$ for points Z_5 to Z_8. Thus, the length of the two regions are the same and equal to infinity. For the general case, it is obvious that the total number of the different correct decision regions is equal to N. However, the length of these regions are not the same. Furthermore, we can classify these regions into two types, i.e., 2 end regions (far left and far right regions) and $N-2$ intermediate regions. The length of these two kinds of regions are infinity and $2d_1$, respectively. In addition, changing the value of M cannot influence the properties of the correct decision regions. To give a more intuitive overview of the two types of regions, an example with 4-PAM for both two users are given in Fig. 3. As we can see from it, the correct decision regions for Z_1 to Z_4 and Z_{13} to Z_{16} are the end regions whose corresponding range are $(2d_1,+\infty)$ and $(-\infty,-2d_1)$. While the intermediate regions $(0,2d_1)$ and $(-2d_1,0)$ are the correct decision region for the constellation points Z_5 to Z_8 and Z_9 to Z_{12}, respectively.

Based on the discussion above, the computation of the SER in the general case can be partitioned into two parts according to the types of correct decision regions for User 1. For the first type of regions, namely, the left end and the right end regions, there are in total $2M$ constellation points located in these two regions, i.e., M points in each of them. As the two regions are y-axis symmetric, we only need to calculate the correct decoding probability for User 1 of the constellation points in one of these regions, i.e., the M constellation points in the far right region, for example, point Z_1 to Z_4 in Fig. 3. Let $\mathcal{S}_1$ denote the set of the constellation points that are located in the far right region, e.g., $\mathcal{S}_1=\{Z_1,Z_2,Z_3,Z_4\}$ in Fig. 3, then the probability of a correct decoding for User 1 with the points in set $\mathcal{S}_1$ is

$$\begin{aligned} P_{\mathcal{S}_1}(U_1=1) &= \sum_{i=1}^{M} P(U_1=1|Z_i\in\mathcal{S}_1)P(Z_i) \\ &= \frac{1}{MN}\sum_{k=1}^{\frac{M}{2}}\left[2-Q\left(\frac{d_1-(2k-1)d_2}{\sqrt{N_1/2}}\right)-Q\left(\frac{d_1+(2k-1)d_2}{\sqrt{N_1/2}}\right)\right], \end{aligned} \tag{10}$$

where the terms in the square brackets represent the correct decoding probabilities of two constellation points, e.g., Z_1 and Z_4 in Fig. 3.

Next, we discuss the intermediate regions. In fact, there are M constellation points in each intermediate region and the position of these points in different intermediate regions have the same pattern. Let $\mathcal{S}_2$ denote the constellation points set of an arbitrary intermediate region, e.g., $\mathcal{S}_2=\{Z_5,Z_6,Z_7,Z_8\}$ in Fig. 3, the probability of a correct decoding for User 1 with set $\mathcal{S}_2$ can be calculated by

$$\begin{aligned} P_{\mathcal{S}_2}(U_1=1) &= \sum_{i=1}^{M} P(U_1=1|Z_i\in\mathcal{S}_2)P(Z_i) \\ &= \frac{1}{MN}\left[2\sum_{k=1}^{\frac{M}{2}}\left(1-Q\left(\frac{d_1-(2k-1)d_2}{\sqrt{N_1/2}}\right)-Q\left(\frac{d_1+(2k-1)d_2}{\sqrt{N_1/2}}\right)\right)\right]. \end{aligned} \tag{11}$$

The derivation of (11) is facilitated by the symmetry property, i.e., we only need to compute the probabilities of half the points (points from 1 to $M/2$ in this decision region). For example, to obtain the correct decoding probabilities of the constellation points from Z_5 to Z_8 in Fig. 3, we only need to calculate the probabilities of points Z_5 and Z_6, then multiply the summation with 2, as Z_5 and Z_6 are equivalent to Z_8 and Z_7 in their probabilities computation, respectively.

Finally, by taking the total number of the two types of regions into account, the overall correct decoding probability for User 1 is obtained by

$$\begin{aligned} P(U_1=1) &= 2P_{\mathcal{S}_1}(U_1=1) + (N-2)P_{\mathcal{S}_2}(U_1=1) \\ &= 1 - \frac{2(N-1)}{MN}\sum_{k=1}^{\frac{M}{2}}\left[Q\left(\frac{d_1-(2k-1)d_2}{\sqrt{N_1/2}}\right)+Q\left(\frac{d_1+(2k-1)d_2}{\sqrt{N_1/2}}\right)\right]. \end{aligned} \tag{12}$$

Therefore, a proposition regarding to the SER of User 1 can be obtained and given in below.

Proposition 1. *For a two-user NOMA system, when N-PAM and M-PAM are utilized by User 1 and User 2, respectively, the SER of User 1 can be expressed as*

$$SER_{UI} = 1 - P(U_1=1). \tag{13}$$

In this subsection, we provided the derivations regarding the exact SER of User 1 in PAM based NOMA systems. In the derivations from Equation (6) to (9), we investigated the exact SER of a simple NOMA system example. Afterwards, by extending the concept and procedure for achieving the SER of this special case to the general one, i.e., from Equation (10) to (13), we obtained the closed-form SER expression of User 1 for the NOMA system with arbitrary PAM constellation orders. In the next subsection, we move on to discuss the SER for User 2.

C. SER of User 2

We start with a case study to illustrates our methodology on addressing the correct decoding probabilities of different types of constellation points. The general case will be analyzed afterwards.

In the case study, we consider an example scenario utilizing 4-PAM for both User 1 and User 2, as illustrated in Fig. 3. Note that the constellation are symmetric regarding to y-axis, i.e., the analysis of either the left half or the right half points in the figure are the same. In other words, we only need to discuss the right half points, i.e., Z_8 to Z_1. By examining and comparing the decision regions and distant parameters, we classify Z_8 to Z_1 into the following three types

- Type 1: Left side points Z_8, Z_4,
- Type 2: Right side points Z_5, Z_1,
- Type 3: Inner part points Z_7, Z_6, Z_3, Z_2.

We first consider the left side point Z_8 in which the carried information symbol for User 2 is C_4^2. And from Fig. 3, we know that Z_4, Z_{12} and Z_{16} convey the same symbol for User 2. Thus, even if the symbol Z_8 is decoded as the other three symbols, User 2 can still retrieve its information correctly. Hence, the probability of a correct decoding for User 2 when transmitting Z_8 is given by

$$\begin{aligned} &P(U_2=1|Z_8) \\ &\quad = P(Z_4|Z_8)+P(Z_8|Z_8)+P(Z_{12}|Z_8)+P(Z_{16}|Z_8). \end{aligned} \tag{14}$$

According to (14), the correct decision region has 4 parts, which are discussed one by one as follows.

For the original decision region, the correct decoding probability is

$$P(Z_8|Z_8) = \frac{1}{\sqrt{\pi N_2}} \int_{-d_1+3d_2}^{d_2} e^{\frac{-r^2}{N_2}} dr, \tag{15}$$

In addition, the probability of decoding as Z_4 when transmitting Z_8 can be obtained by

$$P(Z_4|Z_8) = \frac{1}{\sqrt{\pi N_2}} \int_{2d_1-d_1+3d_2}^{2d_1+d_2} e^{\frac{-r^2}{N_2}} dr, \tag{16}$$

It can be seen that the only difference between (15) and (16) is that there is a right shift with length of $2d_1$ on the integral range. Thus, $P(Z_{12}|Z_8)$ can be obtained with the same method except the integral range shift is to the left. For the last point Z_{16} at the far left, the integral shift patten is no longer holding as it involves a lower bound with infinity. We therefore obtain $P(Z_{16}|Z_8)$ by partitioning the integral range in the following way

$$\begin{aligned} P(Z_{16}|Z_8) &= \frac{1}{\sqrt{\pi N_2}} \int_{-\infty}^{-4d_1+d_2} e^{\frac{-r^2}{N_2}} dr \\ &= \frac{1}{\sqrt{\pi N_2}} \left(\int_{-4d_1-d_1+3d_2}^{-4d_1+d_2} e^{\frac{-r^2}{N_2}} dr + \int_{-\infty}^{-4d_1-d_1+3d_2} e^{\frac{-r^2}{N_2}} dr \right). \end{aligned} \tag{17}$$

Therefore, equation (14) is now further given by

$$\begin{aligned} &P(U_2=1|Z_8) \\ &= \frac{1}{\sqrt{\pi N_2}} \left(\sum_{k=-2}^{1} \int_{2kd_1-d_1+3d_2}^{2kd_1+d_2} e^{\frac{-r^2}{N_2}} dr + \int_{-\infty}^{-4d_1-d_1+3d_2} e^{\frac{-r^2}{N_2}} dr \right) \\ &= \sum_{k=-2}^{1} \left[Q\left(\frac{(2k-1)d_1+3d_2}{\sqrt{N_2/2}} \right) - Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}} \right) \right] + Q\left(\frac{5d_1-3d_2}{\sqrt{N_2/2}} \right). \end{aligned} \tag{18}$$

Next, we consider Z_5 which is a the Type 2 point. It carries the same symbol as Z_8 for User 1, i.e., C_2^1. The same procedure provided from (14) to (18) can be applied to obtain $P(U_2=1|Z_5)$. The only difference (in comparison to the discussion on Z_8) here is that the partition of the integral range is now on the right side. We skip the intermediate derivations and directly provided the final expression of the correct decoding probability of Z_5 for User 2, which is

$$\begin{aligned} &P(U_2=1|Z_5) \\ &= \sum_{k=-2}^{1} \left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}} \right) - Q\left(\frac{(2k+1)d_1-3d_2}{\sqrt{N_2/2}} \right) \right] + Q\left(\frac{3d_1-3d_2}{\sqrt{N_2/2}} \right). \end{aligned} \tag{19}$$

For the Type 3 points carrying information symbol C_2^1, i.e., Z_6 and Z_7, the partition of the integrals at the axes ends are no longer necessary as all the correct decision regions are the closed intervals and have a length of $2d_2$. Meanwhile, Z_6 and Z_7 share the same correct decoding probability as the PDF and the integral range of them are always the same. In fact, this property always holds when the Type 3 constellation points carrying the same information symbol for User 1. Therefore, for Z_6 and Z_7 we have

$$\begin{aligned} P(U_2=1|Z_6) &= P(U_2=1|Z_7) \\ &= \sum_{k=-2}^{1} \left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}} \right) - Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}} \right) \right]. \end{aligned} \tag{20}$$

So far, we have obtained the correct decoding probabilities for the constellation points from Z_5 to Z_8. The property they have in common is that all of these points convey the same information symbol C_2^1 of User 1.

We denote the set of these constellation points as Ω_2. Then, combining (18)-(20), the overall correct decoding probability over this set is given by

$$P_{\Omega_2}(U_2=1)=\sum_{i=5}^{8}P(U_2=1|Z_i\in\Omega_2)P(Z_i). \tag{21}$$

Based on the same derivation steps for the constellation points from Z_8 to Z_5, the correct decoding probabilities for the points from Z_4 to Z_1 can also be calculated. Similarly, denote by Ω_1 the set of points carrying the information symbol C_1^1 for User 1. Thus, the correct decoding probability of the Type 1 point for User 2 in Ω_1 is

$$\begin{aligned}&P(U_2=1|Z_4)\\&=\sum_{k=-3}^{0}\left[Q\left(\frac{(2k-1)d_1+3d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)\right]+Q\left(\frac{7d_1-3d_2}{\sqrt{N_2/2}}\right).\end{aligned} \tag{22}$$

Obviously, although both Z_8 and Z_4 are Type 1 points, their the correct decoding probability expressions are different. In particular, by comparing (22) with (18), we learn that the terms inside of the square bracket are the same. While for the summation part, the lower and upper bounds are changed. And the last terms in the two equation are different due to the change of lower bounds. The same difference is applicable to the relationship between the two Type 2 constellation points Z_5 and Z_1 as well, while the influence on the last term then comes from the change of the upper bounds of the summation. In particular, the correct detection probability of Z_1 can be obtained from (19) by adjusting the upper bounds, which is given by

$$\begin{aligned}&P(U_2=1|Z_1)\\&=\sum_{k=-3}^{0}\left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{(2k+1)d_1-3d_2}{\sqrt{N_2/2}}\right)\right]+Q\left(\frac{d_1-3d_2}{\sqrt{N_2/2}}\right).\end{aligned} \tag{23}$$

For the Type 3 points, from Z_7 and Z_6 to Z_3 and Z_2, the only change in the SER derivation is on the summation bounds, i.e., for Z_3 and Z_2 there is no additional term due to the integral partition. Therefore, based on (20) we have

$$\begin{aligned}P(U_2=1|Z_2)&=P(U_2=1|Z_3)\\&=\sum_{k=-3}^{0}\left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)\right].\end{aligned} \tag{24}$$

According to the discussion above, the correct detection probability over the set Ω_1 can be obtained as

$$P_{\Omega_1}(U_2=1)=\sum_{i=1}^{4}P(U_2=1|Z_i\in\Omega_2)P(Z_i). \tag{25}$$

Considering the constellation points are y-axis symmetric, the correct detection probability of User 2 can be obtained by

$$P(U_2=1)=2P_{\Omega_1}(U_2=1)+2P_{\Omega_2}(U_2=1), \tag{26}$$

which can be calculated and simplified as

$$\begin{aligned}P(U_2=1)=1-\frac{3}{2}Q\left(\frac{d_2}{\sqrt{N_2/2}}\right)+\frac{1}{8}\sum_{k=1}^{3}(4-k)\left[3Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)\right.\\\left.-3Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)+Q\left(\frac{(2k-1)d_1+3d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{(2k-1)d_1-3d_2}{\sqrt{N_2/2}}\right)\right].\end{aligned} \tag{27}$$

$$P(U_2=1)=1-\frac{2(M-1)}{M}Q\left(\frac{d_2}{\sqrt{N_2/2}}\right)+\frac{2}{MN}\sum_{k=1}^{N-1}(N-k)\Bigg\{(M-1)\left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)\right]$$
$$+\left[Q\left(\frac{(2k-1)d_1+(M-1)d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{(2k-1)d_1-(M-1)d_2}{\sqrt{N_2/2}}\right)\right]\Bigg\} \tag{33}$$

Hence, the SER of User 2 in this example case can be obtained, which is $\text{SER}_{\text{U2}}=1-P(U_2=1)$.

So far, we have derived the SER for the example case shown in Fig. 3. The same methodology of derivation process can be applied in the general case with N-PAM for User 1 and M-PAM for User 2. In particular, the constellation points can still be catalogued into three types, namely, left side points, right side points and inner points. In addition, these three types have N, N and $(M-2)N$ points, respectively. As N-PAM is utilized by User 1, therefore, the information symbols for User 1 can be represented as $C_1^1, C_2^1, \ldots, C_N^1$. Each information symbol C_n^1 is associated with M constellation points, and we denote the set of these points as $\Omega_n, \forall n=1,2,\ldots,N$. Furthermore, for the nth set Ω_n, the total number of the sets on its left and right are $N-n$ and $n-1$, respectively. More importantly, these two values not only provide the upper and lower bounds of the summations for the points located in Ω_n, but also determine the additional terms of Type 1 and Type 2 points resulting from the integral partition, e.g., the bounds and the last term in (18).

Therefore, the correct decoding probability of the Type 1 point in set Ω_n for User 2 can be given by

$$\begin{aligned}&P_{\Omega_n}(U_2=1|Type\ 1)\\&=\sum_{k=-(N-n)}^{n-1}\left[Q\left(\frac{(2k-1)d_1+(M-1)d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)\right]\\&\quad+Q\left(\frac{\big(2\,(N-n)+1\big)d_1-(M-1)d_2}{\sqrt{N_2/2}}\right).\end{aligned} \tag{28}$$

Similarly, the correct decoding of the Type 2 constellation point for User 2 in set Ω_n is

$$\begin{aligned}&P_{\Omega_n}(U_2=1|Type\ 2)\\&=\sum_{k=-(N-n)}^{n-1}\left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{(2k+1)d_1-(M-1)d_2}{\sqrt{N_2/2}}\right)\right]\\&\quad+Q\left(\frac{\big(2\,(n-1)+1\big)d_1-(M-1)d_2}{\sqrt{N_2/2}}\right).\end{aligned} \tag{29}$$

Meanwhile, for the Type 3 points in Ω_n, they share the same correct decoding probability expression for User 2, which is

$$\begin{aligned}&P_{\Omega_n}(U_2=1|Type\ 3)\\&\qquad=\sum_{k=-(N-n)}^{n-1}\left[Q\left(\frac{2kd_1-d_2}{\sqrt{N_2/2}}\right)-Q\left(\frac{2kd_1+d_2}{\sqrt{N_2/2}}\right)\right].\end{aligned} \tag{30}$$

Considering that in the set Ω_n, there are 1 Type 1 constellation point, 1 Type 2 point and $M-2$ Type 3 constellation points, the correct decoding probability for User 2 when giving set Ω_n can be calculated by

$$\begin{aligned} P_{\Omega_n}(U_2=1) = (M-2)\,P(U_2=1|Type\ 3)P(Z_i)+ \\ P(U_2=1|Type\ 2)P(Z_i)+P(U_2=1|Type\ 1)P(Z_i). \end{aligned} \tag{31}$$

Recall that User 1 utilizes N-PAM, therefore the number of set Ω_n is N. By employing the y-axis symmetric property of the constellation points, the overall correct decoding probability for User 2 can obtained by

$$P(U_2=1) = 2\sum_{n=1}^{N/2} P_{\Omega_n}(U_2=1). \tag{32}$$

By substituting (28)-(30) into (32), $P(U_2=1)$ is further given by Equation (33) at the top of the next page.

Finally, the proposition associated to the SER of User 2 is obtained and given below.

Proposition 2. *For a two-user NOMA system, when N-PAM and M-PAM are utilized by User 1 and User 2, respectively, the SER of User 2 can be expressed as*

$$SER_{U2} = 1 - P(U_2=1). \tag{34}$$

Section III-C studied the SER expression of User 2 in PAM based NOMA systems. In the first part of this subsection, we conducted a case study with a simple PAM based NOMA system and the derivation for achieving the SER of User 2 in this case is given by Equation (14) to (27). After that, the derivation of this example was extend to a general case. From (28) to (34), the steps for deriving the closed-form SER expression of User 2 for a NOMA system with arbitrary PAM orders were provided.

IV. SERs OF NOMA USERS WITH QAM SYMBOLS

When User 1 and User 2 are encoded with N-ary and M-ary rectangular QAM schemes, respectively, the transmitted superposition symbol becomes a QAM-liked symbol with the constellation order of MN. For example, if 4-QAM is applied to both two users, the transmitted symbol will be a 16-QAM-liked symbol and the construction of it is shown in Fig. 4.

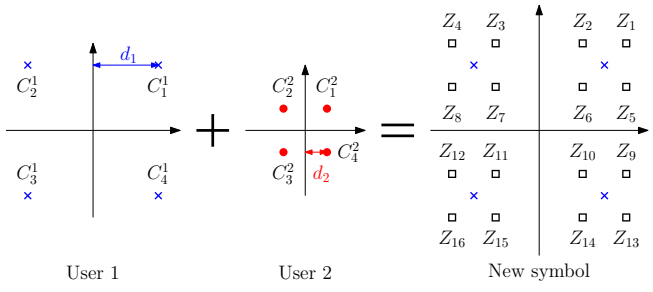

Fig. 4. The construction of the superposed symbol

$$\frac{df_1(\gamma)}{d\gamma} = -\frac{(N-1)}{MN\sqrt{2\pi}} \sum_{k=1}^{\frac{M}{2}} e^{-\frac{\alpha^2\gamma+(2k+1)^2\beta^2(1-\gamma)}{2}} \Big[\alpha\gamma^{-\frac{1}{2}} \left(e^{(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}} + e^{-(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}}\right) + (2k-1)\,\beta\,(1-\gamma)^{-\frac{1}{2}} \left(e^{(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}} - e^{-(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}}\right)\Big] \tag{39}$$

A. Analysis of the superposed QAM symbol

After obtaining the construction pattern of the superposed symbol, we can use the property of the rectangular QAM to further study the NOMA symbol. It is well know that a rectangular QAM symbol can be easily transferred into two independent PAM signals on in-phase and quadrature carriers separately. We denote these two components as x and y, which are located on the x-axis and y-axis accordingly, as shown in Fig. 4. Hence, for User 1 and User 2, their corresponding QAM symbols can be treated as a composition of two $\sqrt{N}$-PAM and two $\sqrt{M}$-PAM symbols, respectively. In this case, we can divide the superposed symbol into two in-phase and quadrature components and each component is a $\sqrt{MN}$-PAM-liked symbol. For example, in Fig. 4, the 4-QAM constellations of the two users are all constructed from two quadrature 2-PAM symbols, and the superposed symbol is a 16-QAM-liked symbol which can be separated as two independent 4-PAM-liked symbols.

From the analysis above, we know that instead of analysing the superposed QAM-liked symbol directly, one can study the corresponding $\sqrt{MN}$-PAM-liked symbol of it first, then reconstruct the NOMA symbol according to its one dimensional PAM-liked symbol. In order to investigate the $\sqrt{MN}$-PAM-liked symbol, the basic parameters of it must be acquired. In particular, the distance parameters associated to this PAM-liked symbol of User 1 and User 2. It is known that the power normalization factor of a rectangular N-QAM symbol is $\sqrt{\frac{3}{2(N-1)}}$. Thus, the distance parameters $\tilde{d}_1$ and $\tilde{d}_2$ for the corresponding PAM symbols of User 1 and User 2 are given as $\tilde{d}_1 = \sqrt{\frac{3p_1}{2(N-1)}}$ and $\tilde{d}_2 = \sqrt{\frac{3p_2}{2(M-1)}}$. For example, as shown in Fig. 4, when 4-QAM is employed by the two users, the distance parameters of them are $\tilde{d}_1 = \sqrt{\frac{p_1}{2}}$ and $\tilde{d}_2 = \sqrt{\frac{p_2}{2}}$, respectively.

B. SER of User 1 and User 2

As discussed before, we know the QAM based NOMA symbol is an extension of the corresponding PAM-liked symbol from one dimension to two dimensions. And this PAM-liked symbol is actually the superposition of the two PAM symbols that are obtained from the QAM symbols of User 1 and User 2. Therefore, the model considered in Section III can be applied to here.

For User 1, the correct decoding probability on dimension x can be calculated by equation (12) and we denote it by $P_x(U1{=}1)$. Notice that the constellation parameter M, N, d_1 and d_2 in (12) must be replaced by $\sqrt{M}$, $\sqrt{N}$, $\tilde{d}_1$ and $\tilde{d}_2$ in this case. Similarly, the correct decoding probability on dimension y for User 1 can be obtained by the same equation, and denoted as $P_y(U1{=}1)$. Therefore, the SER for User 1 can be expressed as

$$\text{SER}_{\text{U1}} = 1 - P_x(U1{=}1)P_y(U1{=}1). \tag{35}$$

Recall that the two components are independent and have identical probability expressions, thus, SER_{U1} can be further simplified. Hence, we can obtain the following proposition.

Proposition 3. *For a NOMA system, when N-QAM and M-QAM are utilized by User 1 and User 2, respectively, the SER of User 1 can be expressed as*

$$\textit{SER}_{U1} = 1 - P_x^2(U1{=}1). \tag{36}$$

Obviously, the same procedure can be conducted on Equation (33) to obtain the SER of User 2, which gives us the proposition in below.

Proposition 4. *For a NOMA system, when N-QAM and M-QAM are utilized by User 1 and User 2, respectively, the SER of User 2 can be expressed as*

$$\textit{SER}_{U2} = 1 - P_x^2(U2{=}1), \tag{37}$$

where $P_x(U2{=}1)$ is the correct decoding probability of User 2 obtained by substituting the replacement parameters to (33)*.*

V. Power allocation design

In this section, we provide power allocation designs to guarantee users' reliabilities for the considered NOMA systems. First, we formulate the optimization problem, which maximize one user's reliability while satisfying the reliability requirement of the other user. Then, we investigate the influence of power allocation on the reliability of each user and propose solutions to the optimization problems for both PAM and QAM based NOMA systems according to the power allocation analysis of each user.

A. Optimization problem

Recall that the user exhibiting the better channel condition is called User 2. In other words, the channel quality of User 1 is relatively more limited and the reliability of the transmission to User 1 needs to be more assured. Hence, we consider the following system design which minimize the SER of User 2 while guaranteeing the SER requirement of the User 1.

According to the SER analysis provided in the previous sections, the SER expressions for both two users can be rewritten as functions of the power allocation parameter γ by substituting $d_1 = \sqrt{\frac{3E_s\gamma}{(N-1)(N+1)}}$ and $d_2 = \sqrt{\frac{3E_s(1-\gamma)}{(M-1)(M+1)}}$ into Equations (13) and (34) for the PAM case and by substituting $\tilde{d}_1 = \sqrt{\frac{3E_s\gamma}{2(N-1)}}$ and $\tilde{d}_2 = \sqrt{\frac{3E_s(1-\gamma)}{2(M-1)}}$ into Equations (36) and (37) for the QAM case. In other words, the SERs are strongly influenced by the decision of the power allocation coefficient $\gamma \in (0.5, 1)$ [1]. Hence, the power allocation problem is formulated as

[1]Recall that we always call the user with a higher transmit power by User 1. Hence, the feasible range of the power allocation coefficient is $0.5 < \gamma < 1$.

$$\frac{df_2(\gamma)}{d\gamma}=\frac{(M-1)\omega}{M\sqrt{2\pi(1-\gamma)}}e^{-\frac{(\omega\sqrt{1-\gamma})^2}{2}}+\frac{2}{MN\sqrt{2\pi}}\sum_{k=1}^{N-1}(N-k)\Bigg\{\frac{(M-1)kv}{\sqrt{\gamma}}\Big(e^{-\frac{(2kv\sqrt{\gamma}-\omega\sqrt{1-\gamma})^2}{2}}-e^{-\frac{(2kv\sqrt{\gamma}+\omega\sqrt{1-\gamma})^2}{2}}\Big)$$
$$+\frac{(M-1)\omega}{2\sqrt{1-\gamma}}\Big(e^{-\frac{(2kv\sqrt{\gamma}-\omega\sqrt{1-\gamma})^2}{2}}+e^{-\frac{(2kv\sqrt{\gamma}+\omega\sqrt{1-\gamma})^2}{2}}-e^{-\frac{\big((2k-1)v\sqrt{\gamma}-(M-1)\omega\sqrt{1-\gamma}\big)^2}{2}}-e^{-\frac{\big((2k-1)v\sqrt{\gamma}+(M-1)\omega\sqrt{1-\gamma}\big)^2}{2}}\Big)$$
$$+\frac{(2k-1)v}{2\sqrt{\gamma}}\Big(e^{-\frac{\big((2k-1)v\sqrt{\gamma}+(M-1)\omega\sqrt{1-\gamma}\big)^2}{2}}-e^{-\frac{\big((2k-1)v\sqrt{\gamma}-(M-1)\omega\sqrt{1-\gamma}\big)^2}{2}}\Big)\Bigg\} \tag{41}$$

$$\frac{d\hat{f}_2(\gamma)}{d\gamma}=\frac{(N-1)(M-1)\omega}{MN\sqrt{2\pi(1-\gamma)}}\Big\{\Big(e^{-\frac{(2v\sqrt{\gamma}-\omega\sqrt{1-\gamma})^2}{2}}-e^{-\frac{(v\sqrt{\gamma}+(M-1)\omega\sqrt{1-\gamma})^2}{2}}\Big)+\Big(e^{-\frac{(\omega\sqrt{1-\gamma})^2}{2}}-e^{-\frac{(v\sqrt{\gamma}-(M-1)\omega\sqrt{1-\gamma})^2}{2}}\Big)+e^{-\frac{(2v\sqrt{\gamma}+\omega\sqrt{1-\gamma})^2}{2}}\Big\}$$
$$+\frac{(N-1)v}{MN\sqrt{2\pi\gamma}}\Big\{2(M-1)\Big(e^{-\frac{(2v\sqrt{\gamma}-\omega\sqrt{1-\gamma})^2}{2}}-e^{-\frac{(2v\sqrt{\gamma}+\omega\sqrt{1-\gamma})^2}{2}}\Big)+e^{-\frac{(v\sqrt{\gamma}+(M-1)\omega\sqrt{1-\gamma})^2}{2}}\Big\}$$
$$+\frac{1}{MN\sqrt{2\pi}}\Big\{\frac{(M-1)\omega}{\sqrt{1-\gamma}}e^{-\frac{(\omega\sqrt{1-\gamma})^2}{2}}-\frac{(N-1)v}{\sqrt{\gamma}}e^{-\frac{(v\sqrt{\gamma}-(M-1)\omega\sqrt{1-\gamma})^2}{2}}\Big\} \tag{42}$$

$$\min_{\gamma\in(0.5,1)}\ \mathrm{SER}_{\mathrm{U2}}$$
$$\text{s.t.}\quad \mathrm{SER}_{\mathrm{U1}}\le\lambda, \tag{38}$$

where λ is the SER constraint of User 1. In the following, we solve the problem under scenarios of users with the PAM symbols and the QAM symbols, respectively.

B. NOMA users with PAM symbols

We first discuss the PAM case and provide the following key proposition for determining γ.

Proposition 5. *For a two-user NOMA system, when N-PAM and M-PAM are utilized by User 1 and User 2, respectively, the SER of User 1 is decreasing in the power allocation coefficient γ within its feasible range.*

Proof. After rewriting the SER of user 1 in Equation (13) as $f_1(\gamma)$, the first derivative of the equation can be derived as (39) with parameter $\alpha=\sqrt{\frac{6E_s}{N_1(N+1)(N-1)}}$ and $\beta=\sqrt{\frac{6E_s}{N_1(M+1)(M-1)}}$. Obviously, the fraction before the summation symbol in the equation is negative. For the part inside the summation, it is clear that the exponential term before the square bracket is always greater than 0. Meanwhile, the first part in the square bracket is positive. In addition, the second part in this square bracket is also positive as $e^{(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}}>1$ and $e^{-(2k-1)\alpha\beta\sqrt{\gamma(1-\gamma)}}<1$. Therefore, the first derivative of SER for User 1 is always negative within the feasible range of γ, which proves that this function is monotonically decreasing. □

After achieving the relationship between the power allocation parameter γ and the SER of User 1 from Proposition 5, we continue to study the influence brought by γ on the SER of User 2 and propose Proposition 6 in the following, which provides a suboptimal solution of (38).

Based on Proposition 5, with certain E_s and N_1, the constraint in Problem (38) is equivalent to $\gamma \geq \gamma_\lambda$, where γ_λ is the solution of equation $f_1(\gamma_\lambda) = \lambda$. Thus, the original optimization problem (38) can be written as

$$\begin{aligned} &\min_{\gamma} \text{SER}_{\text{U2}} \\ &\text{s.t.} \quad \gamma_\lambda \leq \gamma \leq 1. \end{aligned} \tag{40}$$

The closed-form expression of the global optimal solution for this problem is difficult to obtain, especially noting that the objective function SER_{U2} is non-convex in γ. It is true that a suboptimal γ with a sufficient small error (in comparison to the optimal one) can be obtained by exhaustive search, while it requires a large number of searching iterations and provides less analytical insights of the system behaviour. Observing this, in the following, we propose a suboptimal solution of γ with a closed-form expression. We will show in the simulation section that this closed-form solution preserves almost the same SER performance as the optimal one.

Proposition 6. *For the PAM based NOMA system, the suboptimal power allocation coefficient for problem* $\min_{\gamma}$ *SER$_{U2}$ is* $\gamma_{sub} = \frac{M^2(N^2-1)}{M^2N^2-1}$. *There always exists a better power allocation coefficient whose value is smaller than the suboptimal one, i.e.,* $\gamma_{btr} < \gamma_{sub}$, *which can provide a lower SER$_{U2}$ for User 2.*

Proof. Similar to the proof of Proposition 5, the SER expression of User 2 in (34) can be written as a function $f_2(\gamma)$ and the first derivative of it is given in (41), in which $\upsilon = \sqrt{\frac{6E_s}{N_2(N+1)(N-1)}}$ and $\omega = \sqrt{\frac{6E_s}{N_2(M+1)(M-1)}}$. Intuitively, the γ for which $\frac{df_2(\gamma)}{d\gamma} = 0$ holds is the optimal value to provide the minimum SER for User 2. However, it is not easy to obtain the closed form solution for it as the expression of $\frac{df_2(\gamma)}{d\gamma}$ is quite complicated. Therefore, we can try to find a suboptimal solution with a simple closed form expression, which can make the first derivative of $f_2(\gamma)$ has a value relatively close to 0.

Moreover, by examining Equation (33), we can observe that the values of the terms inside of the summation become smaller with the increasing of k. In other words, the influence of these terms with larger k on the SER of User 2 is more insignificant. Thus, we can approximate (34) by only keeping the terms with $k = 1$. Denoting this approximation as $\hat{f}_2(\gamma)$, the first derivative of it with respect to γ can be derived as (42). Instead of finding the suboptimal solution which makes the absolute value of $\frac{df_2(\gamma)}{d\gamma}$ as small as possible, we try to achieve the suboptimal solution based on $\hat{f}_2(\gamma)$.

By comparing all the exponential elements in (42), we notice that $e^{-\frac{(\omega\sqrt{1-\gamma})^2}{2}}$ and $e^{-\frac{(\upsilon\sqrt{\gamma}-(M-1)\omega\sqrt{1-\gamma})^2}{2}}$ have greater values than the remaining exponential elements due to the constraint $d_1 > (M-1)d_2$. Therefore, to make the absolute value of $\frac{df_2(\gamma)}{d\gamma}$ as small as possible, the best choice is to eliminate the terms involving these two exponential elements. Moreover, Equation (42) consists of three parts and only the first and third parts contain the two greater exponential elements. When comparing these two part, the coefficient before $e^{-\frac{(\omega\sqrt{1-\gamma})^2}{2}}$ in the first part is larger than it in the third part, which means the influence brought by the first part is higher than the third one. In other words, we can make the absolute value of $\frac{df_2(\gamma)}{d\gamma}$ relatively small by only eliminating the two greater exponential elements in the first part. By examining this part, it is easy to see that when $\upsilon\sqrt{\gamma} = M\omega\sqrt{1-\gamma}$, the terms in the first and second parentheses within the curly braces can be eliminated. Thus, the suboptimal power allocation coefficient of User 2 can be obtained from this equality, which is $\gamma_{sub} = \frac{M^2(N^2-1)}{M^2N^2-1}$.

After obtaining the suboptimal solution γ_{sub}, we must prove the second statement of Proposition 6. In Appendix A, we show that the first derivative of the SER expression for User 2 is always greater than 0 at this suboptimal γ_{sub} point, i.e., $\frac{df_2(\gamma_{sub})}{\gamma} > 0$ always holds. Thus, $f_2(\gamma)$ is increasing in γ at this specific point. Since $f_2(\gamma)$ is a continuous function on γ, there must exist a γ_{btr} with value smaller than γ_{sub}, which can offer a lower SER for User 2 than γ_{sub}, i.e., $\exists\ \gamma_{btr} < \gamma_{sub} : f_2(\gamma_{btr}) < f_2(\gamma_{sub})$.

□

Therefore, based on Proposition 6, and considering that the SER_{U2} increases in γ when $\gamma \geq \gamma_{sub}$, the suboptimal solution γ^* of Problem (40) is achieved by

$$\gamma^* = \max\left\{\gamma_{sub}, \gamma_{\lambda}\right\}, \tag{43}$$

which is also the suboptimal solution for the original optimization problem given in (38).

C. NOMA users with QAM symbols

Similar to the above case with PAM, we provide also two propositions to show the relationship between the power allocation parameter γ and the SERs of the two users, and to solve the non-convex problem (38) under the QAM case.

Proposition 7. *For a NOMA system, when N-QAM and M-QAM are utilized by User 1 and User 2, respectively, the SER of User 1 is decreasing in the power allocation coefficient γ within its feasible range.*

Proof. The first derivative of SER_{U1} in γ can be calculated, which is quite similar to Equation (39). There are two main differences. First of all, there is an extra coefficient 2 in front of the whole summation in the expression of $\frac{d\tilde{f}_1(\gamma)}{d\gamma}$ due to the power 2 on the term $P_x(U1 = 1)$ of Equation (36). Secondly, the parameters M, N, α and β in (39) are replaced by $\sqrt{M}$, $\sqrt{N}$, $\tilde{\alpha}$ and $\tilde{\beta}$, where $\tilde{\alpha} = \sqrt{\frac{3E_s}{N_1(N-1)}}$ and $\tilde{\beta} = \sqrt{\frac{3E_s}{N_1(M-1)}}$. However, these two differences do change the sign of the equation. Thus, $\frac{d\tilde{f}_1(\gamma)}{d\gamma} < 0$ always holds, which means the SER of User 2 keeps decreasing with the enlargement of γ. □

Proposition 8. *For the QAM based NOMA system, the suboptimal power allocation coefficient for problem* $\min_{\gamma}$ *SER$_{U2}$ is $\gamma_{sub} = \frac{M(N-1)}{MN-1}$. There always exists a better power allocation coefficient whose value is smaller than the suboptimal one, i.e., $\gamma_{btr} < \gamma_{sub}$, which can provide a lower SER$_{U2}$ for User 2.*

Proof. Similar to the proof for Proposition 7, the first derivative of SER_{U2} in γ can be calculated by modifying Equation (41) as well. We can conduct the same procedure for proving Proposition 6, and the condition to make $\frac{d\tilde{f}_2(\gamma)}{d\gamma}$ close to 0 in this case becomes to $\tilde{\upsilon}\sqrt{\gamma} = \sqrt{M}\tilde{\omega}\sqrt{1-\gamma}$, in which $\tilde{\upsilon} = \sqrt{\frac{3E_s}{N_2(N-1)}}$ and $\tilde{\omega} = \sqrt{\frac{3E_s}{N_2(M-1)}}$. Thus, the suboptimal power allocation coefficient of User 2 is $\gamma_{sub} = \frac{M(N-1)}{MN-1}$. Furthermore, $\frac{d\tilde{f}_2(\gamma)}{d\gamma} > 0$ always holds when $\gamma = \gamma_{sub}$. As $\tilde{f}_2(\gamma)$ is a continuous function, there must exist a smaller $\gamma_{btr} < \gamma_{sub}$, which makes the value of $\frac{d\tilde{f}_2(\gamma)}{d\gamma}$ become smaller. In other words, γ_{btr} can offer a lower SER for User 2 than γ_{sub}, i.e., $\exists\ \gamma_{btr} < \gamma_{sub} : f_2(\gamma_{btr}) < f_2(\gamma_{sub})$. □

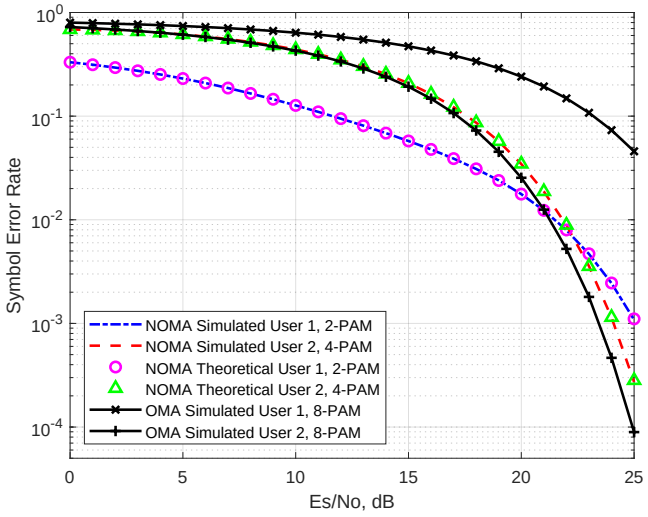

Fig. 5. Comparison of the theoretical and simulated SER of the PAM based NOMA system with $N = 2, M = 4, |h_1| = 0.5$, $|h_2| = 1$ and $\gamma = 0.8$.

Similar to the PAM case, the suboptimal solution of Problem (38) for the QAM case can also be expressed as (43) by substituting the corresponding γ_{sub} and γ_λ into it.

Overall, when $\gamma_{sub} \geq \gamma_\lambda$, Proposition 6 and Proposition 8 provide close-form expressions of suboptimal γ for PAM and QAM cases, respectively. The performance of each suboptimal solution is quite close to the corresponding optimal, which will be validated in the simulation section. Moreover, the closed-form expression allows the source immediately determines per frame the appropriate transmit powers for the users without any searching process.

VI. SIMULATIONS

In this section, we provide our simulation results to validate our analytical models including the SER expressions and propositions, and evaluate the performance of the proposed suboptimal power allocation. Moreover, the corresponding OMA schemes are also included in these simulations to demonstrate the superiority of the NOMA approach. In order to make a fair comparison, we assume that the total bit rate of OMA and NOMA systems are always the same. For example, when 4-QAM is employed for both two users in the NOMA system during two transmission frames, then in the corresponding OMA system, 16-QAM must be utilized for User 1 in the first transmission frame and User 2 in the second frame. In this way, the data rate of the two systems can be consistent, i.e., $2 \times \log_2 4 = \log_2 16 = 4$ bit/frame. In the following simulations, we assume that the additive Gaussian white noise of the two users are identical, namely, $\sigma_1^2 = \sigma_2^2 = N_0$ for the PAM based system and $\sigma_1^2 = \sigma_2^2 = \frac{N_0}{2}$ for one dimension of the QAM based NOMA system.

A. SER Expressions Validation

In the following, we validates our analytical models of PAM and QAM cases, respectively.

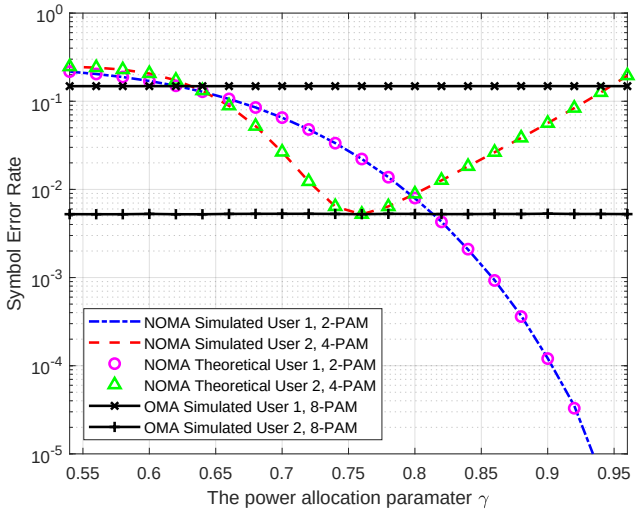

Fig. 6. The SER against the power allocation parameter γ of the PAM based NOMA system with $N = 2, M = 4, |h_1| = 0.5$ and $|h_2| = 1$.

1) NOMA with PAM: In this case, we conduct simulations in a scenario with $N = 2$, $M = 4$ and $\gamma = 0.8$. The channel coefficients for the two users are $|h_1| = 0.5$ and $|h_2| = 1$, respectively. The SER results of User 1 and User 2 are provided in Fig. 5 where plots of an 8-PAM OMA system under the two different channels are also provided for comparison. It is observed that the analytical SERs of both two users match the simulated SERs perfectly, which validates our SER expressions in equation (13) and (33). In addition, the SER of User 1 in the NOMA system is much lower than its SER in the OMA system. On the other hand, the SER of User 2 in the NOMA system is still comparable to that in the OMA system.

Next, to verify Proposition 5 and Proposition 6, another simulation is further provided where we investigate the impact of γ on SERs. In this simulation, we set $N = 2$, $M = 4$ and $\mathrm{Es/N_0} = 22$ dB and keep the channels h_1 and h_2 unchanged. The results are shown in Fig. 6. First, the SER of User 1 is decreasing with the increasing of γ, which matches well with our Proposition 5. In addition, it is observed that the SER of User 2 is quasi-convex in γ. By applying exhaustive search, we found that the minimal value is achieved at the point $\gamma_{min} \approx 0.758$. On the other hand, according to Proposition 6, the suboptimal power allocation parameter in this case is $\gamma_{sub} \approx 0.762$, which is slightly bigger than γ_{min}. Moreover, it can be observed that the bottom of the curve is relatively flat, i.e., the SER of User 2 has little changes in the interval $0.74 < \gamma < 0.78$. Therefore, the SER difference between γ_{sub} and γ_{min} is negligible. The above observations confirms our Proposition 6.

Now, we evaluate the performance of the proposed suboptimal solution γ^* given in (38) to Problem (43). Generally, depending on the feasible range of γ, there are two situations. In the first situation, the coefficient γ_{sub} is located within the feasible range of γ. For example, when assuming that the constraint of User 1 requires the SER of User 1 to be smaller than 10^{-1}, i.e., $\lambda = 10^{-1}$, by referring to Fig. 6, we note that the corresponding power allocation parameter (which makes the SER of User 1 equal to 10^{-1}) is $\gamma_\lambda \approx 0.66$. In addition, as γ grows, i.e., $\gamma > \gamma_\lambda$, the reliability requirement of User 1 is definitely satisfied. This actually defines the feasible range of γ

for the optimization problem and validates that the original problem (38) can be transferred to Problem (40). Then, the remaining work is to find the optimal γ which minimizes the SER of User 2 within its feasible range. Based on the analysis in the previous paragraph and considering Proposition 6, we learn that the suboptimal solution to minimizing the SER of User 2 without considering the feasible range of γ is $\gamma_{sub} \approx 0.762$, which is located within the feasible range of γ. Thus, in the first situation, the suboptimal solution of Problem (38) is $\gamma^* = \gamma_{sub}$, which proves (43) as $\gamma_{sub} > \gamma_\lambda$. Now, we consider the other situation, i.e., γ_{sub} is not in the feasible range of γ. When $\lambda = 10^{-2}$, the corresponding γ to achieve this SER becomes to $\gamma_\lambda \approx 0.8$. Intuitively, the suboptimal solution $\gamma_{sub} \approx 0.762$ is no longer feasible in this case. Considering that the SER of User 2 increases in γ when $\gamma \geq \gamma_{sub}$ and $\gamma_\lambda > \gamma_{sub}$, the optimal solution to Problem (38) with $\lambda = 10^{-2}$ is $\gamma^* = \gamma_\lambda$, which also matches the proposed solution in (43). Therefore, by summarizing the analysis of the two situations, it is validated that $\gamma^* = \max\{\gamma_{sub}, \gamma_\lambda\}$ is the suboptimal solution to minimize the SER of User 2 while ensuring the SER requirement of User 1 in the NOMA system.

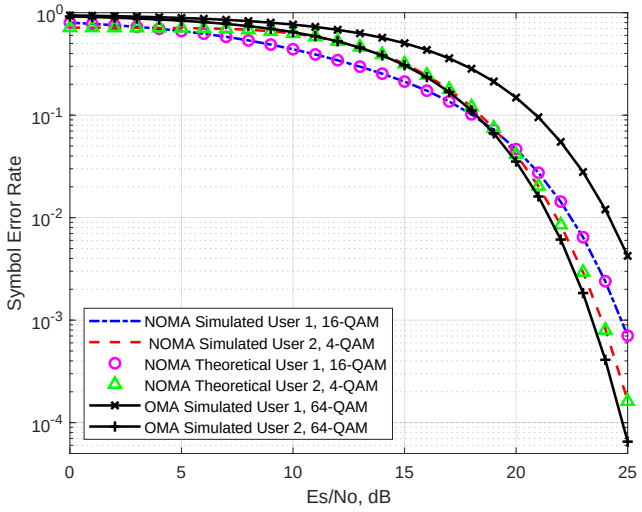

Fig. 7. Comparison of the theoretical and simulated SER of the QAM based NOMA system with $N = 16, M = 4, h_1 = 0.5 + 0.6j$, $h_2 = 0.7 + 0.8j$ and $\gamma = 0.96$.

2) NOMA with QAM: In the simulation for the QAM based NOMA system, we let $N = 16$, $M = 4$ and set the power allocation parameter to 0.96. The channel coefficients for the two users are set to $h_1 = 0.5 + 0.6j$ and $h_2 = 0.7 + 0.8j$. The simulation results are plotted in Fig. 7, where the plots of the Gray-coded 64-QAM OMA system are also included as baselines. Again, we observe that the analytical SER expressions of both users in the NOMA system match the simulation results, which validates (36) and (37). Moreover, the plots of this QAM based NOMA system are very similar to that of the PAM based NOMA (see Fig. 5), i.e., the SER of User 1 is much lower in the NOMA system when compared to the corresponding OMA system, while the SER of User 2 are very close in both NOMA and OMA systems.

Then, we further demonstrate the influence of γ on SERs. The results are shown in Fig. 8, where we set $\mathrm{Es}/\mathrm{N}_0 = 23$ dB. It is observed that the SER of User 1 is decreasing in γ. which confirms our Proposition 7. On the other hand, the SER curve of User 2 is quasi-convex in γ and the minimum value of the curve is at the point

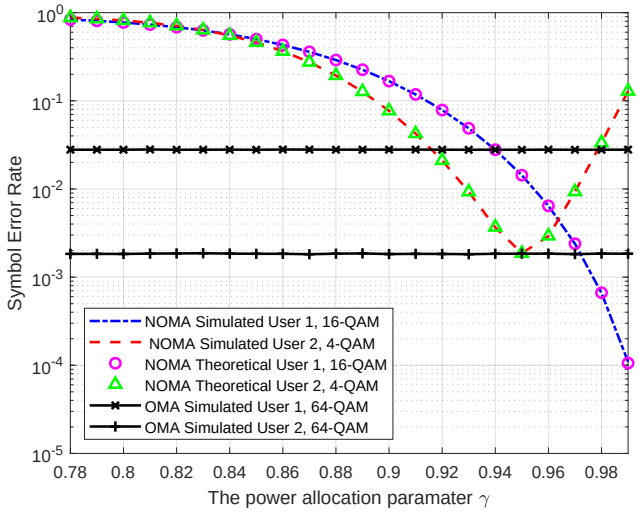

Fig. 8. The SER against the power allocation parameter γ of the QAM based NOMA system with $N = 16, M = 4, h_1 = 0.5 + 0.6j$ and $h_2 = 0.7 + 0.8j$.

$\gamma_{min} \approx 0.95157$ (by an exhaustive search). The proposed suboptimal power allocation parameter can be computed according to Proposition 8, which is $\gamma_{sub} = \frac{20}{21} \approx 0.95238$, Hence, it matches with Proposition 8 in that γ_{sub} is slightly greater than γ_{min} and the performance degradation of γ_{sub} is negligible.

Similar to the PAM based NOMA system, the proposed solution (43) for the optimization problem, i.e., minimizing the SER of User 2 while keeping the reliability performance of User 1, can also be validated in this QAM based NOAM system by referring to Fig. 8.

It is worth noting that the similar simulations with other modulation orders for both PAM and QAM based NOMA systems are also conducted in order to validate our propositions. Due to the page limitation, the results are not provided here.

B. Performance evaluation of the proposed power allocation

In this subsection, we aim at evaluating the proposed closed-form suboptimal power allocation solution in comparison with the optimal one. By examining Equation (43), we notice that when $\gamma_\lambda > \gamma_{sub}$, the equality $\gamma^* = \gamma_\lambda$ always holds, thus, the γ^* is always the optimal solution, i.e., there is no gap between the suboptimal and optimal power allocation parameters. Therefore, in order to investigate the maximal performance error of the proposed suboptimal solution in comparison to the optimal one, we set a sufficient loose reliability constant with respects to User 1 in the following simulations, which corresponds to making γ_λ sufficiently be close to zero.

Moreover, for all simulations in this subsection, we consider the following scenarios, i.e., PAM with $N = 2$ and $M = 4$, PAM with $N = 4$ and $M = 4$, QAM with $N = 4$ and $M = 4$, and QAM with $N = 16$ and $M = 4$, while varying $\mathrm{Es/N_0}$ from 15 dB to 23 dB. The channel coefficients for the PAM case are $h_1 = 0.5$ and $h_2 = 1$, while for the QAM case are $h_1 = 0.5 + 0.6j$ and $h_2 = 0.7 + 0.8j$.

Both the performance difference in term of SER of User 2 and the difference between the choice of γ between the proposed suboptimal approach of the optimal approach are provided in Fig. 9 and Fig. 10, respectively. In particular,

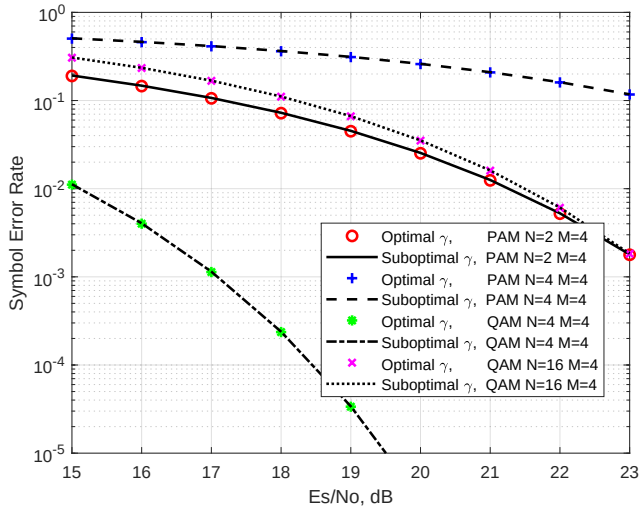

Fig. 9. The SER of User 2 against Es/N_0 with optimal and suboptimal power allocation parameters for PAM and QAM based NOMA systems.

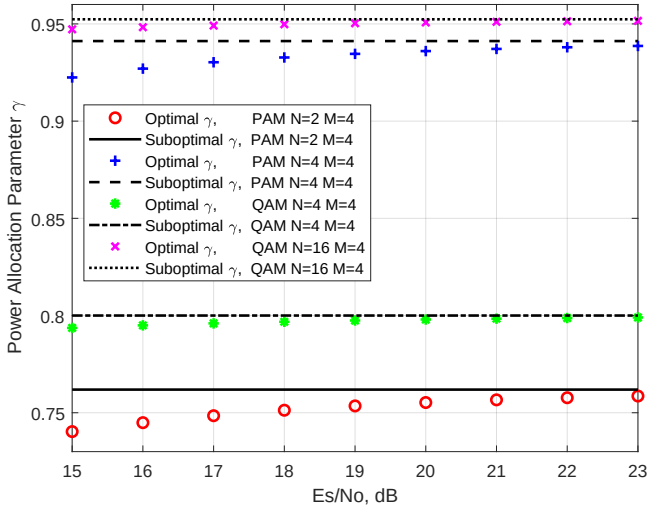

Fig. 10. The optimal and suboptimal values of the power allocation parameter γ against Es/N_0 for PAM and QAM based NOMA systems

the optimal values of the power allocation parameter γ in these simulations are obtained from the analytical SER expressions for User 2 by applying the exhaustive search with a resolution of 0.00001. From Fig. 9, it is observed that the performance difference (between the optimal and suboptimal approaches) under all scenarios are negligible. On the other hand, when we check Fig. 10, it can be seen that there only exist very small gaps between the optimal and suboptimal power allocation parameters. Moreover, as Es/N_0 increases, these gaps become more tiny and invisible. Therefore, the suboptimal approach has an excellent performance, which is a tight approximation of the optimal one.

As seen in the our discussion in the previous sections, it is unlikely to obtain a closed-form expression for the optimal power allocation parameter. The proposed closed-form expression of this suboptimal allows us to investigate how the (approximate) optimal power allocation is influenced by users' constellation order M and N. For example,

$$\frac{df_2(\gamma)}{d\gamma}\Big|_{\gamma=\gamma_{sub}} = \frac{(M-1)\omega}{M\sqrt{2\pi(1-\gamma)}}e^{-\frac{\omega^2(1-\gamma)}{2}} + \frac{2}{MN\sqrt{2\pi}}\sum_{k=1}^{N-1}(N-k)\left\{\frac{(M-1)k\upsilon}{\sqrt{\gamma}}\left(e^{-\frac{(2kM-1)^2\omega^2(1-\gamma)}{2}} - e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}}\right)\right\}$$
$$+\frac{1}{MN\sqrt{2\pi}}\sum_{k=1}^{N-1}(N-k)\left\{\frac{(M-1)\omega}{\sqrt{1-\gamma}}e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}} + \frac{(2k-1)\upsilon}{\sqrt{\gamma}}e^{-\frac{(2kM-1)^2\omega^2(1-\gamma)}{2}}\right\}$$
$$-\frac{1}{MN\sqrt{2\pi}}\sum_{k=1}^{N-1}(N-k)e^{-\frac{(2(k-1)M+1)^2\omega^2(1-\gamma)}{2}}\left\{\frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{(2k-1)\upsilon}{\sqrt{\gamma}}\right\} = g_1 + g_2 + g_3 + g_4 \tag{44}$$

according to Proposition 6 and Proposition 8, we can easily show that for both PAM and QAM cases γ_{sub} is increasing in M but decreasing in N. Then, we can reasonably conclude the same characteristic for the optimal approach.

VII. Conclusion

In this work, we determined SER expressions of a downlink NOMA system under fading channels with arbitrary ordered PAM and QAM symbols. Furthermore, the impact of the power allocation on the reliability, i.e., the SER of each user in the NOMA system, was investigated based their analytical SER expressions. Particularly, an optimization problem regarding the power allocation parameter was proposed, which aims at minimizing the SER of the user with a better channel gain while guaranteeing the reliability demand for the user in a poorer channel condition. By analyzing the derived SER expressions for the NOMA system, we provided a closed-form suboptimal solution to this problem. The accuracy of our analytical model (including both of the derived SER expressions and our propositions) was validated by simulations. Furthermore, the proposed suboptimal power allocation approach was verified as a tight approximation of the optimal one which provides the best SER for the user with the better channel condition.

It should be pointed out that the procedure to obtain the SER expressions in the two users NOMA system also applies to a multiuser NOMA system. In our future work, we will investigate analytical SER expressions of NOMA systems with arbitrary users and study the power allocation schemes based on these expressions. On the other hand, in this paper, we assume that perfect SIC is achievable, which is not always true in practical systems. Therefore, investigating the analytical SER of the NOMA system with imperfect SIC and the corresponding optimal power allocation scheme is considered are interesting topics, which will be considered in our future work.

Appendix A

In this part, we demonstrate that the first derivative of the SER expression for User 2 on the power allocation coefficient γ is always greater than 0 at the proposed suboptimal point γ_{sub}, i.e., $\frac{df_2(\gamma)}{d\gamma}\big|_{\gamma=\gamma_{sub}} > 0$ always holds. As the condition for obtaining γ_{sub} is $\upsilon\sqrt{\gamma} = M\omega\sqrt{1-\gamma}$, therefore, instead of substituting $\gamma_{sub} = \frac{M^2(N^2-1)}{M^2N^2-1}$ into Equation (41), we can obtain $\frac{df_2(\gamma)}{d\gamma}\big|_{\gamma=\gamma_{sub}}$ by replacing the $\upsilon\sqrt{\gamma}$ components with $M\omega\sqrt{1-\gamma}$ for all exponential

terms in (41), which results in (44). Furthermore, Equation (44) can be divided into four parts, correspondingly denoted as g_1, g_2, g_3 and g_4.

Note that in (41), parameters $\gamma, \omega, \upsilon > 0$, M and N are integer modulation orders with $M, N \geq 2$. Intuitively, g_1 and g_3 are always positive. In addition, as $(2kM-1)^2 < (2kM+1)^2$ always holds for all $k = 1, 2, \ldots, N-1$, we can derive that the second part g_2 is greater than 0 as well. However, part g_4 is always negative. Therefore, to prove that $\frac{df_2(\gamma)}{d\gamma}\big|_{\gamma=\gamma_{sub}} > 0$ is always true, we have to merge g_4 with other positive parts and demonstrate that the resulting combinations are always greater than 0. In particular, by combining parts of g_4 with parts of g_3 and merging the remaining parts of g_4 with g_1, we can show that $g_1 + g_3 + g_4 > 0$.

We start with combining g_4 with g_3. Observing the $k-1$ components in the exponential terms of g_4 prohibit us from merging g_4 with g_3, we can rewrite g_4 and combine parts of it with the corresponding parts of g_3. By partitioning the summation of g_4 into $k = 1$ and $k = 2, 3, \ldots, N-1$, and denoting the two parts as $\hat{g}_4$ and $\bar{g}_4$, respectively, g_4 can be expressed as $g_4 = \hat{g}_4 + \bar{g}_4$, in which

$$\hat{g}_4 = -\frac{N-1}{MN\sqrt{2\pi}} e^{-\frac{\omega^2(1-\gamma)}{2}} \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{\upsilon}{\sqrt{\gamma}} \right\}, \tag{45}$$

and

$$\begin{aligned} \bar{g}_4 = -\frac{1}{MN\sqrt{2\pi}} \sum_{k=2}^{N-1} (N-k) e^{-\frac{(2(k-1)M+1)^2\omega^2(1-\gamma)}{2}} \\ \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{(2k-1)\upsilon}{\sqrt{\gamma}} \right\}. \end{aligned} \tag{46}$$

Let $k' = k-1$, then $\bar{g}_4$ can be rewritten as

$$\begin{aligned} \bar{g}_4 = -\frac{1}{MN\sqrt{2\pi}} \sum_{k'=1}^{N-2} (N-k'-1) e^{-\frac{(2k'M+1)^2\omega^2(1-\gamma)}{2}} \\ \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{(2k'+1)\upsilon}{\sqrt{\gamma}} \right\}. \end{aligned} \tag{47}$$

After rewriting g_4, we must modify g_3 so that they can be combined. Observing that $e^{-\frac{(2kM-1)^2\omega^2(1-\gamma)}{2}} > e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}}$ always holds $\forall k = 1, 2, \ldots, N-1$, we can achieve a lower bound g_3' for g_3, which makes $g_3 > g_3' > 0$ always true. And g_3' can be expressed as

$$\begin{aligned} g_3' = \frac{1}{MN\sqrt{2\pi}} \sum_{k=1}^{N-1} (N-k) e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}} \\ \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{(2k-1)\upsilon}{\sqrt{\gamma}} \right\} > 0. \end{aligned} \tag{48}$$

By dividing the summation of (48) into $k = N-1$ and $k = 1, 2, \ldots, N-2$, and denoting the corresponding parts as $\hat{g}_3'(\gamma)$ and $\bar{g}_3'(\gamma)$, we can get $g_3'(\gamma) = \hat{g}_3'(\gamma) + \bar{g}_3'(\gamma)$, in which,

$$\begin{aligned} \hat{g}_3' = \frac{1}{MN\sqrt{2\pi}} e^{-\frac{(2(N-1)M+1)^2\omega^2(1-\gamma)}{2}} \\ \left\{ \frac{(M-1)\,\omega}{\sqrt{1-\gamma}} + \frac{(2\,(N-1)-1)\,\upsilon}{\sqrt{\gamma}} \right\}, \end{aligned} \tag{49}$$

and

$$\bar{g}_3' = \frac{1}{MN\sqrt{2\pi}} \sum_{k=1}^{N-2} (N-k) e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}} \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} + \frac{(2k-1)\upsilon}{\sqrt{\gamma}} \right\}. \tag{50}$$

Intuitively, $\hat{g}_3'(\gamma)$ is always greater than 0. Moreover, we can combine $\bar{g}_3'(\gamma)$ with the expression of $\bar{g}_4(\gamma)$ in Equation (47), which can be simplified as

$$\bar{g}_3' + \bar{g}_4 = \frac{1}{MN\sqrt{2\pi}} \sum_{k=1}^{N-2} e^{-\frac{(2kM+1)^2\omega^2(1-\gamma)}{2}} \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} - \frac{(2N-4k+1)\upsilon}{\sqrt{\gamma}} \right\}. \tag{51}$$

As all the elements outside the curly braces of (51) are positive, thus, the sign of this equation is determined by the terms inside the curly braces. In addition, the value of the last term $\frac{(2N-4k+1)\upsilon}{\sqrt{\gamma}}$ decreases as k increases for $k = 1, 2, \ldots, N-2$, therefore, we can derive that

$$\frac{(M-1)\omega}{\sqrt{1-\gamma}} - \frac{(2N-4k+1)\upsilon}{\sqrt{\gamma}} \geq \frac{(M-1)\omega}{\sqrt{1-\gamma}} - \frac{(2N-3)\upsilon}{\sqrt{\gamma}}. \tag{52}$$

Recall that the condition $\upsilon\sqrt{\gamma} = M\omega\sqrt{1-\gamma}$ must be satisfied to obtain γ_{sub}, therefore, by substituting $\sqrt{1-\gamma} = \frac{\upsilon\sqrt{\gamma}}{M\omega}$, $\upsilon = \sqrt{\frac{6E_s}{N_2(N+1)(N-1)}}$ and $\omega = \sqrt{\frac{6E_s}{N_2(M+1)(M-1)}}$ into the right side of inequality (52), we can acquire

$$\begin{aligned} \frac{(M-1)\omega}{\sqrt{1-\gamma}} - \frac{(2N-3)\upsilon}{\sqrt{\gamma}} &= \frac{1}{\upsilon\sqrt{\gamma}} \left\{ (M-1)M\omega^2 - (2N-3)\upsilon^2 \right\} \\ &= \frac{6E_s}{N_2\upsilon\sqrt{\gamma}} \left\{ 1 - \frac{1}{M+1} - \frac{2}{N+1} + \frac{1}{N^2-1} \right\} > 0, \end{aligned} \tag{53}$$

in which the inequality holds as all the parameters before the curly braces of (53) are positive, and the modulation order is $M, N \geq 2$. Thus, by combining formula (51), (52) and (53), we can prove that

$$\bar{g}_3' + \bar{g}_4 > 0. \tag{54}$$

Now, the only remaining negative part is $\hat{g}_4$. By combining it with g_1 and substituting the expressions of $\sqrt{1-\gamma}$, υ and ω utilized in the derivation of (53) into the combination, we have

$$\begin{aligned} g_1 + \hat{g}_4 &= \frac{1}{MN\sqrt{2\pi}} e^{-\frac{\omega^2(1-\gamma)}{2}} \left\{ \frac{(M-1)\omega}{\sqrt{1-\gamma}} - \frac{(N-1)\upsilon}{\sqrt{\gamma}} \right\} \\ &= \frac{\omega}{MN\upsilon\sqrt{2\pi\gamma}} e^{-\frac{\omega^2(1-\gamma)}{2}} \left\{ (M-1)M\omega^2 - (N-1)\upsilon^2 \right\} \\ &= \frac{6E_s\omega}{N_2MN\upsilon\sqrt{2\pi\gamma}} e^{-\frac{\omega^2(1-\gamma)}{2}} \left\{ 1 - \frac{1}{M+1} - \frac{1}{N+1} \right\} > 0, \end{aligned} \tag{55}$$

in which the inequality holds as the elements outside the curly braces are all positive and the modulation order is $M, N \geq 2$.

According to the analysis above, the sign of $\frac{df_2(\gamma)}{d\gamma}\Big|_{\gamma=\gamma_{sub}}$ given in Equation (44) can be determined, which is

$$\frac{df_2(\gamma)}{d\gamma}\Big|_{\gamma=\gamma_{sub}} = g_1 + g_2 + g_3 + g_4 \tag{56a}$$

$$> g_1 + g_2 + g_3' + g_4 \tag{56b}$$

$$= (g_1 + \hat{g}_4) + g_2 + (\bar{g}_3' + \bar{g}_4) + \hat{g}_3' \tag{56c}$$

$$> 0. \tag{56d}$$

The inequality in (56b) holds as g_3' is the lower bound of g_3, i.e., $g_3 > g_3'$. As $g_3' = \hat{g}_3' + \bar{g}_3'$ and $g_4 = \hat{g}_4 + \bar{g}_4$, (56c) can be derived from (56b). Considering g_2 is always greater than 0 and combining (48), (54) and (55), the inequality in (56d) can be obtained.

REFERENCES

[1] X. Gao, L. Dai, S. Han, C. I and R. W. Heath, "Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.

[2] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106-113, Feb. 2014.

[3] L. Dai, B. Wang, Y. Yuan, S. Han, C. I and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74-81, Sept. 2015.

[4] X. Yue, Z. Qin, Y. Liu, S. Kang and Y. Chen, "A Unified Framework for Non-Orthogonal Multiple Access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5346-5359, Nov. 2018.

[5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

[6] Q. He, Y. Hu and A. Schmeink, "Resource allocation for ultra-reliable low latency communications in sparse code multiple access networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 282-290, Dec. 2018.

[7] M. Ali, H. Tabassum, E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325-6343, Aug. 2016.

[8] X. Yue, Y. Liu, S. Kang and A. Nallanathan, "Performance analysis of NOMA with fixed gain relaying over Nakagami-m fading channels," *IEEE Access*, vol. 5, pp. 5445-5454, Mar. 2017.

[9] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34 no. 4, pp. 938-953, Mar. 2016

[10] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656-1672, Jan. 2017

[11] Z. Ding, F. Adachi and H. V. Poor, "The Application of MIMO to Non-Orthogonal Multiple Access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537-552, Jan. 2016.

[12] L. Lv, J. Chen, Q. Ni and Z. Ding, "Design of Cooperative Non-Orthogonal Multicast Cognitive Multiple Access for 5G Systems: User Scheduling and Performance Analysis," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2641-2656, June 2017.

[13] F. Zhou, Z. Chu, H. Sun, R. Q. Hu and L. Hanzo, "Artificial Noise Aided Secure Cognitive Beamforming for Cooperative MISO-NOMA Using SWIPT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 918-931, April 2018.

[14] L. Lei, D. Yuan, C. K. Ho and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8580-8594, Dec. 2016.

[15] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647-1651, Oct. 2015.

[16] M. Moltafet, P. Azmi, N. Mokari, M. R. Javan and A. Mokdad, "Optimal and fair energy efficient resource allocation for energy harvesting-enabled-PD-NOMA-based HetNets," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2054-2067, Mar. 2018.

[17] C. L. Wang, J. Y. Chen and Y. J. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 532-535, Oct. 2016.

[18] J. Zhu, J. Wang, Y. Huang, S. He, X. You and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744-2757, Dec. 2017.

[19] F. Fang, H. Zhang, J. Cheng and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722-3732, Sept. 2016.

[20] Y. Zhang, H. M. Wang, T. X. Zheng and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852-2857, Mar. 2017.

[21] Q. Sun, S. Han, C. L. I and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, June 2015, pp. 2668-2673.

[22] X. Wang, F. Labeau and L. Mei, "Closed-form BER expressions of QPSK constellation for uplink non-orthogonal multiple access (NOMA)," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2242-2245, Oct. 2017.

[23] H. Huang, J. Wang, J. Wang, J. Yang, J. Xiong and G. Gui, "Symbol error rate performance analysis of non- orthogonal multiple access for visible light communications," *China Commun.*, vol. 14, no. 12, pp. 153-161, December 2017.