Improving Non-Orthogonal Multiple Access by Forming Relaying Broadcast Channels

Jungho So, *Student Member, IEEE*, and Youngchul Sung, *Senior Member, IEEE*

***Abstract*— In this letter, the performance of non-orthogonal multiple access (NOMA) with full-duplex relaying is investigated in multi-user single-input single-output downlink systems. Contrary to the original NOMA, in the proposed NOMA scheme, the base station (BS) and two served receivers within a resource block form a relaying broadcast channel (RBC), where one of the two receivers serves as a relay as well as a receiver. To analyze the performance of the proposed scheme, a simple achievable rate region of an RBC with compress-and-forward (CF) relaying is newly derived based on the noisy network coding (NNC). Furthermore, based on this, an achievable rate region of an RBC with NNC relaying when dirty-paper coding (DPC) is applied at the BS is derived. Numerical results show that the NOMA equipped with the proposed combined coding scheme comprised of DPC and NNC relaying yields significant gain over the originally proposed NOMA.**

***Index Terms*— Non-orthogonal multiple access, relaying broadcast channel, compress-and-forward, dirty-paper coding.**

I. Introduction

NOMA is one of the promising technologies to increase the spectral efficiency for 5G networks [1]. Unlike conventional orthogonalization-based resource allocation with which a point-to-point (P2P) channel is formed between the BS and the single served receiver in each separate resource block, in NOMA the BS and the served users in the same resource block form a broadcast channel (BC) by using superposition coding at the BS with successive interference cancellation (SIC) at some receivers. It is shown that there exists some gain in NOMA over orthogonalization-based sharing schemes when the served users have unequal channel gains. There are some efforts to enhance NOMA by increasing the level of user cooperation and encoding complexity. For example, in [2], the authors considered a half-duplex cooperative NOMA scheme in which the BS sends data to multiple users in the first phase, and the user with good channel helps the other user with bad channel, acting as a half-duplex decode-and-forward (DF) relay in the second phase. However, such half-duplexing requires additional time resource blocks and this limits the performance gain. Recently, full-duplex schemes are gaining increasing attention from research communities due to its advantage in spectral efficiency and recent advances in self-interference cancellation [3]. In this letter, we consider enhancing NOMA by adopting such full-duplex relaying at the user with good channel and several encoding schemes at the BS, and assess the corresponding performance improvement over simple NOMA. When one of the two receivers acts as a full-duplex relay, the 3-node network of the BS, the relaying receiver, and the second receiver is called a *RBC* [4]. For the relaying scheme at the relaying receiver, DF or CF can be used and there exist some rate results on a RBC with DF or CF relaying [4]–[6]. However, it will be seen that just adopting a RBC with DF or CF relaying as the component channel does not improve the performance of NOMA much, but the proposed combination of RBC with CF relaying and DPC at the transmitter enhances the performance of NOMA significantly. To show this, we newly derive a simple achievable rate region for a RBC with CF relaying by extending the simplified CF coding scheme of NNC proposed in [7] to ease the application of DPC in addition to superposition coding at the BS and reduce analysis complexity. Based on this, we assess the performance of NOMA with various encoding schemes.

Manuscript received June 6, 2016; accepted July 14, 2016. Date of publication July 19, 2016; date of current version September 8, 2016. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2A10060852) and in part by The Cross-Ministry Giga KOREA Project of The Ministry of Science, ICT, and Future Planning, Korea [GK14N0100, 5G mobile communication system development based on mmWave]. The associate editor coordinating the review of this letter and approving it for publication was B. Smida.

The authors are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: jhso@kaist.ac.kr; ysung@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/LCOMM.2016.2592901

II. System Model

We consider a single-cell multi-user (MU) SISO downlink network with a single-antenna BS and K single-antenna users. We assume that the overall communication resource is separated into orthogonal blocks based on a certain orthogonalization scheme such as orthogonal frequency-division multiplexing (OFDM) and the BS selects and assigns two users to each resource block in a non-orthogonal manner at each scheduling interval. Under the assumption that resource blocks are orthogonal, each resource block can be considered separately and hence, we consider only one resource block from here on.

Let the two scheduled users be denoted by user 1 and user 2. In the original NOMA scheme [1], the BS and users 1 and 2 form a BC, where there is no communication between users 1 and 2. On the contrary, we here consider a cooperative NOMA scheme in which user 1 serves as a full-duplex relay for user 2 as well as a receiver for its own data, as shown in Fig. 1. In this case, the BS, user 1, and user 2 form a RBC and the received signals at users 1 and 2 are respectively given by

$$Y_1 = h_{01}X_0 + Z_1 \quad \text{and} \quad Y_2 = h_{02}X_0 + h_{12}X_1 + Z_2, \quad (1)$$

where h_{0i} and h_{ij} represent the channel gain from the BS to user i and the channel gain from user i to user j, respectively; X_0 is the transmit signal at the BS with the power constraint $\mathbb{E}\{|X_0|^2\} \leq P_0$; X_1 is the transmit signal at the relaying user 1 with the power constraint $\mathbb{E}\{|X_1|^2\} \leq P_1$; and

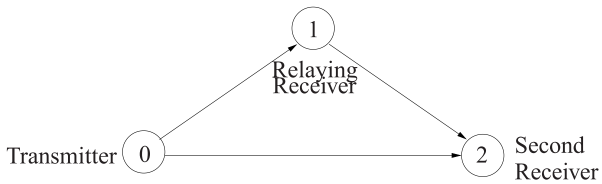

Fig. 1. The relaying broadcast channel.

$Z_1 \sim \mathcal{CN}(0, N_1)$ and $Z_2 \sim \mathcal{CN}(0, N_2)$ are the additive Gaussian noise at users 1 and 2, respectively. We assume that users 1 and 2 are ordered such that $|h_{01}|^2/N_1 \geq |h_{02}|^2/N_2$.

A. Component Channel Analysis

In this subsection, we investigate the achievable rate region of a component RBC capturing the network of the BS and users 1 and 2 under general discrete memoryless and Gaussian channel assumptions for later overall system performance evaluation of NOMA employing RBC component channels. Note that a RBC is different from a conventional relay channel since the transmitter sends two messages: one for the relaying receiver and the other for the second receiver.

A general discrete memoryless RBC is a 3-node discrete memoryless network composed of node 0 (the transmitter), node 1 (the relaying receiver), and node 2 (the second receiver), as shown in Fig. 1, defined by

$$(\mathcal{X}_0 \times \mathcal{X}_1, p(y_1, y_2|x_0, x_1), \mathcal{Y}_1 \times \mathcal{Y}_2), \tag{2}$$

where $\mathcal{X}_0$ and $\mathcal{X}_1$ are the input alphabets of nodes 0 and 1, respectively; $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are the output alphabets of nodes 1 and 2, respectively; and $p(y_1, y_2|x_0, x_1)$ is the channel transition probability mass function (pmf).

Note that by eliminating the link between node 1 and node 2, the considered RBC reduces to a conventional two-user BC, which is the component channel of original NOMA [1]. If the channel is degraded, i.e. $X_0 \to Y_1 \to Y_2$, the capacity region of a degraded BC can be achieved by superposition coding and SIC [8].

In the case of RBC, the achievable rate region of a RBC with DF is well investigated in [4] and thus, we here focus on a RBC with CF relaying at node 1. Achievable rates region of a RBC with CF relaying with and without common information were derived in [5] and [6], respectively. However, for the derivation and the encoding scheme in [5], it is difficult to apply DPC on top of CF, which is required to yield performance gain, as seen later. Hence, we here simplify the problem by adopting the new CF coding scheme of NNC, recently proposed in [7], to ease the additional application of DPC at the BS with a better performance than the scheme in [6]. Although NNC handles the case that the transmitter has only one message common to the two receivers, by extending the NNC scheme with some modification, we obtain a simple achievable rate region for a RBC with NNC CF relaying given in the following theorem:

Theorem 1: The rate pair (R_1, R_2) is achievable for the RBC (2) if

$$R_1 < I(U; Y_1) \tag{3}$$

$$\begin{aligned} R_2 < \min\{&I(V; \hat{Y}_1, Y_2|X_1), \\ &I(V, X_1; Y_2) - I(\hat{Y}_1; Y_1|V, X_1, Y_2)\} \end{aligned} \tag{4}$$

for some joint distribution $p(x_1)p(u)p(v)p(x_0|u, v)$ $p(\hat{y}_1|y_1, x_1)$.

Proof: Proof of Theorem 1 is based on the techniques in [7] with generating the code for node 0 using $p(x_0|u, v)$. Please see [9] for detail. ■

Here, U and V are the random variables related to input messages for node 1 and node 2, respectively, and used to generate the transmit variable X_0 of node 0, as seen in the term $p(x_0|u, v)$ in the generating distribution in Theorem 1. The right-hand side (RHS) of (3) is simply the mutual information between the message variable U for node 1 and the receiver signal variable Y_1 at node 1. The two terms in the minimum in the RHS of (4) are associated with the two cut-set bounds of the 3-node network between node 0 and node 2.

Now consider the Gaussian channel case. In the Gaussian channel case, the received signals at node 1 and node 2 are given by (1). To obtain the achievable rate region determined by Theorem 1 in the Gaussian channel case, we need to set the associated distributions since only the channel $p(y_1, y_2|x_0, x_1)$ is given. We set $X_0 \sim \mathcal{CN}(0, P_0)$, $X_1 \sim \mathcal{CN}(0, P_1)$, and X_0 as a superimposed signal given by $X_0 = U + V$, where U is the signal for node 1 and V is the signal for node 2 with

$$U \sim \mathcal{CN}(0, \alpha P_0), \quad V \sim \mathcal{CN}(0, \bar{\alpha} P_0), \; 0 \leq \alpha \leq 1, \quad \bar{\alpha} = 1 - \alpha. \tag{5}$$

When there is no relaying term $h_{12}X_1$ in (1), the channel reduces to a two-user SISO degraded Gaussian BC (GBC) by the assumption $|h_{01}|^2/N_1 \geq |h_{02}|^2/N_2$. The capacity region of this GBC (simple NOMA) is given by [8]

$$R_1 \leq \log\left(1 + \frac{|h_{01}|^2 \alpha P_0}{N_1}\right), \tag{6}$$

$$R_2 \leq \log\left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0}{|h_{02}|^2 \alpha P_0 + N_2}\right). \tag{7}$$

It can easily be seen that superposition coding $X_0 = U + V$ and SIC achieve the rate-tuple in (6) and (7). In the case of the Gaussian RBC with NNC CF relaying, to apply Theorem 1, we need to further set the distribution for the random variable $\hat{Y}_1$ appearing in (4). We set $\hat{Y}_1$ as

$$\hat{Y}_1 = Y_1 - h_{01}U + \hat{Z} = h_{01}V + Z_1 + \hat{Z}, \tag{8}$$

where $\hat{Z} \sim \mathcal{CN}(0, \hat{N})$. With this setup, after some calculation, we obtain the achievable rate region of RBC with NNC CF relaying given by (9) and (10), as shown at the bottom of the next page. Here, the value of $\hat{N}$, which is the power of the added noise in (8), can be optimized to maximize R_2 in (10) by solving a quadratic equation.

Now let us look into the rates in (9) and (10). First, note that if we set $X_1 = \emptyset$ and $\hat{Y}_1$ as independent of all other random variables, the RHS of (10) reduces to that of (7). Thus, if we set $\hat{Y}_1$ properly, the rate R_2 in (10) of RBC with NNC CF relaying can be made larger than the rate R_2 in (7) of GBC in most cases. However, the rate R_1 in (9) of the relaying receiver is smaller than that of GBC because of the interference caused by the message for the second receiver. To eliminate this interference, we propose to additionally apply DPC [10] at the transmitter to the encoding scheme presented

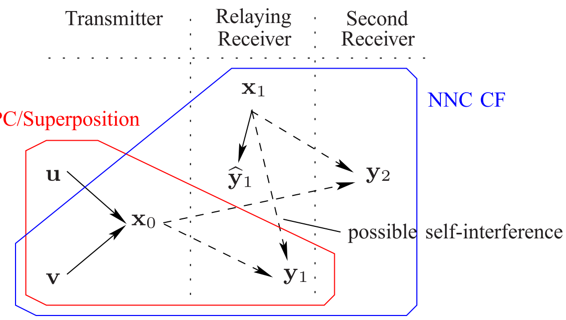

Fig. 2. Markov chain relationship between codewords and received signals (solid arrows: codeword Markov chain and dashed arrows: channel links).

in Theorem 1 since the transmitter knows both messages. Then, the achievable rate of the relaying receiver in this case is given by the pair of equations (6) and (10), which shows that R_2 is the same as (10) of RBC with NNC CF relaying, but R_1 is improved to be the same as (6) of GBC by applying DPC at the transmitter. The proposed encoding scheme based on both DPC/superposition and NNC CF relaying for RBC is described in Fig. 2.

III. User Scheduling and Pairing

In Section II, we described the overall system model and analyzed the achievable rate region of a component RBC composed of the BS and two scheduled users. To examine the impact of forming a RBC for NOMA on the overall system performance, we need to consider user scheduling for selecting two users from the total available K users in the cell to form a RBC together with BS for a given distribution of users in the cell. In this section, we explain the considered user scheduling and pairing methods based on the proportionally fair (PF) principle [11] to evaluate the overall performance of the considered NOMA schemes in a cellular system setting.

We consider two opposite user scheduling and pairing methods; (i) pairing two users so that one user has a good channel gain and the other has a bad channel gain, which is favorable pairing for GBC [12] and (ii) pairing two users close to each other, which is favorable paring for relaying schemes due to the small pass loss from the relaying receiver to the second receiver. For each method, the BS selects two users from the total K users in the cell based on a PF metric for each scheduling interval. We assume that the power allocation parameter α in (5) is fixed, and the BS knows the location of each user and the channel from the BS to each user in the cell.

In the first method, the BS partitions the users into two groups G_1 and G_2 by ordering $|h_{0k}|$, $k = 1, \cdots, K$, so that $|G_1| = |G_1| = K/2$ and $|h_{0i}| \geq |h_{0j}|$ for all $i \in G_1$ and $j \in G_2$. Then, the BS picks two users κ_1 from G_1 as the relaying receiver and κ_2 from G_2 as the second receiver sequentially based on the PF principle as follows:

$$\kappa_1 = \arg\max_{i \in G_1} \frac{R_{1(i)}[t]}{\bar{\mathcal{R}}(i)[t]}, \quad \kappa_2 = \arg\max_{j \in G_2} \frac{R_{2(j|\kappa_1)}[t]}{\bar{\mathcal{R}}(j)[t]}, \tag{11}$$

where $R_{1(i)}[t]$ and $R_{2(j|i)}[t]$ are respectively the rates R_1 and R_2 in Section II-A when user i serves as the relaying receiver and user j is the second receiver paired with the relaying receiver i at scheduling interval t, and $\bar{\mathcal{R}}(i)[t]$ is the average served rate for user i up to scheduling interval t, which is given by

$$\bar{\mathcal{R}}(i)[t+1] := (1-\tau)\bar{\mathcal{R}}(i)[t] + \tau R(i)[t], \tag{12}$$

where τ is the forgetting factor of the averaging filter and $R(i)[t]$ is the served rate of user i at scheduling interval t. The computation of $R_{1(i)}[t]$ is based only on the knowledge of h_{0i}. However, the computation of $R_{2(j|i)}$ is based on the knowledge of the channel gains $|h_{0j}|^2$ and $|h_{ij}|^2$, and the latter is not available at the BS. In this step, we use an estimate for the channel gain from user i to user j based on their locations as [13]

$$\widehat{|h_{ij}|^2} = C_0 d^{-\gamma}, \tag{13}$$

where C_0 is a constant, d is the distance between users i and j, and γ is the path loss exponent.

In the second method, we select one user as the relaying receiver and its nearest neighbor among the users who have lower channel gains than the firstly selected user as the second receiver. Since the second receiver for each relaying user is automatically determined, we can select two users simultaneously based on the PF metric. That is, set $\mathcal{G} = \{1, \cdots, K\}$ and

$$\kappa_1 = \arg\max_{i \in \mathcal{G}} \left(\frac{R_{1(i)}[t]}{\bar{\mathcal{R}}(i)[t]} + \frac{R_{2(\mathcal{N}(i)|i)}[t]}{\bar{\mathcal{R}}(\mathcal{N}(i))[t]} \right) \tag{14}$$

where $\mathcal{N}(i)$ $(= \kappa_2)$ is the index of the nearest neighbor of user i that has a lower channel gain from the BS than user i.

IV. Numerical Results

We first evaluated the achievable rate region of each component channel presented in Section II-A in the Gaussian case. For the evaluation of the performance of component channels, we considered a linear configuration in which the relaying receiver is located at the center of the line between the transmitter and the second receiver. We set $N_1 = N_2 = N = 1$

$$R_1 \leq \log\left(1 + \frac{|h_{01}|^2 \alpha P_0}{|h_{01}|^2 \bar{\alpha} P_0 + N_1}\right), \tag{9}$$

$$R_2 \leq \min\left\{ \log\left(1 + \frac{(|h_{02}|^2 \alpha P_0 + N_2)|h_{01}|^2 \bar{\alpha} P_0 + (N_1 + \hat{N})|h_{02}|^2 \bar{\alpha} P_0}{(N_1 + \hat{N})(|h_{02}|^2 \alpha P_0 + N_2)}\right), \right.$$
$$\left. \log\left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0 + |h_{12}|^2 P_1}{|h_{02}|^2 \alpha P_0 + N_2}\right) - \log\left(1 + \frac{N_1^2 N_2 + N_1^2 |h_{02}|^2 \alpha P_0}{\hat{N} N_1 N_2 + \hat{N} N_2 |h_{01}|^2 \alpha P_0 + \hat{N} N_1 |h_{02}|^2 \alpha P_0 + N_1 N_2 |h_{01}|^2 \alpha P_0}\right) \right\}. \tag{10}$$

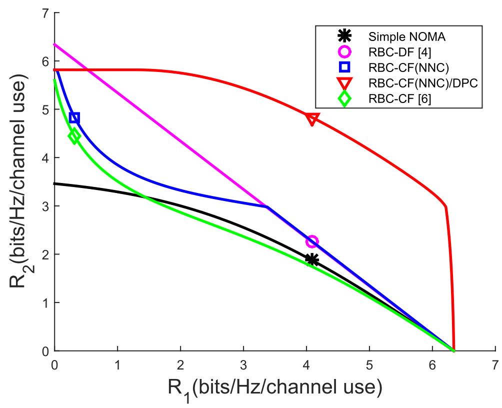

Fig. 3. The achievable rate region for a component channel.

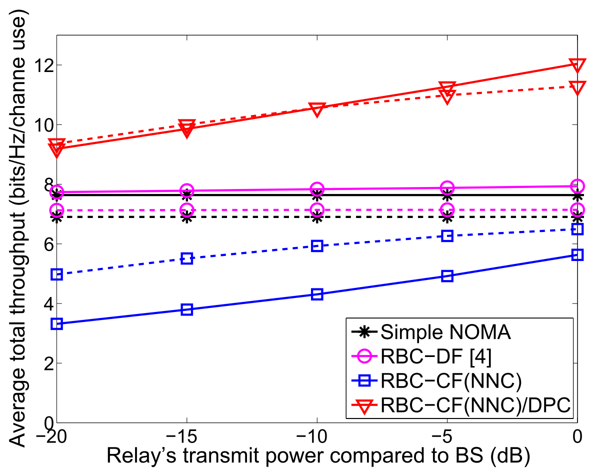

Fig. 4. Total system sum rate: Dashed line - nearest-neighbor pairing and solid line - near-far pairing.

and $(P_0/N, P_1/N) = (10 \text{ dB}, 10 \text{ dB})$. We assumed that the path loss exponent is $\gamma = 3$. In consistence with $\gamma = 3$, we considered a channel gain setup: $|h_{01}|^2 = |h_{12}|^2 = 8$ and $|h_{02}|^2 = 1$. Then, we swept the parameter α in (5) to determine the achievable rate pair (R_1, R_2). Fig. 3 show the rate-tuples of GBC, RBC-DF [4], RBC-CF [6], RBC-CF(NNC) and RBC-CF(NNC)/DPC. The marked points in Fig. 3 are the rate-pair points of $\alpha = 0.2$, a typical power allocation for NOMA. It is seen that either RBC-CF(NNC) or RBF-DF [4] alone does not yield significant gain over GBC at $\alpha = 0.2$. When RBC-CF(NNC) is combined with DPC at the transmitter, the gain over GBC is significant.

We then examined the overall system performance of the considered schemes described in Sections II and III. For the cell topology, we considered a 120° sector of a disk with radius of $D_e = 500m$ and assumed that $K = 40$ users are distributed uniformly over the 120° sector from radius $50m$ to the cell edge. The noise power for each user was set to be 1. The channel gain h_{0k} was modelled as the product of a Rayleigh fading coefficient $f_{0k} \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$ and the large-scale path loss with exponent $\gamma = 3$ as $h_{0k} = f_{0k} \cdot (d_{0k}/D_e)^{-\gamma}$, where d_{0k} is the distance from the BS to user k. The BS transmit power P_0 was set so that the expected received SNR at the cell edge was 10 dB. The transmit power P_1 of the relaying receiver was swept from 0 dB to -20 dB relative to P_0. We averaged the sum rate over 50 independent realizations for user locations. For each realization, we performed the user scheduling and pairing methods described in Section III with $\tau = 0.01$ in (12) for 1000 scheduling intervals, and computed the sum rate normalized by factor 1000 for each scheme. Fig. 4 shows the sum rate of NOMA equipped with four different component channels. It is seen that the performance gain of NOMA with RBC-DF over NOMA with GBC is insignificant and NOMA with RBC-CF(NNC) alone is even worse than simple NOMA with GBC in some cases. However, it is seen that NOMA with RBC-CF(NNC) together with DPC at the transmitter yields significant gain over all other schemes.

V. Conclusion

In this letter, we have considered enhancing NOMA by forming RBC for single-cell MU-SISO cellular downlink systems. Based on the newly derived achievable rate region results for RBC-CF(NNC) and RBC-CF(NNC)/DPC, we have investigated the overall system performance of NOMA equipped with RBC component channels, and have shown that NOMA with RBC-CF(NNC)/DPC yields significant gain over simple NOMA with GBC in a practical system setup. By going beyond GBC and SIC to advanced multi-terminal encoding and cooperation including DPC and CF relaying [14], [15], far larger gains can be achieved for NOMA.

References

[1] A. Benjebbovu, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *Proc. IEEE Globecom*, Dec. 2013, pp. 66–70.

[2] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[3] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C.-B. Chae, "Prototyping real-time full duplex radios," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 56–63, Sep. 2015.

[4] Y. Liang and V. V. Veeravalli, "The impact of relaying on the capacity of broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Chicago, IL, USA, Jun./Jul. 2004, p. 403.

[5] S. I. Bross, "On the discrete memoryless partially cooperative relay broadcast channel and the broadcast channel with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2161–2182, May 2004.

[6] X. Song and Y. Li, "Compress-and-forward for relay broadcast channels without common messages," *Cybern. Inf. Technol.*, vol. 15, pp. 97–110, Jun. 2015.

[7] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[8] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2011.

[9] J. So and Y. Sung. (Oct. 2015). "Enhancing non-orthogonal multiple access by forming relaying broadcast channels." [Online]. Available: http://arxiv.org/abs/1510.07369

[10] M. H. M. Costa, "Writing on dirty paper (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[12] Z. Ding, P. Fan, and V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access downlink transmissions," *IEEE Trans. Veh. Technol.*, to be published.

[13] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*. New York, NY, USA: Wiley, 2001.

[14] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai (Shitz), "Superposition coding for side-information channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.

[15] M. Karzand, "Polar codes for degraded relay channels," in *Proc. Int. Zurich Seminar Commun.*, Zürich, Switzerland, Feb. 2012, pp. 59–62.