On the Degrees of Freedom of MISO Broadcast Channels with Delayed Feedback

Mohammad Ali Maddah-Ali
David Tse

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-122
http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-122.html

September 6, 2010

On the Degrees of Freedom of MISO Broadcast Channels with Delayed Feedback

Mohammad Ali Maddah-Ali and David Tse
Department of Electrical and Computer Engineering,
University of California - Berkeley

Abstract

In information theoretic analysis of communication over MIMO broadcast channels, two extreme assumptions about availability of the channel state information (CSI) at the base station are usually made; either perfect CSI is available at the transmitter, or no CSI is available at all. However, in practical systems, there is usually CSI feedback but the feedback is subject to delays. Conventionally, this issue is alleviated through prediction of the current CSI using the available outdated one. However, as the delay becomes larger or the coherent time becomes smaller, the prediction-based schemes fail and offer no gain beyond no CSI case. This observation supports the popular belief that in such cases, delayed feedback is not useful at all.

In this paper, we disprove this conjecture and show that even when the delay is arbitrary large or the coherent time is arbitrary small, channel feedback can still unboundedly improve the throughput. Indeed, the delayed feedback can increase the degree of freedom (DoF) of the channel. In particular, we focus on a time-varying Gaussian broadcast channels with k transmit antennas and k single-antenna users and assume that users causally have the perfect CSI, but transmitter receives CSI with some delays. We show that even if the channel state varies independently over time, the degrees of freedom of $\frac{k}{1+\frac{1}{2}+\ldots+\frac{1}{k}}$ is achievable. Moreover, we establish that if all users experience CSI, with identical distribution, varying independently over time, then this is the optimal DoF.

I. Introduction

In multiple-antenna broadcast channels, a base station, equipped with more than one antenna, services multiple users simultaneously. The base station relies on the knowledge of channel state information (CSI) to manage or cancel out the cross interference among concurrent data streams. Consequently, the assumption about the knowledge of the base station about CSI would strongly affect the rate region that the base station can support.

In information theoretic analysis of multiple-antenna broadcast channels, it is mainly assumed that the channel is time–invariant and frequency flat, where the perfect CSI is available at the transmitter (CSIT) and all receivers (CSIR). Under such assumptions, the capacity of multiple-antenna Gaussian broadcast channel is well understood. Initially in [1], followed by [2]–[4], it is shown that dirty-paper coding with Gaussian input distribution can achieve the sum-capacity. Later in [5], it is shown that this scheme achieves the entire capacity region [5]. Using the result of [2]–[5], it is easy to see that degrees of freedom (DoF) of the multiple-antenna broadcast channel with perfect CSIT and CSIR is the minimum of the number of transmit antennas, and total number of receive antennas. The full DoF of the channel can be achieved using simple precoding techniques such as zero-forcing schemes.

Another extreme assumption is the case where the perfect CSI is available at the receivers but not available at the base station. In this case, the capacity or even DoF of the channel is generally unknown. In [6], it is shown in a system with two-antenna base station and two single-antenna receivers, where channel realization are time varying, under some general conditions, DoF is upper-bounded by $\frac{4}{3}$. This outer-bound is valid if the base station has a quantized version of the channel state information with non-vanishing quantization error. However, in terms of achievability, no scheme has been reported to attain any DoF greater that one, under the assumptions of [6], and therefore there is an unbounded gap between the best available outer-bound and inner-bound for the throughput of the system. However, in some particular cases, the gap of DoF is closed. The first one is the case where both channel vectors have the same marginal distributions, In this case, the channel is degenerate and the optimal DoF is just one [1]. This result has been extended to a more general isotropic channel in [7]. The second case is compound broadcast channels with an M-antenna base station and K single antenna users, where the channel of each user is selected from a finite set. It is shown that the DoF of that channel is $\frac{MK}{M+K-1}$, almost surely, where the cardinality of each set is greater than M [8], [9]. The scheme of [8], [9] is based on ignoring the cooperation among transmit antennas, and using the idea of interference alignment over rational field proposed in [10] for time-invariant frequency-flat channels.

In practical systems, we do not have any of the above extreme cases. Generally, the CSI is estimated at the side of the receivers, and then fed back to the base station, through a finite-rate feedback channel. Therefore, the access of the base station to the accurate channel state is limited by two factors:

- *Quantization Error:* The limited rate of feedback channel restricts the accuracy of the CSI at the base station. However, if the rate of feedback linearly increases with $\log_2(SNR)$, then zero-forcing scheme can still achieve the full DoF of the channel [11]. Therefore, if the data-rate of feedback is high enough, we can still mitigate interference effectively.
- *Delay:* Generally in wireless systems, the CSI is provided to the base station with some delays. The delay comes from the fact that, the users need some times to receive pilots, estimate CSI, and then feed it back to the base station in a relatively long coding block. Therefore, when the channel information arrives to the base station, the channel state has already changed. Therefore, the base station has always access to the outdated channel information which may not be good enough for interference management. There is considerable number of literature dealing with this problem through exploiting the channel correlation in time to predict the channel state information. However, as the coherent time of the channels becomes smaller, due to higher mobility for example, then the current state of the channel reveals no reliable information about the future states, and therefore the schemes relying on channel prediction offers no gain beyond no–CSIT gain. That observation suggest that in fast fading environment, the delayed channel feedback is not useful at all.

Here in this work, we focus on the second issue and show that, contrary to popular belief, even if the channel state information are completely independent in time, the delayed channel feedback can still be extremely useful. More precisely, the outdated channel information at the base station can change the DoF of the channel.

The capacity of the channels with feedback is first considered by Shannon [12]. He proved that in memoryless point-to-point channel, feedback does not improve the capacity. In [13], Schalkwijk and Kailath proposed a novel approach based on iteratively improving the error at the receiver and showed that feedback can improve the error exponent for the point-to-point memoryless channels. In [14], it is shown that in memoryless physically broadcast degraded channels, feedback cannot enlarge the capacity region. In [15], the scheme of Schalkwijk and Kailath has been extended to show that feedback can increase the capacity region of the single-antenna two-user broadcast channel, but just boundedly. We note that the single-antenna Gaussian channel is statistically, but not physically, degraded. For outer-bound in [15], the original channel is improved by giving the received signal of one user to the other user as a genie and forming a physically degraded broadcast channel. Then as shown in [14] feedback is not helpful for the resultant degraded broadcast channel and thus the capacity region can be computed. In [16], a signaling scheme is proposed for a broadcast channel with two packet erasure receivers where the transmitter receives acknowledgement feedback from both receivers. The signaling scheme is as follows. The transmitter sends packages for each receiver separately. If a packet is received by the intended receiver, then no extra effort is needed for that packet. But if a packet is received by the non-intended receiver, that receiver keeps that packet for later coding opportunity as follows. Let say packet x_1 intended for user one is received by user two, and packet x_2 intended by user two is received by user one. In this case, the transmitter sends (x_1 XOR x_2). Then if user one receives it, it can recover x_1 by subtracting x_2, and if user two receives it, it can recover x_2 by subtracting x_1. The XOR scheme has been initially proposed back in 1993 in [17] for a broadcast channel with several packet erasure receivers and acknowledgement feedback, where all receivers requires all messages. In [18], the outer-bound of [15] is used to show that the scheme of [16] is optimal. In [19], the idea of [16] is extended to more than two users, when all users have identical erasure probability. What we are doing here in this paper has been motivated by the results of [16]–[19] for packet erasure channels. Moreover, these results have some connection with the concept of index coding as well.

Here we assume that the channel state information is changing fast over time. Indeed, the channel realization at each time is statistically independent from the channel realization at other times. In addition, the base station has channel information with some finite delays. Therefore, always the base station has outdated channel information. Still, we will show that we can achieve DoF more than one.

II. Problem Formulation

We consider a complex broadcast channel with M transmit antennas and K receivers, each equipped with a single antenna. In flat fading environment, this channel can be modeled as,

$$y^{[r]}(m) = \mathbf{h}^{[r]\dagger}(m)\mathbf{x}(m) + z^{[r]}(m), \quad r = 1, \ldots, K, \tag{1}$$

where $\mathbf{x}(m) \in \mathbb{C}^{M\times 1}$, $\mathbb{E}(\mathbf{x}^{\dagger}(m)\mathbf{x}(m)) \leq P$, $z_s(m) \sim \mathcal{C}(0,1)$ and the sequences $z^{[r]}(m)$'s are i.i.d. and mutually independent. In addition, $\mathbf{h}^{[r]\dagger}(m) = [h_1^{[r]}(m), \ldots, h_M^{[r]}(m)] \in \mathbb{C}^{1\times M}$. We define $\mathbf{H}(m)$ as $\mathbf{H}(m) = [\mathbf{h}^{[r]}(m), \ldots, \mathbf{h}^{[r]}(m)]$.

We assume that the channel state information of each receiver is available to that receiver at each time, but is available to the base station and other receivers with some finite unit delay. Without loss of generality, we assume that this delay is one unit.

Let us define $\mathcal{E}$ as $\mathcal{E} = \{1, 2, \ldots, K\}$. We assume that for any subset $\mathcal{S}$ of the users, $\mathcal{S} \subset \mathcal{E}$, the base station has a message $W^{[\mathcal{S}]}$ with rate $R^{[\mathcal{S}]}$. For example, message $W^{[\{1,2\}]}$, or simply $W^{[1,2]}$, is for users one and two. We define $d^{[\mathcal{S}]}$, as

$$d^{[\mathcal{S}]} = \lim_{P\to\infty} \frac{R^{[\mathcal{S}]}}{\log_2(P)}. \tag{2}$$

If $|\mathcal{S}| = q$, then we call $W^{[\mathcal{S}]}$ as a degree–q message or a message with degree q. We define q–degrees of freedom, $\overline{\text{DoF}}_q(M,K)$, as

$$\overline{\text{DoF}}_q(M,K) = \lim_{P\to\infty} \max_{\bar{R}\in\mathcal{C}(P)} \sum_{S,|S|\geq q} \frac{R^{[\mathcal{S}]}}{\log_2(P)}, \tag{3}$$

where $\mathcal{C}(P)$ denotes the capacity region of the channel, and $\bar{R} \in \mathbb{R}^{(2^K-1)\times 1}$ denotes the vector of the message rates for each subset of users. It is easy to see that

$$\overline{\text{DoF}}_q(M,K) = \lim_{P\to\infty} \max_{\bar{R}\in\mathcal{C}(P)} \sum_{S,|S|=q} \frac{R^{[\mathcal{S}]}}{\log_2(P)}. \tag{4}$$

We note that $\overline{\text{DoF}}_1(M,K)$ is the well-known notion as the DoF of the channel.

In this paper, we denote the achieved q–degrees of freedom by $\text{DoF}_q(M,K)$. Apparently, $\text{DoF}_q(M,K) \leq \overline{\text{DoF}}_q(M,K)$.

III. Contributions

Theorem 1 *If $M \geq K-j+1$, then $\overline{\text{DoF}}_j(M,K)$ is given by*

$$\frac{K-j+1}{j}\frac{1}{\overline{\text{DoF}}_j(M,K)} = \frac{1}{j}+\frac{1}{j+1}+\ldots+\frac{1}{K}. \quad (5)$$

In particular, if $M=K$, then

$$\overline{\text{DoF}}_1(K,K) = \frac{K}{1+\frac{1}{2}+\ldots+\frac{1}{K}}. \quad (6)$$

For example, $\overline{\text{DoF}}_1(2,2)=\frac{4}{3}$ and $\overline{\text{DoF}}_1(3,3)=\frac{18}{11}$, which are greater than one. Without channel feedback, the channel is degenerate and $\overline{\text{DoF}}_1(K,K)$ is just one. In addition, since the channel states are independent over time, then predicting the channel state and then using conventional schemes like zero-forcing is not an option. The scheme is based on going through some phases, where in each phase, some side-information is provided to the users, which are efficiently exploited in the future phases.

Theorem 2 *In the channel, modeled in Section II, DoF$_j(M,K)$, for $j=1,\ldots,K$, is achievable, where where DoF$_K(M,K)=1$ and*

$$\frac{q_j+1}{j}\frac{1}{\text{DoF}_j(M,K)} = \frac{1}{j}+\frac{q_j}{j+1}\frac{1}{\text{DoF}_{j+1}(M,K)}, \quad (7)$$

where $q_j=\min\{M-1,K-j\}$.

This theorem extends the achievable scheme of 1 to the cases, where $M<K-j+1$.

Theorem 3 *In broadcast channel, modeled in Section II, we have*

$$\frac{\binom{K-1}{j-1}}{\min\{1,M\}}+\frac{\binom{K-2}{j-1}}{\min\{2,M\}}+\ldots+\frac{\binom{j-1}{j-1}}{\min\{K-j+1,M\}} \geq \frac{\binom{n}{j}}{\overline{\text{DoF}}_j(M,K)} \quad (8)$$

This theorem provides an outer-bound for the DoF of the channel. This outer-bound is based on providing specific genie to some users and enhance the channel to a physically degraded broadcast channel. In physically degraded channel, feedback does not improve the capacity and therefore can be ignored. We show that this outer–bound is tight for the case where $M \geq K-j+1$.

IV. Achievable Scheme for Theorem 1

In this section, we explain the achievable scheme for the case that number of transmit antennas are the same with the number of users, i.e. $M=K$. For simplicity, we first explain the idea for two special cases: $M=K=2$ and $M=K=3$.

A. Achievable Scheme for $M=K=2$

In this subsection, we explain the achievable scheme for the case where $M=K=2$. In particular, we show that the DoF of $\frac{4}{3}$ is achievable. We explain the achievable scheme from three different perspectives:

1) Interference Alignment using Outdated CSIT
2) Exploiting Side-Information
3) Generating Higher-Degree Messages

*1) **Interference Alignment using Outdated CSIT**:* Let $u_1^{[r]}$ and $u_2^{[r]}$ be two independently encoded Gaussian codewords, each carries one degree of freedom, intended for user r, $r = 1, 2$. The proposed communication scheme is performed in two phases, which totally take three time slots as follows:

Phase One: Feeding the Receivers: This phase has two sub-phases, each sub-phase includes one time slot.

The first sub-phase, including just one time-slot, at $m = 1$, is dedicated to user one. At this sub-phase, the base station transmits the two data streams, $u_1^{[1]}$ and $u_2^{[1]}$, intended for user one, i.e.

$$\mathbf{x}(1) = \begin{bmatrix} u_1^{[1]} \\ u_2^{[1]} \end{bmatrix}. \tag{9}$$

At receivers, we have:

$$y^{[1]}(1) = h_1^{[1]}(1)u_1^{[1]} + h_2^{[1]}(1)u_2^{[1]} + z^{[1]}(1), \tag{10}$$

$$y^{[2]}(1) = h_1^{[2]}(1)u_1^{[1]} + h_2^{[2]}(1)u_2^{[1]} + z^{[2]}(1). \tag{11}$$

Therefore, user one receives a noisy version of a linear combination from desired signals, while user two overhears a noisy version of an equation from interference data streams $u_1^{[1]}$ and $u_2^{[1]}$. User two saves the overheard equation for later usage, although it only carries interference information.

The second sub-phase of phase one, which includes just time-slot, at $m = 2$, is dedicated to the second user. In this sub-phase, the base station transmits data streams intended for user two, i.e.

$$\mathbf{x}(2) = \begin{bmatrix} u_1^{[2]} \\ u_2^{[2]} \end{bmatrix}. \tag{12}$$

At receivers, we have:

$$y^{[1]}(2) = h_1^{[1]}(2)u_1^{[2]} + h_2^{[1]}(2)u_2^{[2]} + z^{[1]}(2), \tag{13}$$

$$y^{[2]}(2) = h_1^{[2]}(2)u_1^{[2]} + h_2^{[2]}(2)u_2^{[2]} + z^{[2]}(2). \tag{14}$$

Therefore, user two receives a noisy version of an equation from desired signals, while user one overhears an equation from interference data streams of user two. Again, user one saves overheard message for future usage.

Now we have a key observation: If user one has the overheard equation by user two, then it has enough equations to resolve its own messages. Similarly, if user two has the overheard equation by user one, then it has enough equations to resolve its own message. Therefore, the main mission of the second phase is to swap these two overheard equations through the base station.

Phase Two: Swapping Overheard Equations: This phase includes one sub-phase, which takes one time slot, at $m = 3$. In this time, the base station transmits a linear combination of the overheard equations. We note that at this time transmitter is aware of the channel state information at $m = 1$ and $m = 2$, therefore, it can from the overheard equations.

As a particular example, $\mathbf{x}(3)$ is formed as,

$$\mathbf{x}(3) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(h_1^{[2]}(1)u_1^{[1]} + h_2^{[2]}(1)u_2^{[1]} + h_1^{[1]}(2)u_1^{[2]} + h_2^{[1]}(2)u_2^{[2]} \right). \tag{15}$$

At receivers, we have,

$$y^{[1]}(3) = h_1^{[1]}(3) \left(h_1^{[2]}(1)u_1^{[1]} + h_2^{[2]}(1)u_2^{[1]} + h_1^{[1]}(2)u_1^{[2]} + h_2^{[1]}(2)u_2^{[2]} \right) + z^{[1]}(3), \tag{16}$$

$$y^{[2]}(3) = h_1^{[2]}(3) \left(h_1^{[2]}(1)u_1^{[1]} + h_2^{[2]}(1)u_2^{[1]} + h_1^{[1]}(2)u_1^{[2]} + h_2^{[1]}(2)u_2^{[2]} \right) + z^{[2]}(3). \tag{17}$$

Putting together the symbols received by user one over the three time slots, we have,

$$\begin{bmatrix} y^{[1]}(1) \\ y^{[1]}(2) \\ y^{[1]}(3) \end{bmatrix} = \underbrace{\begin{bmatrix} h_1^{[1]}(1) & h_2^{[1]}(1) \\ 0 & 0 \\ h_1^{[1]}(3)h_1^{[2]}(1) & h_1^{[1]}(3)h_2^{[2]}(1) \end{bmatrix}}_{\text{Rank Two}} \begin{bmatrix} u_1^{[2]} \\ u_2^{[1]} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 \\ h_1^{[1]}(2) & h_2^{[1]}(2) \\ h_1^{[1]}(3)h_1^{[1]}(2) & h_1^{[1]}(3)h_2^{[1]}(2) \end{bmatrix}}_{\text{Rank One}} \begin{bmatrix} u_1^{[1]} \\ u_2^{[2]} \end{bmatrix} + \begin{bmatrix} z^{[1]}(1) \\ z^{[1]}(2) \\ z^{[1]}(3) \end{bmatrix}. \tag{18}$$

Then it is easy to see that at user one, the two interference streams $u_1^{[2]}$ and $u_2^{[2]}$ arrived at the same directions, and therefore $u_1^{[2]}$ and $u_2^{[2]}$ are aligned. It means that from the first user's perspective, the two variables $u_1^{[2]}$ and $u_2^{[2]}$ collapse to one variable, which is $h_1^{[1]}u_1^{[2]} + h_1^{[2]}u_2^{[2]}$. Eliminating variable $h_1^{[1]}u_1^{[2]} + h_1^{[2]}u_2^{[2]}$ from (18), we have,

$$\begin{bmatrix} y^{[1]}(1) \\ y^{[1]}(3) - h_1^{[1]}(3)y^{[1]}(2) \end{bmatrix} = \begin{bmatrix} h_1^{[1]}(1) & h_2^{[1]}(1) \\ h_1^{[1]}(3)h_1^{[2]}(1) & h_1^{[1]}(3)h_2^{[2]}(1) \end{bmatrix} \begin{bmatrix} u_1^{[1]} \\ u_2^{[1]} \end{bmatrix} + \begin{bmatrix} z^{[1]}(1) \\ z^{[1]}(3) - h_1^{[1]}(3)z^{[1]}(2) \end{bmatrix}. \tag{19}$$

It is easy to see that as long as $h_1^{[1]}(3) \neq 0$ and $h_1^{[1]}(1)h_2^{[2]}(1) - h_2^{[1]}(1)h_1^{[2]}(1) \neq 0$, then the desired data streams are not aligned at receiver one and can support the two desired data streams. We note that at $h_1^{[1]}(1)h_2^{[2]}(1) - h_2^{[1]}(1)h_1^{[2]}(1)$ is the determinant of the channel matrix $\mathbf{H}(1)$. Indeed, in this scheme, user one borrows the antenna of the second user at time slot $m = 1$ to be able to support two data streams.

Similarly, we have

$$\begin{bmatrix} y^{[2]}(1) \\ y^{[2]}(2) \\ y^{[2]}(3) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ h_1^{[2]}(2) & h_2^{[2]}(2) \\ h_1^{[2]}(3)h_1^{[1]}(2) & h_1^{[2]}(3)h_2^{[1]}(2) \end{bmatrix}}_{\text{Rank Two}} \begin{bmatrix} u_1^{[2]} \\ u_2^{[2]} \end{bmatrix} + \underbrace{\begin{bmatrix} h_1^{[2]}(1) & h_2^{[2]}(1) \\ 0 & 0 \\ h_1^{[2]}(3)h_1^{[2]}(1) & h_1^{[2]}(3)h_2^{[2]}(1) \end{bmatrix}}_{\text{Rank One}} \begin{bmatrix} u_1^{[1]} \\ u_2^{[1]} \end{bmatrix} + \begin{bmatrix} z^{[2]}(1) \\ z^{[2]}(2) \\ z^{[2]}(3) \end{bmatrix}. \tag{20}$$

Again at user two, the two interference streams $u_1^{[1]}$ and $u_2^{[1]}$ arrived at the same directions, and therefore are aligned. Then, we have,

$$\begin{bmatrix} y^{[2]}(2) \\ y^{[2]}(3) - h_1^{[2]}(3)y^{[1]}(1) \end{bmatrix} = \begin{bmatrix} h_1^{[2]}(2) & h_2^{[2]}(2) \\ h_1^{[2]}(3)h_1^{[1]}(2) & h_1^{[2]}(3)h_2^{[1]}(2) \end{bmatrix} \begin{bmatrix} u_1^{[2]} \\ u_2^{[2]} \end{bmatrix} + \begin{bmatrix} z^{[2]}(1) \\ z^{[2]}(3) - h_1^{[2]}(3)z^{[1]}(1) \end{bmatrix}. \tag{21}$$

Similarly, as long as $h_1^{[2]}(3) \neq 0$ and $\det(\mathbf{H}(2)) \neq 0$, then the desired data streams are not aligned at receiver two and can be resolved. Indeed, in this scheme, user two borrows the antenna of the first user at time slot $m = 2$ to be able to support two data streams.

*2) **Exploiting Side-Information**:* So far, we have investigated the proposed solution as an alignment technique using outdated CSI. Now, we revisit the problem as a side information problem.

As explained above, the algorithm has two phases, where phase one has two sub-phases and each sub-phase dedicated to one of the users and takes only one time slot. For simplicity let us define some short-hand notations. In phase p, we denote the sub-phase dedicated to the set of users $\mathcal{S}$ as S-Ph$(p;\mathcal{S})$. For example, the sub-phase of phase one dedicated to user one is denoted by S-Ph$(1;\{1\})$ or simply S-Ph$(1;1)$. In addition, we define $L_t^{[r]}(p;\mathcal{S})$ as the linear combination of transmitted signal received by user r in the t^{th} time slot of the sub-phase S-Ph$(p;\mathcal{S})$. If a sub-phase has only one time slot, then we drop t for simplicity and use $L^{[r]}(p;\mathcal{S})$. For example, the linear combination received by user 2 in the only time slot of S-Ph$(1;1)$ is denoted by $L_1^{[2]}(1;\{1\})$ or simply $L^{[2]}(1;\{1\})$, or even in a more simpler form of $L^{[2]}(1;1)$.

TABLE I

SIGNALING SCHEME FOR $M = K = 2$

Phase	1 (Feeding the Receivers)		2 (Swapping Information)
Sub-Phase	S-Ph(1;1)	S-Ph(1;2)	S-Ph(2; 1 ,2)
m	1	2	3
Tx 1	$u_1^{[1]}$	$u_1^{[2]}$	$L^{[2]}(1;1) + L^{[1]}(1;2)$
Tx 2	$u_2^{[1]}$	$u_2^{[2]}$	0
$y^{[1]}(m) - z^{[1]}(m)$	$L^{[1]}(1;1)$	$L^{[1]}(1;2)$	$L^{[1]}(2;1,2) = h_1^{[1]}(3)(L^{[2]}(1;1) + L^{[1]}(1;2))$
$y^{[2]}(m) - z^{[2]}(m)$	$L^{[2]}(1;1)$	$L^{[2]}(1;2)$	$L^{[2]}(2;1,2) = h_1^{[2]}(3)(L^{[2]}(1;1) + L^{[1]}(1;2))$

The transmission scheme has been summarized in Table I. We note that if $\mathbf{H}(1)$ is full rank, then there is a one-to-one map between $(u_1^{[1]}, u_2^{[1]})$ and $(L^{[1]}(1;1), L^{[2]}(1;1))$. Similarly, if $\mathbf{H}(2)$ is full rank, then there is a one-to-one map between $(u_1^{[2]}, u_2^{[2]})$ and $(L^{[1]}(1;2), L^{[2]}(1;2))$. Therefore, one can call $(L^{[1]}(1;1), L^{[2]}(1;1))$ as the symbols desired by user one. Similarly, one can say that user two requires to resolve $(L^{[1]}(1;2), L^{[2]}(1;2))$, instead of $u_1^{[2]}$ and $u_2^{[2]}$. By the end of the first phase, user one has (noisy version of) desired signal $L^{[1]}(1;1)$, while user two has the (noisy version of) desired signal $L^{[2]}(1;2)$. Meanwhile, user one overhears (a noisy version of) $L^{[1]}(1;2)$, as interference, while user two requires it. Similarly user two overhears (a noisy version of) $L^{[2]}(1;1)$, as interference, while user one needs it. In the second phase, users one and two exploit the availability of side information to swap $L^{[1]}(1;2)$ and $L^{[2]}(1;1)$ in just one time slot. In this phase, base station sends $L^{[1]}(1;2) + L^{[2]}(1;1)$. User one already has a noisy version of $L^{[1]}(1;2)$, therefore, it can cancel out the contribution of $L^{[1]}(1;2)$ from $y^{[1]}(1;3)$, and provide a noisy version of $L^{[2]}(1;1)$. Therefore, it has $L^{[1;1]}(1) + z^{[1]}(1)$ from the first time-slot and $h_1^{[1]}(3)L^{[2]}(1;1) + z^{[1]}(3) - h_1^{[1]}(1)z^{[2]}(1)$ from the last time-slot. We have similar situation for user two.

Remark: In this scheme, we assume that in the first time slot, transmit antenna one sends $u_1^{[1]}$ and transmit antenna two sends $u_2^{[1]}$. However, antenna one and two can sends any random linear combination of $u_1^{[1]}$ and $u_2^{[1]}$. Therefore, for example, we can have

$$\mathbf{x}(1) = \mathbf{A}(1) \begin{bmatrix} u_1^{[1]} \\ u_2^{[1]} \end{bmatrix}, \tag{22}$$

where $\mathbf{A}(1) \in \mathbb{C}^{2\times 2}$ is randomly selected matrices. Similar statement is true for the second time slots. At time slot $m = 3$, we send $L^{[1]}(1;2) + L^{[2]}(1;1)$. However, we can send any combination of combination $L^{[1]}(1;2)$ and $L^{[2]}(1;1)$. In other words,

$$\mathbf{x}(3) = \mathbf{A}(3) \begin{bmatrix} L^{[1]}(1;2) \\ L^{[2]}(1;1) \end{bmatrix}, \tag{23}$$

where $\mathbf{A}(3) \in \mathbb{C}^{2\times 2}$ is randomly selected matrices. However, we can limit the choice of $\mathbf{A}(3)$ to rank one matrices, while rank two matrices also work.

Remark: We note that only the number of independent noisy equations that each user has is important. As long as the variance of the noise of each equation is upper bounded and greater that a constant positive number, it does not affect DoF. Therefore, in what follows, we ignore noise and try to count the number of independent equations.

3) ***Generating Higher Degree Messages****:* We can observe the achievable scheme from another perspective. Remember in the second phase, we send a linear combination of $L^{[1]}(2;2)$ and $L^{[2]}(1;1)$, e.g. $L^{[1]}(2;2) + L^{[2]}(1;1)$, to both users. Indeed, we can consider $L^{[1]}(2;2) + L^{[2]}(1;1)$ as a degree two symbols, required by both receivers. Let us define $u^{[1,2]}$ as $u^{[1,2]} = L^{[1]}(2;2) + L^{[2]}(1;1)$. Clearly, we need $\frac{1}{\mathrm{DoF}_2(2,2)}$ time slots to deliver $u^{[1,2]}$ to both receivers. Therefore, in total, we need $2+\frac{1}{\mathrm{DoF}_2(2,1)}$ to deliver four data streams $u_1^{[1]}$, $u_2^{[1]}$, $u_1^{[2]}$, and $u_2^{[2]}$ to the designated receivers. Thus, we have,

$$\mathrm{DoF}_1(2,2)=\frac{4}{2+\frac{1}{\mathrm{DoF}_2(2,2)}}, \tag{24}$$

or

$$\frac{2}{\mathrm{DoF}_1(2,2)}=1+\frac{1}{\mathrm{DoF}_2(2,2)}. \tag{25}$$

It is easy to see that we can achieve $\mathrm{DoF}_2(2,2)=1$ by simply sending $u^{[1,2]}$ to both receivers. Therefore, $\mathrm{DoF}_1(2,2)=\frac{4}{3}$ is achievable.

Indeed, Phase One deals with two messages with degree one for each receiver. It takes two time slots to deliver one desired equation to each of the receivers. Therefore, each receiver needs one more equation to resolve the desired signals. If we ignore the overheard equations, we need two more time slots to deliver one more equation to each receiver and yield the DoF of one. However, by exploiting the overheard equations, we can form a symbol with degree two. Delivering one symbol with degree two to both receivers takes only one time slot, however, it provides one useful equation to each of the users. Therefore using this scheme, we save one time slot and achieve $\mathrm{DoF}_1(2,2)=\frac{4}{3}$ rather than $\frac{4}{4}$.

B. Three Transmit antennas, Three Receivers

In this section, we show how we achieve DoF of $\frac{3}{1+\frac{1}{2}+\frac{1}{3}}=\frac{18}{11}$ for the channel with three-antenna base station and three single-antenna receivers. As explained in the previous subsection, we can observe the achievable scheme in three different perspective. However, we found the last perspective simpler to follow. Therefore, in the rest of the paper, we just explain the algorithm based on the third perspective.

The achievable scheme has three phases. Phase One takes messages with degree one and generates degree two messages. Phase Two takes degree two symbol and generates degree three messages. The last phase takes degree three messages and deliver them to all three users.

Phase One:

Phase one is based a sub-algorithm with takes three messages for each user and generate three symbols with degree two. Assume that $u_1^{[r]}$, $u_2^{[r]}$, and $u_3^{[r]}$ represent three data streams, independently Gaussian encoded, each carries one DoF for user r, $r=1,2,3$. Therefore, in total, there are 9 data streams. The sub-algorithm has three sub-phases, where each sub-phase takes only one time slot and is dedicated to one of the users. In S-Ph$(1;1)$, the base station sends random linear combinations of $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$ over the three antennas. Similarly, in S-Ph$(1;2)$, the base station sends random linear combinations of $u_1^{[2]}$, $u_2^{[2]}$, and $u_3^{[2]}$ over the three antennas. In S-Ph$(1;3)$, the base station sends random linear combinations of $u_1^{[3]}$, $u_2^{[3]}$, and $u_3^{[3]}$ over the three antennas. In Table II, this phase has been detailed, where to be specific, we assume that in S-Ph$(1;r)$, transmit antenna i transmits $u_i^{[r]}$.

So far the algorithm has taken three time slots and delivered three desired equations to the designated receivers. Therefore, in terms of counting the desired equations, the algorithm delivers one equation per time slot which is natural progress. If we ignore the overheard equations, then we need six more time slots to deliver the messages, which yields DoF of one. However, as seen in the previous example, the overheard equations can help us to improve the degrees of freedom.

Let us focus on the sub-phase dedicated to user one, i.e. S-Ph$(1;1)$. Then, we have the following important observations:

- The three equations $L^{[1]}(1;1)$, $L^{[2]}(1;1)$, and $L^{[3]}(1;1)$ forms three linearly independent equations of $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$, almost surely. In the particular example of Table II, if $\mathbf{H}(1)$ is full rank, then $L^{[1]}(1;1)$, $L^{[2]}(1;1)$, and $L^{[3]}(1;1)$ are linearly independent.

TABLE II

SIGNALING SCHEME FOR $M = K = 3$, PHASE ONE

Phase	1 (Feeding the Receivers)		
Sub-Phase	S-Ph$(1;1)$	S-Ph$(1;2)$	S-Ph$(1;3)$
Tx 1	$u_1^{[1]}$	$u_1^{[2]}$	$u_1^{[3]}$
Tx 2	$u_2^{[1]}$	$u_2^{[2]}$	$u_2^{[3]}$
Tx 3	$u_3^{[1]}$	$u_3^{[2]}$	$u_3^{[3]}$
$y^{[1]}(m) - z^{[1]}(m)$	$L^{[1]}(1;1)$	$L^{[1]}(1;2)$	$L^{[1]}(1;3)$
$y^{[2]}(m) - z^{[2]}(m)$	$L^{[2]}(1;1)$	$L^{[2]}(1;2)$	$L^{[2]}(1;3)$
$y^{[3]}(m) - z^{[3]}(m)$	$L^{[3]}(1;1)$	$L^{[3]}(1;2)$	$L^{[3]}(1;3)$

- If we somehow deliver the overheard equations $L^{[2]}(1;1)$ and $L^{[3]}(1;1)$ to user one, then it has enough equations to resolve $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$.
- The two overheard equations $L^{[2]}(1;1)$ and $L^{[3]}(1;1)$ plus the equation received by user one i.e. $L^{[1]}(1;1)$, fully represent the original data streams. Therefore, the information required to be resolved is already available at the receivers' sides, but not exactly at the desired receiver.

We have similar observations about the received equations in S-Ph$(1;2)$ and S-Ph$(1;3)$. Remember that originally the objective was to deliver $u_1^{[r]}$, $u_2^{[r]}$, and $u_3^{[r]}$ to user r. After these three transmissions, we can redefine the objective. The new objective is to deliver (i) the overheard equations $L^{[2]}(1;1)$ and $L^{[3]}(1;1)$ to user one, (2) the overheard equations $L^{[1]}(1;2)$ and $L^{[3]}(1;2)$ to user two, and (3) the overheard equations $L^{[1]}(1;3)$ and $L^{[2]}(1;3)$ to user three.

Let us define $u^{[1,2]}$ as a random linear combination of $L^{[2]}(1;1)$ and $L^{[1]}(1;2)$. To be specific, let $u^{[1,2]} = L^{[2]}(1;1) + L^{[1]}(1;2)$. Then we have the following observations:

- If user one has $u^{[1,2]}$, then it can use the saved overheard equation $L^{[1]}(1;2)$ to form $u^{[1,2]} - L^{[1]}(1;2)$, which is $L^{[2]}(1;1)$. Remember $L^{[2]}(1;1)$ is a desired equation for user one.
- If user two has $u^{[1,2]}$, then it can used the saved overheard equation $L^{[2]}(1;1)$ to form $u^{[1,2]} - L^{[2]}(1;1)$, which is $L^{[1]}(1;2)$. Again remember $L^{[1]}(1;2)$ is a desired equation for user two.

Therefore, $u^{[1,2]}$ is desired by both users one and two. Similarly, we define $u^{[1,3]}$ as $u^{[1,3]} = L^{[3]}(1;1) + L^{[1]}(1;3)$, which is desired by user one and three. In addition, we define, $u^{[2,3]}$ as $u^{[2,3]} = L^{[3]}(1;2) + L^{[2]}(1;3)$, which is desired by users two and three. We note that if user one has $u^{[1,3]}$ and $u^{[1,3]}$, then it has enough equations to resolve the original data streams $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$. Similarly, it is enough that user two has $u^{[1,2]}$ and $u^{[2,3]}$, and user three has $u^{[1,3]}$ and $u^{[2,3]}$. Therefore, again, we can redefine the objective as delivering $u^{[1,2]}$ to users one and two, $u^{[1,3]}$ to users one and three, and $u^{[2,3]}$ to users two and three. We note that accomplishing this target, i.e. delivering these three degree-two messages takes $\frac{3}{\text{DoF}_2}$ time slots, whatever DoF_2 is. Recall that so far, the algorithm takes three time slots, and needs $\frac{3}{\text{DoF}_2}$ more time slots to deliver the original 9 degree one messages. Therefore, we have

$$\text{DoF}_1(3,3) = \frac{9}{3 + \frac{3}{\text{DoF}_2(3,3)}}, \tag{26}$$

or

$$\frac{3}{\text{DoF}_1(3,3)} = 1 + \frac{1}{\text{DoF}_2(3,3)}, \tag{27}$$

It is trivially easy to achieve $\text{DoF}_2(3,3)$ of one, and therefore achieve $\text{DoF}_1(3,3)$ of $\frac{3}{2}$ which is already greater than one. However, as we will show, we can do better.

Phase Two:

Phase One of the algorithm takes degree–one messages and generates some degree–two symbols to be delivered. Phase Two deals with degree–two symbols, and generates some degree three messages.

Assume that $u_1^{[1,2]}$ and $u_2^{[1,2]}$ represent two symbols that are desired by both users one and two. Similarly, $u_1^{[1,3]}$ and $u_2^{[1,3]}$ are required by both users one and three, and $u_1^{[2,3]}$ and $u_2^{[2,3]}$ are required by both users two and three. Therefore, in total, there are 6 degree–two symbols. We notice that Phase One generates only three degree–two symbols. Two provide 6 degree–two symbols, we can simply repeat Phase One twice. Phase Two has three sub-phases, where each sub-phase takes only one time slot and is dedicated to one pair of the users. More precisely, in S-Ph$(2;\mathcal{S})$, $|\mathcal{S}|=2$, the base station sends random linear combinations of $u_1^{[\mathcal{S}]}$ and $u_2^{[\mathcal{S}]}$ at least over two transmit antennas.

As an example, let us focus on in S-Ph$(2;1,2)$, in which base station sends random linear combinations of $u_1^{[1,2]}$ and $u_2^{[1,2]}$ using at least two of the transmit antennas. To be specific, as shown in Table III, we can assume that transmit antenna i sends $u_i^{[1,2]}$, for $i=1,2$. Again, $L^{[r]}(2;1,2)$ represent the linear combination of transmitted signals received by user r. Referring to Table III, we have the following important observations:

- $L^{[1]}(2;1,2)$ and $L^{[3]}(1;1,2)$ form two linearly independent equations of $u_1^{[1,2]}$ and $u_2^{[1,2]}$, almost surely.
- Similarly, $L^{[2]}(2;1,2)$ and $L^{[3]}(2;1,2)$ form two linearly independent equations of $u_1^{[1,2]}$ and $u_2^{[1,2]}$, almost surely.
- If $L^{[3]}(2;1,2)$ is somehow delivered to both users one and two, then both users have enough equations to resolve $u_1^{[1,2]}$ and $u_2^{[1,2]}$. Therefore, $L^{[3]}(2;1,2)$ which is overheard and saved by user three is simultaneity useful for users one and two.

We have similar observations about the received equations in S-Ph$(2;1,3)$ and S-Ph$(2;2,3)$. Therefore, after these three transmissions, we can redefine the objective of the rest of the algorithm as delivering (i) $L^{[3]}(2;1,2)$ to users one and two, (ii) $L^{[2]}(2;1,3)$ to users one and three, and (iii) $L^{[1]}(2;2,3)$ to users two and three.

Let us define $u_1^{[1,2,3]}$ and $u_2^{[1,2,3]}$ as two random linear combinations of $L^{[3]}(2;1,2)$ and $L^{[2]}(2;1,3)$, and $L^{[1]}(2;2,3)$. For example, let $u_1^{[1,2,3]}=L^{[3]}(2;1,2)+L^{[2]}(2;1,3)+L^{[1]}(2;2,3)$, and $u_2^{[1,2,3]}=L^{[3]}(2;1,2)+1.8L^{[2]}(2;1,3)+2.3L^{[1]}(2;2,3)$. Then, we have the following observations:

- If we somehow deliver $u_1^{[1,2,3]}$ and $u_2^{[1,2,3]}$ to user one, then it uses the saved overheard equation $L^{[1]}(2;2,3)$ to form $u_1^{[1,2,3]}-L^{[1]}(2,3)$ and $u_2^{[1,2,3]}-2.3L^{[1]}(2;2,3)$ as two linearly independent equations of $L^{[3]}(2;1,2)$ and $L^{[2]}(2;1,3)$. Resolving $L^{[3]}(2;1,2)$ and $L^{[2]}(2;1,3)$, user one has enough equations to derive $u_1^{[1,2]}$, $u_2^{[1,2]}$, $u_1^{[1,3]}$, and $u_2^{[1,3]}$.

We have similar observations about user two and three. Therefore, it is enough to deliver $u_1^{[1,2,3]}$ and $u_1^{[1,2,3]}$ to all three users. Delivering these two degree-three symbols to all three users takes $\frac{2}{\mathrm{DoF}_3(3,3)}$, whatever $\mathrm{DoF}_3(3,3)$ is. Recall that Phase Two deals with 6 degree-two messages, takes three time slots, and generates two degree-three messages, therefore, we have

$$\mathrm{DoF}_2(3,3)=\frac{6}{3+\frac{2}{\mathrm{DoF}_3(3,3)}}, \tag{28}$$

or

$$\frac{1}{\mathrm{DoF}_2(3,3)}=\frac{1}{2}+\frac{1}{3\mathrm{DoF}_3(3,3)}. \tag{29}$$

Phase Three:

Phase Three deals with degree three messages. This phase is very simple. Assume that $u^{[1,2,3]}$ is required by all three users. Then, base station can use only one transmit antenna and send $u^{[1,2,3]}$. All three receivers will receive a noisy version of

TABLE III
SIGNALING SCHEME FOR $M = K = 3$, PHASE TWO

Phase	2, (Swapping Side Information)		
Sub-Phase	S-Ph$(2;1,2)$	S-Ph$(2;1,3)$	S-Ph$(2;2,3)$
Tx 1	$u_1^{[1,2]}$	$u_1^{[1,3]}$	$u_1^{[2,3]}$
Tx 2	$u_2^{[1,2]}$	$u_2^{[1,3]}$	$u_2^{[2,3]}$
Tx 3	0	0	0
$y^{[1]}(m) - z^{[1]}(m)$	$L^{[1]}(2;1,2)$	$L^{[1]}(2;1,3)$	$L^{[1]}(2;2,3)$
$y^{[2]}(m) - z^{[2]}(m)$	$L^{[2]}(2;1,2)$	$L^{[2]}(2;1,3)$	$L^{[2]}(2;2,3)$
$y^{[3]}(m) - z^{[3]}(m)$	$L^{[3]}(2;1,2)$	$L^{[3]}(2;1,3)$	$L^{[3]}(2;2,3)$

$u^{[1,2,3]}$. Therefore, we use one time slot to send one degree-three symbols. Therefore,

$$\text{DoF}_3(3,3) = 1. \tag{30}$$

From (26), (28), and (30), we conclude that $\text{DoF}_1(3,3) = \frac{18}{11}$ and $\text{DoF}_2(3,3) = \frac{6}{5}$.

C. $M = k$ transmit antennas and $K = k$ Users

In this section, we explain the achievable scheme for the general case where $M = K = k$, and we prove the total DoF of $\frac{k}{1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{k}}$ is achievable. The algorithm is based on a concatenation of k sub-algorithms. Each sub-algorithm takes symbols with degree j and generates some symbols with degree $j+1$. We repeat this sub-algorithm for $j = 1$ to $j = k-1$. For $j = k$ the sub-algorithm is simple and generates no more messages.

The sub-algorithm takes $(k-j+1)\binom{k}{j}$ symbols with degree j, and gives $j\binom{k}{j+1}$ symbols with degree $j+1$. The sub-algorithm has $\binom{k}{j}$ sub-phases, where each sub-phase takes only one time-slot and is dedicated to a subset of users $\mathcal{S}$, $|\mathcal{S}| = j$. We denote the sub-phase dedicated to the subset $\mathcal{S}$ by S-Ph$(j;\mathcal{S})$. In S-Ph$(j;\mathcal{S})$, the base station sends random linear combinations of the $k-j+1$ symbols $u_1^{[\mathcal{S}]}, u_2^{[\mathcal{S}]}, \ldots, u_{k-j+1}^{[\mathcal{S}]}$, designated for all users in $\mathcal{S}$. The base station utilizes at least $M-j+1$ of the transmit antennas.

The linear combination of the transmitted symbols received by user r is denoted by $L^{[r]}(j;\mathcal{S})$. Let us focus on the linear combinations of the transmitted symbols, received by all users, in S-Ph$(j;\mathcal{S})$. We have the following observations:

- For every $r \in \mathcal{S}$, the $M-j+1$ equations including one equation $L^{[r]}(j;\mathcal{S})$ and the $M-j$ overheard equations $L^{[r']}(j;\mathcal{S})$, $r' \in \mathcal{E}\backslash\mathcal{S}$ are linearly independent equations of $M-j+1$ symbols $u_1^{[\mathcal{S}]}, u_2^{[\mathcal{S}]}, \ldots, u_{k-j+1}^{[\mathcal{S}]}$. This relies on the fact that the base station uses at least $j+1$ transmit antennas.
- For any r, $r \in \mathcal{S}$, if we somehow deliver the $M-j$ equations $L^{[r']}(j;\mathcal{S})$, $r' \in \mathcal{E}\backslash\mathcal{S}$ to user r, then user r has $k-j+1$ linearly independent equations to resolve all $k-j+1$ symbols $u_1^{[\mathcal{S}]}, u_2^{[\mathcal{S}]}, \ldots, u_{k-j+1}^{[\mathcal{S}]}$.
- Having the above two observations, we can say that the overheard equation by user r', $r' \in \mathcal{E}\backslash\mathcal{S}$ is simultaneously useful for all users in $\mathcal{S}$.

After repeating the above transmission scheme for all $\mathcal{S}$, where $\mathcal{S} \subset \mathcal{E}$ and $|\mathcal{S}| = j$, then we have another important observation. Consider a subset $\hat{\mathcal{S}}$ of the users, where $|\hat{\mathcal{S}}| = j+1$. Then each user r', $r' \in \hat{\mathcal{S}}$, has a overheard equation $L^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$, which is simultaneity useful for all users in $\hat{\mathcal{S}}\backslash\{r'\}$. We note that the base station is aware of these overheard equations. For every $\hat{\mathcal{S}} \subset \mathcal{E}$, $|\hat{\mathcal{S}}| = j+1$, the base station forms j random linear combinations of $L^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$, $r' \in \hat{\mathcal{S}}$, denoted by $u_1^{[\hat{\mathcal{S}}]}, u_2^{[\hat{\mathcal{S}}]}, \ldots, u_j^{[\hat{\mathcal{S}}]}$. We note that $u_\tau^{[\hat{\mathcal{S}}]}$, $1 \le \tau \le j$, is simultaneously useful for all users in $\hat{\mathcal{S}}$. Indeed, each r in

$\hat{\mathcal{S}}$, can subtract the contribution of $L^{[r]}(j;\hat{\mathcal{S}}\backslash\{r\})$ from $u_\tau^{[\hat{\mathcal{S}}]}$, $\tau=1,\ldots,j$, and form j linearly independent combinations of $L^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$. $r'\in\hat{\mathcal{S}}\backslash\{r\}$. Using the above procedure, the base stations forms $j\binom{k}{j+1}$ symbols with degree $j+1$. The important observation is that if these $j\binom{k}{j+1}$ symbols are delivered to the designated receivers, the each receivers will have enough equations to resolve all designated messages with degree j. Delivering $j\binom{k}{j+1}$ degree $j+1$ symbols takes $\frac{j\binom{k}{j+1}}{\mathrm{DoF}_{j+1}(K,K)}$. Since the sub-algorithm starts with $(k-j+1)\binom{k}{j}$ symbols with degree j, and takes $\binom{k}{j}$ time slots, and generates $j\binom{k}{j+1}$ symbols with degree $j+1$, then we have

$$\mathrm{DoF}_j(K,K)=\frac{(k-j+1)\binom{k}{j}}{\binom{k}{j}+\frac{j\binom{k}{j+1}}{\mathrm{DoF}_{j+1}(K,K)}},\tag{31}$$

or

$$\frac{k-j+1}{j}\frac{1}{\mathrm{DoF}_j(K,K)}=\frac{1}{j}+\frac{k-j}{j+1}\frac{1}{\mathrm{DoF}_{j+1}(K,K)},\tag{32}$$

It is also easy to see that $\mathrm{DoF}_k(K,K)=1$ is achievable. Therefore, we have

$$\frac{k-j+1}{j}\frac{1}{\mathrm{DoF}_j(K,K)}=\frac{1}{j}+\frac{1}{j+1}+\ldots+\frac{1}{k}.\tag{33}$$

In particular,

$$\frac{k}{\mathrm{DoF}_j(K,K)}=1+\frac{1}{2}+\ldots+\frac{1}{k}.\tag{34}$$

Therefore the achievablity of Theorem 1 is proved.

V. Outer-Bound

In this section, we aim to prove Theorem 3. In this theorem, we focus on the j–DoF of the channel. Therefore, we assume for every subset of users $\mathcal{S}$ with cardinality j, transmitter has a message $W^{[\mathcal{S}]}$, with rate $R^{[\mathcal{S}]}$ and degrees of freedom $d^{[\mathcal{S}]}$.

Remember in Section II, we assume that the channel state information of each user is available to that user causally at each time, while it is available to all other nodes with one delay. We call this channel as original channel and denote its capacity as $\mathcal{C}_{\text{Original}}$. As an outer-bound, we consider the capacity of a channel, referred as *improved channel one* with capacity region $\mathcal{C}^1_{\text{Improved}}$, in which the channel state information of each user is available to *all* receivers causally at each time. Therefore, at time m, user r has $(y_r(t),\mathbf{H}(t))$, $t=1,\ldots,m$, for any r, $1\le r\le K$. On the other hand, in the improved channel one, the base station has the channel state information and received signals with one delay. Therefore, at time m, the base station has $(y_1(t),\ldots,y_K(t),\mathbf{H}(t))$, $t=1,\ldots,m-1$. Obviously, $\mathcal{C}_{\text{Original}}\subset\mathcal{C}^1_{\text{Improved}}$. We upgrade the *improved channel one* even further as follows.

Consider π as a permutation of the set $\mathcal{S}=\{1,2,\ldots,K\}$. We form a K–user broadcast channel from *improved channel one*, by giving the output of the receiver $\pi(i)$ to users $\pi(j)$, $j=i+1,\ldots,K$, for all $i=1,\ldots,K-1$. Therefore, we have an upgraded broadcast channel, referred as *improved channel two* with K receivers as $\left[y_{\pi(1)}(m),\mathbf{H}(m)\right]$, $\left[y_{\pi(1)}(m),y_{\pi(2)}(m),\mathbf{H}(m)\right]$, …, $\left[y_{\pi(1)}(m),y_{\pi(2)}(m),\ldots,y_{\pi(K)}(m),\mathbf{H}(m)\right]$. We denote the capacity of the resultant channel as $\subset\mathcal{C}^2_{\text{Improved}}(\pi)$. Apparently, $\mathcal{C}_{\text{Original}}\subset\mathcal{C}^1_{\text{Improved}}\subset\mathcal{C}^2_{\text{Improved}}(\pi)$. Moreover, it is easy to see that the improved channel two is physically degraded.

In the improved channel two, consider message $W^{[\mathcal{S}]}$, which is required by all j users listed in $\mathcal{S}$. Let i^* be the smallest integer where $\pi(i^*)\in\mathcal{S}$. Then, due to the degradedness of the channel, if $W^{[\mathcal{S}]}$ is decoded by user $\pi(i^*)$, then it can be decoded by all other users in $\mathcal{S}$. Therefore, we can assume that $W^{[\mathcal{S}]}$ is just required by user $\pi(i^*)$. Using this argument, we can reduce the requirements to the following: user $\pi(1)$ requires all messages $W^{[\mathcal{S}]}$, where $\pi(1)\in\mathcal{S}$ and $\mathcal{S}\in\mathcal{E}$. Similarly, user $\pi(2)$ requires all messages $W^{[\mathcal{S}]}$, where $\pi(2)\in\mathcal{S}$ and $\mathcal{S}\subset\mathcal{E}\backslash\{\pi(1)\}$. We follow the same argument for all users.

According to [14], feedback does not improve the capacity of the physically degraded broadcast channels. Consequently, we focus on the capacity region of the improved channel two without feedback. On the other hand, for broadcast channels without feedback, the capacity region is only a function of marginal distributions. Therefore, we can ignoring the coupling between the receivers in the second improved channel. Thus, we have a broadcast channel where user $\pi(i)$ has i antennas, where the distributions of the channels between the transmitter and any of the receive antennas are identical. Moreover user $\pi(i)$ is interested in all messages $W^{[\mathcal{S}]}$, where $\pi(i) \in \mathcal{S}$, $|\mathcal{S}| = j$, and $\mathcal{S} \subset \mathcal{E} \backslash \{\pi(1), \pi(2), \ldots, \pi(i-1)\}$.

Therefore, according to [20], extended by [21], one can conclude that

$$\frac{1}{\min\{1,M\}} \sum_{\substack{|\mathcal{S}|=j \\ \mathcal{S}\subset\mathcal{E} \\ \pi(1)\in\mathcal{S}}} d^{[\mathcal{S}]} + \frac{1}{\min\{2,M\}} \sum_{\substack{|\mathcal{S}|=j \\ \mathcal{S}\subset\mathcal{E}\backslash\{\pi(1)\} \\ \pi(2)\in\mathcal{S}}} d^{[\mathcal{S}]} + \ldots + \frac{1}{\min\{K-j+1,M\}} \sum_{\substack{|\mathcal{S}|=j \\ \mathcal{S}\subset\mathcal{E}\backslash\{\pi(1),\ldots\pi(K-j)\} \\ \pi(K-j+1)\in\mathcal{S}}} d^{[\mathcal{S}]} \leq 1. \tag{35}$$

By applying the same procedure for any permutation of the set $\{1, 2, \ldots, K\}$ and add all of the $K!$ resulting inequalities, the theorem follows.

Corollary 1 *If $M \geq K - j + 1$, then*

$$\frac{K-j+1}{j\overline{\text{DoF}}_j(M,K)} \geq \frac{1}{j} + \frac{1}{j+1} + \ldots + \frac{1}{K}, \tag{36}$$

no matter how large M is.

The proof follows from Theorem 3. This corollary provides the converse for Theorem 3.

VI. Achievable Scheme for Theorem 2

In Section IV, we explain the achievable scheme for $\overline{\text{DoF}}_1(M, K)$, when $M \geq K$. More generally, we derived $\overline{\text{DoF}}_1(M, K)$, when $M \geq K - j + 1$. In this section, we first explain why the achievable scheme of Section IV does not work for degree-one messages where where $M < K$. In general, the scheme of Section IV fails for degree–j messages, where $M < K - j + 1$. Then, we show how we can extend that approach to these cases. We first focus on the case where $M = 2$ and $K = 3$.

A. $M = 2, K = 3$

From Theorem 1, we know $\overline{\text{DoF}}_2(2, 3) = \frac{6}{5}$ and $\overline{\text{DoF}}_2(2, 3) = 1$. However, we do not know what $\overline{\text{DoF}}_1(2, 3)$ is. Clearly, $\text{DoF}_2(2, 3) \geq \overline{\text{DoF}}_1(2, 2) = \frac{4}{3}$, which can be simply achieved by ignoring one the receivers. The question is if we can achieve more than $\frac{4}{3}$. Note that from the outer–bound, we have $\text{DoF}_2(2, 3) \leq \frac{3}{2}$.

First let us try the scheme of Section IV. In the first phase, we have three sub-phases, where each sub-phase is dedicated to one of the receivers. In S-Ph$(1; 1)$, the base station sends linear combinations of three symbols $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$ over the two transmit antennas. Then, we observe that

- The three equations $L^{[1]}(1; 1)$, $L^{[2]}(1; 1)$, and $L^{[3]}(1; 1)$ are not linearly independent. The reason is that we have only two transmit antennas.
- Even if we somehow deliver the overheard equations $L^{[2]}(1; 1)$ and $L^{[3]}(1; 1)$ to user one, then it does not have enough equations to resolve $u_1^{[1]}$, $u_2^{[1]}$, and $u_3^{[1]}$.

These observations shows when $M < K$, then the scheme Section IV for degree–one messages fails. Similarly, if $M < K-j+1$, then that scheme fails for degree–j messages. Now let us modify that scheme by sending linear combinations of just two symbols $u_1^{[1]}$ and $u_2^{[1]}$ in S-Ph$(1; 1)$. Then, in this case, we have the following observations:

- The equations $L^{[1]}(1;1)$ and $L^{[2]}(1;1)$ forms two linearly independent equations of $u_1^{[1]}$ and $u_2^{[1]}$. This is also true for two equations of $L^{[1]}(1;1)$ and $L^{[3]}(1;1)$.
- If we somehow deliver one the the overheard equations S-Ph$(1;1)$, e.g. $L^{[3]}(1;1)$, to user one, then it has enough equations to resolve $u_1^{[1]}$ and $u_2^{[1]}$.

Therefore it seems that, from the two overheard equations $L^{[2]}(1;1)$ and $L^{[3]}(1;1)$, only one of them is useful for user one. From this observation, one may suggest that we should ignore one of the users, and apply the achievable scheme for a system with two transmit antennas and two users and achieve $\text{DoF}_1(2,2)$ of $\frac{4}{3}$. Remember that if we have three transmit antennas and two users the optimal $\text{DoF}_1(3,2)$ is the same as $\text{DoF}_1(2,2) = \frac{4}{3}$, thus, the extra transmit antenna does not improve DoF. On the other hand, we know that in conventional MIMO broadcast channel with perfect CSIR and CSIT, the DoF is determined by $\min\{M,K\}$. Therefore, we may suggest that one extra user does not improve the DoF for channel with delayed feedback, i.e. $\overline{\text{DoF}}_1(2,3)$ is again $\frac{4}{3}$. Here, we show that this suggestion is not true. In other words, unlike MIMO broadcast channel with perfect CSIR and CSIT, the DoF of the MIMO broadcast channel with delayed feedback is not a function of $\min\{M,K\}$. The whole point is that we may not be able to exploit the extra user in the first phase, however we can enjoy the extra user for the next phases.

The achievable scheme is as follows. Again there are thee sub-phases, where each sub-phase is dedicated to one of the users, however each sub-phase includes two time-slots. In S-Ph$(1;r)$, base stations sends random linear combinations of four symbols $u_1^{[r]}$, $u_2^{[r]}$, $u_3^{[r]}$, and, $u_4^{[r]}$. One particular choice has been shown in Table IV. Let us focus on S-Ph$(1;1)$. Then, we have the following observations:

- User one already has two independent linear equations of $u_1^{[1]}$, $u_2^{[1]}$, $u_3^{[1]}$, and, $u_4^{[1]}$. Therefore, it needs two more equations.
- The four overheard equations in S-Ph(1), i.e. $L_1^{[2]}(1;1)$, $L_2^{[2]}(1;1)$, $L_1^{[3]}(1;1)$, and $L_2^{[3]}(1;1)$ are not linearly independent from what user one has already received, i.e. $L_1^{[1]}(1;1)$ and $L_2^{[1]}(1;1)$.
- We can purify the four overheard equations in S-Ph$(1;1)$ and form two equations that are linearly independent with $L_1^{[1]}(1;1)$ and $L_2^{[1]}(1;1)$. For example, user two can form $\hat{L}^{[2]}(1;1)$ as a random linear combination of $L_1^{[2]}(1;1)$ and $L_2^{[2]}(1;1)$. Similarly, user two can form $\hat{L}^{[3]}(1;1)$ as a random linear combination of $L_1^{[3]}(1;1)$ and $L_2^{[3]}(1;1)$. Refer to Table V.
- If somehow deliver $\hat{L}^{[2]}(1;1)$ and $\hat{L}^{[3]}(1;1)$ to user one, then it has enough equations to resolve $u_1^{[1]}$, $u_2^{[1]}$, $u_3^{[1]}$, and, $u_4^{[1]}$.

Similarly, we can purify the overheard equations in S-Ph$(1;2)$ and S-Ph$(1;3)$. The rest of the algorithm is the same as Phase One of the case where we have three transmit antennas and three users. We define $u^{[1,2]}$ as a random linear combination of $\hat{L}^{[2]}(1;1)$ and $\hat{L}^{[1]}(1;2)$. Similarly, we define $u^{[1,3]}$ as a random linear combination of $\hat{L}^{[3]}(1;1)$ and $\hat{L}^{[1]}(1;3)$, and also $u^{[2,3]}$ as a random linear combination of $\hat{L}^{[3]}(1;2)$ and $\hat{L}^{[2]}(1;3)$. Therefore, Phase One starts with 12 degree-one messages, takes 6 time slots, and generates 3 degree-two symbols. Therefore, we have

$$\text{DoF}_1(2,3) = \frac{12}{6 + \frac{3}{\text{DoF}_2(2,3)}}, \tag{37}$$

or

$$\frac{4}{\text{DoF}_1(2,3)} = 2 + \frac{1}{\text{DoF}_2(3,3)}. \tag{38}$$

We can use the algorithm presented in Section IV to deliver degree–two symbols. Remember that the optimal $\overline{\text{DoF}}_2(2,3)$ is $\frac{6}{5}$. Therefore, we can achieve $\text{DoF}_1(2,3) = \frac{24}{17}$, which is greater than $\text{DoF}_1(2,2) = \frac{4}{3}$. Therefore, the extra user still improves

TABLE IV

SIGNALING SCHEME FOR $M = 2$, $K = 3$, PHASE ONE

Phase	1 (Feeding the Receivers)					
Sub-Phase	S-Ph(1)		S-Ph(2)		S-Ph(3)	
Tx 1	$u_1^{[1]}$	$u_3^{[1]}$	$u_1^{[2]}$	$u_4^{[2]}$	$u_1^{[3]}$	$u_3^{[3]}$
Tx 2	$u_2^{[1]}$	$u_4^{[1]}$	$u_2^{[2]}$	$u_4^{[2]}$	$u_2^{[3]}$	$u_4^{[3]}$
$y^{[1]}(m) - z^{[1]}(m)$	$L_1^{[1]}(1;1)$	$L_2^{[1]}(1;1)$	$L_1^{[1]}(1;2)$	$L_2^{[1]}(1;2)$	$L_1^{[1]}(1;3)$	$L_2^{[1]}(1;3)$
$y^{[2]}(m) - z^{[2]}(m)$	$L_1^{[2]}(1;1)$	$L_2^{[2]}(1;1)$	$L_1^{[2]}(1;2)$	$L_2^{[2]}(1;2)$	$L_1^{[2]}(1;3)$	$L_2^{[2]}(1;3)$
$y^{[3]}(m) - z^{[3]}(m)$	$L_1^{[3]}(1;1)$	$L_2^{[3]}(1;1)$	$L_1^{[3]}(1;2)$	$L_2^{[3]}(1;2)$	$L_1^{[3]}(1;3)$	$L_2^{[3]}(1;3)$

TABLE V

SIGNALING SCHEME FOR $M = 2$, $K = 3$, PHASE ONE, PURIFIED OVERHEARD EQUATIONS

<table>
<tr><td>Phase</td><td colspan="6">1 (Feeding the Receivers)</td></tr>
<tr><td>Sub-Phase</td><td colspan="2">S-Ph$(1;1)$</td><td colspan="2">S-Ph$(1;2)$</td><td colspan="2">S-Ph$(1;3)$</td></tr>
<tr><td>Tx 1
Tx 2</td><td>$u_1^{[1]}$
$u_2^{[1]}$</td><td>$u_3^{[1]}$
$u_4^{[1]}$</td><td>$u_1^{[2]}$
$u_2^{[2]}$</td><td>$u_3^{[2]}$
$u_4^{[2]}$</td><td>$u_1^{[3]}$
$u_2^{[3]}$</td><td>$u_3^{[3]}$
$u_4^{[3]}$</td></tr>
<tr><td>$y^{[1]}(m) - z^{[1]}(m)$</td><td>$L_1^{[1]}(1;1)$</td><td>$L_2^{[1]}(1;1)$</td><td colspan="2">$\hat{L}^{[1]}(1;2) = 1.2L_1^{[1]}(1;2) + 1.8L_2^{[1]}(1;2)$</td><td colspan="2">$\hat{L}^{[1]}(1;3) = -1.3L_1^{[1]}(1;3) + 0.5L_2^{[1]}(1;3)$</td></tr>
<tr><td>$y^{[2]}(m) - z^{[2]}(m)$</td><td colspan="2">$\hat{L}^{[2]}(1;1) = -L_1^{[2]}(1;1) + 3.3L_2^{[2]}(1;1)$</td><td>$L_1^{[2]}(1;2)$</td><td>$L_2^{[2]}(1;2)$</td><td colspan="2">$\hat{L}^{[2]}(1;3) = -3L_1^{[2]}(1;3) + 4.1L_2^{[2]}(1;3)$</td></tr>
<tr><td>$y^{[3]}(m) - z^{[3]}(m)$</td><td colspan="2">$\hat{L}^{[3]}(1;1) = 0.4L_1^{[3]}(1;1) + 0.8L_2^{[3]}(1;1)$</td><td colspan="2">$\hat{L}^{[3]}(1;2) = L_1^{[3]}(1;2) + L_2^{[3]}(1;2)$</td><td>$L_1^{[3]}(1;3)$</td><td>$L_2^{[3]}(1;3)$</td></tr>
</table>

the DoF. This is in contrary with the case where there extra transmit antenna with respect to the number users is not useful in terms of DoF. However, we notice that $\text{DoF}_1(2,3) = \frac{24}{17}$ is less that $\frac{3}{2} = \frac{24}{16}$ which is suggested by the outer–bound.

B. Proof of Theorem 2

This achievable scheme is based on a sub-algorithm which takes messages with degree j, and gives messages with degree $j+1$.

Let us define q_j as $q_j = \min\{M-1, K-j\}$. In addition, we define α_j as the largest common factor of q_j and $K-j$. The sub-algorithm has $\binom{K}{j}$ sub-phases, where every subset $\mathcal{S}$ of the users with $|\mathcal{S}| = j$ has a dedicated sub-phase, denoted by denoted by S-Ph$(j;\mathcal{S})$. Each sub-phase takes $\frac{K-j}{\alpha_j}$ time-slots. In S-Ph$(j;\mathcal{S})$, the base station sends random linear combinations of $\beta_j = \frac{(q_j+1)(K-j)}{\alpha_j}$ symbols $u_1^{[\mathcal{S}]}, u_2^{[\mathcal{S}]}, \ldots, u_{\beta_j}^{[\mathcal{S}]}$, designated for all users in $\mathcal{S}$. The base station uses at least q_j+1 of the transmit antennas. The linear combination of the transmitted symbols received by user r, in the t-th time slot of S-Ph$(j;\mathcal{S})$, is denoted by $L_t^{[r]}(j;\mathcal{S})$. Focusing on a particular subset of users $\mathcal{S}$, and the corresponding sub-phase S-Ph$(j;\mathcal{S})$, we have the following observations:

- For every r, $r \in \mathcal{S}$, and t, $t \in \{1, 2, \ldots, \frac{K-j}{\alpha_j}\}$, the $K-j+1$ equations $L_t^{[r']}(j;\mathcal{S})$, $r' \in \{r\} \cup \mathcal{E}\backslash\mathcal{S}$ are NOT linearly independent. The reason is that $|\{r\} \cup \mathcal{E}\backslash\mathcal{S}| = K-j+1$, while the number of transmit antennas M is less than $K-j+1$. Indeed, among the $K-j$ overheard equations $L_t^{[r']}(j;\mathcal{S})$, $r' \in \{r\} \cup \mathcal{E}\backslash\mathcal{S}$, we can only find q_j equations that are simultaneously useful to user r, for any r in $\mathcal{S}$. Therefore, the $\frac{(K-j)^2}{\alpha}$ overheard equations in S-Ph$(j;\mathcal{S})$, are representing only $\frac{q_j(K-j)}{\alpha}$ equations that are useful for any user r, $r \in \mathcal{S}$.
- Therefore, we need to purify the overheard equations. To this end, each user r', $r' \in \mathcal{E}\backslash\mathcal{S}$, forms $\frac{q_j}{\alpha_j}$ random linear combinations of $L_t^{[r']}(j;\mathcal{S})$, $t = 1, \ldots, \frac{K-j}{\alpha}$. The resultant equations are denoted by $\hat{L}_1^{[r']}(j;\mathcal{S}), \hat{L}_2^{[r']}(j;\mathcal{S}), \ldots, \hat{L}_{\frac{q_j}{\alpha_j}}^{[r']}(j;\mathcal{S})$. It is easy to see that for every r, the following $\frac{(q_j+1)(K-j)}{\alpha_j}$ equations are linearly independent: $L_t^{[r]}(j;\mathcal{S})$, $t = 1, \ldots, \frac{K-j}{\alpha_j}$, and $\hat{L}_{\hat{t}}^{[r']}(j;\mathcal{S})$, $r' \in \mathcal{E}\backslash\mathcal{S}$ and $\hat{t} \in 1, \ldots, \frac{q_j}{\alpha_j}$. Therefore, if we give $\hat{L}_{\hat{t}}^{[r']}(j;\mathcal{S})$, $r' \in \mathcal{E}\backslash\mathcal{S}$ and $\hat{t} \in 1, \ldots, \frac{q_j}{\alpha_j}$ to user r, $r \in \mathcal{S}$, then it will have $\beta_j = \frac{(q_j+1)(K-j)}{\alpha_j}$ linearly independent equations to resolve all desired variables $u_1^{[\mathcal{S}]}, u_2^{[\mathcal{S}]}, \ldots, u_{\beta_j}^{[\mathcal{S}]}$. In other words, we have adequacy of overheard equations.

- The purified overheard equations by user r', $r' \in \mathcal{E}\backslash\mathcal{S}$, are simultaneously useful for all users in $\mathcal{S}$.

After repeating the above transmission for all $\mathcal{S}$, $\mathcal{S} \subset \mathcal{E}$, then we have another important property. Consider a subset $\hat{\mathcal{S}}$, $\hat{\mathcal{S}} \subset \mathcal{E}$ and $|\hat{\mathcal{S}}| = j+1$. Then each user r', $r' \in \hat{\mathcal{S}}$, has $\frac{q_j}{\alpha_j}$ purified overheard equation $\hat{L}_t^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$, $t = 1, \ldots, \frac{q_j}{\alpha_j}$ which are simultaneity useful for all users in $\hat{\mathcal{S}}\backslash\{r'\}$. We note that the base station is aware of these purified overheard equations. For every $\hat{\mathcal{S}} \subset \mathcal{E}$, $|\hat{\mathcal{S}}| = j+1$, the base station forms $j\frac{q_j}{\alpha_j}$ random linear combinations of $\hat{L}_t^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$, $r' \in \hat{\mathcal{S}}$, $t = 1, \ldots, \frac{q_j}{\alpha_j}$, denoted by $u_1^{[\hat{\mathcal{S}}]}, u_2^{[\hat{\mathcal{S}}]}, \ldots, u_{j\frac{q_j}{\alpha_j}}^{[\hat{\mathcal{S}}]}$. We note that $u_\tau^{[\hat{\mathcal{S}}]}$, $1 \le \tau \le j\frac{q_j}{\alpha_j}$, is simultaneously useful for all users in $\hat{\mathcal{S}}$. Indeed, each user r, $r \in \hat{\mathcal{S}}$, can subtract the contributions of $\hat{L}_t^{[r]}(j;\hat{\mathcal{S}}\backslash\{r\})$, $t = 1, \ldots, \frac{q_j}{\alpha_j}$ from $u_\tau^{[\hat{\mathcal{S}}]}$, $\tau = 1, \ldots, j\frac{q_j}{\alpha_j}$, and form $j\frac{q_j}{\alpha_j}$ linearly independent combinations of $\hat{L}_t^{[r']}(j;\hat{\mathcal{S}}\backslash\{r'\})$, $r' \in \hat{\mathcal{S}}\backslash\{r\}$, $t = 1, \ldots, \frac{q_j}{\alpha_j}$. Using the above procedure, the base stations forms $j\frac{q_j}{\alpha_j}\binom{K}{j+1}$ symbols with degree $j+1$. The important observation is if these $j\frac{q_j}{\alpha_j}\binom{K}{j+1}$ symbols are delivered to the designated receivers, then each receivers will have enough equations to resolve all designated messages with degree j. Delivering $j\frac{q_j}{\alpha_j}\binom{K}{j+1}$ degree $j+1$ symbols takes $\frac{j\frac{q_j}{\alpha_j}\binom{K}{j+1}}{\text{DoF}_{j+1}(M,K)}$. Since the sub-algorithm starts with $(K-j)\frac{q_j+1}{\alpha_j}\binom{K}{j}$ symbols with degree j, and takes $\frac{K-j}{\alpha_j}\binom{K}{j}$ time slots, and generates $j\frac{q_j}{\alpha_j}\binom{K}{j+1}$ messages with degree $j+1$, then we have

$$\text{DoF}_j(M,K) = \frac{(K-j)\frac{q_j+1}{\alpha_j}\binom{K}{j}}{\frac{K-j}{\alpha_j}\binom{K}{j} + \frac{j\frac{q_j}{\alpha_j}\binom{K}{j+1}}{\text{DoF}_{j+1}(M,K)}}, \tag{39}$$

or

$$\frac{q_j+1}{j}\frac{1}{\text{DoF}_j(M,K)} = \frac{1}{j} + \frac{q_j}{j+1}\frac{1}{\text{DoF}_{j+1}(M,K)}. \tag{40}$$

Therefore Theorem 2 has been proven.

VII. Improved Scheme for $M = 2$

Recall that the scheme of Section VI achieves DoF_1 of $\frac{24}{17}$ for $M = 2$ and $K = 3$. The achieved DoF is greater that $\text{DoF}_2(2,2) = \frac{4}{3}$, which shows that we could exploit the extra number of users with respect to the number of transmit antennas. However, it is sill smaller than $\frac{3}{2}$ which is suggested by the outer–bound. Now the question is whether the achievable scheme or the outer–bound is loose.

First let us have an interesting observation. In Table IV, let us hypothetically assume that $L^{[2]}(1;1) \equiv L^{[3]}(1;1)$. Of course by no mean this assumption is valid. Similarly let us assume that $L^{[1]}(1;2) \equiv L^{[3]}(1;2)$ and $L^{[1]}(1;3) \equiv L^{[2]}(1;3)$. Then, we could define $u^{[1,2,3]}$ as a random linear combination of $L^{[2]}(1;1)$, $L^{[1]}(1;2)$, and $L^{[1]}(1;3)$, e.g., $L^{[2]}(1;1) + L^{[1]}(1;2) + L^{[1]}(1;3)$, Then it is easy to see that $u^{[1,2,3]}$ was simultaneously useful for all three users.For example, if we deliver $u^{[1,2,3]}$ to user one, then it can use the saved overheard equations $L^{[1]}(1;2)$ and $L^{[1]}(1;3)$ to form $u^{[1,2,3]} - L^{[1]}(1;2) - L^{[1]}(1;3)$ which is $L^{[2]}(1;1)$, and then resolve $u_1^{[1]}$ and $u_2^{[1]}$. Similar statement is valid for users two and three. Therefore, this phase would start with 6 degree–one messages, takes 3 time-slots, and generates one degree–one symbol. Then, we had $\text{DoF}_3(2,3) = \frac{6}{4+\frac{1}{\text{DoF}_3(2,3)}}$. Since $\text{DoF}_3(2,3)$ is optimally one, then we could achieve $\text{DoF}_3(2,3) = \frac{6}{4} = \frac{3}{2}$ and meet the outer–bound. This observation supports this conjecture that the outer–bound is tight under some unrealistic assumptions, and therefore it is loose in general. However, in what follows, we show that for $M = 2$ and $K = 3$, the outer–bound is tight and the achievable scheme of Section VI is loose.

A. Alternative Scheme for $M = K = 2$

Here first, we explain an alternative optimal solution for $M = 2$ and $K = 2$. Again Phase One of the algorithm deals with degree-one messages. Let us assume that the base station has $u_1^{[1]}$ and $u_2^{[1]}$ for user one and $u_1^{[2]}$ and $u_2^{[2]}$ for user two.

TABLE VI

SIGNALING SCHEME FOR $M = K = 2$

Phase	1 (Feeding the Receivers)
Sub-Phase	S-Ph(1;1,2)
Tx 1	$u_1^{[1]} + u_1^{[2]}$
Tx 2	$u_2^{[1]} + u_2^{[2]}$
$y^{[1]}(m) - z^{[1]}(m)$	$L^{[1]}(1;1,2) = L^{[1]}(1;1,2;1) + L^{[1]}(1;1,2;2)$
$y^{[2]}(m) - z^{[2]}(m)$	$L^{[2]}(1;1,2) = L^{[2]}(1;1,2;1) + L^{[2]}(1;1,2;2)$

However, Phase One has only one sub-phase, which is dedicated to both users. We denote this sub-phase of the phase one with S-Ph$(1;1,2)$. In S-Ph$(1;1,2)$, we send random linear combinations of all four symbols $u_1^{[1]}$, $u_2^{[1]}$, $u_1^{[2]}$ and $u_2^{[2]}$. The linear combination of transmitted symbols received by user r is denoted by $L^{[r]}(1;1,2)$. We note that $L^{[r]}(1;1,2)$ is a summation of two terms, the first term is a linear combination of $u_1^{[1]}$ and $u_2^{[1]}$, denoted by $L^{[r]}(1;1,2;1)$ and the second term is a linear combination of $u_1^{[2]}$ and $u_2^{[2]}$, denoted by $L^{[r]}(1;1,2;2)$. Therefore, $L^{[r]}(1;1,2) = L^{[r]}(1;1,2;1) + L^{[r]}(1;1,2;2)$ (See Table VI).

Then, we have the following observation:

- If we somehow give $L^{[1]}(1;1,2;2)$ and $L^{[2]}(1;1,2;1)$ to user one, then user one has two linearly independent equations $L^{[1]}(1;\{1,2\}) - L^{[1]}(1;1,2;2)$ and $L^{[2]}(1;1,2;1)$. Noting that user one already has $L^{[1]}(1;1,2) = L^{[1]}(1;1,2;1) + L^{[1]}(1;1,2;2)$, then it is enough to give it $L^{[1]}(1;1,2;2)$ and $L^{[2]}(1;1,2;1)$.
- If user two has $L^{[1]}(2;1,2;2)$ and $L^{[2]}(1;1,2;1)$, then it has enough equations to resolve $u_1^{[2]}$ and $u_2^{[2]}$. Noting that user two already has $L^{[2]}(1;1,2;1) + L^{[2]}(1;1,2;2)$, then it is enough to give user two the two equations $L^{[1]}(1;1,2;2)$ and $L^{[2]}(1;1,2;1)$.

In other words, both users one and two want $L^{[1]}(1;1,2;2)$ and $L^{[2]}(1;1,2;1)$. Therefore, we can define degree-two symbols $u_1^{[1,2]}$ and $u_2^{[1,2]}$ as

$$u_1^{[1,2]} = L^{[1]}(1;1,2;2), \tag{41}$$

$$u_2^{[1,2]} = L^{[2]}(1;1,2;1). \tag{42}$$

In summary, this sub-algorithm starts with 4 degree–one messages, takes one time-slot, and provides two degree–two symbols. Therefore, we have

$$\text{DoF}_1(2,2) = \frac{4}{1 + \frac{2}{\text{DoF}_2(2,2)}}. \tag{43}$$

We know how to achieve $\text{DoF}_2(2,2) = 1$. Therefore, we again achieve $\text{DoF}_1(2,2) = \frac{4}{3}$.

B. Alternative Scheme for First Phase of $M = 2$, $K = k$

Here, we explain a sub-algorithm for the systems with 2 transmit antennas, which takes degree one messages and gives degree two symbols. This sub-algorithm leads to an optimal scheme for systems with $M = 2$ and $K = 3$.

This sub-algorithm has $\binom{k}{2}$ sub-phases, where each sub-phase takes only one time slots and is dedicated to a pair of users. Let us focus on the sub-phase dedicated to users one and two, denoted by S-Ph$(1;1,2)$. This sub-phase is exactly the same as Phase One of the algorithm explained in Subsection VI. It takes four symbols of degree one and generate two message of degree two. Note that in this algorithm we ignore the overheard equations by user 3 to K in S-Ph$(1;1,2)$. We apply similar

TABLE VII

SIGNALING SCHEME FOR $M=2$, $K=4$, $j=2$

Phase	2
Sub-Phase	S-Ph(2;1,2,3)
Tx 1	$u_1^{[1,2]}+u_1^{[1,3]}$
Tx 2	$u_2^{[1,2]}+u_2^{[1,3]}$
$y^{[1]}(m)-z^{[1]}(m)$	$L^{[1]}(2;1,2,3)=L^{[1]}(1;1,2,3;1,2)+L^{[1]}(1;1,2,3;1,3)$
$y^{[2]}(m)-z^{[2]}(m)$	$L^{[2]}(2;1,2,3)=L^{[2]}(1;1,2,3;1,2)+L^{[2]}(1;1,2,3;1,3)$
$y^{[3]}(m)-z^{[3]}(m)$	$L^{[3]}(2;1,2,3)=L^{[3]}(1;1,2,3;1,2)+L^{[3]}(1;1,2,3;1,3)$
$y^{[4]}(m)-z^{[4]}(m)$	$L^{[4]}(2;1,2,3)=L^{[4]}(1;1,2,3;1,2)+L^{[4]}(1;1,2,3;1,3)$

scheme for any other pair of users. Therefore, in total, this scheme takes $4\binom{K}{2}$ data streams with degree one, takes $\binom{K}{2}$ time slots, and generates $2\binom{K}{2}$ symbols with degree two. Therefore, we have

$$\text{DoF}_1(2,K)=\frac{4\binom{K}{2}}{\binom{K}{2}+\frac{2\binom{K}{2}}{\text{DoF}_2(2,K)}}, \tag{44}$$

or

$$\frac{4}{\text{DoF}_1(2,K)}=1+\frac{2}{\text{DoF}_2(2,K)}. \tag{45}$$

In Section IV, it is shown that the optimal $\overline{\text{DoF}}_2(2,3)=\frac{6}{5}$ is achievable. Then using (43), we can see that $\text{DoF}_2(2,3)=\frac{3}{2}$ is achievable which meets the outer–bound. This result show that the scheme of Section IV is not optimal.

C. Infinite–Horizon Scheme $M=2$, $K=k$

In Subsection VII-A, we presented a scheme which achieves the optimal DoF for $M=2$ and $K=3$. Now let us focus on the general case with K number of users. As a first example, let us focus on the system with $M=2$ and $K=4$. For this channel, in Section IV, it is shown that the optimal $\overline{\text{DoF}}_3(2,4)=\frac{8}{7}$ and $\overline{\text{DoF}}_4(2,4)=1$. Now the question is what the optimal $\overline{\text{DoF}}_2(2,4)$ and $\overline{\text{DoF}}_1(2,4)$ are. The outer–bound suggests that $\overline{\text{DoF}}_2(2,4)\leq\frac{4}{3}$ and $\overline{\text{DoF}}_1(2,4)\leq\frac{8}{5}$. The interesting fact is that if we can achieve $\text{DoF}_2(2,4)$ of $\frac{4}{3}$, then we can use the scheme of Subsection VII-A to achieve $\text{DoF}_1(2,4)$ of $\frac{8}{5}$.

In this section, we explain a scheme which achieves the $\text{DoF}_2(2,4)$ of $\frac{14}{11}$. This is still strictly less than the outer–bound of $\frac{4}{3}$. However, it is greater than $\frac{24}{19}$ which is what we can achieve using the scheme of Section VI. The important difference is that this scheme is infinite-horizon.

Let us assume that $u_1^{[1,2]}$ and $u_2^{[1,2]}$ for users one and two and $u_1^{[1,3]}$ and $u_2^{[1,3]}$ for users one and three. In sub-phase S-Ph$(2;1,2,3)$, we send linear combination of all these four messages. Refer to Table VII for specific choices of transmission.

Then, if (i) users one, two and three have $L^{[2]}(1;1,2,3;1,3)$ and $L^{[3]}(1;1,2,3;1,2)$, (ii) user one has $L^{[4]}(2;1,2,3)=L^{[4]}(1;1,2,3;1,2)+L^{[4]}(1;1,2,3;1,3)$ which is available at user four, then user one and two can decode $u_1^{[1,2]}$ and $u_2^{[1,2]}$ and user one and three can decode $u_1^{[1,3]}$ and $u_2^{[1,3]}$. We can define $L^{[2]}(1;1,2,3;1,3)$ and $L^{[3]}(1;1,2,3;1,2)$ as two degree–three symbols. In addition, following the ideas of Section IV, we can treat $L^{[4]}(2;1,2,3)$ as a half of a degree–two symbol.

Therefore, this sub-algorithm takes 4 degree-two symbols, takes one time slots, and generates two degree–three symbols, and half degree–two symbol. We can repeat this scheme iteratively over many degree-two symbols, and divide the number of degree–two messages by 8 in each iteration. As a fixed point, this scheme achieves $\text{DoF}_2(2,4)$ as the solution of the following equation.

$$\text{DoF}_2(2,4)=\frac{4}{1+\frac{2}{\text{DoF}_3(2,4)}+\frac{0.5}{\text{DoF}_2(2,4)}}. \tag{46}$$

From Section IV, we know $\overline{\mathrm{DoF}}_3(2,4) = \frac{8}{7}$. Then, the fixed point of the above equation is $\mathrm{DoF}_2(2,4) = \frac{14}{11}$. Then, using the sub-algorithm of Subsection VII-B, we can achieve $\mathrm{DoF}_2(2,4) = \frac{14}{9}$. This scheme can be extended to degree j messages for a system with two transmit antennas and K receiver.

References

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

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

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

[4] W. Yu and J. Cioffi, "Sum capacity of vector Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-50, pp. 1875 – 1892, Sept. 2004, submitted for Publication.

[5] S. Shamai(Shitz) H. Weingarten, Y. Steinberg, "The capacity region of the gaussian mimo broadcast channel," *in Proc. IEEE Int. Symp. on Information Theory*, p. 174, Jun./Jul. 2004.

[6] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of Fading MIMO broadcast channels with imperfect transmitter side-information," *arXiv:cs/0605079v1*, May 2006.

[7] S.A. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels:the scalar upper bound and loss in degrees of freedom," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 848 – 857, March 2005.

[8] T. Gou, S. A. Jafar, and C. Wang, "On the Degrees of Freedom of Finite State Compound Wireless Networks - Settling a Conjecture by Weingarten et. al," *Arxiv preprint arXiv:0909.4177*, September 2009.

[9] M. A. Maddah-Ali, "The degrees of freedom of the compound MIMO broadcast channels with finite states," *http://arxiv.org/abs/0909.5006*, Sept. 2009.

[10] A. S. Motahari, M.A. Maddah-Ali S. Oveis Gharan, and A. K. Khandani, "Forming Pseudo-MIMO by Embedding Infinite Rational Dimensions Along a Single Real Line: Removing Barriers in Achieving the DOFs of Single Antenna Systems," *Arxiv preprint arXiv:0908.2282*, Aug. 2009.

[11] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *Information Theory, IEEE Transactions on*, vol. 52, no. 11, pp. 5045 – 5060, Nov. 2006.

[12] C. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IEEE Transactions on*, vol. 2, no. 3, pp. 8 – 19, Sept. 1956.

[13] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *Information Theory, IEEE Transactions on*, vol. 12, no. 2, pp. 172 – 182, Apr. 1966.

[14] A. El Gamal, "The feedback capacity of degraded broadcast channels," *Information Theory, IEEE Transactions on*, vol. 24, no. 3, pp. 379 – 381, Apr. 1978.

[15] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *Information Theory, IEEE Transactions on*, vol. 30, no. 4, pp. 667 – 671, Jul 1984.

[16] P. Larsson and N. Johansson, "Multi-user arq," in *IEEE 63rd Vehicular Technology Conference,*, Melbourne, May 2006, pp. 2052 – 2057.

[17] M. Jolfaei, S. Martin, and J. Mattfeldt, "A new efficient selective repeat protocol for point-to-multipoint communication," in *IEEE International Conference on Communications (ICC'93),*, Geneva, May 1993, pp. 1113 – 1117.

[18] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback - capacity and algorithms," in *Workshop on Network Coding, Theory, and Applications*, Lausanne, June 1993, pp. 54 – 61.

[19] Chih Chun Wang, "Capacity of 1-to-k broadcast packet erasure channel with output feedback," *Unpublished Report*, 2010.

[20] C. Huang, S. A. Jafar, and S. Shamai (Shitz), "Multiuser MIMO degrees of freedom with no CSIT," in *Information Theory and Application,*, San Diego, CA, 2009.

[21] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *http://arxiv.org/abs/0909.5424*, Oct. 2009.