On the Power Allocation for a Practical Multiuser Superposition Scheme in NOMA Systems

Jinho Choi

***Abstract*—In this letter, we study nonorthogonal multiple access (NOMA) systems with a linear multiuser superposition transmission (MUST) scheme. To perform the power allocation or the determination of the transmission power ratio for linear MUST with practical modulation schemes of finite constellation sizes, we consider the mutual information. An approach based on Monte Carlo simulations is devised to obtain the mutual information. While the sum rate maximization (based on the capacity) has a trivial solution, numerical examples show that there is a nontrivial solution to the power allocation problem for maximizing total mutual information. Thus, for certain given conditions (e.g., the total transmission power and modulation schemes), we are able to decide the optimal power (and rate) allocation for the linear MUST scheme to maximize the total mutual information.**

***Index Terms*—Nonorthogonal multiple access, code division multiple access, hierarchical modulation.**

I. Introduction

VARIOUS multiuser superposition transmission (MUST) schemes are considered for non-orthogonal multiple access (NOMA) systems in wireless standards [1]. Among those, a linear MUST scheme is based on the notion of superposition coding (SC) [2] that can achieve the capacity of multiple access channel (MAC) using successive interference cancellation (SIC) [3]. The linear MUST scheme is also similar to the hierarchical modulation scheme that is widely used for broadcast [4].

NOMA is also studied for various systems including multiple input multiple output (MIMO) systems with the linear MUST scheme. In [5], NOMA is studied for downlink coordinated two-point systems. In the case that the base station (BS) is equipped with multiple antennas, beamforming can be used for NOMA downlink transmissions as in [6], [7]. A performance analysis is presented in [8] and a power allocation problem for NOMA is studied in [9]. For open-loop MIMO downlink transmissions, in [10], the ergodic capacity of MIMO-NOMA with two users is derived and the power allocation is carried out (i.e., the BS allocates the power to two users' signals based on statistical channel state information (CSI)).

In general, the power allocation is important for NOMA and can be carried out by dividing a total power to two symbols in the linear MUST scheme for two users. The optimal power allocation to maximize the sum rate in downlink transmissions becomes trivial in the case that one user is close to the BS (this user is referred to as user 1) and the other user is far away from the BS (this user is referred to as user 2). In this case, all the power is to be assigned to user 1 to maximize the sum rate. Since this case is a typical one in NOMA where the power domain is to be exploited, we may need to consider constraints for individual users (in terms of power or rate) or different criteria (e.g., the fairness [9]) to avoid this trivial solution to the power allocation.

In this letter, we consider the power allocation for the linear MUST scheme in NOMA downlink transmissions with two users when practical modulation schemes are employed. Due to practical modulation schemes of finite constellation sizes, we use the mutual information rather than capacity for the objective function in the power allocation problem. Since a closed-form expression for the mutual information is not available, a numerical approach based on Monte Carlo simulations is proposed. It is shown that the solution to the optimal power allocation problem (to maximize the total mutual information with practical modulation schemes) is not trivial and depends on the modulation schemes employed.

II. NOMA System With Two Users

A. Linear Multiuser Superposition Transmissions

Suppose that a BS is to transmit two symbols to two users over the same radio resource block. In the linear MUST scheme [1], the signal to be transmitted is given by

$$x = s_1 + s_2, \tag{1}$$

where $s_k \in \mathcal{S}_k$ denotes the symbol to user k and $\mathcal{S}_k$ is the associated signal constellation. The resulting symbol, x, can be seen as a symbol of hierarchical modulation. It is assumed that

$$\mathbb{E}[s_k] = 0 \text{ and } \mathbb{E}[|s_k|^2] = P_k, \tag{2}$$

where P_k is the signal power to symbol s_k or user k and $\mathbb{E}[\cdot]$ denotes the statistical expectation. Let $P_T = P_1 + P_2$ be the total transmission power. According to [1], the transmission power ratio for the MUST-near user is defined as $\alpha = \frac{P_1}{P_T}$.

Denote by h_k the channel coefficient from the BS to user k. Then, the received signals at users 1 and 2, denoted by y and z, respectively, are given by

$$y = h_1 x + n_1 \text{ and } z = h_2 x + n_2, \tag{3}$$

where $n_k \sim \mathcal{CN}(0, 1)$. Here $\mathcal{CN}(\mu, \sigma^2)$ represents the probability density function (pdf) of a circularly symmetric complex Gaussian random variable with mean μ and variance σ^2. Note that the noise variance is normalized for convenience.

Manuscript received November 11, 2015; accepted January 4, 2016. Date of publication January 14, 2016; date of current version March 8, 2016. The associate editor coordinating the review of this paper and approving it for publication was Z. Ding.

The author is with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju 500 712, South Korea (e-mail: jchoi0114@gist.ac.kr).

Digital Object Identifier 10.1109/LCOMM.2016.2518165

B. Maximum Sum Rate Power Allocation

Let $\kappa_k = |h_k|^2$. Furthermore, it is assumed that s_1 and s_2 are Gaussian in this subsection.

At user 1, s_2 is to be detected and decoded first. Once s_2 is decoded and recovered, it can be removed from y in decoding s_1. Thus, the achievable rate for s_2, denoted by R_2, is bounded by $R_2 \le \log_2\left(1 + \frac{\kappa_1 P_2}{\kappa_1 P_1 + 1}\right)$. Since s_2 can be removed, the achievable rate for s_1, denoted by R_1, becomes $R_1 \le C_1 = \log_2(1 + \kappa_1 P_1)$, where C_1 is the capacity of the channel from the BS to user 1 without s_2.

At user 2, it is assumed that the signal to user 1 is negligible if $P_1 < P_2$. Thus, s_2 is to be detected and decoded in the presence of the interference due to s_1. The resulting achievable rate for s_2 becomes

$$R_2 \le C_2 = \log_2\left(1 + \frac{\beta P_2}{\beta P_1 + 1}\right), \tag{4}$$

where $\beta = \min\{\kappa_1, \kappa_2\}$ and C_2 is the capacity of the channel from the BS to user 2 in the presence of the interference s_1. In general, it is expected that $\kappa_1 > \kappa_2$. In this case, the sum rate becomes

$$\begin{aligned} R_1 + R_2 \le \log_2(1 + \kappa_2 P_T) + \log_2(1 + \kappa_1 P_1) \\ - \log(1 + \kappa_2 P_2). \end{aligned} \tag{5}$$

Note that in orthogonal multiple access (OMA), the sum rate becomes

$$R_1 + R_2 \le \frac{1}{2}\left(\log_2(1 + \kappa_1 P_T) + \log_2(1 + \kappa_2 P_T)\right), \tag{6}$$

which is generally lower than the sum rate of NOMA.

From (5), for a given total power P_T, we can see that the sum rate is maximized if $P_1 = P_T$ and $P_2 = 0$ (or $\alpha = 1$). That is, the optimal solution to the following problem,

$$\begin{aligned} \{P_1, P_2\} = \underset{P_1 + P_2 \le P_T}{\operatorname{argmax}} \ \log_2(1 + \kappa_2 P_T) \\ + \log_2(1 + \kappa_1 P_1) - \log(1 + \kappa_2 P_2), \end{aligned} \tag{7}$$

becomes $P_1 = P_T$ and $P_2 = 0$. In this power allocation, we have assumed that *i)* the channel coefficients are known to the BS; *ii)* Gaussian codebooks are used to achieve the capacity. While the first assumption (i.e., the known CSI at the BS) might be valid due to the channel reciprocity in time division duplex (TDD) mode or CSI feedback in frequency division duplex (FDD) mode, the second assumption may not be true when conventional modulation schemes (e.g., M-ary QAM) are used. As a result, we have a different power allocation result, which depends on the employed modulation scheme.

III. Mutual Information With Practical Modulation Schemes

With practical modulation schemes (of finite constellation sizes) for linear MUST, it is desirable to consider the mutual information to decide the code rate and power allocation rather than the capacity (as practical modulation schemes cannot achieve the capacity). In this section, we explain how the mutual information of linear MUST can be obtained for given signal constellations of s_1 and s_2, and consider for the power allocation.

Suppose that the number of elements in $\mathcal{S}_k$ is finite. Let $M_k = |\mathcal{S}_k|$. Furthermore, assume that s_k is equally likely. Thus, we have $\Pr(s_k) = \frac{1}{M_k},\ s_k \in \mathcal{S}_k$. For a finite signal constellation, the mutual information without any interference is widely studied and called coded modulation (CM) capacity [11] [12]. In particular, the mutual information between y and s_1 for given s_2 is well-known and given by

$$\mathsf{I}(y; s_1 \,|\, s_2) = \mathsf{I}(\tilde{y}; s_1) = \mathbb{E}\left[\log_2 \frac{f(\tilde{y} \,|\, s_1)}{f(\tilde{y})}\right], \tag{8}$$

where $\tilde{y} = y - h_1 s_2 = h_1 s_1 + n_1,\ f(\tilde{y} \,|\, s_1)$ is the conditional pdf of $\tilde{y}$ for given s_1, and $f(\tilde{y})$ is the pdf of $\tilde{y}$.

Assuming that $\kappa_1 > \kappa_2$, we can decide the rate for s_2 at user 2 using the mutual information. The mutual information between z and s_2 is given by

$$\begin{aligned} \mathsf{I}(z; s_2) &= \mathbb{E}\left[\log_2 \frac{f(z \,|\, s_2)}{f(y)}\right] \\ &= \sum_{s_2 \in \mathcal{S}_2} \int f(z, s_2) \log_2 \frac{f(z \,|\, s_2)}{f(z)} dz. \end{aligned} \tag{9}$$

The conditional pdf of z for given s_2 is given by

$$f(z \,|\, s_2) = \sum_{s_1 \in \mathcal{S}_1} f(z \,|\, s_1, s_2) \Pr(s_1), \tag{10}$$

where $f(z \,|\, s_1, s_2) = \frac{1}{\pi} \exp\left(-|z - h_2(s_2 + s_1)|^2\right)$. In general, it is not easy to derive a closed-form expression for $\mathsf{I}(z; s_2)$ in (9). Thus, we may use Monte-Carlo simulations to find an estimate of $\mathsf{I}(v; s_2)$. In this case, the following expression becomes useful:

$$\begin{aligned} \mathsf{I}(z; s_2) &= \log_2 M_2 \\ &\quad - \frac{1}{M_2} \sum_{s_2 \in \mathcal{S}_2} \mathbb{E}\left[\log_2 \left(\frac{\sum_{s_2' \in \mathcal{S}_2} f(z|s_2')}{f(z|s_2)}\right) \middle| \, s_2\right] \\ &= \log_2 M_2 - \frac{1}{M_2} \sum_{s_2 \in \mathcal{S}_2} \mathbb{E}\left[A(z, s_2)|s_2\right], \end{aligned} \tag{11}$$

where

$$\begin{aligned} A(z, s_2) &= \log_2 \frac{\sum_{s_2' \in \mathcal{S}_2} f(z|s_2')}{f(z|s_2)} \\ &= \log_2 \frac{\sum_{s_2' \in \mathcal{S}_2} \sum_{s \in \mathcal{S}_1} e^{-|z - h_2(s_1 + s_2')|^2}}{\sum_{s \in \mathcal{S}_1} e^{-|z - h_2(s_1 + s_2)|^2}}. \end{aligned} \tag{12}$$

Then, we have

$$\begin{aligned} \mathbb{E}[A(z, s_2)|s_2] &= \mathbb{E}\left[\mathbb{E}[A(z, s_2)|s_1, s_2]\right] \\ &= \sum_{s_1 \in \mathcal{S}_1} \mathbb{E}[A(z, s_2)|s_1, s_2] \Pr(s_1). \end{aligned} \tag{13}$$

In the term on the right-hand side (RHS) of the last equality in (13), the expectation can be replaced with the sample mean of

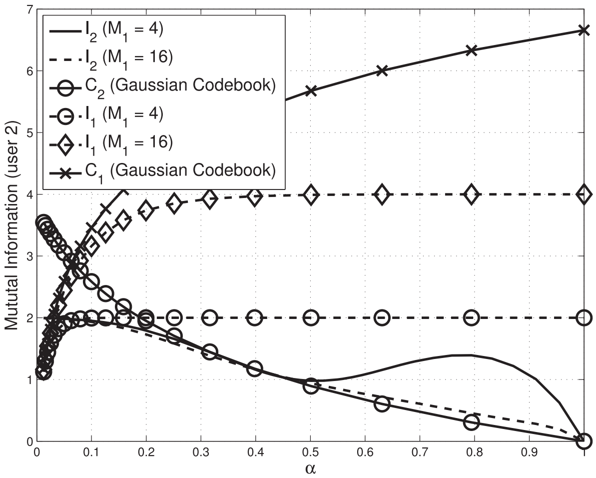

Fig. 1. Mutual information versus α when $P_T = 20$ dB and $M_2 = 4$ (I_1 and I_2 stand for $\mathsf{I}(y; s_1|s_2)$ and $\mathsf{I}(z; s_2)$, respectively).

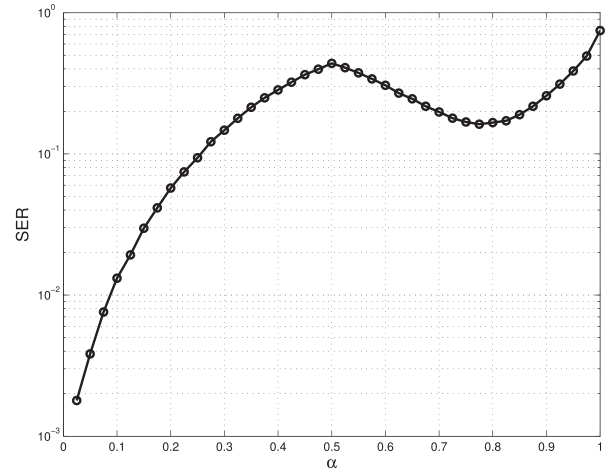

Fig. 2. Symbol error rate versus α when $P_T = 20$ dB and $M_1 = M_2 = 4$.

a number of realizations of z generated from the distribution, $f(z|s_1, s_2)$, i.e.,

$$\mathbb{E}[A(z, s_2)|s_1, s_2] = \int A(z, s_2) f(z|s_1, s_2) dz \approx \frac{1}{N} \sum_{n=1}^{N} A(z_{(n)}, s_2), \tag{14}$$

where $z_{(n)} \sim f(z|s_1, s_2)$ and N is the number of samples.

Note that if $\kappa_2 > \kappa_1$, we need to find $\mathsf{I}(y; s_2)$, which can also be obtained by the Monte Carlo simulation approach mentioned in above.

The power allocation problem with practical modulation schemes can be given by

$$\underset{P_1+P_2 \le P_T}{\operatorname{argmax}} \ \min\{\mathsf{I}(y; s_2), \mathsf{I}(z; s_2)\} + \mathsf{I}(y; s_1). \tag{15}$$

Since we do not have closed-form expressions for $\mathsf{I}(y; s_2)$, $\mathsf{I}(z; s_2)$, and $\mathsf{I}(y; s_1)$, the above numerical approach based on Monte Carlo simulations allows us to estimate the values of $\mathsf{I}(y; s_2)$, $\mathsf{I}(z; s_2)$, and $\mathsf{I}(y; s_1)$ and to decide P_1 and P_2 in (15) (note that once P_1 and P_2 are decided, the code rates are also decided according to the mutual information in (8) and (9).

IV. Numerical Results and Discussions

In this section, we consider few examples for the power allocation in (15) with M-ary QAM, where M is a power of an even number (e.g., 2 or 4 for 4-QAM or 16-QAM, respectively).

Suppose that $\kappa_1 = 1$ and $\kappa_2 = \frac{1}{d^\eta}$, where d is the normalized distance between the BS and user 2 and η is the path loss exponent (we only consider large-scale fading for convenience). With $d = 2$, $\eta = 3$, and $P_T = 20$ dB, the mutual informations, $\mathsf{I}(z; s_2)$ and $\mathsf{I}(y; s_1|s_2)$ are obtained by the Monte Carlo method with $N = 10,000$ samples. For s_1, we consider both 4-QAM and 16-QAM (i.e., $M_1 \in \{4, 16\}$), while $M_2 = 4$ (i.e., 4-QAM is employed for s_2). The results are shown in Fig. 1. In Fig. 1, $\mathsf{I}(z; s_1|s_2)$ is shown for different values of P_1, where we can see typical behaviors of the mutual information. The maximum value of $\mathsf{I}(z; s_1|s_2)$ is $\log_2 M_1$ and increases with P_1. At a higher P_1, certainly, a higher order modulation is desirable to approach the capacity. That is, if $P_1 > 4$ dB, $M_1 = 16$ is preferable to $M_1 = 4$. Furthermore, the mutual information is always lower than or equal to the capacity, $C_1 = \log_2(1 + \kappa_1 P_1)$, which can be achieved with a Gaussian codebook.

Since $P_2 = P_T - P_1$, when P_1 is small, $\mathsf{I}(z; s_2)$ can achieve the maximum value, which is $\log_2 M_2 = 2$, as shown in Fig. 1 (by the solid lines with square and cross markers). Unlike $\mathsf{I}(y; s_1|s_2)$, however, we can see few different behaviors. For example, the mutual information can be higher than $C_2 = \log_2\left(1 + \frac{\kappa_2 P_2}{\kappa_2 P_1 + 1}\right)$, when P_1 is large or P_2 is small. Since C_2 is obtained under the assumption that the interference s_1 is Gaussian, it is not the maximum achievable rate for s_2 when s_1 is a symbol from a finite constellation $\mathcal{S}_1$. Thus, $\mathsf{I}(z; s_2)$ can be larger than C_2. Another interesting observation is that $\mathsf{I}(z; s_2)$ can increase when P_2 increases (and P_1 decreases), which has been observed with the curve of $M_1 = 4$ in Fig. 1 (i.e., the solid line with square markers). This behavior can be explained from the signal detection point of view. At user 2, the maximum likelihood (ML) detection can be considered to detect s_2. From (10), we have

$$\begin{aligned}\hat{s}_2 &= \underset{s_2 \in \mathcal{S}_2}{\operatorname{argmax}} \ f(z \mid s_2) \\ &= \underset{s_2 \in \mathcal{S}_2}{\operatorname{argmax}} \sum_{s_1 \in \mathcal{S}_1} \exp\left(-|z - h_2(s_1 + s_2)|^2\right). \end{aligned} \tag{16}$$

If P_1 is sufficiently large, the term of the true s_1 might be dominant in the sum, which can effectively cancel the interference from s_1 and result in a good detection performance for s_2. However, if P_1 approaches P_T, the signal power to s_2 decreases, which results in a bad detection performance. The simulation result of the symbol error rate (SER) for s_2 at user 2 is shown in Fig. 2, where we can observe that the SER can decrease with P_1 when P_1 is large and then increase again as P_1 approaches P_T.

Fig. 3 shows the total mutual information, $\mathsf{I}(y; s_1 \mid s_2) + \mathsf{I}(z; s_2)$, when $P_T = 20$ dB. While the power allocation to the sum rate maximization in NOMA system is trivial ($P_1 = P_T$ and $P_2 = 0$ from Subsection II-B), as shown in Fig. 3, the power allocation is not trivial if practical modulation schemes

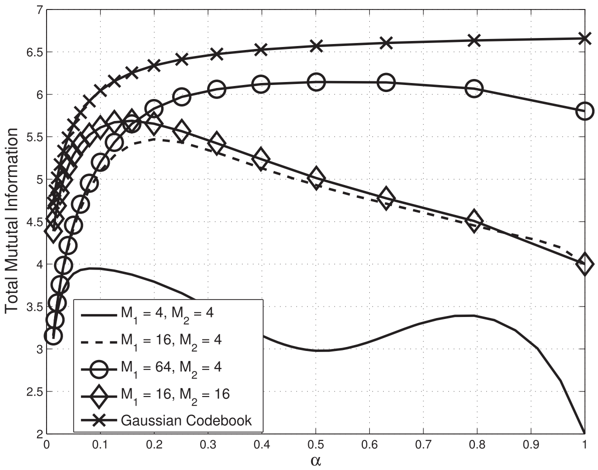

Fig. 3. Total mutual information versus α when $P_T = 20$ dB.

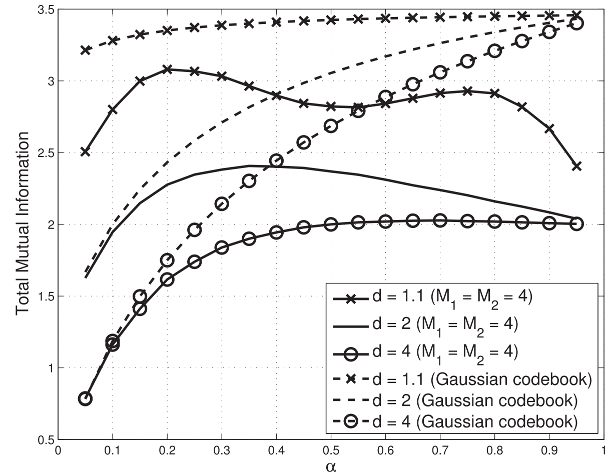

Fig. 4. Total mutual information versus α for different values of d when $M_1 = M_2 = 4$ and $P_T = 10$ dB.

are considered. It is shown that the maximum total mutual information can be obtained at different values of P_1 depending on (M_1, M_2) for a fixed P_T. With $(M_1, M_2) = (4, 4)$, for the maximum total mutual information, we need $P_1 = 10$ dB (or the transmission power ratio for the MUST-near user is $\alpha = 0.1$). On the other hand, with $(M_1, M_2) = (16, 4)$, we need $P_1 = 13$ dB (or $\alpha = 0.2$). With $(M_1, M_2) = (64, 4)$, the maximum total mutual information can be achieved at about $P_1 = 17$ dB (or $\alpha = 0.5$). In general, as M_1 increases, the maximum total mutual information increases for a fixed M_2 and the optimal power to s_1, P_1, also increases.

It is also shown that the increase of M_2 results in a higher total mutual information. If $(M_1, M_2) = (16, 16)$, the maximum total mutual information is achieved when P_1 is about 12 dB (or $\alpha = 0.15$) and higher than that with $(M_1, M_2) = (16, 4)$.

In general, the larger M_1 and M_2, the higher total mutual information the linear MUST scheme can achieve with the optimal power allocation. However, the increase of the total mutual information becomes limited once M_1 and M_2 are sufficiently large, while the complexity of (ML) detection grows exponentially with M_1 and M_2. Thus, for practical implementations, small M_1 and M_2 would be desirable. For example, in Fig. 3, $(M_1, M_2) = (16, 4)$ can provide a reasonably high total mutual information by the optimal power allocation with relatively low complexity in signal detection.

In Fig. 4, we consider the case of $M_1 = M_2 = 4$, which might be one of practical choices for the linear MUST scheme for a moderate value of total transmission power, e.g., $P_T = 10$ dB. In order to see the impact of the normalized distance d, three different values of d are considered: $d \in (1.1, 2, 4)$. For a fixed power allocation, the sum rate of NOMA with Gaussian codebook can increase with $\kappa_2 = \frac{1}{d^\eta}$ (or decrease with d) according to (5). In addition, the maximum sum rate is maximized as α approaches 1. However, when 4-QAM is employed for the linear MUST scheme, we have different results with the mutual information as shown in Fig. 4. While a small d provides a better performance, the optimal power allocation to maximize the total mutual information depends on d. In general, as d increases, the optimal value of α increases.

V. Conclusions

In this letter, we studied the power allocation for the NOMA system of two users when practical modulation schemes are employed for the linear MUST scheme. For the power allocation, we derived a numerical approach to compute the mutual information for given modulation schemes of finite constellation sizes. Unlike the power allocation with ideal capacity, it was shown that the solution to the power allocation problem for maximizing the total mutual information is not trivial and depends on the employed modulation schemes.

References

[1] 3GPP, "TP for classification of MUST schemes," 3GPP R1-154999, 2015.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[3] R. Gallager, "An inequality on the capacity region of multiple access multipath channels," in *Communications and Cryptography: Two Sides and One Tapestry*, R. E. Blahut, D. J. Costello, U. Maurer and T. Mittelholzer, Eds. Norwell, MA, USA: Kluwer, 1994, pp. 129–139.

[4] ETSI, "Framing structure, channel coding and modulation for satellite services to handheld devices (SH) below 3 GHz," ETSI EN 302 583 V1.1.1 (2008-03), Digital Video Broadcasting (DVB), 2008.

[5] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[6] B. Kim *et al.*, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278–1283.

[7] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98.B, no. 3, pp. 403–414, 2015.

[8] Z. Ding, Z. Yang, P. Fan, and H. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[9] 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.

[10] Q. Sun, S. Han, C. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[11] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," in *Foundations and Trends on Communications and Information Theory*, vol. 5, Delft, The Netherlands: Now Publishers, 2008, pp. 1–153.

[12] E. Biglieri, *Coding for Wireless Channels*. New York, NY, USA: Springer, 2005.