

Completely Stale Transmitter Channel State Information is Still Very Useful

Mohammad Ali Maddah-Ali and David Tse

Abstract—Transmitter channel state information (CSIT) is crucial for the multiplexing gains offered by advanced interference management techniques such as multiuser multiple-input multiple-output (MIMO) and interference alignment. Such CSIT is usually obtained by feedback from the receivers, but the feedback is subject to delays. The usual approach is to use the fed back information to predict the current channel state and then apply a scheme designed assuming perfect CSIT. When the feedback delay is large compared to the channel coherence time, such a prediction approach completely fails to achieve any multiplexing gain. In this paper, we show that even in this case, the completely stale CSI is still very useful. More concretely, we show that in an MIMO broadcast channel with K transmit antennas and K receivers each with 1 receive antenna, $\frac{K}{1+\frac{1}{2}+\dots+\frac{1}{K}} (> 1)$ degrees of freedom is achievable even when the fed back channel state is completely independent of the current channel state. Moreover, we establish that if all receivers have independent and identically distributed channels, then this is the optimal number of degrees of freedom achievable. In the optimal scheme, the transmitter uses the fed back CSI to learn the side information that the receivers receive from previous transmissions rather than to predict the current channel state. Our result can be viewed as the first example of feedback providing a degree-of-freedom gain in memoryless channels.

Index Terms—Feedback delay, interference alignment, multiple-antenna channels, network coding, output feedback, side information, vector Gaussian broadcast channels.

I. INTRODUCTION

IN wireless communication, transmitter knowledge of the channel state information (CSIT) can be very important. While in point-to-point channels, CSIT only provides power gains via waterfilling, in multiuser channels, it can also provide *multiplexing gains*. For example, in a multiple-input multiple-output (MIMO) broadcast channel, CSIT can be used to

send information along multiple beams to different receivers simultaneously. In interference channels, CSIT can be used to align the interference from multiple receivers to reduce the aggregate interference footprint [1], [2].

In practice, it is not easy to achieve the theoretical gains of these techniques. In the high SNR regime, where the multiplexing gain offered by these techniques is particularly significant, the performance of these techniques is very sensitive to inaccuracies of the CSIT. However, it is hard to obtain accurate CSIT. This is particularly so in frequency-division duplex systems, where the channel state has to be measured at the receiver and fed back to the transmitter. This feedback process leads to two sources of inaccuracies.

- 1) *Quantization Error*: The limited rate of the feedback channel restricts the accuracy of the CSI at the transmitter.
- 2) *Delay*: There is a delay between the time the channel state is measured at the receiver and the time when the information is used at the transmitter. The delay comes from the fact that the receivers need some time to receive pilots, estimate CSI, and then feed it back to the transmitter in a relatively long coding block. In time-varying wireless channels, when the channel information arrives at the transmitter, the channel state has already changed.

Much work in the literature has focused on the first issue. The general conclusion is that the rate of the feedback channel needed to achieve the perfect CSIT multiplexing gain scales well with the SNR. For example, for the MIMO broadcast channel, it was shown in [3] that the rate of feedback should scale linearly with \log_2 SNR. Since the capacity of the MIMO broadcast channel also scales linearly with \log_2 SNR, this result says that the overhead from feedback will not overwhelm the capacity gains.

We now focus on the second issue, the issue of feedback delay. The standard approach of dealing with feedback delay is to exploit the time correlation of the channel to predict the current channel state from the delayed measurements [4]. The predicted channel state is then used in place of the true channel state in a scheme designed assuming perfect CSIT is available. However, as the coherence time of the channel becomes shorter compared to the feedback delay, due to higher mobility, for example, the delayed feedback information reveals no information about the current state, and a prediction-based scheme can offer no multiplexing gain.

In this paper, we raise the question: is this a fundamental limitation imposed by feedback delay, or is this just a limitation of the prediction-based approach? In other words, is there another way to use the delayed feedback information to achieve nontrivial multiplexing gains? We answer the question in the affirmative.

Manuscript received September 26, 2011; revised February 23, 2012; accepted March 27, 2012. Date of publication April 03, 2012; date of current version June 12, 2012. This work was supported in part by a gift from Qualcomm, Inc. and by the Air Force Office of Scientific Research under Grant FA9550-09-1-0317. An initial version of this paper has been reported as Technical Report No. UCB/EECS-2010-122 at the University of California–Berkeley, September 6, 2010. The material in this paper was presented in part at the 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2010.

M. A. Maddah-Ali was with Wireless Foundations, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA. He is now with Bell-Labs Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: maddah.ali.ee@gmail.com).

D. Tse is with Wireless Foundations, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

Communicated by S. Tatikonda, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2012.2193116

For concreteness, we focus on a channel which has received significant attention in recent years: the MIMO broadcast channel. In particular, we focus on a system where the transmitter has M antennas and there are K receivers each with a single receive antenna. The transmitter wants to send an independent data stream to each receiver. To model completely outdated CSI, we allow the channel state to be independent from one symbol time to the next, and the CSI is available to both the transmitter and the receivers one symbol time later. This means that by the time the feedback reaches the transmitter, the current channel is already *completely* different. We also assume that the overall $M \times K$ channel matrix is full rank at each time.

Our main result is that, for $M \geq K$, one can achieve a total of

$$\frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}}$$

degrees of freedom per second per Hz in this channel. In other words, we can achieve a sum rate that scales like

$$\frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}} \log_2 \text{SNR} + o(\log_2 \text{SNR}) \quad \text{bits/s/Hz}$$

as the SNR grows. Moreover, we show that under the further assumption that all receivers have independent and identically distributed (i.i.d.) channels, this is the optimal number of degrees of freedom achievable.

It is instructive to compare this result with the case when there is no CSIT and the case when there is perfect CSIT. While the capacity or even the number of degrees of freedom is unknown for general channel statistics when there is no CSIT, in the case when all receivers have identically distributed channels, it is easy to see that the total number of degrees of freedom is only 1. Since $\frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}} > 1$ for any $K \geq 2$, we see that, at least in that case, there is a multiplexing gain achieved by exploiting completely outdated CSI. However, the multiplexing gain is not as good as K , the number of degrees of freedom achieved in the perfect CSIT case. On the other hand, when K is large

$$\frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}} \approx \frac{K}{\ln K}$$

almost linear in K .

Why is outdated CSIT useful? When there is perfect CSIT, information intended for a receiver can be transmitted to that receiver without other receivers overhearing it (say by using a zero-forcing precoder), so that there is no cross-interference. When the transmitter does not know the current channel state, this cannot be done and information intended for a receiver will be overheard by other receivers. This overheard *side information* is in the form of a linear combination of data symbols, the coefficients of which are the channel gains at the time of the transmission. Without CSIT at all, this side information will be wasted since the transmitter does not know what the coefficients are and hence does not know what side information was received in previous transmissions. With outdated CSIT, however, the transmitter can exploit the side information already received at the various receivers to create future transmissions which are simultaneously useful for more than one receiver and can, therefore, be efficiently transmitted. Note that there is no

such overheard side-information in simpler scenarios such as point-to-point and multiple access channels, where there is only a single receiver. Indeed, it is shown in [5] and [6] that for such channels, the only role of delayed CSIT is to predict the current state, and when the delayed CSIT is independent of the current state, the delayed CSIT provides no capacity gains.

The rest of this paper is structured as follows. In Section II, the problem is formulated and the main results are stated precisely. Sections III, VI, and VII describe the proposed schemes, and Section IV describes the converse. In Section V, the DoF region for the case of $M = K$ is characterized. The connection between our results and those for the packet erasure broadcast channel is explained in Section VIII. Some follow-up results to the conference version of this paper are discussed in Section IX. We conclude with a discussion of our result in the broader context of the role of feedback in communication in Section X.

II. PROBLEM FORMULATION AND MAIN RESULTS

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

$$y_r[n] = \mathbf{h}_r^\dagger[n] \mathbf{x}[n] + z_r[n], \quad r = 1, \dots, K \quad (1)$$

where \dagger denotes transpose-conjugate operation, $\mathbf{x}[n] \in \mathbb{C}^{M \times 1}$, $\mathbb{E}[\mathbf{x}^\dagger[n] \mathbf{x}[n]] \leq \text{SNR}$, $z_r[n] \sim \mathcal{CN}(0, 1)$ and the sequences $z_r[n]$'s are i.i.d. and mutually independent. In addition, $\mathbf{h}_r^\dagger[n] = [h_{r1}^\dagger[n], \dots, h_{rM}^\dagger[n]] \in \mathbb{C}^{1 \times M}$. We define $\mathbf{H}[n]$ as $\mathbf{H}[n] = [\mathbf{h}_1[n], \dots, \mathbf{h}_K[n]]$.

We assume that $\mathbf{H}[n]$ is available at the transmitter and all receivers with one unit delay¹

Let us define \mathcal{E} as $\mathcal{E} = \{1, 2, \dots, K\}$. We assume that for any subset \mathcal{S} of the receivers, $\mathcal{S} \subset \mathcal{E}$, the transmitter has a message $W_{\mathcal{S}}$ with rate $R_{\mathcal{S}}$ bits/s/Hz. For example, message $W_{\{1,2\}}$ is a common message for receivers one and two. Similarly, $W_{\{1\}}$, or simply W_1 , is a message for receiver one. We define $d_{\mathcal{S}}$ as

$$d_{\mathcal{S}} = \lim_{\text{SNR} \rightarrow \infty} \frac{R_{\mathcal{S}}}{\log_2 \text{SNR}}. \quad (2)$$

If $|\mathcal{S}| = j$, then we call $W_{\mathcal{S}}$ an order- j message or a message of order j . We define degrees of freedom order j , $\text{DoF}_j^*(M, K)$ as

$$\text{DoF}_j^*(M, K) = \lim_{\text{SNR} \rightarrow \infty} \max_{\mathbf{R} \in \mathcal{C}} \sum_{\mathcal{S}, |\mathcal{S}|=j} \frac{R_{\mathcal{S}}}{\log_2 \text{SNR}} \quad (3)$$

where \mathcal{C} denotes the capacity region of the channel, and $\mathbf{R} \in \mathbb{R}^{(2^K - 1) \times 1}$ denotes the vector of the message rates for each subset of receivers. We note that $\text{DoF}_1^*(M, K)$ is the well-known notion of the degrees of freedom of the channel.

In this paper, we establish the following results.

Theorem 1: As long as $\mathbf{H}[n]$ is full rank almost surely for each n , and $\{\mathbf{H}[n]\}$ is stationary and ergodic, then for $M \geq K$

$$\text{DoF}_1^*(M, K) \geq \frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}}. \quad (4)$$

¹All our achievable results hold regardless of what the delay is, since they do not depend on the temporal statistics of the channel. Hence, for convenience, we will just normalize the delay to be 1 symbol time.

More generally, as long as $M \geq K - j + 1$, then

$$\text{DoF}_j^*(M, K) \geq \frac{K - j + 1}{j} \frac{1}{\frac{1}{j} + \frac{1}{j+1} + \dots + \frac{1}{K}}. \quad (5)$$

For example, $\text{DoF}_1^*(2, 2) \geq \frac{4}{3}$ and $\text{DoF}_1^*(3, 3) \geq \frac{18}{11}$, which are greater than one. Note that this achievability result holds under very weak assumptions about the channel statistics. Hence, even when $\{\mathbf{H}[n]\}$ is an i.i.d. process over time, delayed CSIT is still useful in achieving a degree-of-freedom gain.

The following theorem gives a tight converse under specific assumptions on the channel process.

Theorem 2: If the channel matrices $\{\mathbf{H}[n]\}$ is an i.i.d. process over time and the channels are also i.i.d. across the receivers, then

$$\text{DoF}_j^*(M, K) \leq \frac{\binom{K}{j}}{\frac{\binom{K-1}{j-1}}{\min\{1, M\}} + \frac{\binom{K-2}{j-1}}{\min\{2, M\}} + \dots + \frac{\binom{j-1}{j-1}}{\min\{K-j+1, M\}}}. \quad (6)$$

The equality between the expressions in (5) and (6) in the case of $M \geq K - j + 1$ can be verified using the identity (47), proved in Appendix A, thus yielding the following corollary.

Corollary 1: If the channel matrices $\{\mathbf{H}[n]\}$ is an i.i.d. process over time and is also i.i.d. across the receivers, then the lower bounds in Theorem 1 are tight.

In addition, the region of order-1 DoF for the case $M = K$ is characterized as follows.

Theorem 3: If the channel matrices $\{\mathbf{H}[n]\}$ is an i.i.d. process over time and is also i.i.d. across the receivers, then the DoF region for the case $M = K$ is characterized as all positive K -tuples (d_1, d_2, \dots, d_K) satisfying

$$\sum_{i=1}^K \frac{d_{\pi(i)}}{i} \leq 1 \quad (7)$$

for all permutations π of the set $\{1, \dots, K\}$.

The achievability result for Theorem 1 holds for $M \geq K - j + 1$. We have the following achievability result for general M , K , and j .

Theorem 4: Assume that $\mathbf{H}[n]$ is full rank almost surely for each n , and $\{\mathbf{H}[n]\}$ is stationary and ergodic. If $\text{DoF}_{j+1}(M, K)$ is achievable for order- $(j+1)$ symbols, then $\text{DoF}_j(M, K)$ is achievable for order- j symbols, where

$$\text{DoF}_j(M, K) = \frac{\frac{q_j+1}{j}}{\frac{1}{j} + \frac{q_j}{j+1} \frac{1}{\text{DoF}_{j+1}(M, K)}} \quad (8)$$

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

Starting from $\text{DoF}_K^*(M, K) = 1$, which is simply achievable, one can use iterative (8) to derive an achievable $\text{DoF}_j(M, K)$ with the following closed form:

$$\text{DoF}_j(M, K) = \frac{\frac{M}{j}}{\sum_{i=j}^{K-M} \frac{1}{i} \left(\frac{M-1}{M}\right)^{i-j} + \left(\frac{M-1}{M}\right)^{K-M-j+1} \left(\sum_{i=K-M+1}^K \frac{1}{i}\right)} \quad (9)$$

for the case of $M < K - j + 1$. Unlike the case of $M \geq K - j + 1$, however, the expression in (9) does not match the upper bound in Theorem 2. In particular, this means that Theorem 4 does not allow us to characterize the degrees of freedom $\text{DoF}_1^*(M, K)$ when the number of users K is greater than the number of transmit antennas M . On the other hand, it is easy to verify that the achievable $\text{DoF}_1(M, K)$ in Theorem 4 is increasing with K , even when $K > M$. Therefore, unlike the situation with full CSIT, the degrees of freedom under delayed CSIT is not determined by the minimum of the number of transmit antennas and the number of receivers.

For the special case of $M = 2$ and $K = 3$, we obtain an exact characterization of the degrees of freedom.

Theorem 5: Assume that $\mathbf{H}[n]$ is full rank almost surely for each n , and $\{\mathbf{H}[n]\}$ is stationary and ergodic; then $\text{DoF}_1^*(2, 3) = \frac{3}{2}$.

III. ACHIEVABLE SCHEME FOR THEOREM 1

In this section, we explain the achievable scheme for Theorem 1. The key is to understand the square case when $M = K$. For simplicity, we start with the cases $M = K = 2$ and $M = K = 3$.

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

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

- 1) exploiting side-information;
- 2) generating higher order messages;
- 3) interference alignment using outdated CSIT.

For notational clarity, in this section, we will use A and B to denote the two receivers instead of 1 and 2.

1) Exploiting Side-Information: Let u_r and v_r be symbols from two independently encoded Gaussian codewords intended for receiver r . The proposed communication scheme is performed in two phases, which take three time-slots in total.

Phase One—Feeding the Receivers: This phase has two time-slots.

The first time slot is dedicated to receiver A . The transmitter sends the two symbols, u_A and v_A , intended for receiver A , i.e.,

$$\mathbf{x}[1] = \begin{bmatrix} u_A \\ v_A \end{bmatrix}. \quad (10)$$

At the receivers, we have

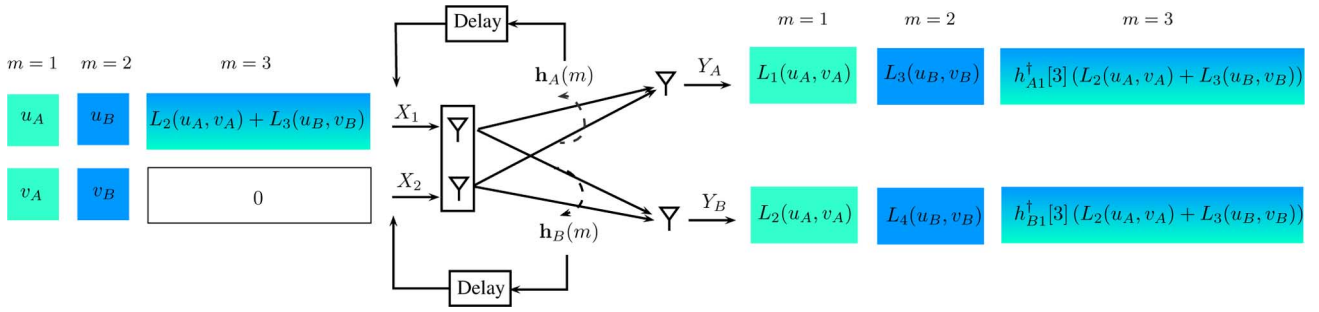
$$y_A[1] = h_{A1}^\dagger[1]u_A + h_{A2}^\dagger[1]v_A + z_A[1] \quad (11)$$

$$y_B[1] = h_{B1}^\dagger[1]u_A + h_{B2}^\dagger[1]v_A + z_B[1]. \quad (12)$$

Both receivers A and B receive noisy versions of linear combinations of u_A and v_A . Receiver B saves the overheard equation for later usage, although it only carries information intended for receiver A .

The second time-slot of phase one is dedicated to the second receiver. In this time-slot, the transmitter sends symbols intended for receiver B , i.e.,

$$\mathbf{x}[2] = \begin{bmatrix} u_B \\ v_B \end{bmatrix}. \quad (13)$$

Fig. 1. Achievable scheme for $M = K = 2$.

At receivers, we have

$$y_A[2] = h_{A1}^\dagger[2]u_B + h_{A2}^\dagger[2]v_B + z_A[2] \quad (14)$$

$$y_B[2] = h_{B1}^\dagger[2]u_B + h_{B2}^\dagger[2]v_B + z_B[2]. \quad (15)$$

Receiver A saves the overheard equation for future usage, although it only carries information intended for receiver B .

Let us define short-hand notations

$$\begin{aligned} L_1(u_A, v_A) &= h_{A1}^\dagger[1]u_A + h_{A2}^\dagger[1]v_A \\ L_2(u_A, v_A) &= h_{B1}^\dagger[1]u_A + h_{B2}^\dagger[1]v_A \\ L_3(u_B, v_B) &= h_{A1}^\dagger[2]u_B + h_{A2}^\dagger[2]v_B \\ L_4(u_B, v_B) &= h_{B1}^\dagger[2]u_B + h_{B2}^\dagger[2]v_B. \end{aligned}$$

The transmission scheme is summarized in Fig. 1. In this figure, for simplicity, we drop the thermal noise from the received signals. We note that, assuming $\mathbf{H}[1]$ is full rank, there is a one-to-one map between (u_A, v_A) and $(L_1(u_A, v_A), L_2(u_A, v_A))$. If receiver A has the equation overheard by receiver B , i.e., $L_2(u_A, v_A)$, then it has enough equations to solve for its own symbols u_A , and v_A . Similarly, assuming $\mathbf{H}[2]$ is full rank, there is a one-to-one map between (u_B, v_B) and $(L_3(u_B, v_B), L_4(u_B, v_B))$. If receiver B has the equation overheard by receiver A , i.e., $L_3(u_B, v_B)$, then it has enough equations to solve for its own symbols u_B and v_B .

Therefore, the main mission of the second phase is to swap these two overheard equations through the transmitter.

Phase Two—Swapping Overheard Equations: This phase takes only one time-slot at $n = 3$. At this time, the transmitter sends a linear combination of the overheard equations, i.e., $L_2(u_A, v_A)$ and $L_3(u_B, v_B)$. We note that at this time the transmitter is aware of the CSI at $n = 1$ and $n = 2$; therefore, it can form the overheard equations $L_2(u_A, v_A)$ and $L_3(u_B, v_B)$.

For example, $\mathbf{x}[3]$ can be formed as

$$\mathbf{x}[3] = \begin{bmatrix} L_2(u_A, v_A) + L_3(u_B, v_B) \\ 0 \end{bmatrix}. \quad (16)$$

At receivers, we have

$$y_A[3] = h_{A1}^\dagger[3](L_2(u_A, v_A) + L_3(u_B, v_B)) + z_A[3] \quad (17)$$

$$y_B[3] = h_{B1}^\dagger[3](L_2(u_A, v_A) + L_3(u_B, v_B)) + z_B[3]. \quad (18)$$

Remember that receiver A already has (a noisy version of) $L_3(u_B, v_B)$. Thus, together with $y_A[3]$, it can solve for its two symbols u_A, v_A . We have a similar situation for receiver B .

Remark: In this scheme, we assume that in the first time-slot, transmit antenna one sends u_A and transmit antenna two sends v_A . However, antenna one and two can send any random linear combination of u_A and v_A . Therefore, for example, we can have

$$\mathbf{x}[1] = \mathbf{A}[1] \begin{bmatrix} u_A \\ v_A \end{bmatrix} \quad (19)$$

where $\mathbf{A}[1] \in \mathbb{C}^{2 \times 2}$ is a randomly selected matrix. Similar statement is true for the second time-slot. At time-slot $n = 3$, we send $L_3(u_B, v_B) + L_2(u_A, v_A)$. However, we can send any combination of $L_3(u_B, v_B)$ and $L_2(u_A, v_A)$. In other words

$$\mathbf{x}[3] = \mathbf{A}[3] \begin{bmatrix} L_3(u_B, v_B) \\ L_2(u_A, v_A) \end{bmatrix} \quad (20)$$

where $\mathbf{A}[3] \in \mathbb{C}^{2 \times 2}$ is a randomly selected matrix. However, we can limit the choice of $\mathbf{A}[3]$ to rank 1 matrices.

Remark: We note that only the number of independent noisy equations that each receiver has is important. As long as the variance of the noise of each equation is bounded, the DoF is not affected. Therefore, in what follows, we ignore noise and just focus on the number of independent equations available at each receiver.

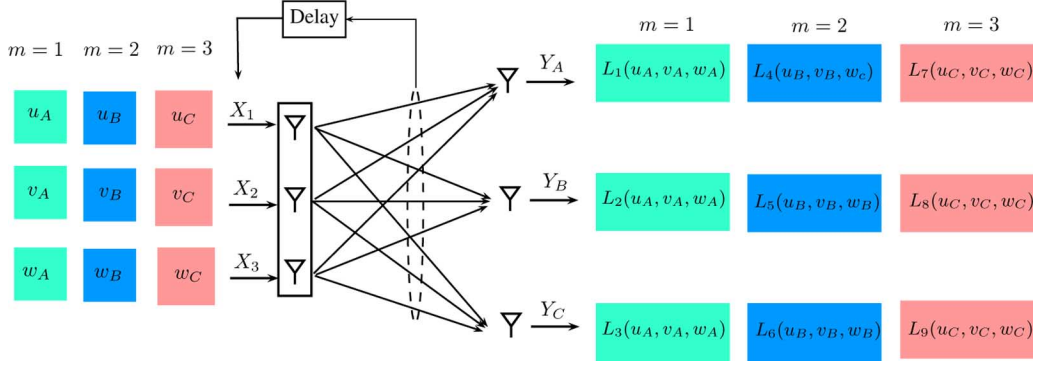
Remark: Note that if the transmitter has $2N$ transmit antennas, and each of the receivers has N antennas, then we can follow the same scheme and achieve DoF of $\frac{4N}{3}$.

2) Generating Higher Order Symbols: We can observe the achievable scheme from another perspective. Remember in the second phase, we send a linear combination of $L_2(u_A, v_A)$ and $L_3(u_B, v_B)$, e.g., $L_2(u_A, v_A) + L_3(u_B, v_B)$, to both receivers. We can consider $L_2(u_A, v_A) + L_3(u_B, v_B)$ as an order-2 common symbol, required by both receivers. Let us define $u_{AB} = L_2(u_A, v_A) + L_3(u_B, v_B)$. If we have an algorithm which achieves the degrees of freedom of DoF_2 for order-2 common symbols, then we need $\frac{1}{\text{DoF}_2(2,2)}$ time-slots to deliver the common symbol u_{AB} to both receivers. Therefore, in total, we need $2 + \frac{1}{\text{DoF}_2(2,2)}$ to deliver four symbols u_A, v_A, u_B , and v_B to the designated receivers. Thus, we have

$$\text{DoF}_1(2,2) = \frac{4}{2 + \frac{1}{\text{DoF}_2(2,2)}}. \quad (21)$$

It is easy to see that we can achieve $\text{DoF}_2(2,2) = 1$ by simply sending u_{AB} to both receivers in one time-slot. Therefore, $\text{DoF}_1(2,2)$ of $\frac{4}{3}$ is achievable.

In summary, phase one takes as input two order-1 symbols for each receiver. It takes two time-slots to deliver one desired equation to each of the receivers. Therefore, each receiver needs one

Fig. 2. Achievable Scheme for $K = 3$: Phase One.

more equation to resolve the desired symbols. If the transmitter ignores the overheard equations, we need two more time-slots to deliver one more equation to each receiver and yield the DoF of 1. However, by exploiting the overheard equations, we can form a common symbol of order 2. Delivering one common symbol of order 2 to both receivers takes only one time-slot but it simultaneously provides one useful equation to each of the receivers. Therefore, using this scheme, we save one time-slot and achieve $\text{DoF}_1(2, 2) = \frac{4}{3}$ rather than $\frac{4}{4}$.

3) *Interference Alignment Using Outdated CSIT*: Putting together the symbols received by receiver A over the three time-slots, we have (22), shown at the bottom of the page. From (22), it is easy to see that at receiver A, the two interference streams u_B and v_B arrived from the same directions $[0, h_{A1}[2], h_{A1}[3]h_{A1}[2]]^\dagger$, and therefore, u_B and v_B are aligned. Note that the alignment is done using outdated CSIT. By making the interference data symbols aligned at receiver A, the two symbols u_B and v_B collapse into one symbol $h_{A1}^\dagger[2]u_B + h_{A1}^\dagger[3]h_{A1}[2]v_B$. Eliminating the variable $h_{A1}^\dagger[2]u_B + h_{A1}^\dagger[3]h_{A1}[2]v_B$ from (22), we have (23), shown at the bottom of the page, which is an equation set of the two desired symbols u_A and v_A . It is easy to see that as long as $h_{A1}^\dagger[3] \neq 0$ and $h_{A1}^\dagger[1]h_{B2}^\dagger[1] - h_{A2}^\dagger[1]h_{B1}^\dagger[1] \neq 0$, then the desired data symbols are not aligned at receiver A and they can be solved for. We note that at $h_{A1}^\dagger[1]h_{B2}^\dagger[1] - h_{A2}^\dagger[1]h_{B1}^\dagger[1]$ is the determinant of the channel matrix $\mathbf{H}[1]$. Indeed, in this scheme, receiver A borrows the antenna of the second receiver at time-slot $n = 1$ to be able to solve for the two symbols.

B. Achievable Scheme for $M = K = 3$

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 a three-antenna transmitter and three

single-antenna receivers. As explained in the previous section, we can observe the achievable scheme from three different perspectives. However, we find the second perspective simpler to follow. Therefore, in the rest of this paper, we just explain the algorithm based on the second perspective.

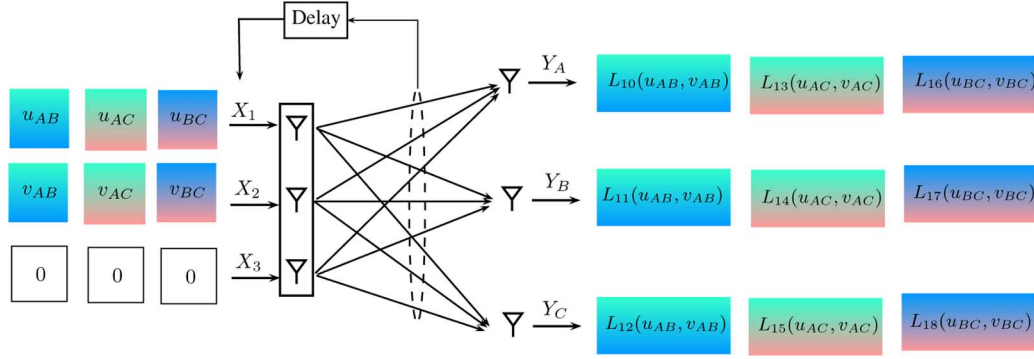
The achievable scheme has three phases. Phase one takes order-1 symbols and generates order-2 common symbols. Phase two takes order-2 common symbols and generates order-3 common symbols. The last phase takes order-3 common symbols and deliver them to all three receivers.

Phase One: This phase is similar to phase one for the 2×2 case. It takes three independent symbols for each receiver and generates three symbols of order 2. Assume that u_r , v_r , and w_r represent three symbols, independently Gaussian encoded, for receiver r , $r = A, B, C$. Therefore, in total, there are nine data symbols. This phase has three time-slots, where each time-slot is dedicated to one of the receivers. In the time-slot dedicated to receiver A, the transmitter sends random linear combinations of u_A , v_A , and w_A over the three antennas. Similarly, in the time-slot dedicated to receiver B, the transmitter sends random linear combinations of u_B , v_B , and w_B over the three antennas. In the time-slot dedicated to receiver C, the transmitter sends random linear combinations of u_C , v_C , and w_C over the three antennas. Refer to Fig. 2 for details.

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 for a system without CSIT. If we ignore the overheard equations, then we need six more time-slots to successfully deliver the nine data streams, which yield the DoF of 1. However, as described in the

$$\begin{bmatrix} y_A[1] \\ y_A[2] \\ y_A[3] \end{bmatrix} = \underbrace{\begin{bmatrix} h_{A1}^\dagger[1] & h_{A2}^\dagger[1] \\ 0 & 0 \\ h_{A1}^\dagger[3]h_{B1}^\dagger[1] & h_{A1}^\dagger[3]h_{B2}^\dagger[1] \end{bmatrix}}_{\text{Rank 2}} \begin{bmatrix} u_A \\ v_A \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 \\ h_{A1}^\dagger[2] & h_{A2}^\dagger[2] \\ h_{A1}^\dagger[3]h_{A1}^\dagger[2] & h_{A1}^\dagger[3]h_{A2}^\dagger[2] \end{bmatrix}}_{\text{Rank 1}} \begin{bmatrix} u_B \\ v_B \end{bmatrix} + \begin{bmatrix} z_A[1] \\ z_A[2] \\ z_A[3] \end{bmatrix} \quad (22)$$

$$\begin{bmatrix} y_A[1] \\ y_A[3] - h_{A1}^\dagger[3]y_A[2] \end{bmatrix} = \begin{bmatrix} h_{A1}^\dagger[1] & h_{A2}^\dagger[1] \\ h_{A1}^\dagger[3]h_{B1}^\dagger[1] & h_{A1}^\dagger[3]h_{B2}^\dagger[1] \end{bmatrix} \begin{bmatrix} u_A \\ v_A \end{bmatrix} + \begin{bmatrix} z_A[1] \\ z_A[3] - h_{A1}^\dagger[3]z_A[2] \end{bmatrix}. \quad (23)$$

Fig. 3. Achievable Scheme for $K = 3$: Phase Two.

2×2 case, the overheard equations can help us to improve the degrees of freedom.

Let us focus on the time-slot dedicated to receiver A . Then, we have the following observations.

- 1) The three equations $L_1(u_A, v_A, w_A)$, $L_2(u_A, v_A, w_A)$, and $L_3(u_A, v_A, w_A)$ form three linearly independent equations of u_A , v_A , and w_A , almost surely.
- 2) If we somehow deliver the overheard equations $L_2(u_A, v_A, w_A)$ and $L_3(u_A, v_A, w_A)$ to receiver A , then it has enough equations to solve for u_A , v_A , and w_A .
- 3) The two overheard equations $L_2(u_A, v_A, w_A)$ and $L_3(u_A, v_A, w_A)$ plus the equation received by receiver A i.e., $L_1(u_A, v_A, w_A)$, fully represent the original data symbols. Therefore, sufficient information to solve for the data symbols is already available at the receivers, but not exactly at the desired receiver.

We have similar observations about the equations received in the time-slots dedicated to receivers B and C . Remember that originally the objective was to deliver u_r , v_r , and w_r to receiver r . After these three transmissions, we can redefine the objective. The new objective is to deliver:

- 1) the overheard equations $L_2(u_A, v_A, w_A)$ and $L_3(u_A, v_A, w_A)$ to receiver A ,
- 2) the overheard equations $L_4(u_B, v_B, w_B)$ and $L_6(u_B, v_B, w_B)$ to receiver B , and
- 3) the overheard equations $L_7(u_C, v_C, w_C)$ and $L_8(u_C, v_C, w_C)$ to receiver C .

Let us define u_{AB} as a random linear combination of $L_2(u_A, v_A, w_A)$ and $L_4(u_B, v_B, w_B)$. To be specific, let $u_{AB} = L_2(u_A, v_A, w_A) + L_4(u_B, v_B, w_B)$. Then, we have the following observations.

- 1) If receiver A has u_{AB} , then it can use the saved overheard equation $L_4(u_B, v_B, w_B)$ to obtain $L_2(u_A, v_A, w_A)$. Remember $L_2(u_A, v_A, w_A)$ is a desired equation for receiver A .
- 2) If receiver B has u_{AB} , then it can use the saved overheard equation $L_2(u_A, v_A, w_A)$ to obtain $L_4(u_B, v_B, w_B)$. Remember $L_4(u_B, v_B, w_B)$ is a desired equation for receiver B .

Therefore, u_{AB} is desired by both receivers A and B . Similarly, we define $u_{AC} = L_3(u_A, v_A, w_A) + L_7(u_C, v_C, w_C)$, which is desired by receivers A and C , and define $u_{BC} = L_6(u_B, v_B, w_B) + L_8(u_C, v_C, w_C)$, which is desired by receivers B and C . We note that if receiver A has u_{AB} and u_{AC} ,

then it has enough equations to solve the original data symbols u_A , v_A , and w_A . Similarly, it is enough that receiver B has u_{AB} and u_{BC} , and receiver C has u_{AC} and u_{BC} . Therefore, again, we can redefine the objective as delivering u_{AB} to receivers A and B , u_{AC} to receivers A and C , and u_{BC} to receivers B and C . Suppose now we have an algorithm that can achieve $\text{DoF}_2(3, 3)$ degrees of freedom for order-2 common symbols. Then, the total time to deliver the original nine data symbols is the initial three time-slots of sending linear combinations of the nine symbols plus $\frac{3}{\text{DoF}_2(3, 3)}$ time-slots to deliver the three order-2 symbols generated. Therefore, the overall DoF to send the order-1 symbols is given by

$$\text{DoF}_1(3, 3) = \frac{9}{3 + (3/\text{DoF}_2(3, 3))}. \quad (24)$$

It is trivially easy to achieve $\text{DoF}_2(3, 3) = 1$, which yields $\text{DoF}_1(3, 3)$ of $\frac{3}{2}$. However, as we will elaborate in the following, we can do better.

Phase Two: Phase one of the algorithm takes order-1 symbols and generates order-2 symbols to be delivered. Phase two takes order-2 symbols and generates order-3 symbols. Phases two and three together can also be viewed as an algorithm which delivers order-2 common symbols.

Assume that u_{AB} and v_{AB} represent two symbols that are desired by both receivers A and B . Similarly, u_{AC} and v_{AC} are required by both receivers A and C , and u_{BC} and v_{BC} are required by both receivers B and C . Therefore, in total, there are six order-2 symbols. We notice that phase one generates only three order-2 symbols. To provide six order-2 symbols, we can simply repeat phase one twice with new input symbols. Phase two takes three time-slots, where each time-slot is dedicated to one pair of the receivers. In the time-slot dedicated to receivers A and B , the transmitter sends random linear combinations u_{AB} and v_{AB} from two of the transmit antennas. We have analogous transmissions in the other two time-slots. For details, see Fig. 3.

In Fig. 3, we focus on the first time-slot dedicated to both users A and B . Then, we have the following important observations.

- 1) $L_{10}(u_{AB}, v_{AB})$ and $L_{12}(u_{AB}, v_{AB})$ form two linearly independent equations of u_{AB} and v_{AB} , almost surely.
- 2) Similarly, $L_{11}(u_{AB}, v_{AB})$ and $L_{12}(u_{AB}, v_{AB})$ form two linearly independent equations of u_{AB} and v_{AB} , almost surely.
- 3) If $L_{12}(u_{AB}, v_{AB})$ is somehow delivered to both receivers A and B , then both receivers have enough equations to

solve for u_{AB} and v_{AB} . Therefore, $L_{12}(u_{AB}, v_{AB})$, which is overheard and saved by receiver C , is *simultaneously* useful for receivers A and B .

We have similar observations about the received equations in the other two time-slots. Therefore, after these three time-slots, we can redefine the objective of the rest of the algorithm as delivering

- 1) $L_{12}(u_{AB}, v_{AB})$ to receivers A and B ,
- 2) $L_{14}(u_{AC}, v_{AC})$ to receivers A and C , and
- 3) $L_{16}(u_{BC}, v_{BC})$ to receivers B and C .

Let us define u_{ABC} and v_{ABC} as any two linearly independent combinations of $L_{12}(u_{AB}, v_{AB})$ and $L_{14}(u_{AC}, v_{AC})$, and $L_{16}(u_{BC}, v_{BC})$:

$$\begin{aligned} u_{ABC} &= \alpha_1 L_{12}(u_{AB}, v_{AB}) + \alpha_2 L_{14}(u_{AC}, v_{AC}) \\ &\quad + \alpha_3 L_{16}(u_{BC}, v_{BC}), \\ v_{ABC} &= \beta_1 L_{12}(u_{AB}, v_{AB}) + \beta_2 L_{14}(u_{AC}, v_{AC}) \\ &\quad + \beta_3 L_{16}(u_{BC}, v_{BC}) \end{aligned}$$

where the constants α_i and β_i , $i = 1, 2, 3$, have been shared with receivers. If we somehow deliver u_{ABC} and v_{ABC} to receiver A , then together with its saved overheard equation $L_{16}(u_{BC}, v_{BC})$, receiver A has three linearly independent equations to solve for $L_{12}(u_{AB}, v_{AB})$ and $L_{14}(u_{AC}, v_{AC})$. Then, it has enough equations to solve for u_{AB} , v_{AB} , u_{AC} , and v_{AC} . We have the similar situation for receivers B and C . Therefore, it is enough to deliver u_{ABC} and v_{ABC} to all three receivers. If we have an algorithm that can provide $\text{DoF}_3(3, 3)$ degrees of freedom to deliver order-3 common symbols, then the total time to deliver the original six order-2 common symbols is $3 + \frac{2}{\text{DoF}_3(3, 3)}$, taking into account the first three transmissions (described in Fig. 3). Therefore, we have

$$\text{DoF}_2(3, 3) = \frac{6}{3 + \frac{2}{\text{DoF}_3(3, 3)}}. \quad (25)$$

Phase Three: Phase Three transmits order-3 common symbols. This phase is very simple. Assume that u_{ABC} is required by all three receivers. Then, the transmitter can use only one transmit antenna and send u_{ABC} . All three receivers will receive a noisy version of u_{ABC} . Therefore, we use one time-slot to send one order-3 symbol. Therefore, $\text{DoF}_3(3, 3) = 1$. Then, from (24) and (25), we conclude that $\text{DoF}_1(3, 3) = \frac{18}{11}$ and $\text{DoF}_2(3, 3) = \frac{6}{5}$.

C. General Proof of Achievability for Theorem 1

In this section, we explain the achievable scheme for the general case in Theorem 1.

First, we focus on the general $M = K$ square case. The algorithm is based on a concatenation of K phases. Phase j takes symbols of order j and generates symbols of order $j + 1$. For $j = K$, the phase is simple and generates no more symbols. For each j , we can also view phases $j, j + 1, \dots, K$ together, as an algorithm whose job is to deliver common symbols of order j to the receivers.

The j th phase takes $(K - j + 1)\binom{K}{j}$ common symbols of order j , and yields $j\binom{K}{j+1}$ symbols of order $j + 1$. This phase has $\binom{K}{j}$ time-slots, with each time-slot dedicated to a

subset \mathcal{S} of receivers, $|\mathcal{S}| = j$. We denote the time-slot dedicated to the subset \mathcal{S} by $t_{\mathcal{S}}$. In this time-slot, the transmitter sends random linear combinations of the $K - j + 1$ symbols $u_{\mathcal{S},1}, u_{\mathcal{S},2}, \dots, u_{\mathcal{S},K-j+1}$, desired by all the receivers in \mathcal{S} . The transmitter utilizes $K - j + 1$ of the transmit antennas.

The linear combination of the transmitted symbols received by receiver r is denoted by $L_{\mathcal{S},r}$. Let us focus on the linear combinations of the transmitted symbols received by all receivers, in time-slot $t_{\mathcal{S}}$. We have the following observations.

- 1) For every $r \in \mathcal{S}$, the $K - j + 1$ equations consisting of one equation $L_{\mathcal{S},r}$ and the $K - j$ overheard equations: $\{L_{\mathcal{S},r'} : r' \in \mathcal{E} \setminus \mathcal{S}\}$ are linearly independent equations of the $K - j + 1$ symbols $u_{\mathcal{S},1}, u_{\mathcal{S},2}, \dots, u_{\mathcal{S},K-j+1}$. This relies on the fact that the transmitter uses $K - j + 1$ transmit antennas.
- 2) For any $r, r \in \mathcal{S}$, if we somehow deliver the $K - j$ equations $\{L_{\mathcal{S},r'} : r' \in \mathcal{E} \setminus \mathcal{S}\}$ to receiver r , then receiver r has $K - j + 1$ linearly independent equations to solve for all $K - j + 1$ symbols $u_{\mathcal{S},1}, u_{\mathcal{S},2}, \dots, u_{\mathcal{S},K-j+1}$.
- 3) Having the aforementioned two observations, we can say that the overheard equation by receiver r' , $r' \in \mathcal{E} \setminus \mathcal{S}$, is *simultaneously* useful for all receivers in \mathcal{S} .

After repeating the aforementioned transmission for all \mathcal{S} , where $\mathcal{S} \subset \mathcal{E}$ and $|\mathcal{S}| = j$, we have another important observation. Consider any subset \mathcal{T} of receivers, where $|\mathcal{T}| = j + 1$. Then, each receiver $r, r \in \mathcal{T}$, has an overheard equation $L_{\mathcal{T} \setminus \{r\},r}$, which is simultaneously useful for all the receivers in $\mathcal{T} \setminus \{r\}$. We note that the transmitter is aware of these overheard equations. For every $\mathcal{T} \subset \mathcal{E}$, $|\mathcal{T}| = j + 1$, the transmitter forms j random linear combinations of $L_{\mathcal{T} \setminus \{r\},r}$, $r \in \mathcal{T}$, denoted by $u_{\mathcal{T},1}, u_{\mathcal{T},2}, \dots, u_{\mathcal{T},j}$. We note that $u_{\mathcal{T},\xi}$, $1 \leq \xi \leq j$, is simultaneously useful for all receivers in \mathcal{T} . Indeed, each receiver r in \mathcal{T} can subtract the contribution of $L_{\mathcal{T} \setminus \{r\},r}$ from $u_{\mathcal{T},\xi}$, $\xi = 1, \dots, j$, and form j linearly independent combinations of $L_{\mathcal{T} \setminus \{r\},r}$, $r \in \mathcal{T} \setminus \{r\}$. Using the previous procedure, the transmitter generates $j\binom{K}{j+1}$ symbols of order $j + 1$. The important observation is that if these $j\binom{K}{j+1}$ symbols are delivered to the designated receivers, then each receiver will have enough equations to solve for all of the original common symbols of order j . Delivering $j\binom{K}{j+1}$ order— $(j + 1)$ symbols takes $\frac{j\binom{K}{j+1}}{\text{DoF}_{j+1}(K, K)}$ using an algorithm that provides $\text{DoF}_{j+1}(K, K)$ degrees of freedom for order- $(j + 1)$ symbols. Since the phase starts with $(K - j + 1)\binom{K}{j}$ symbols of order j , and takes $\binom{K}{j}$ time-slots, and generates $j\binom{K}{j+1}$ symbols with order $j + 1$, we have

$$\text{DoF}_j(K, K) = \frac{(K - j + 1)\binom{K}{j}}{\binom{K}{j} + \frac{j\binom{K}{j+1}}{\text{DoF}_{j+1}(K, K)}} \quad (26)$$

or

$$\frac{K - j + 1}{j} \frac{1}{\text{DoF}_j(K, K)} = \frac{1}{j} + \frac{K - j}{j + 1} \frac{1}{\text{DoF}_{j+1}(K, K)}. \quad (27)$$

It is also easy to see that $\text{DoF}_K(K) = 1$ is achievable. Solving the recursive equation, we have

$$\text{DoF}_j(K, K) = \frac{K - j + 1}{j} \frac{1}{\frac{1}{j} + \frac{1}{j+1} + \dots + \frac{1}{K}}. \quad (28)$$

In particular

$$\text{DoF}_1(K, K) = \frac{K}{1 + \frac{1}{2} + \dots + \frac{1}{K}}. \quad (29)$$

Therefore, the achievability of Theorem 1 in the square case has been established.

Now observe that in the aforementioned algorithm, phase j only requires the use of $K - j + 1$ transmit antennas, not all K of the transmit antennas. Moreover, common symbols of order j are delivered using phases $j, j + 1, \dots, K$. Hence, we conclude that the degree of freedom of order- j messages achieved above in the square system can actually be achieved in a system with less transmit antennas as long as $M \geq K - j + 1$. This proves Theorem 1 in the rectangular case as well.

Remark: We note that if the transmitter has KM transmit antennas, and each of the K receivers has N receive antennas, then the DoF_1 of

$$\frac{KN}{1 + \frac{1}{2} + \dots + \frac{1}{K}}$$

is achievable. More generally, in this channel, for order- j symbols, the DoF_j of

$$\frac{K - j + 1}{j} \frac{N}{\frac{1}{j} + \dots + \frac{1}{K}}$$

is achievable.

D. Implementation Issues

For simplicity, the proposed scheme has been presented in a symbol-by-symbol based format. However, this scheme can be implemented in a block-by-block fashion as well. This would allow us to exploit the coherence of the channel over time and frequency to reduce channel training and feedback overhead. To be specific, let us again focus on the case of $M = K = 2$.

Consider a block of time-frequency resources, consecutive in time and frequency. Let us assume that in the first phase of the scheme, we dedicate half of these resources to receiver A and the other half to receiver B . To start the second phase, the transmitter needs to know channel coefficients during the first phase. For example, if the lengths of the block in time and frequency are respectively less than coherent time and bandwidth of the channel, then during the first phase the channel coefficients are (almost) constant. Therefore, to start the second phase, the transmitter needs only to know the four channel coefficients. Let us denote the coherent time and bandwidth by T_c and W_c , respectively. Then, for each $T_c W_c$ time-frequency resources, the transmitter needs to dedicate at least two time-frequency resources to send orthogonal pilot signals and learn four coefficients through feedback. Then, the transmitter uses the remaining resources to send $2T_c W_c - 2$ order-1 symbols. Remember that the transmitter

is also required to report the channel coefficients of each receiver to the other receiver. Since each receiver knows its own CSI, the transmitter can exploit that and send to both receivers the two symbols of $h_{A1}[1] + h_{B1}[1]$ and $h_{A2}[1] + h_{B2}[1]$, as the symbols of order-2 in the second phase. Therefore, the second phase takes $\frac{2T_c W_c - 2}{4} + 2$ resource units for order-2 messages. Following the previous argument, the scheme can achieve DoF of $\frac{2T_c W_c - 2}{\frac{3}{2}T_c W_c + 0.5}$. If $T_c W_c \gg 1$, as in most wireless channels, then the degree of freedom is close to $4/3$.

IV. OUTER-BOUND

In this section, we aim to prove Theorem 2. In this theorem, we focus on the degrees of freedom of the channel for order- j messages. Therefore, we assume for every subset \mathcal{S} with cardinality j of receivers, the 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 CSI is available to all nodes with one time-unit delay. As an outer-bound, we consider the capacity of a channel in which the CSI at time n is available to all receivers instantaneously at time n . Therefore, at time n , receiver r has $(y_r[t], \mathbf{H}[t])$, $t = 1, \dots, m$, for any r , $1 \leq r \leq K$. On the other hand, the transmitter has not only the CSI, but also received signals, both with one unit delay. Therefore, at time n , the transmitter has $(y_1[t], \dots, y_K[t], \mathbf{H}[t])$, $t = 1, \dots, m - 1$. Now, we improve the resultant channel even further as follows.

Consider a permutation π of the set $\mathcal{E} = \{1, 2, \dots, K\}$. We form a K -receiver broadcast channel, by giving the output of the receiver $\pi(i)$ to the receivers $\pi(j)$, $j = i + 1, \dots, K$, for all $i = 1, \dots, K - 1$. Therefore, we have an upgraded broadcast channel, referred to as *improved channel* with K receivers as shown at the bottom of the page. We denote the capacity of the resultant channel as $\mathcal{C}_{\text{Improved}}(\pi)$. Denoting the capacity of the original channel with \mathcal{C} , we obviously have $\mathcal{C} \subset \mathcal{C}_{\text{Improved}}(\pi)$. Moreover, it is easy to see that the improved channel is physically degraded.

In the improved channel, consider message $W_{\mathcal{S}}$, which is required by all j receivers 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 receiver $\pi(i^*)$, then it can be decoded by all other receivers in \mathcal{S} . Therefore, we can assume that $W_{\mathcal{S}}$ is just required by receiver $\pi(i^*)$. Using this argument, we can simplify the messages requirements from order- j common messages to pure private messages as follows: receiver $\pi(1)$ requires all messages $W_{\mathcal{S}}$, where $\pi(1) \in \mathcal{S}$ and $\mathcal{S} \in \mathcal{E}$. Similarly, receiver $\pi(2)$ requires all messages $W_{\mathcal{S}}$, where $\pi(2) \in \mathcal{S}$ and $\mathcal{S} \subset \mathcal{E} \setminus \{\pi(1)\}$. We follow the same argument for all receivers.

According to [7], feedback does not improve the capacity of the physically degraded broadcast channels. Consequently, we focus on the capacity region of the improved channel without feedback, and with the new private message set. On the other

$$(y_{\pi(1)}[n], \mathbf{H}[n]), (y_{\pi(1)}[n], y_{\pi(2)}[n], \mathbf{H}[n]), \dots, (y_{\pi(1)}[n], y_{\pi(2)}[n], \dots, y_{\pi(K)}[n], \mathbf{H}[n])$$

hand, for broadcast channels without feedback, the capacity region is only a function of marginal distributions. Therefore, we can ignore the coupling between the receivers in the improved channel. Thus, we have a broadcast channel where receiver $\pi(i)$ has i antennas, and the distributions of the channels between the transmitter and any of the receive antennas are identical. Moreover receiver $\pi(i)$ is interested in all messages W_S , where $\pi(i) \in S$, $|S| = j$, and $S \subset \mathcal{E} \setminus \{\pi(1), \pi(2), \dots, \pi(i-1)\}$.

Therefore, according to Huang *et al.* [8], extended by Vaze and Varanasi [9], one can conclude that

$$\sum_{i=1}^{K-j+1} \frac{1}{\min\{i, M\}} \sum_{\substack{|S|=j \\ S \subset \mathcal{E} \setminus \{\pi(1), \dots, \pi(i-1)\} \\ \pi(i) \in S}} d_S \leq 1. \quad (30)$$

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

V. DoF REGION FOR $K = M$

In this section, we prove Theorem 3 which characterizes the DoF region of the channel for the case $M = K$.

We note that the region of Theorem 3 is the polyhedron proposed by the outer-bound (30) for order-1 messages where $M = K$. Here, we show by induction on K that the region is achievable. The hypothesis is clearly true for $K = 1$. Now assume that the hypothesis is true for $K = 1, \dots, k-1$. Consider the case when $K = k$. First we argue that any point $(\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_k)$ in the polyhedron such that $\tilde{d}_i > 0$ for all i and $\tilde{d}_i \neq \tilde{d}_j$ for some i, j cannot be a corner point of the polyhedron. Without loss of generality, we can assume that the coordinates of such a point is ordered in a nondecreasing order, since the polyhedron is invariant to permutation of coordinates. Let i_1, i_2 be such that either $0 < \tilde{d}_1 = \dots = \tilde{d}_{i_1} < \tilde{d}_{i_1+1} = \dots = \tilde{d}_{i_2} < \tilde{d}_{i_2+1}$, or $0 < \tilde{d}_1 = \dots = \tilde{d}_{i_1} < \tilde{d}_{i_1+1} = \dots = \tilde{d}_{i_2}$ and $i_2 = k$. Now a direct calculation shows that π is a permutation of $\{1, \dots, k\}$ which maximizes

$$\sum_{i=1}^k \frac{\tilde{d}_{\pi(i)}}{i}$$

among all permutations if and only if $\tilde{d}_{\pi(i)} > \tilde{d}_{\pi(j)}$ whenever $i < j$. This means that the only constraints, if any, of the polyhedron that $(\tilde{d}_1, \dots, \tilde{d}_k)$ satisfies with equality correspond to permutations satisfying $\pi(j) \in \{1, \dots, i_1\}$ for all $j \in \{k - i_1 + 1, \dots, k\}$ and $\pi(j) \in \{i_1 + 1, \dots, i_2\}$ for all $j \in \{k - i_2 + 1, \dots, k - i_1\}$. All other constraints are satisfied with strict inequality. We define vector (e_1, \dots, e_k) as

$$\begin{cases} e_i = \frac{\epsilon}{\sum_{i=k-i_1+1}^k \frac{1}{i}}, & \text{for } i = 1, \dots, i_1 \\ e_i = \frac{-\epsilon}{\sum_{i=k-i_2+1}^{k-i_1} \frac{1}{i}}, & \text{for } i = i_1 + 1, \dots, i_2 \\ e_i = 0, & \text{otherwise.} \end{cases} \quad (31)$$

An explicit calculation shows that for any $\epsilon > 0$, both the point

$$(\tilde{d}_1, \dots, \tilde{d}_k) + (e_1, \dots, e_k)$$

and the point

$$(\tilde{d}_1, \dots, \tilde{d}_k) - (e_1, \dots, e_k)$$

continue to satisfy the tight inequalities with equality. Moreover, for ϵ sufficiently small, the constraints that are not tight on $(\tilde{d}_1, \dots, \tilde{d}_k)$ remain not tight on these two points. Hence, both these points lie in the polyhedron, and hence $(\tilde{d}_1, \dots, \tilde{d}_k)$, which is the average between these points, cannot be a corner point.

Thus, the only point in the strict positive quadrant that can be a corner point of the polyhedron is the point:

$$\frac{1}{\sum_{i=1}^k \frac{1}{i}} (1, 1, \dots, 1).$$

This point is achievable by Theorem 1. Any other point in the polyhedron is a convex combination of this point and points for which some of the coordinates are zero. Each one of these latter points is in fact in the polyhedron for some smaller value of $K = k' < k$. By the induction hypothesis, each of these points is achievable. Hence, by time sharing, any point in the polyhedron for $K = k$ is achievable.

VI. ACHIEVABLE SCHEME FOR THEOREM 4

In Section III, we explained an algorithm to achieve $\text{DoF}_1^*(M, K)$, when $M \geq K$. More generally, we characterized $\text{DoF}_j^*(M, K)$, when $M \geq K - j + 1$. In this section, we extend the optimal achievable scheme of Section III and develop a suboptimal algorithm for the case that $M < K - j + 1$ for order- j messages. We first focus on the case $M = 2$ and $K = 3$.

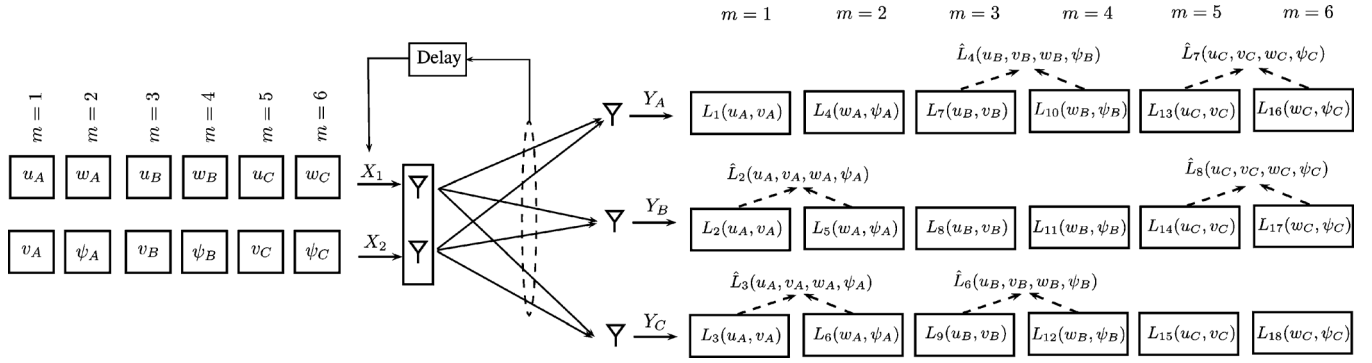
A. Achievable Scheme for $M = 2, K = 3$

From Theorems 1 and 2, we have $\text{DoF}_2^*(2, 3) = \frac{6}{5}$ and $\text{DoF}_3^*(2, 3) = 1$. However, for order-1 messages, we only know from the outer-bound that $\text{DoF}_1^*(2, 3) \leq \frac{3}{2}$. On the other hand, in terms of achievability, it is easy to see that $\text{DoF}_1^*(2, 3) \geq \text{DoF}_1^*(2, 2) = \frac{4}{3}$ which can be achieved by simply ignoring one of the receivers. Now the question is whether $\text{DoF}_1^*(2, 3)$ is indeed the same as $\text{DoF}_1^*(2, 2)$ or the extra receiver can be exploited to achieve DoF beyond $\text{DoF}_1^*(2, 2)$. Here, we propose an algorithm to show that $\text{DoF}_1^*(2, 3) > \text{DoF}_1^*(2, 2)$.

The achievable scheme is as follows. Let u_r, v_r, w_r , and ψ_r be four symbols for receiver r , $r = A, B, C$. The first phase of the scheme has six time-slots. The first two time-slots are dedicated to receiver A. In these two time-slots, the transmitter sends four random linear combinations of u_A, v_A, w_A , and ψ_A through the two transmit antennas. As a particular example, in the first time slot, the transmitter sends u_A and v_A , and in the second time slot, it sends w_A and ψ_A . Refer to Fig. 4 for details. Similarly, in time-slots 3 and 4, the transmitter sends four random linear combinations of u_B, v_B, w_B , and ψ_B . In time-slots 5 and 6, the transmitter sends four random linear combinations of u_C, v_C, w_C , and ψ_C .

Referring to Fig. 4, we have the following observations.

- 1) Receiver A already has two independent linear equations $L_1(u_A, v_A)$ and $L_4(w_A, \psi_A)$ of u_A, v_A, w_A , and ψ_A . Therefore, it needs two more equations.

Fig. 4. Suboptimal Scheme for $M = 2$ and $K = 3$, The First Phase.

- 2) The four overheard equations in $L_2(u_A, v_A)$, $L_3(u_A, v_A)$, $L_5(w_A, \psi_A)$, and $L_6(w_A, \psi_A)$ are not linearly independent from what receiver A has already received, i.e., $L_1(u_A, v_A)$ and $L_4(w_A, \psi_A)$.
- 3) We can purify the four overheard equations and form two equations that are linearly independent with $L_1(u_A, v_A)$ and $L_4(w_A, \psi_A)$. For example, receiver B can form $\hat{L}_2(u_A, v_A, w_A, \psi_A)$ as a random linear combination of $L_2(u_A, v_A)$ and $L_5(w_A, \psi_A)$. Similarly, receiver C can form $\hat{L}_3(u_A, v_A, w_A, \psi_A)$ as a random linear combination of $L_3(u_A, v_A)$ and $L_6(w_A, \psi_A)$. The coefficients of these linear combinations have been preselected and shared among all nodes.
- 4) It is easy to see that almost surely, $\hat{L}_2(u_A, v_A, w_A, \psi_A)$ and $\hat{L}_3(u_A, v_A, w_A, \psi_A)$ are linearly independent of $L_1(u_A, v_A)$ and $L_4(w_A, \psi_A)$.
- 5) If somehow we deliver $\hat{L}_2(u_A, v_A, w_A, \psi_A)$ and $\hat{L}_3(u_A, v_A, w_A, \psi_A)$ to receiver A , then it has enough equations to solve for u_A, v_A, w_A , and ψ_A .

Similarly, as shown in Fig. 4, we can purify the overheard equations in time-slots dedicated to receivers B and C . Now, the available side information and the requirements are the same as those we had after phase one for the case of $M = K = 3$ (see Section III-B). Equations $\hat{L}_2(u_A, v_A, w_A, \psi_A)$ and $\hat{L}_3(u_A, v_A, w_A, \psi_A)$ are available at receivers B and C , respectively, and are needed by receiver A , equations $\hat{L}_4(u_B, v_B, w_B, \psi_B)$ and $\hat{L}_6(u_B, v_B, w_B, \psi_B)$ are available at receivers A and C , respectively, and are needed by receiver B , and equations $\hat{L}_7(u_C, v_C, w_C, \psi_C)$ and $\hat{L}_8(u_C, v_C, w_C, \psi_C)$ are available at receivers A and B , respectively, and are needed by receiver C . We define

$$u_{AB} = \hat{L}_2(u_A, v_A, w_A, \psi_A) + \hat{L}_4(u_B, v_B, w_B, \psi_B) \quad (32)$$

$$u_{AC} = \hat{L}_3(u_A, v_A, w_A, \psi_A) + \hat{L}_7(u_C, v_C, w_C, \psi_C) \quad (33)$$

$$u_{BC} = \hat{L}_6(u_B, v_B, w_B, \psi_B) + \hat{L}_8(u_C, v_C, w_C, \psi_C). \quad (34)$$

Considering the available overheard equations at each receiver, one can easily conclude that u_{AB} is needed by both receivers A and B , u_{AC} is needed by both receivers A and C , and u_{BC} is needed by both receivers B and C . The transmitter needs $\frac{3}{\text{DoF}_2^*(2,3)}$ time-slots to deliver these three order-2 symbols, where according to Theorem 1, $\text{DoF}_2^*(2,3) = \frac{6}{5}$. In summary, phase one starts with 12 order-1 messages, takes six time-slots, and generates 3 order-2 symbols. Therefore, we achieve

$$\text{DoF}_1(2,3) = \frac{12}{6 + \frac{3}{\text{DoF}_2^*(2,3)}} = \frac{24}{17} \quad (35)$$

which is strictly greater than $\text{DoF}_1^*(2,2) = \frac{4}{3}$. Therefore, the proposed achievable scheme exploits the extra receiver to improve DoF. However, we notice that the achieved $\text{DoF}_1(2,3)$ of $\frac{24}{17}$ is still less than $\frac{3}{2} = \frac{24}{16}$ which is suggested by the outer-bound.

B. General Proof for Theorem 4

Here, we explain a general version of the proposed algorithm. Again the algorithm includes $K - j + 1$ phases. Phase j takes symbols of order j (meaning that it is needed by j receivers simultaneously), and generates symbols of order $j + 1$. For $j = K$, the phase is simple and generates no more symbols.

Let us define q_j as

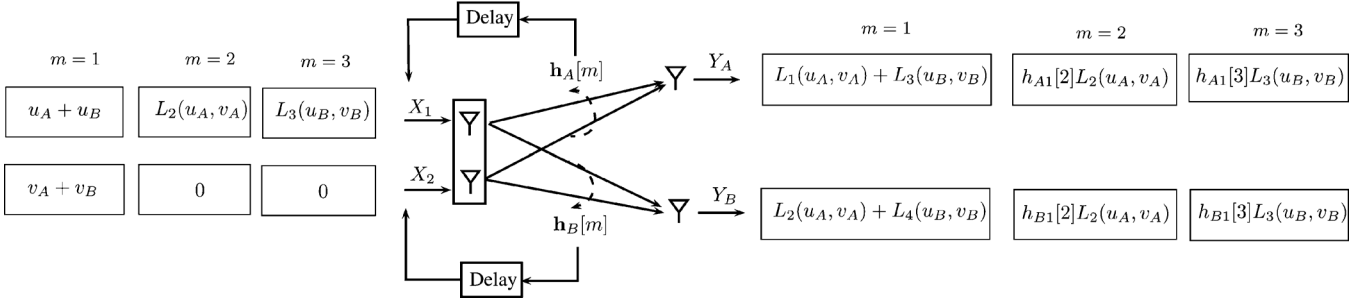
$$q_j = \min\{M - 1, K - j\}. \quad (36)$$

In addition, we define η_j as the greatest common factor of q_j and $K - j$, i.e.,

$$\eta_j = \text{gcf}\{q_j, K - j\}. \quad (37)$$

Phase j takes $(K - j) \frac{q_j + 1}{\eta_j} \binom{K}{j}$ symbols of order j and yields $j \frac{q_j}{\eta_j} \binom{K}{j+1}$ symbols with order $j + 1$. This phase has $\binom{K}{j}$ subphases, where each subphase is dedicated to a subset \mathcal{S} of the receivers, $|\mathcal{S}| = j$. The subphase dedicated to subset \mathcal{S} is denoted by S-Ph(\mathcal{S}). Each subphase takes $\frac{K-j}{\eta_j}$ time-slots. In S-Ph(\mathcal{S}), the transmitter sends random linear combinations of $\beta_j = \frac{(q_j+1)(K-j)}{\eta_j}$ symbols $u_{\mathcal{S},1}, u_{\mathcal{S},2}, \dots, u_{\mathcal{S},\beta_j}$, desired by all receivers in \mathcal{S} . The transmitter uses at least $q_j + 1$ of the transmit antennas. The linear equation of the transmitted symbols received by receiver r , in the t th time-slot of S-Ph(\mathcal{S}), is denoted by $L_{\mathcal{S},r}(t)$. Let us focus on the equations of the transmitted symbols received by all receivers in S-Ph(\mathcal{S}). We have the following observations.

- 1) For every r , $r \in \mathcal{S}$, and t , $t \in \{1, 2, \dots, \frac{K-j}{\eta_j}\}$, the $K - j + 1$ equations $\{L_{\mathcal{S},r'}(t), r' \in \{r\} \cup \mathcal{E} \setminus \mathcal{S}\}$ are not necessarily linearly independent. The reason is that $|\{r\} \cup \mathcal{E} \setminus \mathcal{S}| = K - j + 1$, while the number of transmit antennas is M which can be less than $K - j + 1$. Indeed, among the $K - j$ overheard equations $L_{\mathcal{S},r'}(t)$, $r' \in \mathcal{E} \setminus \mathcal{S}$,

Fig. 5. Alternative Achievable Scheme for $M = K = 2$.

we can only form q_j overheard equations that are simultaneously useful to receiver r , for any r in \mathcal{S} . Therefore, among $\frac{(K-j)^2}{\eta_j}$ overheard equations in $\mathcal{S}\text{-Ph}(\mathcal{S})$, we can form only $\frac{q_j(K-j)}{\eta_j}$ overheard equations that are useful for any receiver $r, r \in \mathcal{S}$.

- 2) We purify the overheard linear combinations. To this end, receiver $r', r' \in \mathcal{E} \setminus \mathcal{S}$, forms $\frac{q_j}{\eta_j}$ linear combinations of $L_{\mathcal{S}, r'}(t), t = 1, \dots, \frac{K-j}{\eta_j}$. The resultant equations are denoted by $\hat{L}_{\mathcal{S}, r'}(1), \hat{L}_{\mathcal{S}, r'}(2), \dots, \hat{L}_{\mathcal{S}, r'}\left(\frac{q_j}{\eta_j}\right)$. The coefficients of the linear combinations have been preselected and shared among all nodes. It is easy to see that for every r , the following $\frac{(q_j+1)(K-j)}{\eta_j}$ equations are linearly independent: $L_{\mathcal{S}, r}(t), t = 1, \dots, \frac{K-j}{\eta_j}$, and $\hat{L}_{\mathcal{S}, r'}(\hat{t}), r' \in \mathcal{E} \setminus \mathcal{S}$ and $\hat{t} \in 1, \dots, \frac{q_j}{\eta_j}$. Therefore, if we somehow deliver $\hat{L}_{\mathcal{S}, r'}(\hat{t}), r' \in \mathcal{E} \setminus \mathcal{S}$ and $\hat{t} \in 1, \dots, \frac{q_j}{\eta_j}$ to receiver $r, r \in \mathcal{S}$, then it will have $\beta_j = \frac{(q_j+1)(K-j)}{\eta_j}$ linearly independent equations to solve for all desired symbols $u_{\mathcal{S}, 1}, u_{\mathcal{S}, 2}, \dots, u_{\mathcal{S}, \beta_j}$.
- 3) Having the previous two observations, we note that the purified linear combinations by receiver $r', r' \in \mathcal{E} \setminus \mathcal{S}$, are simultaneously useful for all receivers in \mathcal{S} .

After repeating the previous transmission for all \mathcal{S} , where $\mathcal{S} \subset \mathcal{E}$ and $|\mathcal{S}| = j$, we have another important property. Consider a subset \mathcal{T} of the receivers, where $|\mathcal{T}| = j+1$. Then, each receiver $r, r \in \mathcal{T}$, has $\frac{q_j}{\eta_j}$ purified linear combination $\hat{L}_{\mathcal{T} \setminus \{r\}, r}(t), t = 1, \dots, \frac{q_j}{\eta_j}$, which are simultaneously useful for all receivers in $\mathcal{T} \setminus \{r\}$. We note that the transmitter is aware of these purified equations through delayed CSIT. For every $\mathcal{T} \subset \mathcal{E}, |\mathcal{T}| = j+1$, the transmitter forms $j \frac{q_j}{\eta_j}$ random linear combinations of $\hat{L}_{\mathcal{T} \setminus \{r\}, r}(t), r \in \mathcal{T}, t = 1, \dots, \frac{q_j}{\eta_j}$, denoted by $u_{\mathcal{T}, 1}, u_{\mathcal{T}, 2}, \dots, u_{\mathcal{T}, j \frac{q_j}{\eta_j}}$. We note that $u_{\mathcal{T}, \xi}, 1 \leq \xi \leq j \frac{q_j}{\eta_j}$, is simultaneously useful for all receivers in \mathcal{T} . The reason is that each receiver $r, r \in \mathcal{T}$, can subtract the contributions of $\hat{L}_{\mathcal{T} \setminus \{r\}, r}(t), t = 1, \dots, \frac{q_j}{\eta_j}$, from $u_{\mathcal{T}, \xi}, \xi = 1, \dots, j \frac{q_j}{\eta_j}$, and form $j \frac{q_j}{\eta_j}$ linearly independent combinations of $\hat{L}_{\mathcal{T} \setminus \{r'\}, r'}(t), r' \in \mathcal{T} \setminus \{r\}, t = 1, \dots, \frac{q_j}{\eta_j}$. Therefore, using the aforementioned procedure, the transmitter forms $j \frac{q_j}{\eta_j} \binom{K}{j+1}$ symbols with order $j+1$. The important observation is that if these $j \frac{q_j}{\eta_j} \binom{K}{j+1}$ symbols are delivered to the designated receivers, then each receiver will have enough equations to solve for all designated messages with order j .

In summary, this phase takes $(K-j) \frac{q_j+1}{\eta_j} \binom{K}{j}$ symbols of order j , takes $\frac{K-j}{\eta_j} \binom{K}{j}$ time-slots, and yields $j \frac{q_j}{\eta_j} \binom{K}{j+1}$ symbols of order $j+1$. If we have a scheme which achieves $\text{DoF}_{j+1}(M, K)$ for order $(j+1)$ symbols, then we achieve $\text{DoF}_j(M, K)$

$$\text{DoF}_j(M, K) = \frac{(K-j) \frac{q_j+1}{\eta_j} \binom{K}{j}}{\frac{K-j}{\eta_j} \binom{K}{j} + \frac{j \frac{q_j}{\eta_j} \binom{K}{j+1}}{\text{DoF}_{j+1}(M, K)}} \quad (38)$$

or

$$\frac{q_j+1}{j} \frac{1}{\text{DoF}_j(M, K)} = \frac{1}{j} + \frac{q_j}{j+1} \frac{1}{\text{DoF}_{j+1}(M, K)}. \quad (39)$$

VII. IMPROVED SCHEME FOR $M = 2$

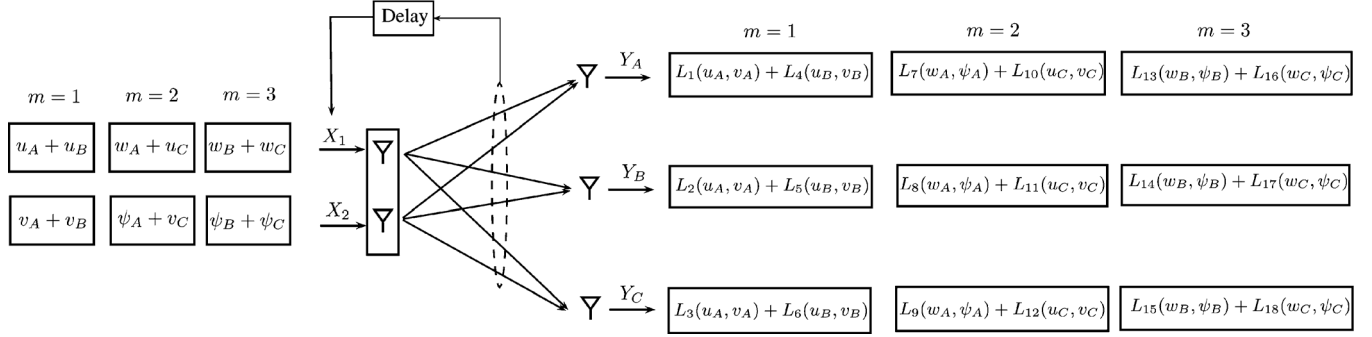
Recall that the scheme of Section VI achieves $\text{DoF}_1(2, 3)$ of $\frac{24}{17}$. The achieved DoF is greater than $\text{DoF}_1^*(2, 2) = \frac{4}{3}$, which shows that we could exploit the extra receiver with respect to the number of transmit antennas. However, it is still smaller than $\frac{3}{2}$ which is suggested by the outer-bound. Now the question is whether the achievable scheme or the outer-bound is loose.

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. Before that, we explain an alternative solution for a system with $M = K = 2$. The idea of the alternative solution is the key to achieve the optimal DoF for the systems with $M = 2$ and $K = 3$.

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

Phase one of the algorithm takes order-1 messages. Let us assume that the transmitter has u_A and v_A for receiver A and u_B and v_B for receiver B. Here, phase one takes only one time-slot which is dedicated to both receivers. In this time-slot, the transmitter sends random linear combinations of all four symbols u_A and v_A, u_B , and v_B . Refer to Fig. 5 to see the details of particular examples for the linear combinations. Receiver A receives a linear combination of all four symbols. We denote this linear combination by $L_1(u_A, v_A) + L_3(u_B, v_B)$, where $L_1(u_A, v_A)$ represents the contribution of u_A and v_A , and $L_3(u_B, v_B)$ represents the contribution of u_B and v_B . Similarly, receiver B receives a linear combination of all four symbols denoted by $L_2(u_A, v_A) + L_4(u_B, v_B)$.

Then, we have the following observations.

Fig. 6. Optimal Scheme for a System with $M = 2$ and $K = 3$, The First Phase.

- 1) If we somehow give $L_3(u_B, v_B)$ to receiver A, then receiver A can compute $L_1(u_A, v_A)$ by subtracting $L_3(u_B, v_B)$ from what it already has. Then if we also give $L_2(u_A, v_A)$ to receiver A, then it has two equations to solve for u_A and v_A .
- 2) If we somehow give $L_2(u_A, v_A)$ to receiver B, then receiver B can compute $L_4(u_B, v_B)$ by subtracting $L_2(u_A, v_A)$ from what it already has. Then if we also give $L_3(u_B, v_B)$ to receiver B, then it has two equations to solve for u_B and v_B .

In other words, both receivers A and B want $L_2(u_A, v_A)$ and $L_3(u_B, v_B)$. Therefore, we can define two order-2 symbols u_{AB} and v_{AB} as

$$u_{AB} = L_2(u_A, v_A) \quad (40)$$

$$v_{AB} = L_3(u_B, v_B). \quad (41)$$

In summary, this phase starts with four order-1 symbols, takes one time-slot, and provides two order-2 symbols. Two order-2 symbols take $\frac{2}{\text{DoF}_2^*(2,2)}$ time-slots to deliver. Therefore, we achieve

$$\text{DoF}_1(2, 2) = \frac{4}{1 + \frac{2}{\text{DoF}_2^*(2,2)}}. \quad (42)$$

Since $\text{DoF}_2^*(2, 2) = 1$, this scheme achieves $\text{DoF}_1^*(2, 2) = \frac{4}{3}$.

B. Optimal Scheme for $M = 2$ and $K = 3$

Here, we explain an algorithm for the systems with $M = 2$ and $K = 3$. The first phase of this algorithm takes 12 order-1 messages, takes three time-slots, and gives 6 order-2 symbols. This subalgorithm leads to an optimal scheme for systems with $M = 2$ and $K = 3$.

Let u_r, v_r, w_r , and ψ_r be four symbols for receiver $r, r = A, B, C$. In the first time slot, which is dedicated to receivers A and B, the transmitter sends random linear combinations of four symbols u_A and v_A, u_B , and v_B . Refer to Fig. 6 to see the details of particular realizations for the linear combinations. Receiver A receives a linear combination of all four symbols denoted by $L_1(u_A, v_A) + L_4(u_B, v_B)$. Receivers B and C also receive linear combinations of all four symbols denoted by $L_2(u_A, v_A) + L_5(u_B, v_B)$ and $L_3(u_A, v_A) + L_6(u_B, v_B)$, respectively. In the second time slot, which is dedicated to receivers A and C, the transmitter sends random linear combinations of four symbols w_A and ψ_A, u_C , and v_C . In the third

time slot, which is dedicated to receivers B and C, the transmitter sends random linear combinations of four symbols w_B, ψ_B, w_C , and ψ_C .

By referring to Fig. 6, it is easy to see that for each receiver to solve for all four desired symbols, it is enough that

- 1) receiver A has $L_2(u_A, v_A), L_4(u_B, v_B), L_9(w_A, \psi_A)$, and $L_{10}(u_C, v_C)$.
- 2) receiver B has $L_2(u_A, v_A), L_4(u_B, v_B), L_{15}(w_B, \psi_B)$, and $L_{17}(w_C, \psi_C)$.
- 3) receiver C has $L_9(w_A, \psi_A), L_{10}(u_C, v_C), L_{15}(w_B, \psi_B)$, and $L_{17}(w_C, \psi_C)$.

Therefore, the transmitter needs to deliver

- 1) $L_2(u_A, v_A)$ and $L_4(u_B, v_B)$ to both receivers A and B.
- 2) $L_9(w_A, \psi_A)$ and $L_{10}(u_C, v_C)$ to both receivers A and C.
- 3) $L_{15}(w_B, \psi_B)$ and $L_{17}(w_C, \psi_C)$ to both receivers B and C.

Therefore, we have six order-2 symbols as

$$u_{AB} = L_2(u_A, v_A), \quad v_{AB} = L_4(u_B, v_B) \quad (43)$$

$$u_{AC} = L_9(w_A, \psi_A), \quad v_{AC} = L_{10}(u_C, v_C) \quad (44)$$

$$u_{BC} = L_{15}(w_B, \psi_B), \quad v_{BC} = L_{17}(w_C, \psi_C). \quad (45)$$

Therefore, the transmitter needs $\frac{6}{\text{DoF}_2^*(2,3)}$ more time-slots to deliver these six order-2 symbols. Thus, we have

$$\text{DoF}_1(2, 3) = \frac{12}{3 + \frac{6}{\text{DoF}_2^*(2,3)}} = \frac{3}{2} \quad (46)$$

where we used Theorem 1 to set $\text{DoF}_2^*(2, 3) = \frac{6}{5}$. Note the outer-bound in Theorem 2 yields $\text{DoF}_1^*(2, 3) \leq \frac{3}{2}$, and therefore, this algorithm meets the outer-bound. This result shows that the scheme of Section VI is in general suboptimal.

VIII. CONNECTIONS WITH THE PACKET ERASURE BROADCAST CHANNEL

The schemes we proposed in this paper are inspired by schemes designed for the packet erasure broadcast channel, where each receiver observes the same transmitted packet but with a probability of erasure, and acknowledgement feedback is received by the transmitter from both receivers. Here, the delayed CSI that is fed back to the transmitter is the *erasure states* of the previous transmissions.

The goal of these packet erasure broadcast schemes is to exploit the fact that a packet intended for a receiver may be erased at that receiver but received at other receivers. These overheard packets become side information that can be exploited later. The

basic scheme, initially proposed by [10] for unicast setting, and then by [11] for multicasting setting, in the two-receiver case, works as follows. The transmitter sends packets intended 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 nonintended receiver, and not received by intended receiver, that receiver keeps that packet for later coding opportunity. Let us say packet x_A intended for receiver A is received by receiver B , and packet x_B intended for receiver B is received by receiver A . In this case, the transmitter sends $(x_A \text{ XOR } x_B)$. Then, if receiver A receives it, it can recover x_A by subtracting x_B , and if receiver B receives it, it can recover x_B by subtracting x_A . In [12], the outer-bound of [13] is used to show that the scheme of [11] is optimal. In [14] and [15], this two-receiver scheme is extended to more than two receivers, when all receivers have identical erasure probability. The scheme we proposed in this paper for the MIMO broadcast channel can be viewed as the counterpart to this scheme for the packet erasure broadcast channel.

IX. FOLLOW-UP RESULTS

After the conference version of this paper has appeared in [16], the problem of exploiting outdated CSIT in networks have been investigated in several pieces of work. In [17], it is shown that for three-user interference channels and two-user X channels, outdated CSIT can be used to achieve DoF more than one. In [18], for two-user X channels, the result of [17] has been improved and for three-user case, an achievable DoF has been proposed. In [19], an achievable DoF for K -user single-antenna interference channels has been derived. In [20]–[22], the DoF regions of two-user and three-user MIMO broadcasts channels and two-user MIMO interference channels with delayed CSIT are studied. In [23], the load of feedback to implement the proposed scheme is evaluated. It is shown that for a wide and practical range of channel parameters, the scheme of this paper outperforms zero-forcing precoding and also single-user transmission.

X. CONCLUSION

From the point of view of the role of feedback in information theory, this work provides yet another example that feedback can be useful in increasing the capacity of *multiuser* channels, even when the channels are memoryless. This is in contrast to Shannon's pessimistic result that feedback does not increase the capacity of memoryless *point-to-point* channels [24]. In the specific context of broadcast channels, Ozarow [13] has in fact already shown that feedback can increase the capacity of Gaussian scalar nonfading broadcast channels. However, the nature of the gain is unclear, as it was shown numerically. Moreover, the gain is quite limited. We argue that the MIMO fading broadcast channel considered in this paper provides a much more interesting example of the role of feedback. The nature of the gain is very clear. In contrast to the Gaussian scalar nonfading broadcast channel, the main uncertainty from the point of the view of the transmitter is the channel direction rather than the additive noise, particularly in the high SNR regime. This means that although the MIMO channel has intrinsically multiple degrees

of freedom, the transmitter cannot segregate it into multiple orthogonal channels, one for each receiver. Hence, when transmitting information for one receiver, significant part of that information is overheard at other receivers. This overheard information becomes side information that can be exploited in future transmissions. The role of feedback is to provide the channel directions to the transmitter *after* the transmission to allow the transmitter to determine the side information that was received at the receivers. Overall, feedback leads to a much more efficient use of the intrinsic multiple degrees of freedom in the MIMO channel, yielding a multiplexing gain over the nonfeedback case.

APPENDIX IDENTITY

In this appendix, we prove that for any j , $1 \leq j \leq K$

$$\frac{1}{\binom{K}{j-1}} \sum_{i=1}^{K-j+1} \frac{\binom{K-i}{j-1}}{i} = \sum_{i=j}^K \frac{1}{i}. \quad (47)$$

We define LHS of (47) as $f(j)$

$$f(j) = \frac{1}{\binom{K}{j-1}} \sum_{i=1}^{K-j+1} \frac{\binom{K-i}{j-1}}{i}. \quad (48)$$

Then it is easy to see that $f(K) = \frac{1}{K}$. In what follows, we prove that for any j , $1 \leq j \leq K-1$

$$f(j) = \frac{1}{j} + f(j+1)$$

which yields identity (47).

We have

$$\begin{aligned} & f(j) - f(j+1) \\ &= \frac{1}{\binom{K}{j-1}} \sum_{i=1}^{K-j+1} \frac{\binom{K-i}{j-1}}{i} - \frac{1}{\binom{K}{j}} \sum_{i=1}^{K-j} \frac{\binom{K-i}{j}}{i} \\ &= \frac{(j-1)!(K-j)!}{K!} \\ &\quad \times \left\{ \sum_{i=1}^{K-j+1} \frac{(K-j+1)\binom{K-i}{j-1}}{i} - \sum_{i=1}^{K-j} \frac{j\binom{K-i}{j}}{i} \right\} \\ &= \frac{(j-1)!(K-j)!}{K!} \\ &\quad \times \left\{ \sum_{i=1}^{K-j} \frac{\binom{K-i}{j-1}[(K-j+1) - (K-i-j+1)]}{i} + 1 \right\} \\ &= \frac{(j-1)!(K-j)!}{K!} \sum_{i=1}^{K-j+1} \binom{K-i}{j-1} \\ &= \frac{(j-1)!(K-j)!}{K!} \sum_{l=j-1}^{K-1} \binom{l}{j-1} \\ &\stackrel{(a)}{=} \frac{(j-1)!(K-j)!}{K!} \binom{K}{j} \\ &= \frac{1}{j} \end{aligned}$$

where (a) follows from the identity that

$$\sum_{l=p}^q \binom{l}{p} = \binom{q+1}{p+1}, \quad 0 \leq p \leq q. \quad (49)$$

Equation (49) can simply be proved by induction.

REFERENCES

- [1] M. A. Maddah-Ali, S. A. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.
- [2] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k -user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [4] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
- [5] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.
- [6] U. Basher, A. Shirazi, and H. Permuter, "Capacity region of finite state multiple-access channel with delayed state information at the transmitters Jan. 2011," arxiv.org/abs/1101.2389.
- [7] A. E. Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 379–381, Apr. 1978.
- [8] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without channel state information at transmitters," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 849–857, Feb. 2012.
- [9] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT Oct. 2009," arxiv.org/abs/0909.5424.
- [10] M. Jolfaei, S. Martin, and J. Mattfeldt, "A new efficient selective repeat protocol for point-to-multipoint communication," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1113–1117.
- [11] P. Larsson and N. Johansson, "Multi-user ARQ," in *Proc. IEEE 63rd Veh. Technol. Conf.*, Melbourne, Vic., Australia, May 2006, pp. 2052–2057.
- [12] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback—Capacity and algorithms," in *Proc. Workshop Network Coding, Theory, Appl.*, Lausanne, Switzerland, Jun. 2009, pp. 54–61.
- [13] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 667–671, Jul. 1984.
- [14] C.-C. Wang, "On the capacity of 1-to- k broadcast packet erasure channels with channel output feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931–956, Feb. 2012.
- [15] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback—Capacity and algorithms," in *Arxiv. Org.*, 2010, arxiv.org/abs/1009.1254.
- [16] M. A. Maddah-Ali and D. N. T. Tse, "Completely stale transmitter channel state information is still very useful," presented at the presented at the 48th Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, Sep. 2010.
- [17] H. Maleki, S. A. Jafar, and S. Shamai, "Retrospective interference alignment over interference networks," *IEEE J. Selected Topics Signal Process.*, Issue on Signal Processing in Heterogeneous Networks for Future Broadband Wireless Systems, vol. 6, no. 3, pp. 228–240, Jun. 2012.
- [18] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "On the degrees of freedom of X channel with delayed CSIT," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint-Petersburg, Russia, Jul. 2011, pp. 909–912.
- [19] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, "On the degrees of freedom of k -user SISO interference and X channels with delayed CSIT," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2011, pp. 625–632.
- [20] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, "On the degrees of freedom of three-user MIMO broadcast channel with delayed CSIT," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint-Petersburg, Russia, Jul. 2011, pp. 341–345.
- [21] C. S. Vaze and M. K. Varanasi, "The degrees of freedom region of the two-user MIMO broadcast channel with delayed CSI Dec. 2010," arxiv.org/abs/1101.0306.
- [22] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "Interference alignment for the MIMO interference channel with delayed local CSIT Feb. 2011," arxiv.org/abs/1102.5673.
- [23] J. Xu, J. G. Andrews, and S. A. Jafar, "Broadcast channels with delayed finite-rate feedback: Predict or observe? May 2011," arxiv.org/abs/1105.3686.
- [24] C. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.

Mohammad Ali Maddah-Ali received the B.Sc. degree from Isfahan University of Technology, Isfahan, Iran, in 1997 and the M.A.Sc. degree from the University of Tehran, Tehran, Iran, in 2000, both in electrical engineering with highest rank in classes. From 2002 to 2007, he was with the Coding and Signal Transmission Laboratory (CST Lab), Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, working toward the Ph.D. degree. From March 2007 to December 2007, he was a Postdoctoral Fellow at the Wireless Technology Laboratories, Nortel Networks, Ottawa, ON, Canada, in a joint program between CST Lab and Nortel Networks. From January 2008 to August 2010, he was a Postdoctoral Fellow at the Wireless Foundations Center, the Department of Electrical Engineering and Computer Sciences in the University of California at Berkeley. Since September 2010, he has been at Bell Laboratories, Alcatel-Lucent, Holmdel, NJ, as a communication network research scientist. His research interests include wireless communications and multiuser information theory. Dr. Maddah-Ali received several awards including Natural Science and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship.

David Tse received the B.A.Sc. degree in systems design engineering from University of Waterloo in 1989, and the M.S. and Ph.D. degrees in electrical engineering from Massachusetts Institute of Technology in 1991 and 1994 respectively. From 1994 to 1995, he was a postdoctoral member of technical staff at AT&T Bell Laboratories. Since 1995, he has been at the Department of Electrical Engineering and Computer Sciences in the University of California at Berkeley, where he is currently a Professor. He received a 1967 NSERC graduate fellowship from the government of Canada in 1989, a NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, the Information Theory Society Paper Award in 2003, the 2009 Frederick Emmons Terman Award from the American Society for Engineering Education, and a Gilbreth Lectureship from the National Academy of Engineering in 2012. He has given plenary talks at international conferences such as ICASSP in 2006, MobiCom in 2007, CISS in 2008, and ISIT in 2009. He was the Technical Program co-chair of the International Symposium on Information Theory in 2004, and was an Associate Editor of the IEEE Transactions on Information Theory from 2001 to 2003. He is a coauthor, with Pramod Viswanath, of the text *Fundamentals of Wireless Communication*, which has been used in over 60 institutions around the world.