An Efficient Signaling Scheme for MIMO Broadcast Systems: Design and Performance Evaluation

Mohammad Ali Maddah-Ali, Mehdi Ansari Sadrabadi, and Amir K. Khandani

Coding & Signal Transmission Laboratory

Department of Electrical & Computer Engineering

University of Waterloo

Waterloo, Ontario, Canada, N2L 3G1

Technical Report UW-E&CE 2005-9

July, 2005

An Efficient Signaling Scheme for MIMO Broadcast Systems: Design and Performance Evaluation

Mohammad Ali Maddah-Ali, Mehdi Ansari, and Amir K. Khandani
Coding & Signal Transmission Laboratory (www.cst.uwaterloo.ca)
Dept. of Elec. and Comp. Eng., University of Waterloo
Waterloo, ON, Canada, N2L 3G1
Tel: 519-725-7338, Fax: 519-888-4338
e-mail: {mohammad, mehdi, khandani}@cst.uwaterloo.ca

Abstract

A simple signaling method for broadcast channels with multiple transmit multiple receive antennas is proposed. In this method, for each user, the direction in which the user has the maximum gain is determined. The best user in terms of the largest gain is selected. The corresponding direction is used as the modulation vector (MV) for the data stream transmitted to the selected user. The algorithm proceeds in a recursive manner where in each step, the search for the best direction is performed in the null space of the previously selected MVs. It is demonstrated that with the proposed method, each selected MV has no interference on the previously selected MVs. Dirty paper coding is used to cancel the remaining interference. To analyze the performance of the scheme, an upper-bound on the outage probability of each sub-channel is derived which is used to establish the diversity order and the asymptotic sum-rate of the scheme. It is shown that the diversity order of the j^{th} data stream, $1 \leq j \leq M$, is equal to $N(M-j+1)(K-j+1)$, where M, N, and K indicate the number of transmit antennas, the number of receive antennas, and the number of users, respectively. Furthermore, it is proven that the throughput of this

The material in this paper was presented in part at the 42nd Allerton Conference, Monicello, IL, Sep. 2004 and the 9th Canadian Workshop on Information Theory (CWIT), Montreal, Quebec, June 2005.

This work is financially supported by Communications and Information Technology Ontario (CITO), Nortel Networks, and National Sciences and Engineering Research Council of Canada (NSERC).

scheme scales as $M \log\log(K)$ and asymptotically ($K \longrightarrow \infty$) tends to the sum-capacity of the MIMO broadcast channel. The simulation results indicate that the achieved sum-rate is close to the sum-capacity of the underlying broadcast channel.

Index Terms

Multiuser system, multiple-antenna arrays, MIMO broadcast channel, dirty paper coding, space-division-multiple-access, multiuser diversity, wireless communications.

I. Introduction

Recently, multiple input multiple output (MIMO) systems have received considerable attention as a promising solution to provide reliable and high data rate communication [1]–[3]. More recently, the work on MIMO systems has been extended to MIMO multi-user channels [4]–[7]. In [4], [5], a duality between the broadcast channel and the multiple access channel is introduced. This duality is applied to characterize the sum-capacity of the broadcast channel as a convex optimization problem. In [6], a reformulation of the sum-capacity as a min-max optimization problem is introduced and a signaling method which achieves the sum-capacity is presented. It is shown that in an optimal signaling (maximizing the sum-rate), the power is allocated to, at most, M^2 uses (active users), where M is the number of transmit antennas [8]. In practical systems, the number of users is large. In this case, finding the set of active users by solving the optimization problem is a complex operation. In addition, to perform such a computation, all the channel state information is required at the base station which necessitates a high data rate feedback link.

The duality and signaling method introduced in [4]–[6] are based on a result, known as *dirty paper coding*, on cancelling known interference at the transmitter [9]. Dirty paper coding states that in an AWGN channel with interference, if the transmitter non-causally knows the interference, the capacity of the channel is the same as the capacity of the channel without interference. A method for approximate implementation of the dirty paper coding is presented in [10], [11].

A number of research works have focused on practical methods for signaling over MIMO broadcast channels. In [12], a simple method that supports one user at a given time is presented. This method exploits a special kind of diversity, *multiuser diversity*, which is available in the multiuser system with independent channels [13]. To exploit multiuser diversity, the transmission resources are allocated to the user(s) which result in the highest throughput for the given channel condition. Unlike [12], the signaling method presented in other related works support multiple users at a given time. In [14], a variation of channel inversion method is used, where the inverse of the channel matrix is regularized and the data is perturbed to reduce the energy of the transmitted signal. However, in this method, the pre-coding matrix depends on the data, and therefore, the method is computationally extensive.

In addition, no method for selecting active users is suggested. In [15], a signaling method based on the QR decomposition and dirty paper coding is introduced. The QR decomposition converts the channel matrix, and consequently the interference matrix[1], to a lower triangular form. Dirty paper coding eliminates the remaining interference. By modifying the QR decomposition, a greedy method for selecting active users which exploits multiuser diversity is presented in [16]. References [14]–[16] present methods to support M simultaneous users, each with one receive antenna.

When there is more than one antenna at the receiver, a generalized version of the zero forcing method is utilized in [17], [18]. However, the methods of [17], [18] are highly restrictive in the sense that the number of transmit antennas must be greater than the total number of the receive antennas. In addition, similar to the conventional zero forcing, the method presented in [17], [18] degrades the signal-to-noise-ratio (SNR).

In this paper, an efficient sub-optimum method for selecting the set of active users and signaling over such users is proposed. This method converts the interference matrix – but not necessarily the channel matrix – to a lower-triangular form. This is in contrast to the earlier method proposed in [15], [16] which uses QR decomposition to triangularis the channel matrix. In the proposed method, first, the direction in which each user has the maximum gain is determined. The base station selects the

[1]The entry (p, q) of the interference matrix denotes the interference of user p over user q

best user in terms of the largest maximum gain, where the corresponding direction is used as the modulation vector (MV) for that user. The algorithm proceeds in a recursive manner where in each step, the search for the best direction is performed in the null space of the previously selected MVs. Finally, the transmitted signal is formed as a linear combination over the selected MVs. It is shown that in this method, data stream j has no interference on data stream i, $i = 1, \cdots, j-1$. Dirty paper coding is used to eliminate the remaining interference. Thus, the underlying sub-channels can be treated independent of each other in terms of encoding/decoding and provision of QoS. In addition, this method offers other desirable features such as: (i) accommodating users with different number of receive antennas, (ii) exploiting multi-user diversity, and (iii) requiring low feedback rate. It is easy to see that for the special case of $N = 1$, the proposed algorithm is the same as the methods presented in [15], [16].

To analyze the performance of the scheme, an upper-bound on the outage probability of each sub-channel is derived which is used to establish the diversity order and the asymptotic sum-rate of the scheme. It is shown that the diversity order of the j^{th} data stream, $1 \leq j \leq M$, is equal to $N(M-j+1)(K-j+1)$. Furthermore, it is proven that the throughput of this scheme scales as $M \log\log(K)$ and asymptotically ($K \longrightarrow \infty$) tends to the sum-capacity of the MIMO broadcast channel. The simulation results indicate that the achieved sum-rate is close to the sum-capacity of the underlying broadcast channel.

The rest of the paper is organized as follows: In Section II, the system model and the proposed signaling method are presented. In Section III, an algorithm to select the active users and the corresponding MVs is developed. The performance analysis of the system is presented in Section IV. In this section, an upper-bound on the outage probability of each sub-channel is derived which is used to establish the diversity order and the asymptotic sum-rate of the scheme. In Section V, the simulation results and comparisons with the sum-capacity of the MIMO broadcast are discussed. Some concluding remarks are provided in Section VI.

Notation: All boldface letters indicate vectors (lower case) or matrices (upper case). $[a_{(p,q)}]_{(p,q)}^{m\times n}$ represents an $m \times n$ matrix where $a_{(p,q)}$ is entry (p,q). $(.)^\dagger$ denotes transpose conjugate operation, and $\mathcal{C}$ represents the set of complex numbers. $F_{j:K}(.)$ denotes the cumulative distribution function (CDF) of the j^{th} largest variable (represented by $z^{(j)} = j^{th} \max\{z_1, \ldots, z_K\}$) among K variables.

II. Preliminaries

Consider a MIMO broadcast channel with M transmit antennas and K users, where the r^{th} user is equipped with N_r receive antennas. In a flat fading environment, the baseband model of this system is given by,

$$\mathbf{y}_r = \mathbf{H}_r \mathbf{s} + \mathbf{w}_r, \quad 1 \leq r \leq K, \tag{1}$$

where $\mathbf{H}_r \in \mathcal{C}^{N_r \times M}$ denotes the channel matrix from the base station to user r, $\mathbf{s} \in \mathcal{C}^{M\times 1}$ represents the transmitted vector, and $\mathbf{y}_r \in \mathcal{C}^{N_r \times 1}$ signifies the received vector by user r. The vector $\mathbf{w}_r \in \mathcal{C}^{N_r\times 1}$ is white Gaussian noise with a zero-mean and unit-variance.

The base station supports M simultaneous data streams, distributed among at most M users (*active users*), indexed by $\pi(j)$, $j = 1, \ldots, M$. The transmitted vector $\mathbf{s}$ is equal to:

$$\mathbf{s} = \sum_{j=1}^{M} d_j \mathbf{v}_j, \tag{2}$$

where $\mathbf{v}_j \in \mathcal{C}^{M\times 1}$, $j = 1, \ldots, M$, is the modulation vector (MV) corresponding to user $\pi(j)$, $\pi(j) \in \{1, 2, \ldots, K\}$, and d_j contains the information for user $\pi(j)$. Note that with this formulation, a given user may receive multiple data streams. Vectors $\mathbf{v}_j$, $j = 1, \ldots, M$, form an orthonormal set. Dirty-paper coding is used such that for $i > j$, the interference of data stream i over data stream j is canceled.

To detect the data stream j, user $\pi(j)$ multiplies the received vector by a demodulation vector $\mathbf{u}_j^\dagger$. In the next section, we propose a method to select the set of active users $\{\pi(1), \pi(2), \ldots, \pi(M)\} \subset \{1, 2, \ldots, K\}$, modulation vectors $\mathbf{v}_j$, and demodulation vectors $\mathbf{u}_j$, for $j = 1, \ldots, M$.

III. Selecting active users, modulation, and demodulation vectors

Assuming channel state information (CSI) is available at the base station, the proposed algorithm works as follows. First, for each user, the maximum *gain* and the corresponding *direction* are determined[2]. Next, the best user in terms of the largest gain is chosen as an active user. The MV for the selected user is along the corresponding direction. These steps are repeated recursively until the M MVs and the set of active users are determined. In each step, the search for the best direction is performed in the null space of the previously selected MVs. It is shown that in this manner, any given MV has no interference over the previously selected MVs. In the following, the proposed algorithm is presented in details.

1) Set $j = 1$ and $\mathbf{\Xi} = [0]_{M \times M}$.
2) Find σ_j^2, where

$$\begin{aligned} \sigma_j^2 = \quad & \max_r \max_{\mathbf{x}} \mathbf{x}^\dagger \mathbf{H}_r^\dagger \mathbf{H}_r \mathbf{x}. \\ s.t. \quad & \qquad \mathbf{x}^\dagger \mathbf{x} = 1 \\ & \qquad \mathbf{\Xi}^\dagger \mathbf{x} = 0. \end{aligned} \tag{3}$$

   Set $\pi(j)$ and $\mathbf{v}_j$ equal to the optimizing parameters r and $\mathbf{x}$, respectively.
3) Set

$$\mathbf{u}_j = \frac{1}{\sigma_j} \mathbf{H}_{\pi(j)} \mathbf{v}_j. \tag{4}$$

4) Substitute $\mathbf{v}_j$ in column j of matrix $\mathbf{\Xi}$.
5) Set $j \leftarrow j + 1$. If $j \leq M$, move to step two; otherwise, stop.

In Step 2 of the algorithm, maximization over r selects the best user, and therefore, exploits the multiuser diversity. Maximization over $\mathbf{x}$ determines the best MV for each user, and at the same time converts the interference matrix to a lower triangular form, implying that data stream j has no interference over data stream i, $i = 1, \ldots, j - 1$. This property has been proven in the following theorem.

[2]The gain of the channel $\mathbf{H}$ along the direction (unit vector) $\mathbf{x}$ is defined as the square root of $\mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{x}$.

Theorem 1 *Consider the following optimization problem:*

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{x}, \\ s.t. \quad & \mathbf{x}^\dagger \mathbf{x} = 1 \\ & \mathbf{\Xi}^\dagger \mathbf{x} = 0, \end{aligned} \tag{5}$$

where $\mathbf{H}$ and $\mathbf{\Xi} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_\varrho]$ are complex matrices. Let $\mathbf{v}$ be the vector that maximizes (5) and σ^2 be the corresponding optimum value. Define vector $\mathbf{u}$ as follows:

$$\mathbf{u} = \frac{\mathbf{H}\mathbf{v}}{\sigma}. \tag{6}$$

If there exists a vector $\widehat{\mathbf{v}}$ such that $\mathbf{\Xi}^\dagger \widehat{\mathbf{v}} = 0$ and $\mathbf{v}^\dagger \widehat{\mathbf{v}} = 0$, then

$$\mathbf{u}^\dagger \mathbf{H} \widehat{\mathbf{v}} = 0. \tag{7}$$

Proof: According to (6),

$$\mathbf{u}^\dagger \mathbf{H} \widehat{\mathbf{v}} = \left(\frac{\mathbf{H}\mathbf{v}}{\sigma} \right)^\dagger \mathbf{H} \widehat{\mathbf{v}} = \frac{1}{\sigma} \mathbf{v}^\dagger \mathbf{H}^\dagger \mathbf{H} \widehat{\mathbf{v}}. \tag{8}$$

To optimize the cost function in (5), Lagrange multipliers technique is adopted.

$$L(\mathbf{x}, \lambda, \mathbf{\Theta}) = -\mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{x} + \lambda(\mathbf{x}^\dagger \mathbf{x} - 1) + \mathbf{\Theta} \ \mathbf{\Xi}^\dagger \mathbf{x}, \tag{9}$$

where λ and $\mathbf{\Theta} = \mathrm{diag}([\theta_1, \theta_2, \ldots, \theta_\varrho])$ are Lagrange multipliers. The gradient of $L(\mathbf{x}, \lambda, \mathbf{\Theta})$, corresponding to the vector $\mathbf{x}$, is

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mathbf{\Theta}) = -2\mathbf{H}^\dagger \mathbf{H} \mathbf{x} + 2\lambda \mathbf{x} + \sum_{\tau=1}^{\varrho} \theta_\tau \boldsymbol{\xi}_\tau. \tag{10}$$

Since $\mathbf{v}$ maximizes the cost function, $\mathbf{v}$ satisfies (10). Therefore,

$$\nabla_{\mathbf{x}} L(\mathbf{v}, \lambda, \mathbf{\Theta}) = -2\mathbf{H}^\dagger \mathbf{H} \mathbf{v} + 2\lambda \mathbf{v} + \sum_{\tau=2}^{\varrho} \theta_\tau \boldsymbol{\xi}_\tau = 0. \tag{11}$$

Multiplying both sides of (11) by $\widehat{\mathbf{v}}^\dagger$ results in

$$\widehat{\mathbf{v}}^\dagger \nabla_{\mathbf{x}} L(\mathbf{v}_1, \lambda, \mathbf{\Theta}) = -2\widehat{\mathbf{v}}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{v} + 2\lambda \widehat{\mathbf{v}}^\dagger \mathbf{v} + \widehat{\mathbf{v}}^\dagger \sum_{\tau=2}^{\varrho} \theta_\tau \boldsymbol{\xi}_\tau = 0. \tag{12}$$

If $\widehat{\mathbf{v}}^\dagger \mathbf{v} = 0$ and $\widehat{\mathbf{v}}^\dagger \boldsymbol{\xi}_\tau = 0$ for $\tau = 1, \ldots, \varrho$ are substituted into in (12),

$$\widehat{\mathbf{v}}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{v} = 0. \tag{13}$$

Finally, (8) and (13) result in

$$\mathbf{u}^{\dagger}\mathbf{H}\widehat{\mathbf{v}} = 0. \tag{14}$$

■

The interference of data stream i over data stream j is equal to $\mathbf{u}_j^{\dagger}\mathbf{H}_{\pi(j)}\mathbf{v}_i$. Noting (3) which derive $\mathbf{v}_j$ and according to $\mathbf{v}_j^{\dagger}\mathbf{v}_i = 0$, Theorem 1 implies that $\mathbf{u}_j^{\dagger}\mathbf{H}_{\pi(j)}\mathbf{v}_i = 0$, for $i > j$. This means that data stream i has no interference over data stream j, $j = 1, \ldots, i-1$. Note that if $i < j$, the interference of data stream i over data stream j is canceled by dirty paper coding. Therefore, the MIMO broadcast channel is effectively reduced to a set of parallel sub-channels with gains σ_j, $j = 1, \ldots, M$. As a result, the sum-rate of the system is equal to

$$R = \sum_{j=1}^{M} \log(1 + \sigma_j^2 P_j) \quad \text{Nat/Sec/Hz}, \tag{15}$$

where P_j is the power allocated to data stream j, and $\sum_{j=1}^{M} P_j \leq P$. Note that (15) is based on the channel model (1), where the power of the noise is normalized. To maximize (15), the power can be allocated using water-filling [19].

In the proposed algorithm, it is assumed the CSI is available at the transmitter which necessitates a high-data-rate feedback link. In Appendix I, the proposed algorithm is modified to reduce the rate of the feedback at the cost of adding some hand-shaking steps to the algorithm.

IV. Performance Analysis

In this section, the performance of the proposed algorithm is investigated. To simplify the analysis, we assume: (i) available power P is divided equally among the active users, (ii) at most one data stream is assigned to each user. To impose the second restriction, we can simply eliminate a user, whenever that user is allocated one data stream in Step 2 of the algorithm. It is apparent that the sum-rate of the system with these two restrictions lower-bounds the maximum sum-rate achievable by the proposed algorithm. Although these assumptions simplify the derivations, it is shown that the results dealing with the asymptotic sum-rate remain valid even if we relax these restrictive assumptions.

To study the performance of the system, we first derive an upper-bound on the outage probability of each sub-channel. Using the derived upper-bound, we study the diversity and asymptotic sum-rate achieved by the proposed algorithm. In this study, it is assumed that all users are equipped with N receive antennas.

A. Outage Probability

The outage probability of sub-channel j is defined as $\Pr(\sigma_j^2 < z)$, $j = 1, \ldots, M$, for a given z. For σ_1^2, the derivation of the outage probability $\Pr(\sigma_1^2 < z)$ is straitforward. In the proposed algorithm, for $j = 1$, we have $\mathbf{\Xi} = [0]_{M \times M}$. From (3), we have

$$\begin{aligned} \sigma_1^2 = \quad & \max_{1 \le r \le K} \max_{\mathbf{x}} \mathbf{x}^\dagger \mathbf{H}_r^\dagger \mathbf{H}_r \mathbf{x}. \\ s.t. \quad & \mathbf{x}^\dagger \mathbf{x} = 1 \end{aligned} \tag{16}$$

Referring to [20], $\max_{\mathbf{x}} \mathbf{x}^\dagger \mathbf{H}_r^\dagger \mathbf{H}_r \mathbf{x}$ subject to $\mathbf{x}^\dagger \mathbf{x} = 1$ is equal to the maximum eigenvalue of the matrix $\mathbf{H}_r^\dagger \mathbf{H}_r$. Therefore, (16) can be written as

$$\sigma_1^2 = \max \left\{ \lambda_{\max}(\mathbf{H}_1^\dagger \mathbf{H}_1), \ldots, \lambda_{\max}(\mathbf{H}_K^\dagger \mathbf{H}_K) \right\}, \tag{17}$$

where $\lambda_{\max}(\mathbf{H}_r^\dagger \mathbf{H}_r)$ denotes the maximum eigenvalue of $\mathbf{H}_r^\dagger \mathbf{H}_r$. By assuming Rayleigh fading channel, the entries of $\mathbf{H}_r$, $r = 1, \ldots, K$, have independent normal distribution with zero-mean and unit-variance. Therefore, $\mathbf{H}_r^\dagger \mathbf{H}_r$ follows a Wishart distribution [21]. The distribution of the maximum eigenvalue of a Wishart matrix is formulated in the following lemma.

Lemma 1 *[21], [22] Assume that the entries of $A \in \mathcal{C}^{\tilde{m} \times \tilde{n}}$ have a zero mean, unit variance Gaussian distribution; then, the cumulative distribution function (CDF) of the maximum eigenvalue of the matrix $A^\dagger A$ is equal to*

$$G_{\tilde{m},\tilde{n}}(z) = \Pr \left\{ \lambda_{max}(A^\dagger A) \le z \right\} = \frac{1}{\prod_{k=1}^{n} \Gamma(m-k+1)\Gamma(n-k+1)} \det(\mathbf{\Psi}), \tag{18}$$

where $n = \min\{\tilde{m}, \tilde{n}\}$, $m = \max\{\tilde{m}, \tilde{n}\}$, and $\mathbf{\Psi}$ is an $n \times n$ Hankel matrix which is a function of $z \in [0, \infty)$ defined as

$$\mathbf{\Psi} = \left[\gamma(m - n + p + q - 1, z) \right]_{(p,q)}^{n \times n}, \qquad p, q = 1, \ldots, n, \tag{19}$$

and γ is incomplete gamma function

$$\gamma(n+1,z) = n!\left(1 - e^{-z}\sum_{k=1}^{n}\frac{z^k}{k!}\right). \tag{20}$$

Since $\lambda_{\max}(\mathbf{H}_r^\dagger\mathbf{H}_r)$ for different r's, $1 \le r \le K$, are i.i.d random variables, using (17) and Lemma 5 in Appendix II, we obtain

$$\Pr(\sigma_1^2 \le z) = G_{N,M}^K(z). \tag{21}$$

Unlike $\Pr(\sigma_1^2 \le z)$, the derivation of the outage probability for σ_j^2, $j = 2, \ldots, M$, is not simple. Alternatively, we derive an upper-bound for the outage probability of each sub-channel using the CDF of the axillary variables $\widehat{\sigma}_j^2$, $j = 2, \ldots, M$, defined as follows. Let us order the values of $\max_{\mathbf{x}} \mathbf{x}^\dagger\mathbf{H}_r^\dagger\mathbf{H}_r\mathbf{x}$, $r = 1, \ldots, K$, subject to $\mathbf{x}^\dagger\mathbf{x} = 1$ and $\widehat{\mathbf{\Xi}}_j^\dagger\mathbf{x} = 0$, where $\widehat{\mathbf{\Xi}}_j$ is a unitary matrix, $j = 2, \ldots, M$, selected randomly from $\mathcal{A}^{M\times(j-1)}$, the set of $M \times (j-1)$ complex unitary matrices. $\widehat{\sigma}_j^2$ is selected as the j^{th} largest element at this ordered set, i.e.

$$\begin{aligned}\widehat{\sigma}_j^2 = j^{\text{th}} \max_{r,\, 1\le r\le K} \quad & \max_{\mathbf{x}} \mathbf{x}^\dagger\mathbf{H}_r^\dagger\mathbf{H}_r\mathbf{x}. \\ s.t. \quad & \mathbf{x}^\dagger\mathbf{x} = 1 \\ & \widehat{\mathbf{\Xi}}_j^\dagger\mathbf{x} = 0\end{aligned} \tag{22}$$

Lemma 2 *The outage probability of the sub-channel j is upper-bounded by the CDF of $\widehat{\sigma}_j^2$. In other word,*

$$\Pr(\sigma_j^2 \le z) \le \Pr(\widehat{\sigma}_j^2 \le z). \tag{23}$$

Proof: Assume that users $\pi(1), \ldots, \pi(M)$ corresponding to the MVs $\mathbf{v}_1, \ldots, \mathbf{v}_M$ have been selected. According to the proposed algorithm, $\mathbf{v}_j$, $j = 1, \ldots, M$, is in the $M - j + 1$ dimensional hyperplane Ω_j which is the intersection of the null spaces of the previously selected MVs, i.e.

$$\mathbf{v}_j \in \Omega_j = \left\{\mathbf{x} \quad | \quad \mathbf{v}_1^\dagger\mathbf{x} = 0, \ldots, \mathbf{v}_{j-1}^\dagger\mathbf{x} = 0\right\} \tag{24}$$

Fix the hyperplane Ω_j, and multiply the channel matrix $\mathbf{H}_{\pi(i)}$, for $i = 1, \ldots, j-1$, with a unitary matrix $\widetilde{\mathbf{\Phi}}_i$ selected randomly and uniformly from $\mathcal{A}^{M\times M}$, the set of $M \times M$ complex unitary matrices.

$$\widetilde{\mathbf{H}}_{\pi(i)} = \mathbf{H}_{\pi(i)}\widetilde{\mathbf{\Phi}}_i, \quad i = 1, \ldots, j-1. \tag{25}$$

It is apparent that $\widetilde{\mathbf{H}}_{\pi(i)}$ has the gain of σ_i^2 in the direction $\widetilde{\mathbf{v}}_i = \widetilde{\mathbf{\Phi}}_i^\dagger \mathbf{v}_i$.

Let us define $\bar{\sigma}_j^2$ as follows:

$$\begin{aligned} \bar{\sigma}_j^2 = j^{\text{th}} \max_{r,\, 1 \le r \le K} \quad & \max_{\mathbf{x}} \mathbf{x}^\dagger \widetilde{\mathbf{H}}_r^\dagger \widetilde{\mathbf{H}}_r \mathbf{x}. \\ s.t. \quad & \mathbf{x}^\dagger \mathbf{x} = 1 \\ & \mathbf{x} \in \Omega_j \end{aligned} \tag{26}$$

where $\widetilde{\mathbf{H}}_r = \mathbf{H}_r$, for $r = 1, \ldots, K$, $r \notin \{\pi(1), \ldots, \pi(j-1)\}$.

Let us define the set $\mathcal{D}$, with cardinality of $K - j + 1$, of: $\max_{\mathbf{x}} \mathbf{x}^\dagger \mathbf{H}_r^\dagger \mathbf{H}_r \mathbf{x}$ subject to $\mathbf{x}^\dagger \mathbf{x} = 1$, $\mathbf{x} \in \Omega_j$ for $r = 1, \ldots, K$, $r \notin \{\pi(1), \ldots, \pi(j-1)\}$. Similarly, let us define the set $\overline{\mathcal{D}}$, with cardinality of K, of: $\max_{\mathbf{x}} \mathbf{x}^\dagger \widetilde{\mathbf{H}}_r^\dagger \widetilde{\mathbf{H}}_r \mathbf{x}$ subject to $\mathbf{x}^\dagger \mathbf{x} = 1$, $\mathbf{x} \in \Omega_j$ for $r = 1, \ldots, K$. Regarding (3) and (26), we have $\sigma_j^2 = \max \mathcal{D}$, and $\bar{\sigma}_j^2 = j^{th} \max \overline{\mathcal{D}}$. Since $\widetilde{\mathbf{H}}_r = \mathbf{H}_r$ for $r \notin \{\pi(1), \ldots, \pi(j-1)\}$, the set $\overline{\mathcal{D}}$ is equal to the union of $\mathcal{D}$ and $j - 1$ values of $\max_{\mathbf{x}} \mathbf{x}^\dagger \widetilde{\mathbf{H}}_r^\dagger \widetilde{\mathbf{H}}_r \mathbf{x}$ subject to $\mathbf{x}^\dagger \mathbf{x} = 1$, $\mathbf{x} \in \Omega_j$ for $r \in \{\pi(1), \ldots, \pi(j-1)\}$. It follows that $\sigma_j^2 \ge \bar{\sigma}_j^2$. Consequently, for a given real number z, $\Pr(\sigma_j^2 \le z) \le \Pr(\bar{\sigma}_j^2 \le z)$.

We claim that $\bar{\sigma}_j^2$ in (26) has the same distribution as $\widehat{\sigma}_j^2$ in (22). As mentioned before, Ω_j is the intersection of the null spaces of $\mathbf{v}_i$, $i = 1, \ldots, j-1$. Since the channel matrices $\mathbf{H}_{\pi(i)}$, $i = 1, \ldots, j-1$, are randomized using the unitary random matrices $\widetilde{\mathbf{\Phi}}_i^\dagger$, the vector space Ω_j is a random and independent hyperplane with respect to $\widetilde{\mathbf{H}}_r$, $r = 1, \ldots, K$. Furthermore, since the channel matrices $\mathbf{H}_r$, $r = 1, \ldots, K$, are multiplied with unitary matrices ($\widetilde{\mathbf{\Phi}}_i^\dagger$, for $\mathbf{H}_{\pi(i)}$, $i = 1, \ldots, j-1$, and identity matrix for the rest), the entries of $\widetilde{\mathbf{H}}_r$, $r = 1, \ldots, K$, have the same distribution as the entries of $\mathbf{H}_r$ (normal i.i.d distribution with zero mean and unit variance). Therefore, in both (22) and (26), we have K matrices with the same distribution while the inner maximization is performed in an $M - j + 1$ dimensional hyperplane which is random and independent of the channel matrices. Thus, each realization in problem (22) corresponds to a realization in problem (26) with the same probability. Consequently, $\bar{\sigma}_j^2$ and $\widehat{\sigma}_j^2$ have the same distribution. $\blacksquare$

The following lemma helps to derive $\Pr(\widehat{\sigma}_j^2 \le z)$.

Lemma 3 *Consider a vector space* $\widehat{\Omega}$ *defined by*

$$\widehat{\Omega} = \{\mathbf{x} \quad | \quad \mathbf{x} \in \mathcal{C}^{M\times 1}, \quad \widehat{\mathbf{\Xi}}^{\dagger}\mathbf{x} = 0\}, \tag{27}$$

where $\widehat{\mathbf{\Xi}}$ *is a complex matrix. Assume that* $\widehat{\Omega}$ *is spanned by a set of orthogonal vectors* $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_\nu\}$*, where* $\nu \leq M$*. Then, given the complex matrix* $\mathbf{H}$*, the result of the following optimization,*

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{x}, \\ s.t. \quad & \mathbf{x}^{\dagger}\mathbf{x} = 1 \\ & \mathbf{x} \in \widehat{\Omega}, \end{aligned} \tag{28}$$

is equal to $\lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}})$*, the maximum eigenvalue of matrix* $\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}$*, where*

$$\widehat{\mathbf{H}} = \mathbf{H}\mathbf{\Phi} \tag{29}$$

and

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_\nu]. \tag{30}$$

Proof: $\lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}})$, the maximum eigenvalue of matrix $\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}$ is equal to [20],

$$\begin{aligned} \lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}) = \quad & \max_{\mathbf{y}} \mathbf{y}^{\dagger}\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}\mathbf{y}. \\ s.t. \quad & \mathbf{y}^{\dagger}\mathbf{y} = 1 \end{aligned} \tag{31}$$

If (29) is substituted into (31), we obtain

$$\begin{aligned} \lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}) = \quad & \max_{\mathbf{y}} \mathbf{y}^{\dagger}\mathbf{\Phi}^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{\Phi}\mathbf{y}. \\ s.t. \quad & \mathbf{y}^{\dagger}\mathbf{y} = 1. \end{aligned} \tag{32}$$

Let $\mathbf{x} = \mathbf{\Phi}\mathbf{y} = y_1\boldsymbol{\phi}_1 + \ldots + y_\nu\boldsymbol{\phi}_\nu$. Since $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_\nu\}$ is an orthogonal vector set, then $\mathbf{y}^{\dagger}\mathbf{y} = \mathbf{x}^{\dagger}\mathbf{x}$. Also, $\mathbf{x}$ is a linear combination of vectors $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_\nu\}$; therefore, $\mathbf{x} \in \widehat{\Omega}$. Consequently,

$$\begin{aligned} \lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}) = \quad & \max_{\mathbf{x}} \mathbf{x}^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{x}. \\ s.t. \quad & \mathbf{x}^{\dagger}\mathbf{x} = 1 \\ & \mathbf{x} \in \widehat{\Omega}. \end{aligned} \tag{33}$$

■

According to Lemma 3, $\widehat{\sigma}_j^2$ in (22) is equal to

$$\widehat{\sigma}_j^2 = j^{\text{th}} \max \left\{ \lambda_{\max}(\widehat{\mathbf{H}}_{1,j}^{\dagger} \widehat{\mathbf{H}}_{1,j}), \ldots, \lambda_{\max}(\widehat{\mathbf{H}}_{K,j}^{\dagger} \widehat{\mathbf{H}}_{K,j}) \right\}, \tag{34}$$

where $\lambda_{\max}(\widehat{\mathbf{H}}_{r,j}^{\dagger} \widehat{\mathbf{H}}_{r,j})$ is the maximum eigenvalue of $\widehat{\mathbf{H}}_{r,j}^{\dagger} \widehat{\mathbf{H}}_{r,j}$, $\widehat{\mathbf{H}}_{r,j} = \mathbf{H}_r \mathbf{\Phi}_j$, and $\mathbf{\Phi}_j$ is a matrix with orthogonal columns which span the complex vector space $\Omega_j = \{\mathbf{x} | \mathbf{x} \in \mathcal{C}^{M \times 1}, \quad \widehat{\mathbf{\Xi}}_j^{\dagger} \mathbf{x} = 0\}$. Note that in (22), $\widehat{\mathbf{\Xi}}_j$ has $j-1$ non-zero orthogonal columns. Therefore, the dimension of the complex vector space Ω_j is $M - (j-1)$, resulting in $\mathbf{\Phi}_j \in \mathcal{C}^{M \times (M-j+1)}$. Since the columns of $\mathbf{\Phi}_j$ are orthonormal and the entries of $\mathbf{H}_r$ have independent unit variance Gaussian distributions (Rayleigh channel), the entries of $\widehat{\mathbf{H}}_{r,j} \in \mathcal{C}^{N \times (M-j+1)}$ have independent unit variance Gaussian distributions. Furthermore, it is easy to see that $\widehat{\mathbf{H}}_{r,j}$, $r = 1, \ldots, K$, are independent for different r. Consequently, according to the definition, $\widehat{\mathbf{H}}_{r,j}^{\dagger} \widehat{\mathbf{H}}_{r,j}$, $r = 1, \ldots, K$, have Wishart distribution. Therefore, by using Lemma 1, we obtain

$$\Pr \left\{ \lambda_{\max}(\widehat{\mathbf{H}}_{r,j}^{\dagger} \widehat{\mathbf{H}}_{r,j}) \leq z \right\} = G_{N,M-j+1}(z). \tag{35}$$

Using (34), (35), Lemma 5 in Appendix II, and regarding the independency of $\widehat{\mathbf{H}}_{r,j}$ for different r's, we obtain

$$\Pr(\widehat{\sigma}_j^2 \leq z) = \sum_{i=K-j+1}^{K} \binom{K}{i} G_{N,M-j+1}^{i}(z) [1 - G_{N,M-j+1}(z)]^{K-i}. \tag{36}$$

By using (21), (36), and Lemma 2, we have

Theorem 2

$$\Pr(\sigma_j^2 \leq z) \leq \sum_{i=K-j+1}^{K} \binom{K}{i} G_{N,M-j+1}^{i}(z) [1 - G_{N,M-j+1}(z)]^{K-i} \tag{37}$$

with equality if $j = 1$.

Theorem 2 provides a lower-bound on the performance of the proposed method. In the following, we use the above result to investigate the achieved diversity and the asymptotic sum-rate.

B. Diversity Analysis

The diversity order in a wireless channel is equal to the asymptotic slope ($z \to 0$) of the outage probability curve. This quantity determines the asymptotic slope of the curve of the symbol error rate versus signal-to-noise-ratio. In the following theorem, we use this definition to establish the diversity order of the j^{th} data stream.

Theorem 3 *Sub-channel j achieves the diversity order at least equal to $(K - j + 1)(M - j + 1)N$.*

Proof: To derive the minimum diversity of the sub-channel j, we first obtain the limiting function ($z \to 0$) of the introduced upper-bound on $\Pr(\sigma_j^2 \leq z)$. In Appendix III, it is shown that

$$\lim_{z \to 0} G_{\tilde{m},\tilde{n}}(z) = c_{\tilde{m},\tilde{n}} z^{\tilde{m}\tilde{n}} \left(1 + O(z)\right), \tag{38}$$

where $c_{\tilde{m},\tilde{n}}$ is defined in (97). Using (38) and (37), we have,

$$\lim_{z \to 0} \Pr(\sigma_j^2 < z) \leq \binom{K}{K - j + 1} c_{N,M-j+1}^{K-j+1} z^{(K-j+1)N(M-j+1)} \left(1 + O(z)\right), \tag{39}$$

where $c_{N,M-j+1}$ is equal to $c_{\tilde{m},\tilde{n}}$ by substituting N for $\tilde{m}$ and $M - j + 1$ for $\tilde{n}$ in (97).

Using (39), we conclude that the sub-channel j, $1 \leq j \leq M$, achieves the minimum diversity order of $(K - j + 1)N(M - j + 1)$. ■

Theorem (3) states that the diversity of all the sub-channels is proportional to the number of users K and number of receive antennas N. This means that the proposed method exploits both multiuser and receive diversities. In addition, the transmit diversity of sub-channel j is equal to $M - j + 1$.

C. Asymptotic sum-Rate Analysis

By using (15) and Theorem 2, a lower-bound on the average sum-rate of the proposed method can be computed. However, an examination of the asymptotic behavior ($K \to \infty$) of the sum-rate provides insight into the performance of the proposed algorithm. For this investigation, we apply some results from theory of extreme order statistics. Appendix II contains some theorems that will be used in our following discussion.

As mentioned in (17), σ_1^2 is equal to the maximum of K i.i.d random variables with common CDF of $G_{N,M}(z)$. Similarly, $\widehat{\sigma}_j^2$, $j = 2, \ldots, M$, in (34) is equal to the j^{th} largest of K i.i.d random variables with common CDF of $G_{N,M-j+1}(z)$. In general, the behavior of the j^{th} largest of K i.i.d random variables with common CDF $F(z)$ depends on the tail of the $F(z)$ (large z). In Appendix IV, it is shown that

$$G_{\tilde{m},\tilde{n}}(z) = 1 - \frac{e^{-z} z^{\widetilde{m}+\widetilde{n}-2}}{\Gamma(\widetilde{m})\Gamma(\widetilde{n})} \left(1 + O(z^{-1}e^{-z})\right), \tag{40}$$

which has the form of $F(z)$ in (67) for large z. Using (40) and applying Lemma 6 from Appendix II with $\alpha = M + N - 2$ and $\beta = \Gamma(M)\Gamma(N)$ for σ_1^2, we obtain

$$\Pr\left\{\eta_1 - \log\log(\sqrt{K}) \leq \sigma_1^2 \leq \eta_1 + \log\log(\sqrt{K})\right\} \geq 1 - O\left(\frac{1}{\log K}\right), \tag{41}$$

where

$$\eta_1 = \log\left(\frac{K}{\Gamma(M)\Gamma(N)}\right) - (M + N - 2)\log\log\left(\frac{K}{\Gamma(M)\Gamma(N)}\right). \tag{42}$$

Similarly, using (40) and applying Lemma 6 with $\alpha = M + N - j - 1$ and $\beta = \Gamma(M - j + 1)\Gamma(N)$ for $\widehat{\sigma}_j^2$, $j = 2, \ldots, M$, we obtain

$$\Pr\left\{\eta_j - \log\log(\sqrt{K}) \leq \widehat{\sigma}_j^2 \leq \eta_j + \log\log(\sqrt{K})\right\} \geq 1 - O\left(\frac{1}{\log K}\right), \tag{43}$$

where

$$\eta_j = \log\left(\frac{K}{\Gamma(M - j + 1)\Gamma(N)}\right) - (M+N-1-j)\log\log\left(\frac{K}{\Gamma(M - j + 1)\Gamma(N)}\right). \tag{44}$$

Lemma 4 *For σ_j^2, $j = 1, \ldots, M$, we have,*

$$\Pr\left\{\eta_j - \log\log(\sqrt{K}) \leq \sigma_j^2 \leq \eta_1 + \log\log(\sqrt{K})\right\} \geq 1 - O\left(\frac{1}{\log K}\right). \tag{45}$$

Proof: For $j = 1$, (45) is the same as (41). For $j = 2, \ldots, M$, the proof is as follows. From (43), we have

$$\Pr\left\{\eta_j - \log\log(\sqrt{K}) \leq \widehat{\sigma}_j^2\right\} \geq 1 - O\left(\frac{1}{\log K}\right). \tag{46}$$

Using (46) and Lemma 2, we obtain

$$\Pr\left\{\eta_j - \log\log(\sqrt{K}) \leq \sigma_j^2\right\} \geq 1 - O\left(\frac{1}{\log K}\right). \tag{47}$$

On the other hand, from (41), we have

$$\Pr\left\{\sigma_1^2 \leq \eta_1 + \log\log(\sqrt{K})\right\} \geq 1 - O\left(\frac{1}{\log K}\right). \tag{48}$$

It is easy to see that

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_M^2. \tag{49}$$

Using (48) and (49), we have,

$$\Pr\left\{\sigma_j^2 \leq \eta_1 + \log\log(\sqrt{K})\right\} \geq 1 - O\left(\frac{1}{\log K}\right). \tag{50}$$

Equations (47) and (50) result in (45). This conclusion comes from the fact that if A and B are two events with $\Pr(A) \geq 1 - \epsilon_1$ and $\Pr(B) \geq 1 - \epsilon_2$, then $\Pr(A \bigcap B) \geq 1 - \epsilon_1 - \epsilon_2$. ■

Using Lemma 4, we can prove the following theorem (refer to Appendix V).

Theorem 4

$$\lim_{K\to\infty} \frac{R}{M \log[\frac{P}{M}\log(K)]} = 1, \tag{51}$$

with probability one, where R is the sum-rate of the proposed method. In addition,

$$\lim_{K\to\infty} R_{\textit{Sum-Capacity}} - R \longrightarrow 0, \tag{52}$$

with probability one, where $R_{\textit{Sum-Capacity}}$ indicates the sum-capacity of the MIMO broadcast channel.

Equation (51) indicates that the average sum-rate of the proposed method increases linearly with the number of transmit antennas. Furthermore, the increase with the number of users K is proportional to $\log\log(K)$. In addition, Theorem 4 states that for large K, the proposed method achieves the sum-capacity of the MIMO broadcast channel. Note that these results are derived with two assumptions of equal power distribution among active users (no water-filling) and allocation of at most one data stream to each user. Apparently, Theorem 4 remains valid even if these two restrictive assumptions are relaxed.

V. Simulation Results

In this section, the outage probability and the sum-rate of the proposed method are simulated and compared with the bounds derived by the Theorem 2 and with the sum-capacity. In these simulations, the perfect channel state information is assumed to be available at the base station.

Figures 1 and 2 show the outage probability of each individual sub-channel as compared with the upper-bound CDFs introduced in Theorem 2.

Figures 3 and 4 show the sum-rate of the proposed method in comparison with the sum-capacity and the derived lower-bound on sum-rate. In the simulation of the sum-rate, the power is optimally allocated to active users by using the water-filling method, while in the simulation of the lower-bound, the power is divided equally among the sub-channels. To compute the sum-capacity of the MIMO broadcast channel, the algorithm presented in [23] is used.

Figure 3 depicts the average sum-rate of the proposed method, the derived lower-bound, and the average sum-capacity versus K (number of users) for different number of receive antennas. This figure shows that the sum-rate of the proposed method is very close to the sum-capacity, even when the number of users is small. Based on this result, we conclude that the major part of the sum-capacity is achieved with only M data streams, regardless of the number of receive antennas. In addition, Fig. 3 shows that the derived lower-bound provides an accurate estimate of the sum-rate over a wide range of values for K.

Figure 4 shows the sum-rate of the proposed method in comparison with the sum-capacity as well as the derived lower-bound versus the transmit power. It can be seen that the sum-rate of the proposed scheme is very close to the sum-capacity. In addition, Fig. 4 shows that the derived lower-bound provides an accurate estimate of the sum-rate for the different power levels.

Figure 5 shows the sum-rate of the proposed method versus the values of M (number of transmit antennas). It can be seen that the average sum-rate increases linearly with the number of transmit antennas.

VI. Conclusion

In this paper, a simple signaling method for a multi-antenna broadcast channel is proposed. This method reduces the MIMO broadcast system to a set of parallel channels. The proposed scheme has several desirable features in terms of: (i) accommodating users with different number of receive antennas, (ii) exploiting multi-user diversity, and (iii) requiring low feedback rate. The simulation results indicate that the achieved sum-rate is close to the sum-capacity of the underlying broadcast channel. To analyze the performance of the scheme, an upper-bound on the outage probability of each sub-channel is derived which is used to establish the diversity order and the asymptotic sum-rate of the scheme. It is shown that the diversity order of the j^{th} data stream, $1 \leq j \leq M$ is equal to $N(M-j+1)(K-j+1)$. Furthermore, it is proven that the throughput of this scheme scales as $M \log\log(K)$ and asymptotically $(K \longrightarrow \infty)$ tends to the sum-capacity of the MIMO broadcast channel.

Appendix I
Reducing the Feedback Rate

In this appendix, we modify the proposed algorithm to reduce the rate of the feedback at the cost of adding some hand-shaking steps to the algorithm. As mentioned in Section III, one part of the algorithm is to find the direction in which each user has maximum gain. This part of the processing can be accomplished at the receiver and then if the maximum gain of the user is larger than a given threshold, the gain and the corresponding direction are reported to the transmitter. The base station selects the best user in terms of the largest gain. By using this technique, the complete channel state information is not required at the transmitter and the rate of the feedback is significantly reduced. The details of the algorithm are presented in the following.

1) Set $j = 1$ and $\mathbf{\Xi} = [0]_{M \times M}$.
2) The user r, $r = 1, \ldots, K$, calculates $\tilde{\sigma}^2_{r(j)}$, defined as follows:

$$\begin{aligned} \tilde{\sigma}^2_{r(j)} = \quad & \max_{\mathbf{x}} \mathbf{x}^\dagger \mathbf{H}_r^\dagger \mathbf{H}_r \mathbf{x}. \\ s.t. \quad & \mathbf{x}^\dagger \mathbf{x} = 1 \\ & \mathbf{\Xi}^\dagger \mathbf{x} = 0. \end{aligned} \tag{53}$$

   $\tilde{\mathbf{v}}_{r(j)}$ represents the optimizing parameter $\mathbf{x}$.
3) The user r, $r = 1, \ldots, K$, calculates

$$\tilde{\mathbf{u}}_{r(j)} = \frac{1}{\tilde{\sigma}_{r(j)}} \mathbf{H}_r \tilde{\mathbf{v}}_{r(j)}. \tag{54}$$

4) The user r, $r = 1, \ldots, K$, sends $\tilde{\sigma}^2_{r(j)}$ and $\tilde{\mathbf{v}}_{r(j)}$ to the base station, if $\tilde{\sigma}^2_{r(j)} \geq \mathrm{th}(j)$. $\mathrm{th}(j)$ is a threshold which is predetermined by the base station.
5) The base station selects the user with the largest $\tilde{\sigma}^2_{r(j)}$, namely $\pi(j)$. σ^2_j, $\mathbf{v}_j$, and $\mathbf{u}_j$ are the gain, the corresponding MV, and the demodulation vector of the user $\pi(j)$, respectively.
6) The $\pi(j)^{th}$ user sends $\mathbf{u}_j \mathbf{H}_{\pi(j)} \mathbf{v}_i$, $i = 1, \ldots, j-1$, to the base station. This information is required for dirty paper coding.
7) The base station sends $\mathbf{v}_j$ to all the users. Each user substitutes $\mathbf{v}_j$ in the j^{th} column of $\boldsymbol{\Xi}$.
8) Set $j \leftarrow j+1$. If $j \leq M$ move to step two; otherwise stop.

The performance of this method is the same as that of the first algorithm (assume that the gain of at least one user is larger than the threshold $\mathrm{th}(j)$). However, the rate of the feedback required in the modified algorithm is significantly reduced as compared to that of the first algorithm.

Threshold $\mathrm{th}(j)$ is determined such that with high probability there exists at least one user with gain larger than $\mathrm{th}(j)$. Refereing to Lemma 4, we conclude that when K is large, with probability one the largest gain is greater than $\eta_j - \log\log(\sqrt{K})$. Consequently, for large k, an appropriate choice for $\mathrm{th}(j) = \eta_j - \log\log(\sqrt{K})$, where η_j is defined in (42).

Appendix II
Some Results on Order Statistics

Let $z_1, z_2, \ldots, z_K$ be i.i.d random variables with a common CDF $F(.)$ and probability density function $f(.)$. Let $F_{j:K}(.)$ denote the CDF of the j^{th} largest variable, $z^{(j)} = j^{th} \max\{z_1, \ldots, z_K\}$. Then, we have the following lemmas and theorems.

Lemma 5 *[24, Chapter 2, Page 8]*

$$F_{j:K}(z) = \Pr(z^{(j)} \leq z) = \sum_{i=K-j+1}^{K} \binom{K}{i} F^i(z)[1-F(z)]^{K-i}. \tag{55}$$

When $K \longrightarrow \infty$, the following theorem characterizes the limiting distribution of $F_{j:K}(.)$.

Theorem 5 *[25, Smirnov, 1949] Assume that there exists the sequence of normalizing constants $a_i > 0$ and b_i, $i = 1, \ldots, K$, such that*

$$\lim_{K \longrightarrow \infty} F_{j:K}(a_K z + b_K) = \Upsilon^{(j)}(z). \tag{56}$$

Then, $\Upsilon^{(j)}(z)$ has the following form:

$$\Upsilon^{(j)}(z) = \Lambda(z) \sum_{i=0}^{j-1} \frac{\{-\log[\Lambda(z)]\}^i}{i!}, \tag{57}$$

where $\Lambda(z)$ belongs to one of the following three types of functions:

$$\textit{Type (i)} \qquad \Lambda_1(z) = \begin{cases} 0 & z \leq 0 \\ \exp(-z^{-\epsilon}) & z > 0, \epsilon > 0 \end{cases} \tag{58}$$

$$\textit{Type (ii)} \qquad \Lambda_2(z) = \begin{cases} \exp(-(-z)^{\epsilon}) & z \leq 0, \epsilon > 0 \\ 1 & z > 0 \end{cases} \tag{59}$$

$$\textit{Type (iii)} \qquad \Lambda_3(z) = \exp(-e^{-z}). \tag{60}$$

The following theorem gives the necessary and sufficient condition for distribution $F(z)$ to belong to the domain of attraction of one of the three limiting forms.

Theorem 6 *[25] Suppose $a_K > 0$ and b_K are sequences of real numbers. For distribution function $F_{j:K}$ and $\Lambda_l(z)$, where j is a fixed natural number, we have*

$$\lim_{K \longrightarrow \infty} F_{j:K}(a_K z + b_K) = \Upsilon_l^{(j)}(z) = \Lambda_l(z) \sum_{i=0}^{j-1} \frac{\{-\log[\Lambda_l(z)]\}^i}{i!}, \tag{61}$$

if and only if

$$\lim_{K \longrightarrow \infty} K\left[1 - F(a_K z + b_K)\right] = -\log[\Lambda_l(z)]. \tag{62}$$

The following theorem determines the rate of the convergence to the limiting distributions.

Theorem 7 *[26, Dziubdziela, 1974] Assume $F(z)$ with normalizing sequences a_K and b_K is in the domain of attraction of type l limiting distribution, $l \in \{1, 2, 3\}$. If*

$\frac{1}{2} < F(a_K z + b_K) < 1$ *and* $-\log[\Lambda_l(z)] < \infty$, *then for natural number* j,

$$\left| F_{j:K}(a_K z + b_K) - \Upsilon_l^{(j)}(z) + \frac{1}{2} K\delta_K^2 g(j, K\delta_K) \right| \leq$$
$$\frac{1}{2}\pi \exp(2K\delta_K) K\delta_K^3 \left[\frac{4}{3(1-2\delta_K)} + \left(\frac{16}{9} K\delta_K^3 \frac{1}{(1-2\delta_K)^2} + \right.\right.$$
$$\left.\left. + \frac{8}{3} K\delta_K^2 \frac{1}{1-2\delta_K} + K\delta_K \right) \exp\left(K\delta_K^2 \left\{ 1 + \frac{4}{3}\delta_K \frac{1}{1-2\delta_K} \right\} \right) \right] + \Theta(z), \quad (63)$$

where

$$\Theta(z) = \left| \frac{1}{(j-1)!} \int_{K\delta_K}^{-\log[\Lambda_l(z)]} \varpi^{j-1} \exp(-\varpi) d\varpi \right|, \quad (64)$$

and

$$\delta_K(z) = 1 - F(a_K z + b_K), \quad (65)$$

and

$$g(z, \vartheta) = \begin{cases} 0 & z \leq 0 \\ \exp(-\vartheta) & 0 < z \leq 1 \\ \exp(-\vartheta)\left(\frac{\vartheta^{\mu+1}}{(\mu+1)!} - \frac{\vartheta^{\mu}}{(\mu)!} \right) & \mu + 1 < z \leq \mu + 2 \qquad \mu = 0, 1, 2, \ldots \end{cases} \quad (66)$$

In the following lemma, we apply the above theorems for a specific distribution which is used throughout the paper.

Lemma 6 *Let* $z_1, z_2, \ldots, z_K$ *be* K *i.i.d random variables with a common CDF*

$$F(z) = 1 - \frac{1}{\beta} z^{\alpha} e^{-z} \quad \alpha \geq 0, \beta > 0, \quad (67)$$

then,

- *Distribution function* $F(z)$ *is in the domain of attraction of type (iii) limiting distribution with normalizing sequences*

$$a_K = 1, \quad (68)$$
$$b_K = \log\left(\frac{K}{\beta}\right) - \alpha \log\log\left(\frac{K}{\beta}\right). \quad (69)$$

- *If* $z^{(j)}$ *denotes the* j^{th} *largest random variable, then,*

$$\Pr\left\{ b_K - \log\log(\sqrt{K}) \leq z^{(j)} \leq b_K + \log\log(\sqrt{K}) \right\} \geq 1 - O\left(\frac{1}{\log K}\right). \quad (70)$$

Proof:

Part One:

Using a_K and b_K, defined in (68) and (69), we have

$$\lim_{K\longrightarrow\infty} K(1-F(a_Kz+b_K)) = K\frac{1}{\beta}(a_Kz+b_K)^{-\alpha}\exp(-a_Kz-b_K) =$$
$$\lim_{K\longrightarrow\infty} K\frac{1}{\beta}\left[z+\log\left(\frac{K}{\beta}\right)-\alpha\log\log\left(\frac{K}{\beta}\right)\right]^{-\alpha}\exp\left[-z-\log\left(\frac{K}{\beta}\right)+\alpha\log\log\left(\frac{K}{\beta}\right)\right] =$$
$$\lim_{K\longrightarrow\infty}\exp(-z)\left[z+\log\left(\frac{K}{\beta}\right)-\alpha\log\log\left(\frac{K}{\beta}\right)\right]^{-\alpha}\left[\log\left(\frac{K}{\beta}\right)\right]^{\alpha} =$$
$$= \exp(-z) = -\log[\Lambda_3(z)]. \tag{71}$$

Therefore, regarding Theorem 6, the distribution (67) is in the domain of attraction of type (iii) limiting distribution.

Part Two:

Substituting $\log\log(\sqrt{K})$ and $-\log\log(\sqrt{K})$ in $\Upsilon_3^{(j)}(z)$, defined in (61), we obtain

$$\Upsilon_3^{(j)}\left(\log\log\sqrt{K}\right) = \exp\left(-\frac{1}{\log\sqrt{K}}\right)\sum_{i=0}^{j-1}\frac{1}{i!\log^i\sqrt{K}} = 1-O\left(\frac{1}{\log K}\right), \tag{72}$$

and

$$\Upsilon_3^{(j)}\left(-\log\log\sqrt{K}\right) = \frac{1}{\sqrt{K}}\sum_{i=0}^{j-1}\frac{\log^i\sqrt{K}}{i!} = O\left(\frac{\log^j\sqrt{K}}{\sqrt{K}}\right). \tag{73}$$

Therefore,

$$\Upsilon_3^{(j)}\left(\log\log\sqrt{K}\right) - \Upsilon_3^{(j)}\left(-\log\log\sqrt{K}\right) \geq 1-O\left(\frac{1}{\log K}\right). \tag{74}$$

In the following, we apply Theorem 7 to find out how $F_{j:K}(z)$ is close to limiting distribution $\Upsilon_3^{(j)}(z)$ at $z=\log\log\sqrt{K}$ and $z=-\log\log\sqrt{K}$. To simplify the derivation, we first calculate some terms appeared in Theorem 7 at these two points.

Using (65), (68), and (69), for $F(z)$ in (67), we obtain

$$\delta_K(z) = 1-F(a_Kz+b_K) = \frac{e^{-z}}{K}\left[1+O\left(\frac{1}{\log K}\right)\right]. \tag{75}$$

Therefore,

$$\delta_K\left(\log\log\sqrt{K}\right) = \frac{1}{K\log\sqrt{K}}\left[1+O\left(\frac{1}{\log K}\right)\right], \tag{76}$$

and

$$\delta_K\left(-\log\log\sqrt{K}\right) = \frac{\log\sqrt{K}}{K}\left[1+O\left(\frac{1}{\log K}\right)\right]. \tag{77}$$

It is easy to show that $\Theta(z)$ in (64) is equal to,

$$\Theta(z) = \left| \exp[-K\delta_K(z)] \sum_{i=0}^{j-1} \frac{[K\delta_K(z)]^i}{i!} - \Upsilon_3^{(j)}(z) \right|. \tag{78}$$

On the other hand, using (76), we obtain

$$\exp[-K\delta_K(\log\log\sqrt{K})] = 1 - O\left(\frac{1}{\log K}\right), \tag{79}$$

and using (76), we obtain

$$\exp[-K\delta_K(-\log\log\sqrt{K})] = O\left(\frac{1}{\sqrt{K}}\right). \tag{80}$$

Consequently, using (72), (76), (78) and (79), we have

$$\Theta\left(\log\log\sqrt{K}\right) = O\left(\frac{1}{\log K}\right), \tag{81}$$

and using (73), (77), (78) and (80), we have

$$\Theta\left(-\log\log\sqrt{K}\right) = O\left(\frac{\log^j\sqrt{K}}{\sqrt{K}}\right). \tag{82}$$

Regarding (66), we have,

$$g(j, K\delta_K(z)) = \begin{cases} \exp(-K\delta_K(z)) & j = 1 \\ \exp(-K\delta_K(z))\left(\frac{[K\delta_K(z)]^{j-1}}{(j-1)!} - \frac{[K\delta_K(z)]^{j-2}}{(j-2)!}\right) & j \geq 2 \end{cases} \tag{83}$$

Therefore, using (76), we obtain,

$$K\delta_K^2(\log\log\sqrt{K})g\left(j, K\delta_K(\log\log\sqrt{K})\right) = o\left(\frac{1}{K}\right), \tag{84}$$

and using (77), we obtain,

$$K\delta_K^2(-\log\log\sqrt{K})g\left(j, K\delta_K(-\log\log\sqrt{K})\right) = o\left(\frac{1}{K}\right). \tag{85}$$

Applying Theorem 7 for $z = \log\log\sqrt{K}$, and using (76), (81), and (84), we have

$$\left| F_{j:K}(\log\log\sqrt{K} + b_K) - \Upsilon_3^{(j)}(\log\log\sqrt{K}) + o\left(\frac{1}{K}\right) \right| \leq O\left(\frac{1}{\log K}\right). \tag{86}$$

Similarly, Applying Theorem 7 for $z = -\log\log\sqrt{K}$, and using (77), (82), and (85), we have

$$\left| F_{j:K}(-\log\log\sqrt{K} + b_K) - \Upsilon_3^{(j)}(-\log\log\sqrt{K}) + o\left(\frac{1}{K}\right) \right| \leq O\left(\frac{\log^j\sqrt{K}}{\sqrt{K}}\right). \tag{87}$$

Using (74), (87), and (86), we obtain

$$\left| F_{j:K}\left(\log\log\sqrt{K} + b_K\right) - F_{j:K}\left(-\log\log\sqrt{K} + b_K\right) \right| \geq 1 - O\left(\frac{1}{\log K}\right). \tag{88}$$

Since $F_{j:K}(.)$ denotes CDF of $z^{(j)}$, (88) results in (70). ∎

Appendix III
$G_{\tilde{m},\tilde{n}}(z)$ for Small z

By substituting the Taylor expansion of e^z and e^{-z} into (20),

$$\begin{aligned}
\gamma(n+1,z) = \\
& n!\left(1-e^{-z}\sum_{m=1}^{n}\frac{z^m}{m!}\right) = n!e^{-z}\left(e^z-\sum_{m=1}^{n}\frac{z^m}{m!}\right) \\
&= n!\left(1-z+\frac{z^2}{2!}-\frac{z^3}{3!}+\cdots\right)\left(\sum_{m=1}^{\infty}\frac{z^m}{m!}-\sum_{m=1}^{n}\frac{z^m}{m!}\right) \\
&= n!\left(1-z+\frac{z^2}{2!}-\frac{z^3}{3!}+\cdots\right)\left(\sum_{m=n+1}^{\infty}\frac{z^m}{m!}\right) \\
&= \frac{z^{n+1}}{n+1}(1+O(z)). \qquad (89)
\end{aligned}$$

Substituting (89) in (19), we have,

$$\mathbf{\Psi} = \left[\frac{z^{m-n+p+q-1}}{m-n+p+q-1}(1+O(z))\right]_{(p,q)}^{n\times n}. \qquad (90)$$

It is known that if a column or row of a matrix is multiplied by variable z, the determinant of the resulting matrix is z times the determinant of the original matrix. Using this property, first, we factor z^{m-n+q} from column q, $0 \le q \le n$ of the $\mathbf{\Psi}$, and then we factor z^{p-1} from row p, $1 \le p \le n$. The remaining matrix is equal to $\left[\frac{1}{m-n+p+q-1}(1+O(z))\right]_{(p,q)}^{n\times n}$, and the power of z outside the determinant is equal to

$$\sum_{q=1}^{n}(m-n+q)+\sum_{p=1}^{n}(p-1) = mn. \qquad (91)$$

Therefore,

$$\det(\mathbf{\Psi}) = z^{mn}\det\left(\left[\frac{1}{m-n+p+q-1}(1+O(z))\right]_{(p,q)}^{n\times n}\right). \qquad (92)$$

By substituting (92) into (18), we have

$$\begin{aligned}
G_{\tilde{m},\tilde{n}}(z) = \\
\frac{z^{mn}}{\prod_{k=1}^{n}\Gamma(m-k+1)\Gamma(n-k+1)}\det\left(\left[\frac{1}{m-n+p+q-1}(1+O(z))\right]_{(p,q)}^{n\times n}\right) \qquad (93)
\end{aligned}$$

According to (93), the coefficient of the smallest degree of z is equal to

$$c = \frac{1}{\prod_{k=1}^{n} \Gamma(m-k+1)\Gamma(n-k+1)} \det\left(\left[\frac{1}{m-n+p+q-1}\right]_{(p,q)}^{n\times n}\right). \tag{94}$$

We apply the following formula to calculate the determinant in (94) (see [27, p. 92, Problem 3])

$$\det\left(\left[\frac{1}{x_p+y_q}\right]_{(p,q)}\right) = \frac{\prod_{q>p}(x_q-x_p)(y_q-y_p)}{\prod_{p,q}(x_p+y_q)}, \tag{95}$$

where x_p and y_q depend only on p and q, respectively. Substituting $x_p = m-n+p-1$ and $y_q = q$ in (95), we compute the determinant term in (94), resulting in

$$\lim_{z\to 0} G_{\tilde{m},\tilde{n}}(z) = c_{\tilde{m},\tilde{n}} z^{\tilde{m}\tilde{n}}, \tag{96}$$

where $c_{\tilde{m},\tilde{n}}$ is equal to

$$c_{\tilde{m},\tilde{n}} = \frac{\Pi_{\zeta=1}^{n-1}(n-\zeta)!}{\prod_{k=1}^{n}(m-k)!\Pi_{\zeta=1}^{n}(m-n+\zeta)^{\zeta}(m+n-\zeta)^{\zeta}}, \tag{97}$$

where $n = \min\{\tilde{m},\tilde{n}\}$ and $m = \max\{\tilde{m},\tilde{n}\}$.

Appendix IV

$G_{\tilde{m},\tilde{n}}(z)$ for Large z

By using (20), the determinant of matrix $\mathbf{\Psi}$ in (19) has the following structure:

$$\begin{aligned} \det(\mathbf{\Psi}) = & \\ & \det\left([\gamma(m-n+p+q-1,z)]_{(p,q)}^{n\times n}\right) = \\ & \varphi_0 + \varphi_1(z)e^{-z} + \varphi_2(z)e^{-2z} + \cdots + \varphi_n(z)e^{-nz}, \end{aligned} \tag{98}$$

where φ_0 is a constant number, and $\varphi_i(z)$, $i = 1, \cdots, n$ are polynomials. Therefore, when $z \to \infty$,

$$\det(\mathbf{\Psi}) \to \varphi_0 + \kappa z^{\iota} e^{-z}, \tag{99}$$

where ι is the degree of $\varphi_1(z)$, and κ is the coefficient of z^{ι} in $\varphi_1(z)$. In the following, we determine φ_0, κ, and ι.

Computing φ_0: Using the expansion (20), it is easy to verify that

$$\lim_{z\to\infty} \gamma(m-n+p+q-1,z) = (m-n+p+q-2)!. \tag{100}$$

Regarding (98) and (100) ,

$$\varphi_0 = \lim_{z\to\infty} \det(\mathbf{\Psi}) = \det\left([(m-n+p+q-2)!]_{(p,q)}^{n\times n}\right). \tag{101}$$

On the other hand, since $G_{\tilde{m},\tilde{n}}(z)$ is the CDF of a random variable, $\lim_{z\to\infty} G_{\tilde{m},\tilde{n}}(z) = 1$. Substituting (101) in (18), we have

$$\lim_{z\to\infty} G_{\tilde{m},\tilde{n}}(z) = \frac{\det\left([(m-n+p+q-2)!]_{(p,q)}^{n\times n}\right)}{\prod_{k=1}^{n}\Gamma(m-k+1)\Gamma(n-k+1)} = 1. \tag{102}$$

Considering (101) and (102), we obtain

$$\varphi_0 = \prod_{k=1}^{n}\Gamma(m-k+1)\Gamma(n-k+1). \tag{103}$$

Computing κ, and ι: Applying the method of *expansion by minors*, we expand the determinant of $\mathbf{\Psi}$ in (19), based on the last row of the matrix. It is evident that the largest power of z in $\varphi_1(z)$ is determined by $\Psi(n,n)$, multiplied by the constant term of its cofactor. By using (19) and (20), it is easy to show that this term is equal to

$$\det\left([(m-n+p+q-2)!]_{(p,q)}^{(n-1)\times(n-1)}\right)\gamma(m+n-1,z), \tag{104}$$

where $\gamma(m+n-1,z)$ is entry (n,n) of matrix $\mathbf{\Psi}$, and $\det\left([(m-n+p+q-2)!]_{(p,q)}^{(n-1)\times(n-1)}\right)$ is the constant part of its cofactor. Using (20), we obtain

$$\gamma(m+n-1,z) = (m+n-2)! - z^{m+n-2}e^{-z}\left(1+O(z^{-1})\right). \tag{105}$$

By rewriting (102), we obtain

$$\det\left([(m-n+p+q-2)!]_{(p,q)}^{n\times n}\right) = \prod_{k=1}^{n}\Gamma(m-k+1)\Gamma(n-k+1). \tag{106}$$

By substituting $m-1$ for m and $n-1$ for n in (106),

$$\det\left([(m-n+p+q-2)!]_{(p,q)}^{(n-1)\times(n-1)}\right) = \prod_{k=1}^{n-1}\Gamma(m-k)\Gamma(n-k). \tag{107}$$

Considering (104), (105), and (107), we have

$$\kappa = -\prod_{k=1}^{n-1}\Gamma(m-k)\Gamma(n-k), \tag{108}$$

and,

$$\iota = m+n-2. \tag{109}$$

Using (99), (103), (108), (109), and (18), we have,

$$G_{\tilde{m},\tilde{n}}(z) = 1 - \frac{e^{-z} z^{m+n-2}}{(m-1)!(n-1)!} \left(1 + O(z^{-1}e^{-z})\right), \tag{110}$$

Since $m = \max\{\tilde{m}, \tilde{n}\}$ and $n = \min\{\tilde{m}, \tilde{n}\}$, we have

$$G_{\tilde{m},\tilde{n}}(z) = 1 - \frac{e^{-z} z^{\tilde{m}+\tilde{n}-2}}{(\tilde{m}-1)!(\tilde{n}-1)!} \left(1 + O(z^{-1}e^{-z})\right), \tag{111}$$

Appendix V
Asymptotic Sum-Rate

Since log(.) is an increasing function and using (45), we have,

$$\begin{aligned} \Pr\Bigg\{\log\left(1 + \frac{P}{M}[\eta_j - \log\log\sqrt{K}]\right) \\ \leq \log(1 + \frac{P}{M}\sigma_j^2) \\ \leq \log\left(1 + \frac{P}{M}[\eta_1 + \log\log\sqrt{K}]\right)\Bigg\} \\ \geq 1 - O\left(\frac{1}{\log K}\right). \end{aligned} \tag{112}$$

Consequently,

$$\begin{aligned} \lim_{K\to\infty} \Pr\left\{\frac{\log\left(1 + \frac{P}{M}[\eta_j - \log\log\sqrt{K}]\right)}{\log[\frac{P}{M}\log(K)]}\right. \\ \leq \frac{\log(1 + \frac{P}{M}\sigma_j^2)}{\log[\frac{P}{M}\log(K)]} \\ \left.\leq \frac{\log_2\left(1 + \frac{P}{M}[\eta_1 + \log\log\sqrt{K}]\right)}{\log[\frac{P}{M}\log(K)]}\right\} \\ \geq 1 - O\left(\frac{1}{\log K}\right). \end{aligned} \tag{113}$$

Using (42) and (44), we conclude that the left hand side and the right hand side of the inequalities inside Pr in (113) tend to the same value of one as $K \to \infty$, therefore

$$\lim_{K\to\infty} \frac{\log(1 + \frac{P}{M}\sigma_j^2)}{\log(\frac{P}{M}\log K)} = 1, \tag{114}$$

with probability one.

Equation (114) indicates that the rate of each sub-channel attains $\log[\frac{P}{M}\log(K)]$, when $K \to \infty$. Using (15), the sum-rate of the proposed method achieves $M\log[\frac{P}{M}\log(K)]$.

On the other hand, according to (112),

$$\Pr\left\{\log\left(1+\frac{P}{M}\left[\eta_j-\log\log\sqrt{K}\right]\right)\leq\log\left(1+\frac{P}{M}\sigma_j^2\right)\right\}\geq 1-O\left(\frac{1}{\log K}\right). \quad (115)$$

In [28], it is shown that,

$$\Pr\left\{\frac{R_{\text{Sum-Capacity}}}{M}\leq\log\left(1+\frac{P}{M}\left[\log(KN)+O(\log\log[K])\right]\right)\right\}\geq 1-O\left(\frac{1}{\log^2 K}\right). \quad (116)$$

As mentioned before, if A and B are two events with $\Pr(A)\geq 1-\epsilon_1$ and $\Pr(B)\geq 1-\epsilon_2$, then $\Pr(A\bigcap B)\geq 1-\epsilon_1-\epsilon_2$. Therefore, the probability that the inequalities inside Pr in (115) and (116) are both valid is greater than $1-O\left(\frac{1}{\log K}\right)-O\left(\frac{1}{\log^2 K}\right)$. Subtracting these two inequalities, we obtain

$$\begin{aligned}\Pr\Bigg\{&\log(1+\frac{P}{M}\sigma_j^2)-\frac{R_{\text{Sum-Capacity}}}{M}\geq\\ &\log\left(1+\frac{P}{M}\left[\eta_j-\log\log\sqrt{K}\right]\right)-\quad\log\left(1+\frac{P}{M}[\log(KN)+O(\log\log[K])]\right)\Bigg\}\\ &\geq 1-O\left(\frac{1}{\log K}\right)-O\left(\frac{1}{\log^2 K}\right).\end{aligned} \quad (117)$$

Using (44), we conclude that the right side of the inequality inside Pr in (117) tends to zero as $K\to\infty$. Consequently, for large K, with probability one, we have

$$0\leq\log(1+\frac{P}{M}\sigma_j^2)-\frac{R_{\text{Sum-Capacity}}}{M},\quad j=1,\ldots,M. \quad (118)$$

Using (118), we obtain that when $K\to\infty$, $R\geq R_{\text{Sum-Capacity}}$. Since $R_{\text{Sum-Capacity}}$ provides an upper bound on the sum-rate of any algorithm , we obtain

$$\lim_{K\to\infty}R_{\text{Sum-Capacity}}-R=0, \quad (119)$$

with probability one.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, pp. 585–595, Nov. 1999.

[2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time code for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1988.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

[4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[5] P. Viswanath and D.N.C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1912 – 1921, Aug. 2003.

[6] W. Yu and J. Cioffi, "Sum capacity of vector Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, submitted for Publication.

[7] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," in *38th Annual Conference on Information Sciences and Systems (CISS 2004)*, Princeton, NJ, March 2004.

[8] W. Yu and W. Rhee, "Degrees of freedom in multi-user spatial multiplex systems with multiple antennas," *IEEE Transactions on Communications*, 2004, submitted for Publication.

[9] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[10] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," *IEEE Trans. Inform. Theory*, August 2002, submitted for Publication.

[11] U. Erez and S. ten Brink, "Approaching the dirty paper limit in canceling known interference," in *Allerton Conf. Commun., Contr., and Computing*, Oct. 2003.

[12] P. Viswanath, D.N.C Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[13] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE Int. Computer Conf. (ICC 95)*, Seattle, WA, June 1995.

[14] C. Peel, B. Hochwald, and L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication—parts I and II," *IEEE Trans. on Commun.*, June 2003, submitted for Publication.

[15] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, July 2003.

[16] Z. Tu and R.S. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Communications Letters*, vol. 7, pp. 370–372, Aug. 2003.

[17] L.U. Choi and R.D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Communications*, vol. 3, pp. 20–24, Jan. 2004.

[18] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 52, pp. 461–471, Feb. 2004.

[19] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[20] R.G. Horn and C.A. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[21] M. Kang and M.-S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO mrc systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 418–426, April 2003.

[22] C. G. Khatri, "Distribution of the largest or the smallest characteristic root under null hyperthesis concerning complex multivariate normal populations," *Ann. Math. Stat.*, vol. 35, pp. 1807–1810, 1964.

[23] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple access channel sum

capacity problem," in *37th Annual Conference on Information Sciences and Systems (CISS)*, March 2003.

[24] H. A. David, *Order Statistics*, John Wiley and Sons, Inc, New York, second edition, 1980.

[25] N. V. Smirnov, "Limit distributions for the terms of a variational series," in *Trudy Mat. Inst.*, 1949, vol. 25.

[26] W. Dziubdziela, "On convergence rates in the limit laws of extreme order statistics," in *Trans. 7th Prague Conference and 1974 European Meeting of Statisticians*, 1974, vol. B, pp. 119–127.

[27] G. Polya and G. Szego, *Problems and Theorems in Analysis*, vol. 2, Springer-Verlag, Heidelberg, 1976.

[28] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and bemforming for MIMO broadcast channels with many users," *IEEE Trans. on Comm.*, 2004, Submitted for Publication.

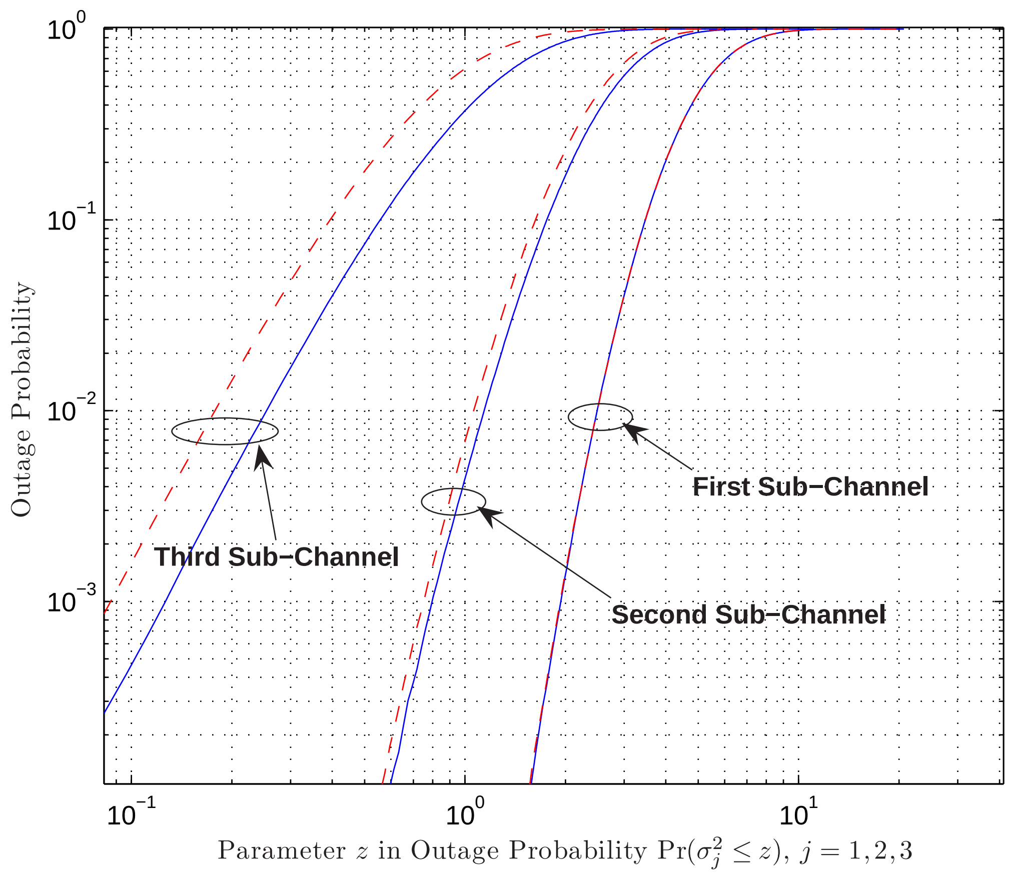

Fig. 1. Outage Probability for the Sub-Channels (Solid Curves) and the Upper-Bound for Outage Probability (Dashed Curves) – $K = 6$, $M = 3$, $N = 1$.

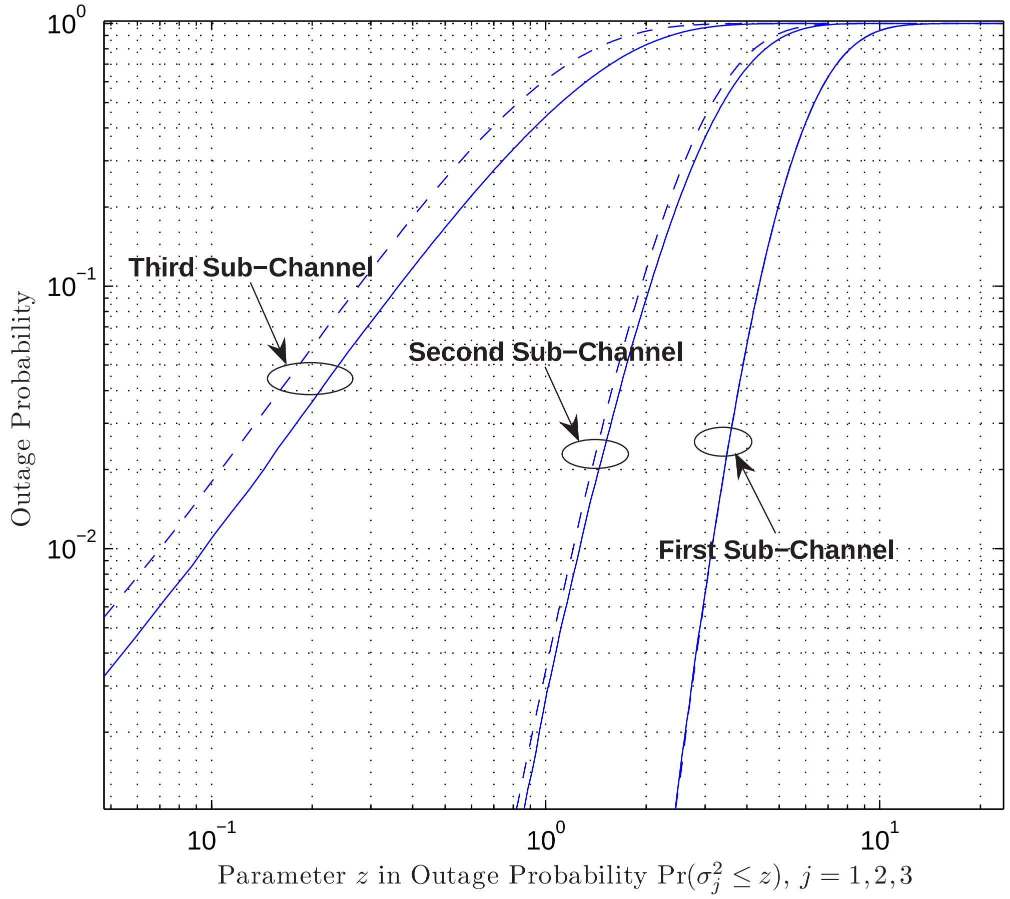

Fig. 2. Outage Probability for the Sub-Channels (Solid Curves) and the Upper-Bound for Outage Probability (Dashed Curves) – $K = 3,\ M = 3,\ N = 2$.

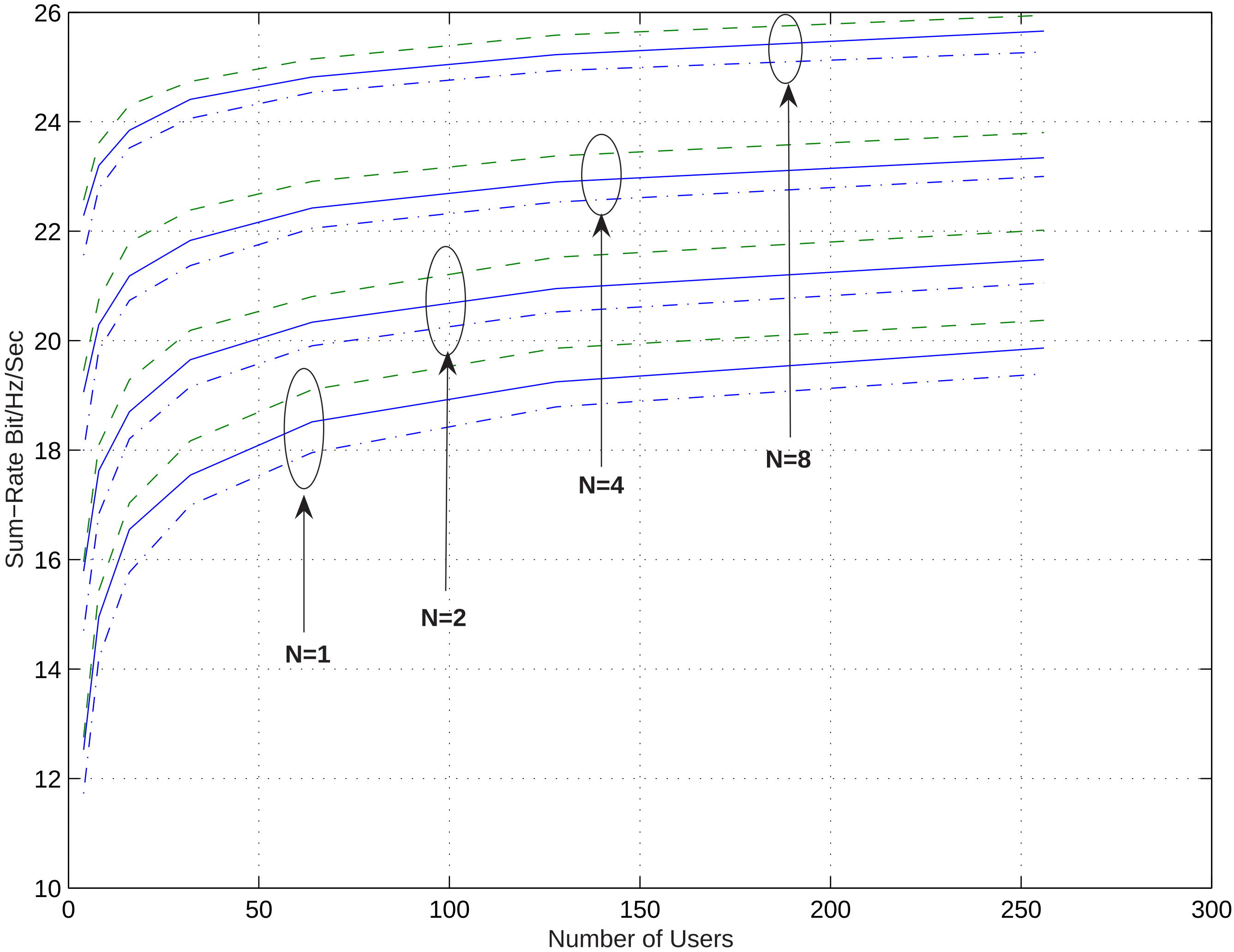

Fig. 3. Average Sum-Rate of the Proposed Method (Solid Curves), Average Sum-Capacity (Dashed Curves), and Lower-Bound on the Sum-Rate of the Proposed Method (Dash-Dot Curves) – $M = 4$, $P = 15$.

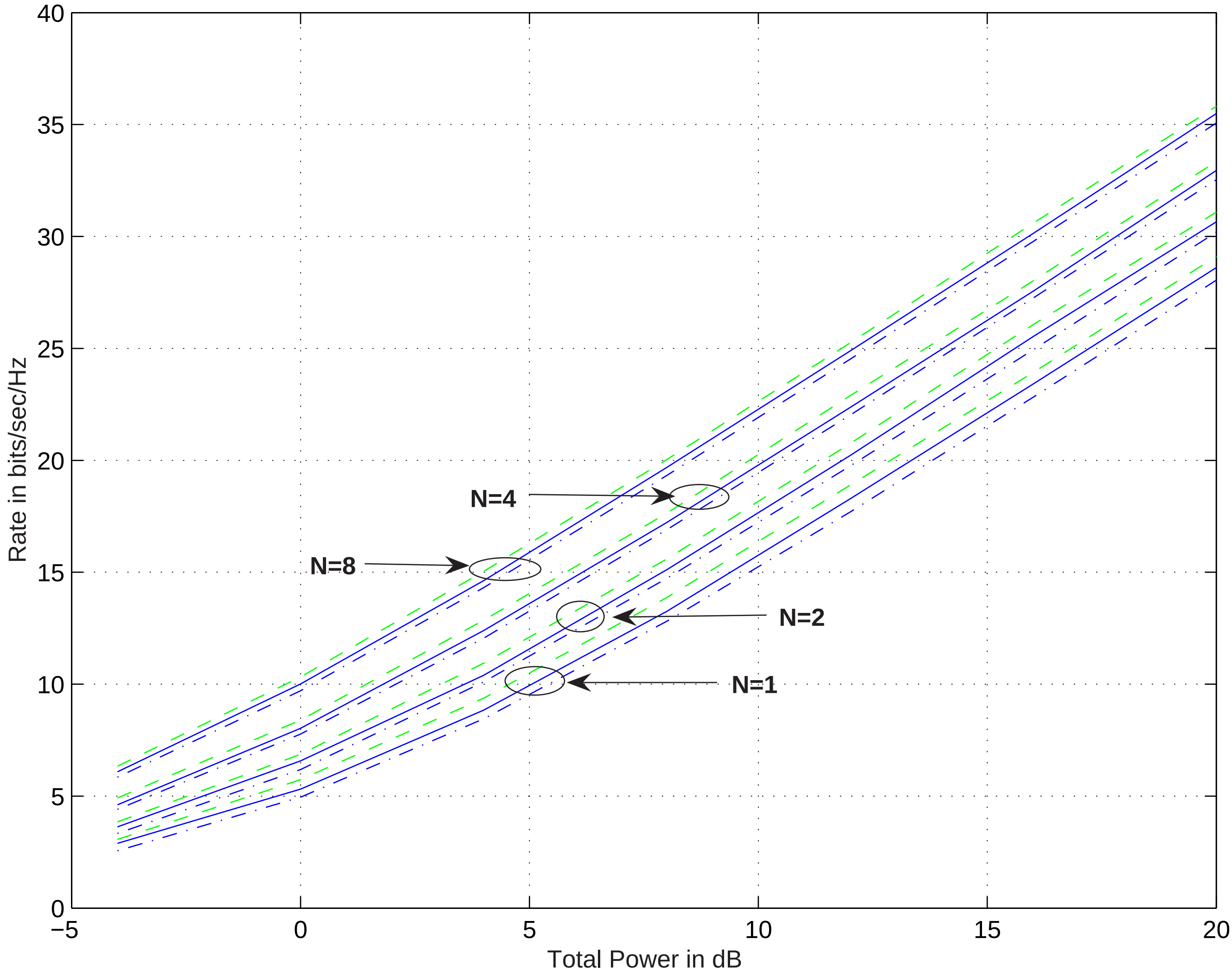

Fig. 4. Average Sum-Rate of the Proposed Method (Solid Curves), Average Sum-Capacity (Dashed Curves), and Lower-Bound on the Sum-Rate of the Proposed Method (Dash-Dot Curves) – $M = 4$, $K = 40$.

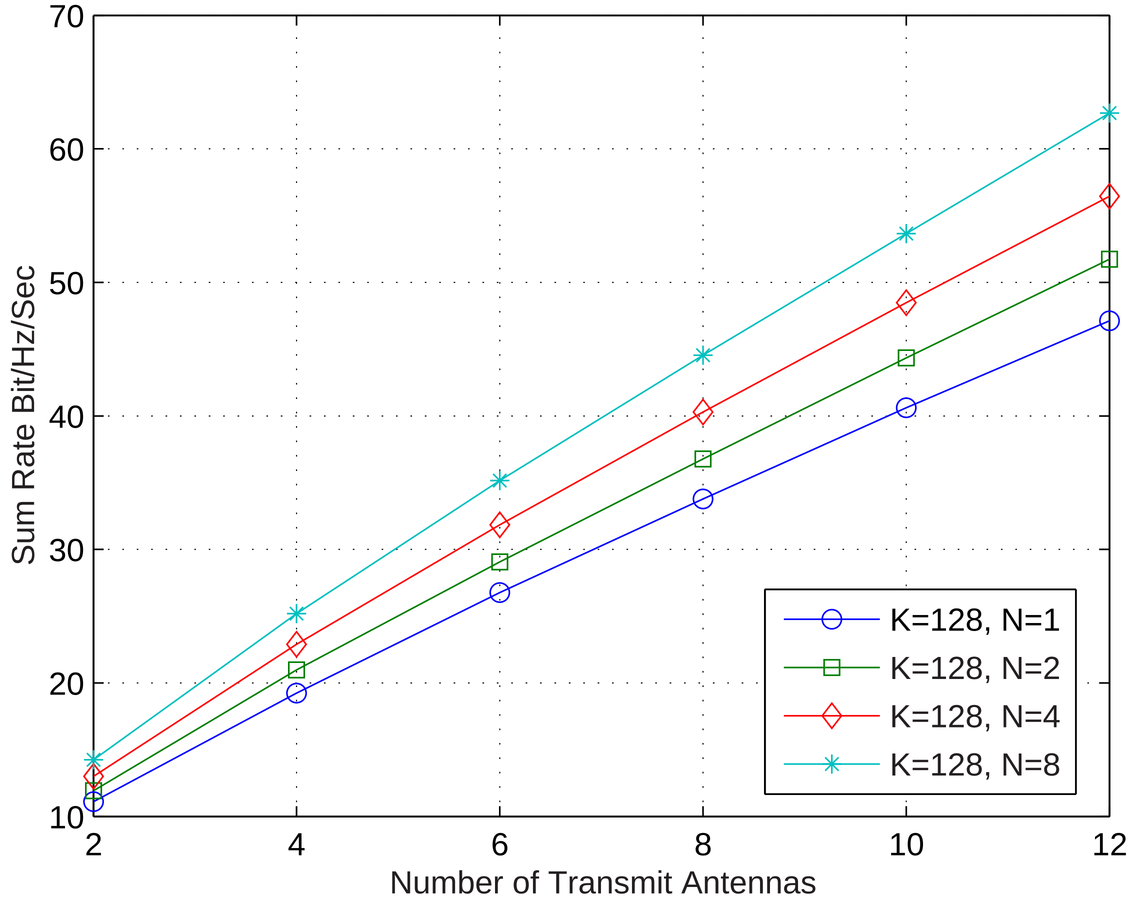

Fig. 5. Average Sum-Rate of the Proposed Method versus the Number of Transmit Antennas