Cooperative Non-Orthogonal Multiple Access in 5G Systems

Zhiguo Ding, *Member, IEEE*, Mugen Peng, *Senior Member, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA) has recently received considerable attention as a promising candidate for 5G systems. A key feature of NOMA is that users with better channel conditions have prior information about the messages of the other users. This prior knowledge is fully exploited in this paper, where a cooperative NOMA scheme is proposed. Outage probability and diversity order achieved by this cooperative NOMA scheme are analyzed, and an approach based on user pairing is also proposed to reduce system complexity in practice.**

I. Introduction

Non-orthogonal multiple access (NOMA) is fundamentally different from conventional orthogonal multiple access (MA) schemes, where multiple users are encouraged to transmit at the same time, code and frequency, but with different power levels [1]. In particular, NOMA allocates less power to the users with better channel conditions, and these users can decode their own information by applying successive interference cancellation [2]. Consequently the users with better channel conditions will know the messages intended to the others; however, such prior information has not been exploited by the existing works about NOMA [3] and [4].

In this paper, a cooperative NOMA transmission scheme is proposed by fully exploiting prior information available in NOMA systems. In particular, the use of the successive detection strategy at the receivers means that users with better channel conditions need to decode the messages for the others, and therefore these users can be used as relays to improve the reception reliability for the users with poor connections to the base station. Local short-range communication techniques, such as bluetooth and ultra-wideband (UWB), can be used to deliver messages from the users with better channel conditions to the ones with poor channel conditions. The outage probability and diversity order achieved by this cooperative NOMA scheme are analyzed, and these analytical results demonstrate that cooperative NOMA can achieve the maximum diversity gain for all the users. In practice, inviting all users in the network to participate in cooperative NOMA might not be realistic due to two reasons. One is that a large amount of system overhead will be consumed to coordinate multi-user networks, and the other is that user cooperation will consume extra short-range communication resources. User pairing is a promising solution to reduce system complexity, and we demonstrate that grouping users with high channel quality does not necessarily yield a large performance gain over orthogonal MA. Instead, it is more preferable to pair users whose channel conditions are more distinctive.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Computing and Communications, Lancaster University, LA1 4WA, UK. M. Peng is also with the Institute of Telecommunications, Beijing University of Posts and Telecommunications, Beijing, China.

II. System Model

Consider a broadcast channel with one base station (the source), and K users (the destinations). Cooperative NOMA consists of two phases, as described in the following.

A. Direct Transmission Phase

During this phase, the base station sends K messages to the destinations based on the NOMA principle, i.e., the base station sends $\sum_{m=1}^{K} p_m s_m$, where s_m is the message for the m-th user, and p_m is the power allocation coefficient. The observation at the k-th user is given by

$$y_{1,k} = \sum_{m=1}^{K} h_k p_m s_m + n_k, \tag{1}$$

where h_k denotes the Rayleigh fading channel gain from the base station to the k-th user and n_k denote the additive Gaussian noise. Without loss of generality, consider that the users are ordered based on their channel quality, i.e.,

$$|h_1|^2 \leq \cdots \leq |h_K|^2. \tag{2}$$

The use of NOMA implies $|p_1|^2 \geq \cdots \geq |p_K|^2$, with $\sum_{m=1}^{K} p_m^2 = 1$. Successive detection will be carried out at the K-th user at the end of this phase. The receiving signal to noise ratio (SNR) for the K-th ordered user to detect the k-th user's message, $1 \leq k < K$, is given by

$$SNR_{K,k} = \frac{|h_K|^2 |p_k|^2}{\sum_{m=k+1}^{K} |h_K^H p_m|^2 + \frac{1}{\rho}}, \tag{3}$$

where ρ is the transmit SNR. After these users' messages are decoded, the K-th user can decode its own information at the following SNR

$$SNR_{K,K} = \rho |h_K|^2 |p_k|^2. \tag{4}$$

Therefore the conditions under which the K-th user can decode its own information are given by

$$\log(1 + SNR_{K,k}) > R_k, \quad \forall 1 \leq k \leq K,$$

where R_k denotes the targeted data rate for the k-th user.

B. Cooperative Phase

During this phase, the users cooperate with each other via short range communication channels. Particularly the second phase consists of $(K-1)$ time slots. During the first time slot, the K-th user broadcasts the combination of the $(K-1)$ messages with the coefficients $\mathbf{q}_K$, i.e., $\sum_{m=1}^{K-1} q_{K,m} s_m$, where $\sum_{m=1}^{K-1} q_{K,m}^2 = 1$. The k-th user observes the following

$$y_{2,k} = \sum_{m=1}^{K-1} g_{K,k} q_{K,m} s_m + n_{2,k}, \tag{5}$$

for $k < K$, where $g_{K,k}$ denotes the inter-user channel gain. The $(K-1)$-th user uses maximum ratio combining to combine the observations from both phases, and the SNR for this user to decode the k-th user's message, $k < (K-1)$, is given by

$$SNR_{K-1,k} = \frac{|h_{K-1}|^2 p_k^2}{|h_{K-1}|^2 \sum_{m=k+1}^{K} p_m^2 + \frac{1}{\rho}} + \frac{|g_{K,K-1}|^2 q_{K,k}^2}{|g_{K,K-1}|^2 \sum_{m=k+1}^{K-1} q_{K,m}^2 + \frac{1}{\rho}}. \quad (6)$$

After the $(K-1)$-th user decodes the other users' messages, it can decode its own information with the following SNR

$$SNR_{K-1,K-1} = \frac{|h_{K-1}|^2 p_{K-1}^2}{|h_{K-1}|^2 p_K^2 + \frac{1}{\rho}} + |g_{K,K-1}|^2 q_{K,K-1}^2. \quad (7)$$

Similarly at the n-th time slot, $1 \leq n \leq (K-1)$, the $(K-n+1)$-th user broadcasts the combination of the $(K-n)$ messages with the coefficients $q_{K-n+1,m}$, i.e., $\sum_{m=1}^{K-n} q_{K-n+1,m} s_m$. The k-th user, $k < (K-n+1)$, observes

$$y_{2,k} = \sum_{m=1}^{K-n} g_{K-n+1,k}^H q_{K-n+1,m} s_m + n_{n+1,k}. \quad (8)$$

Combining the observations from both phases, the $(K-n)$-th user can decode the k-th user's message, $1 \leq k < (K-n)$, with the following SNR

$$SNR_{K-n,k} = \frac{|h_{K-n}|^2 p_k^2}{|h_{K-n}|^2 \sum_{m=k+1}^{K} p_m^2 + \frac{1}{\rho}} + \sum_{i=1}^{n} \frac{|g_{K-i+1,K-n}|^2 q_{K-i+1,k}^2}{|g_{K-i+1,K-n}|^2 \sum_{m=k+1}^{K-i} q_{K-i+1,m}^2 + \frac{1}{\rho}}, \quad (9)$$

and it can decode its own information with the following SNR

$$SNR_{K-n,K-n} = \frac{|h_{K-n}|^2 p_{K-n}^2}{|h_{K-n}|^2 \sum_{m=K-n+1}^{K} p_m^2 + \frac{1}{\rho}} + \sum_{i=1}^{n-1} \frac{|g_{K-i+1,K-n}|^2 q_{K-i+1,K-n}^2}{|g_{K-i+1,K-n}|^2 \sum_{m=K-n+1}^{K-i} q_{K-i+1,m}^2 + \frac{1}{\rho}} + \rho |g_{K-n+1,K-n}|^2 q_{K-n+1,K-n}^2. \quad (10)$$

Recall that, without cooperation, the SNR at the k-th user is $\frac{|h_{K-n}|^2 p_{K-n}^2}{|h_{K-n}|^2 \sum_{m=K-n+1}^{K} p_m^2 + \frac{1}{\rho}}$. Compared it to (10), one can find out that the use of cooperation can boost reception reliability.

III. Performance Analysis

Provided that the $(n-1)$ best users can achieve reliable detection, the outage probability for the $(K-n)$-th user can be expressed as follows:

$$\mathrm{P}_o^{K-n} \triangleq \mathrm{P}(SNR_{K-n,k} < \epsilon_k, \forall k \in \{1, \cdots, K-n\}), \quad (11)$$

where $\epsilon_k = 2^{R_k} - 1$. Note that the use of local short-range communications does not reduce the data rate. For notational simplicity, define $a_{k,i}^{K-n} = q_{K-i+1,k}^2$ and $b_{k,i}^{K-n} = \sum_{m=k+1}^{K-i} q_{K-i+1,m}^2$, where $1 \leq k \leq (K-n)$ and $1 \leq i \leq n$ with the special case of $a_{K-n,n}^{K-n} = q_{K-n+1,K-n}^2$ and $b_{K-n,n}^{K-n} = 0$. In addition, define $a_{k,0}^{K-n} = p_k^2$ and $b_{k,0}^{K-n} = \sum_{m=k+1}^{K} p_m^2$, for $1 \leq k \leq (K-n)$. By using the definition of the outage probability, we can have the following proposition for the diversity order achieved by the proposed cooperative NOMA scheme.

Proposition 1. *Assume that the $(n-1)$ best users can achieve reliable detection. The proposed cooperative NOMA scheme can ensure that the $(K-n)$-th ordered user experiences a diversity order of K, conditioned on $\epsilon_k < \frac{a_{k,i}^{K-n}}{b_{k,i}^{K-n}}$, for $1 \leq k \leq (K-n)$ and $0 \leq i \leq n$.*

Proof: For notational simplicity, define $z_{k,i}^{K-n} = \frac{|g_{K-i+1,K-n}|^2 q_{K-i+1,k}^2}{|g_{K-i+1,K-n}|^2 \sum_{m=k+1}^{K-i} q_{K-i+1,m}^2 + \frac{1}{\rho}}$, where $1 \leq k \leq (K-n)$ and $1 \leq i \leq n$, except $z_{K-n,n}^{K-n} = \rho |g_{K-n+1,K-n}|^2 q_{K-n+1,K-n}^2$. In addition, define $z_{k,0}^{K-n} = \frac{|h_{K-n}|^2 p_k^2}{|h_{K-n}|^2 \sum_{m=k+1}^{K} p_m^2 + \frac{1}{\rho}}$. The SNRs can be expressed as follows:

$$SNR_{K-n,k} = z_{k,0}^{K-n} + \sum_{i=1}^{n} z_{k,i}^{K-n}, \quad (12)$$

for $1 \leq k \leq (K-n)$. Therefore the outage probability can be rewritten as follows:

$$\mathrm{P}_o^{K-n} = \mathrm{P}\left(z_{k,0}^{K-n} + \sum_{i=1}^{n} z_{k,i}^{K-n} < \epsilon_k, \forall k \in \{1, \cdots, K-n\} \right) \leq \sum_{k=1}^{K-n} \mathrm{P}\left(z_{k,0}^{K-n} + \sum_{i=1}^{n} z_{k,i}^{K-n} < \epsilon_k \right).$$

Because channel gains are independent, the outage probability can be further bounded as follows:

$$\mathrm{P}_o^{K-n} \leq \sum_{k=1}^{K-n} \prod_{i=0}^{n} \mathrm{P}\left(z_{k,i}^{K-n} < \epsilon_k \right), \quad (13)$$

All the elements in (12) except $z_{k,0}^{K-n}$ and $z_{K-n,n}^{K-n}$ share the same structure as follows:

$$z_{k,i}^{K-n} = \frac{a_{k,i}^{K-n} x}{b_{k,i}^{K-n} x + \frac{1}{\rho}}. \quad (14)$$

When x is exponentially distributed, the cumulative density function (CDF) of $z_{k,i}^{K-n}$ is given by

$$\mathrm{P}_{z_{k,i}^{K-n}}(Z < z) = \begin{cases} 1, & if \;\; z \geq \frac{a_{k,i}^{K-n}}{b_{k,i}^{K-n}} \\ 1 - e^{-\frac{z}{\rho\left(a_{k,i}^{K-n} - b_{k,i}^{K-n} z\right)}}, & otherwise \end{cases},$$

where the definitions for the coefficients $a_{k,i}^{K-n}$ and $b_{k,i}^{K-n}$ are given in the proposition.

At high SNR, $\frac{\epsilon_k}{\rho\left(a_{k,i}^{K-n} - b_{k,i}^{K-n} z\right)} \to 0$, and the probability for the event, $z_{k,i}^{K-n} < \epsilon_k$, can be approximated as follows:

$$\mathrm{P}\left(z_{k,i}^{K-n} < \epsilon_k \right) = 1 - e^{-\frac{\epsilon_k}{\rho\left(a_{k,i}^{K-n} - b_{k,i}^{K-n} \epsilon_k\right)}} \approx \frac{\epsilon_k}{\rho a_{k,i}^{K-n}}, \quad (15)$$

which is conditioned on $\epsilon_k < \frac{a_{k,i}^{K-n}}{b_{k,i}^{K-n}}$.

The density functions of the two special cases, $z_{k,0}^{K-n}$ and $z_{K-n,n}^{K-n}$, can be obtained as follows. Note that the source-user channels are sorted according to their quality. By applying order statistics [5], the CDF of $z_{k,0}^{K-n}$ can be found as follows:

$$\mathrm{P}_{z_{k,0}^{K-n}}(Z<z)= \begin{cases} 1, & if \quad z \geq \frac{a_{k,0}^{K-n}}{b_{k,0}^{K-n}} \\ \int_0^{\frac{\frac{z}{\rho}}{a_{k,0}^{K-n}-b_{k,0}^{K-n}z}} \frac{e^{-x}}{(K-n-1)!}x^{K-n-1}dx, & otherwise \end{cases}.$$

Again applying the high SNR approximation, the probability, $\mathrm{P}(z_{k,0}^{K-n}<\epsilon_k)$, can be approximated as follows:

$$\mathrm{P}\left(z_{k,0}^{K-n}<\epsilon_k\right) = \int_0^{\frac{\epsilon_k}{\rho\left(a_{k,i}^{K-n}-b_{k,i}^{K-n}\epsilon_k\right)}} \frac{x^{K-n-1}e^{-x}}{(K-n-1)!}dx \approx \frac{\epsilon_k^{K-n}}{(K-n)!\left(a_{k,i}^{K-n}\right)^{K-n}\rho^{K-n}}, \tag{16}$$

conditioned on $\epsilon_k < \frac{a_{k,0}^{K-n}}{b_{k,0}^{K-n}}$. Similarly the probability for the event $z_{K-n,n}^{K-n}<\epsilon_k$ can be approximated as follows:

$$\mathrm{P}(z_{K-n,n}^{K-n}<\epsilon_{K-n}) \approx \frac{\epsilon_{k-n}}{q_{K-n+1,K-n}^2\rho}. \tag{17}$$

Combining (13), (15), (16) and (17), the diversity order achieved by the cooperative NOMA scheme can be obtained, which completes the proof. ■

The overall system outage event is defined as the event that any user in the system cannot achieve reliable detection, which means the overall outage probability is defined as follows:

$$\mathrm{P}_o \triangleq 1-\prod_{k=1}^{K}\left(1-\mathrm{P}_o^k\right). \tag{18}$$

By using Proposition 1 and the fact that the source-destination channels are independent, the following lemma can be obtained straightforwardly.

Lemma 1. *The proposed cooperative NOMA scheme can ensure that the n-th best user, $1 \leq n \leq K$, experiences a diversity order of K, conditioned on $\epsilon_k < \frac{a_{k,i}^{K-n}}{b_{k,i}^{K-n}}$, for $1 \leq k \leq (K-n)$ and $0 \leq i \leq n$.*

Note that K is the maximum diversity order to the addressed scenario. For example, the user with the worst channel condition gets help from the other $(K-1)$ users, in addition to its own direct channel to the source, which implies that the maximum diversity for this scenario is K. As can be seen from Lemma 1, encouraging user cooperation can ensure that the maximum diversity order of K is achievable to all users, regardless of the quality of their direct link to the base station, whereas a non-cooperative NOMA can achieve only a diversity order of n for the n-th ordered user [4].

Reducing System Complexity via User Pairing

Inviting all the users to participate in NOMA may not be preferable in practice because of the extra system overhead to coordinate multiple users. Furthermore, the more users participate into cooperation, the more short-range communication bandwidth resource is consumed. A promising solution to this issue is to reduce the number of users for cooperation. Without loss of generality, we focus on the case to select only two users. An important question to be answered here is which two users should be grouped together.

Consider that the users are ordered as (2), and the m-th and n-th users are paired together, $m<n$. The conventional TDMA can achieve the following rates

$$\bar{R}_m = \frac{1}{2}\log\left(1+\rho|h_m|^2\right), \quad \bar{R}_n = \frac{1}{2}\log\left(1+\rho|h_n|^2\right). \tag{19}$$

The rates achieved by cooperative NOMA is quite complicated, so we first consider conventional NOMA which can achieve the following rates

$$R_m = \log\left(1+\frac{\rho|h_m|^2p_m^2}{\rho|h_m|^2p_n^2+1}\right), \tag{20}$$

and

$$R_n = \log\left(1+\rho p_n|h_n|^2\right), \tag{21}$$

where R_n is achievable since $\log\left(1+\frac{|h_n|^2p_m^2}{|h_n|^2p_n^2+1}\right) \geq R_m$.

The gap between the two sum rates achieved by TDMA and conventional NOMA can be expressed as follows:

$$\begin{aligned} &R_m+R_n-\bar{R}_m-\bar{R}_n \\ \approx&\log\left(1+\frac{p_m^2}{p_n^2}\right)+\log\rho p_n^2|h_n|^2-\frac{\log\rho|h_m|^2}{2}-\frac{\log\rho|h_n|^2}{2} \\ =&\frac{\log|h_n|^2}{2}-\frac{\log|h_m|^2}{2}, \end{aligned} \tag{22}$$

where the approximation is obtained at high SNR. It is interesting to observe that the gap is not a function of power allocation coefficients p_m. By applying order statistics, the average of the gap can be calculated as follows:

$$\begin{aligned} &\mathcal{E}\{R_m+R_n-\bar{R}_m-\bar{R}_n\} \\ \approx&\int_0^\infty \frac{\log x}{2}\frac{e^{-x}x^{n-1}}{(n-1)!}dx-\int_0^\infty \frac{\log x}{2}\frac{e^{-x}x^{m-1}}{(m-1)!}dx. \end{aligned} \tag{23}$$

By using Eq. (4.352.1) in [6], the averaged gap is given by

$$\begin{aligned} &\mathcal{E}\{R_m+R_n-\bar{R}_m-\bar{R}_n\} \\ \approx&\frac{\log e}{2}\left(\psi(n)-\psi(m)\right)=\sum_{i=m}^{n-1}\frac{1}{i}, \end{aligned} \tag{24}$$

where $\psi(x)$ denotes the Euler's integral, and the last equation is due to the property of $\psi(x)$, i.e., $\psi(x+1)=\psi(x)+\frac{1}{x}$.

Therefore to conventional NOMA, the worst choice of m and n is $n=m+1$, and it is ideal to group two uses who experience significantly different channel fading. This observation is also valid to cooperative NOMA. Particularly an important observation from (3) is that the data rate for the m-th user is bounded as $R_m \leq \log\left(1+\frac{\rho|h_n|^2 p_m^2}{\rho|h_n|^2 p_n^2+1}\right)$, although R_m can be as large as $\log\left(1+\frac{\rho|h_m|^2 p_m^2}{\rho|h_m|^2 p_n^2+1}+\rho|g_{n,m}|^2\right)$, where the bound is due to the fact that the n-th user needs to decode the m-th user's information. Since $\log\left(1+\frac{\rho|h_n|^2 p_m^2}{\rho|h_n|^2 p_n^2+1}\right) \approx \log\left(1+\frac{p_m^2}{p_n^2}\right)$, the conclusion obtained for conventional NOMA can also be applied to cooperative NOMA.

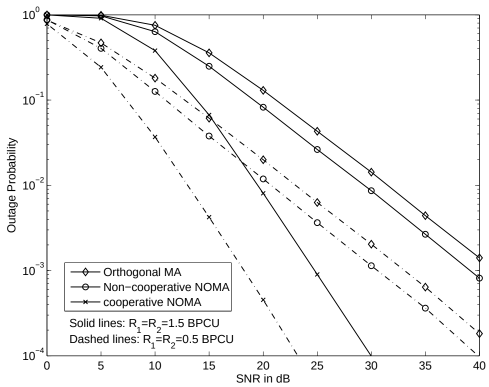

Fig. 1. Outage probability achieved by cooperative NOMA.

IV. Numerical Studies

In this section, the performance of cooperative NOMA is evaluated by using computer simulations. In Fig. 1, the outage probability achieved by the three schemes, e.g., the orthogonal MA scheme, non-coopertive NOMA, and cooperative NOMA, is shown as a function of SNR, with $K=2$. As can be seen from the figure, cooperative NOMA outperforms the other two schemes, since it can ensure that the maximum diversity gain is achievable to all the users as indicated by Lemma 1. In Fig. 2, the outage capacity achieved by the three schemes is demonstrated, by setting $R_1 = R_2$. With 10% outage probability, the orthogonal MA scheme can achieve a rate of 0.7 bits per channel use (BPCU), non-cooperative NOMA can support 0.95 BPCU, and cooperative NOMA can support 1.7 BPCU, much larger than the other two schemes.

In Fig. 3, the impact of user pairing is investigated by studying the difference between the sum rates achieved by the orthogonal MA scheme and NOMA, i.e., $\mathcal{E}\{R_m+R_n-\bar{R}_m-\bar{R}_n\}$ as in (23). Particularly, consider that the K-th ordered user, i.e., the user with the best channel condition, is scheduled, and Fig. 3 demonstrates how large a sum rate gain can be obtained by pairing it with different users. As discussed in Section III, it is helpful to increase $\mathcal{E}\{R_m+R_n-\bar{R}_m-\bar{R}_n\}$ by scheduling two users whose channel connections to the source are more distinctive. Such a conclusion is confirmed by the results shown in Fig. 3, where pairing the K-th user with the first user, i.e., the user with the worst channel condition, can yield a significant gain. This observation is consistent to the motivation of NOMA in [1] which is to schedule two users, one close to the cell edge and the other close to the BS.

V. Conclusions

In this paper, we have proposed a cooperative NOMA transmission scheme which fully uses the fact that some users in NOMA systems have prior information about the

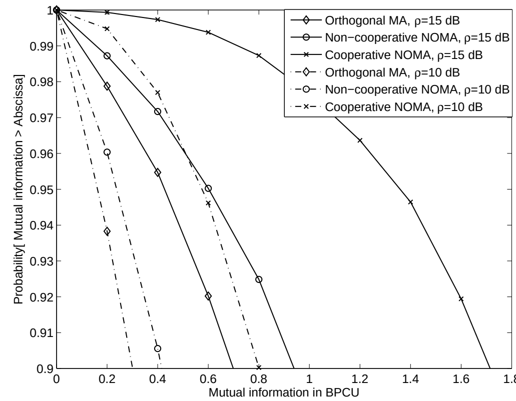

Fig. 2. Outage capacity achieved by cooperative NOMA.

others' messages. Analytical results have been developed to demonstrate that cooperative NOMA can achieve the maximum diversity gain for all users. User pairing and its impact on system throughput have also been discussed in order to implement cooperative NOMA with low system complexity. One promising future direction is to apply game theoretic algorithms for opportunistic user pairing/grouping, where users can form coalitions in a distributed manner [7].

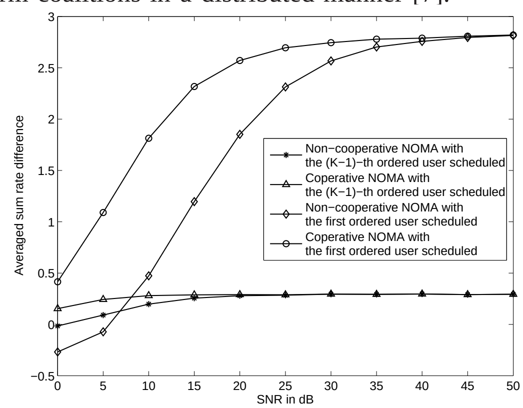

Fig. 3. The impact of user pairing on the sum rate.

References

[1] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annual Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, UK, Sept. 2013.

[2] T. Cover and J. Thomas, *Elements of Information Theory*, 6th ed. Wiley and Sons, New York, 1991.

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

[4] 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 Processing Letters*, vol. 21, no. 12, pp. 1501–1505, Dec 2014.

[5] H. A. David and H. N. Nagaraja, *Order Statistics*. John Wiley, New York, 3rd ed., 2003.

[6] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic Press, 2000.

[7] Z. Ding, I. Krikidis, B. Sharif, and H. V. Poor, "Wireless information and power transfer in cooperative networks with spatially random relays," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4440–4453, Aug. 2014.