A Novel Transmission Scheme for the K -User Broadcast Channel With Delayed CSIT

Chao He, Sheng Yang, Pablo Piantanida

A Novel Transmission Scheme for the K-user Broadcast Channel with Delayed CSIT

Chao He, *Student Member,* Sheng Yang, *Member,*
and Pablo Piantanida, *Senior Member*

Abstract

The state-dependent K-user memoryless Broadcast Channel (BC) with state feedback is investigated. We propose a novel transmission scheme and derive its corresponding achievable rate region, which, compared to some general schemes that deal with feedback, has the advantage of being relatively simple and thus is easy to evaluate. In particular, it is shown that the capacity region of the symmetric erasure BC with an arbitrary input alphabet size is achievable with the proposed scheme. For the fading Gaussian BC, we derive a symmetric achievable rate as a function of the signal-to-noise ratio (SNR) and a small set of parameters. Besides achieving the optimal degrees of freedom at high SNR, the proposed scheme is shown, through numerical results, to outperform existing schemes from the literature in the finite SNR regime.

Index Terms

Broadcast channel; Erasure channel; Fading Gaussian channel; State feedback.

I. Introduction

With the dramatic growth of the number of mobile devices, modern wireless communication networks have become interference limited. As such, the interference mitigation problem has attracted a surge of interest in recent years. In a downlink Broadcast Channel (BC), for instance, it is well known that interference can be efficiently mitigated through precoding, provided that timely Channel State Information (CSI) is available at the transmitter side (CSIT) (see, e.g., [1] and the references therein). While timely CSIT may not be available in mobile communications,

The authors are with the Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506), CentraleSupélec - CNRS - Université Paris-Sud, 3, rue Joliot-Curie, 91190, Gif-sur-Yvette, France. ({chao.he, sheng.yang, pablo.piantanida}@centralesupelec.fr).

it has been revealed in [2]–[4] that delayed CSIT is still very useful and can strictly enlarge the capacity region of a BC.

In particular, the capacity region of the erasure BC (also referred to as the EBC) with delayed CSIT was fully determined for up to three users and partially characterized for the case with more users [2], [3]. The main idea behind their proposed schemes in [2] and [3] is fundamentally the same: the source first sends out the source message packets, then generates according to the state feedback some adequate *linear combinations* of the packets that are erased at certain receivers but overheard by some others. Such linear combinations are then multicast to a group of users in later phases. Their schemes are carefully designed such that at the end of the transmission a sufficient number of linearly independent combinations are available to each receiver for the decoding of the original message packet. However, the schemes in [2], [3] are limited to packet erasure channels for which the input alphabet size can only be 2^q with $q \in \mathbb{N}$ being the number of bits per packet. In addition, there is an extra constraint, $2^q \geq K$, to guarantee the existence of a desired number of linearly independent vectors in the corresponding vector space in finite field. As such, the capacity region is still open for the general EBC with arbitrary alphabet sizes.

For the multi-antenna fading Gaussian BC (also referred to as the GBC) with delayed CSIT, Maddah-Ali and Tse proposed a linear scheme that achieves the optimal Degree of Freedom (DoF) for the K-user Multiple-Input-Single-Output (MISO) case. The authors showed that with delayed CSIT the sum-DoF can still scale almost linearly with the number of users. Remarkably, there is a striking similarity between the Maddah-Ali-Tse (MAT) scheme and the schemes from [2], [3]. Namely, based on the CSI feedback, the transmitter can create and transmit useful linear combinations of the past received signals by the users. The intended group of users receive such linear combinations and use them to decode the message together with the previous observations. Note that the MAT scheme in [4] has a fixed structure designed based on a dimension counting argument. Although such a structure guarantees the DoF optimality at high Signal-to-Noise-Ratio (SNR), it may not be efficient at finite SNR due to its inflexibility.

As a matter of fact, there are only a small number of works on the performance gain with delayed CSIT in the finite SNR regime. In [5], the authors developed two linear precoding methods that attempt to balance the interference and the useful received signal. For $K = 2$ and 3, performance gain over MAT was revealed when a specific type of decoder is used. To adjust the multicast cost in the MAT scheme, the authors of [6] proposed to transmit a quantized version of the linear combinations. For the K-user Rayleigh fading case, they demonstrated that a gap

between the corresponding inner bound and a genie-aided outer bound, in terms of the symmetric rate, is upper bounded by $2\log_2(K+2)$ which scales sublinearly with K. More recently, the work [7] studied a scenario where both the CSI statistics and the feedback of the channel realizations are available at the transmitter. It was shown by numerical examples that statistics of CSI can enlarge the rate region for temporally correlated Rayleigh fading GBC. The authors of [8] investigated the outage performance for GBC with an adapted MAT scheme. It is worth mentioning that these schemes are variants of the linear MAT scheme, i.e., they applied either linear coding or linear coding with quantization with the same fixed frame structure of MAT. Although the rate performance of the MAT-like schemes is rather convincing in the medium-to-high SNR regime, their performance in the medium-to-low SNR regime is still questionable since it can be strictly dominated by the simple time-division multiple access (TDMA) strategy [9].

Instead of imposing the linear structure, we can tackle the problem directly from the information-theoretic perspective. To that end, we formulate the setup as a K-user state-dependent memoryless BC with state feedback. This formulation includes both the EBC and the GBC as special cases. In the two-user case, Shayevitz and Wigger studied such BC with generalized feedback and derived a general achievable rate region using information-theoretic tools [10]. Later on, Kim *et al.* demonstrated in [11] that in the two-user symmetric setting, the Shayevitz-Wigger (SW) region, actually includes the MAT region. Similar recent works on the two-user case have been reported in [12], [13]. In this work, we are interested in the general K-user case. The main contributions are summarized as follows.

- We propose a novel scheme for the general K-user channel and derive the corresponding achievable rate region. The novelty of this scheme lies in the proper combination of two main ingredients: coded time-sharing and joint source-channel coding (JSC) with side information at the decoder. We refer to our scheme in short as the JSC scheme. As compared to the existing schemes, e.g., the Shayevitz-Wigger scheme (which is limited to two users) [10], our scheme is conceptually simpler in the sense that neither block-Markov coding nor Marton coding is required. Such simplicity, at the cost of a slight loss of generality, allows us to derive the K-user rate region with a reasonable number of parameters. To the best of our knowledge, the JSC scheme is the first information-theoretic scheme for the K-user BC with state feedback for $K \geq 3$.
- The general rate region is then evaluated for both the EBC and fading GBC. First, we show that our scheme achieves the capacity of a symmetric EBC with an arbitrary input alphabet

size, whereas the previous schemes in [2], [3] only apply to packet erasure channels. Second, for the symmetric fading GBC, we derive the achievable symmetric rate as a function of SNR and a set of $K-1$ compression noise variances. At high SNR, we show analytically that the proposed scheme achieves the optimal DoF under the same setting as in [4]. At finite SNR, we perform numerical optimization over the set of $K-1$ variances. The results show that in the two- and three-user cases, the JSC scheme outperforms the existing schemes at *all* SNR.

The remainder of the paper is organized as follows. We introduce the system model formally in Section II. Then we begin with the two-user case in Section III, before presenting the general K-user scheme in a more abstract way in Section IV. The general region is applied to the erasure BC and fading Gaussian BC in Section V. In Section VI, numerical results are provided for the two-user and three-user fading Gaussian BC where we compare the JSC scheme to some baseline schemes from the literature. The paper is concluded in Section VII. Although most of the derivations are provided in the main text, some more technical details are deferred to the appendices.

Notation: First, for random quantities, we use upper case letters, e.g., X, for scalars, upper case letters with bold and non-italic fonts, e.g., $\mathbf{V}$, for vectors, and upper case letter with bold and sans serif fonts, e.g., M, for matrices. Deterministic quantities are denoted in a rather conventional way with italic letters, e.g., a scalar x, a vector $\boldsymbol{v}$, and a matrix $\boldsymbol{M}$. Logarithms are in base 2. Calligraphic letters are used for sets. In particular, we let $\mathcal{K} \triangleq \{1,\ldots,K\}$ be the set of all users. To denote subset of users, we use $\mathcal{I}$ and $\mathcal{J}$ for some subsets with implicit size constraints $|\mathcal{I}|=i$ and $|\mathcal{J}|=j$, respectively. The constraints are made explicit when necessary. $\mathcal{U}$ is also used as subset of users but without size constraint. Hence, $\{V_{\mathcal{I}}\}_{\mathcal{I}} \equiv \{V_{\mathcal{I}} : \mathcal{I} \subseteq \mathcal{K}, |\mathcal{I}|=i\}$ and $\{V_{\mathcal{U}}\}_{\mathcal{U}} \equiv \{V_{\mathcal{U}} : \mathcal{U} \subseteq \mathcal{K}\}$. The complement of $\mathcal{I}$ in $\mathcal{J}$ is denoted by $\mathcal{J} \setminus \mathcal{I}$. We use $\bar{\mathcal{U}}$ to denote the complement of the set $\mathcal{U}$ in $\mathcal{K}$, i.e., $\bar{\mathcal{U}} = \mathcal{K} \setminus \mathcal{U}$.

II. System Model

We consider a K-user state-dependent memoryless BC in which the source wishes to communicate, in n channel uses, K independent messages to the K receivers, respectively. The channel can be described by the joint probability mass function (pmf),

$$p(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_K|\boldsymbol{x},\boldsymbol{s})p(\boldsymbol{s}) = \prod_{i=1}^{n} p(y_{1i},\ldots,y_{Ki}|x_i,s_i)p(s_i) \tag{1}$$

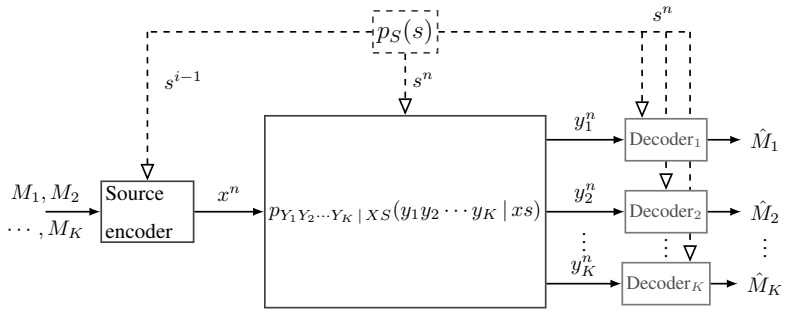


Figure 1: General system model of K-user BC with state feedback.

where $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{y}_k \in \mathcal{Y}_k^n$, and $\boldsymbol{s} \in \mathcal{S}^n$ are the sequences of the channel input, the channel output at the k-th receiver, and the channel state, respectively. The channel state information (CSI) is known instantaneously to all the receivers. At transmitter's side, the channel state is known strictly causally without error via a noiseless feedback link from the receivers. For simplicity, we assume that the CSI is provided at the transmitter with one slot delay and the channel itself is temporally i.i.d. The channel model is illustrated in Fig. 1.

Let the message for user k, M_k, be uniformly distributed in the message set $\mathcal{M}_k \triangleq [1 : 2^{nR_k}]$, for $k \in \mathcal{K}$. We say that the rate tuple $(R_1, \ldots, R_K)$ is achievable if there exist

- a sequence of encoding functions $\{f_i : \ \mathcal{M}_1 \times \cdots \times \mathcal{M}_K \times \mathcal{S}^{i-1} \to \mathcal{X}\}_{i=1}^n$, and
- K decoding functions $\{g_k : \ \mathcal{Y}_k^n \times \mathcal{S}^n \to \mathcal{M}_k\}_{k=1}^K$,

such that $\max_k \mathrm{P}\big(g_k(Y_k^n, S^n) \neq M_k\big) \to 0$ when $n \to \infty$. The symmetric rate R_{sym} is achievable if the rate tuple $(R_{\text{sym}}, \ldots, R_{\text{sym}})$ is achievable. In particular, we are interested in the following two specific channels.

A. Fading Gaussian Broadcast Channel

The fading GBC with n_t transmit antennas and $n_{r,k}$ receive antennas at user k, $k \in \mathcal{K}$, is defined by

$$\mathbf{Y}_k = \mathbf{H}_k\mathbf{X} + \mathbf{Z}_k, \quad k \in \mathcal{K}, \tag{2}$$

where $\mathbf{X} \in \mathbb{C}^{n_{\mathrm{t}} \times 1}$ is the input vector, $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{r},k} \times 1}$ is the output vector at receiver k, $\mathbf{Z}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{n_{\mathrm{r},k}})$ is the additive white Gaussian noise (AWGN), and $\mathbf{H}_k \in \mathbb{C}^{n_{\mathrm{r},k} \times n_{\mathrm{t}}}$ is the channel matrix to receiver k. The channel input is subject to the power constraint as $\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{x}_i\|^2 \leq P$ for any input sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. The SNR is defined as $\mathsf{snr} \triangleq \frac{P}{n_{\mathrm{t}}\sigma^2}$. We assume that both

the channel matrices and the AWGN are independent across users. We use $\mathbf{H}_{\mathcal{U}}$ to denote a matrix from a vertical concatenation of the channel matrices of receivers in $\mathcal{U}$, i.e., $\{\mathbf{H}_k\}_{k\in\mathcal{U}}$, same notation applies for $\mathbf{Y}_{\mathcal{U}}$ and $\mathbf{Z}_{\mathcal{U}}$. Hence, it follows that $\mathbf{Y}_{\mathcal{U}} = \mathbf{H}_{\mathcal{U}}\mathbf{X} + \mathbf{Z}_{\mathcal{U}}$. The matrix $\mathbf{H}_{\mathcal{K}}$ corresponds to the channel state S in the general formulation.

B. Erasure Broadcast Channel

The EBC is a state-dependent deterministic channel in which

$$Y_k = \begin{cases} X, & S_k = 1, \\ ?, & S_k = 0, \end{cases} \tag{3}$$

for $k \in \mathcal{K}$. Here the input alphabet $\mathcal{X}$ is arbitrary and finite with size $|\mathcal{X}|$; , "?" stands for erasure, and the output alphabet is $\mathcal{Y} = \mathcal{X} \cup \{?\}$. The distribution of the channel state is characterized by the set of probabilities

$$\phi_{\mathcal{U},\bar{\mathcal{U}}} \triangleq \mathrm{P}(S_{\mathcal{U}} = \mathbf{0},\ S_{\bar{\mathcal{U}}} = \mathbf{1}), \quad \mathcal{U} \subseteq \{1,\ldots,K\} \tag{4}$$

with $\sum_{\mathcal{U}} \phi_{\mathcal{U},\bar{\mathcal{U}}} = 1$. Throughout the paper, we use $S_{\mathcal{U}} = \mathbf{0}$ (*resp.* $S_{\mathcal{U}} = \mathbf{1}$) to define the event that $S_k = 0$ (*resp.* $S_k = 1$), $\forall\ k \in \mathcal{U}$. For simplicity, we use $\delta_{\mathcal{F}}$ to denote $\mathrm{P}(S_{\mathcal{F}} = \mathbf{0})$, and use $\phi_{\mathcal{F},\mathcal{T}}$ to denote $\mathrm{P}(S_{\mathcal{F}} = \mathbf{0}, S_{\mathcal{T}} = \mathbf{1})$ for any $\mathcal{F}$ and $\mathcal{T}$ that satisfy $\mathcal{F} \cap \mathcal{T} = \phi$ and $\mathcal{F} \cup \mathcal{T} \subseteq \mathcal{K}$. For notational brevity, $\delta_{\{k\}}$ is written as δ_k. The K-tuple $S_{\mathcal{K}}$ corresponds to the channel state S in the general formulation.

III. The Two-User Case

Before presenting the main results for the general K-user channel, we provide a description of the two-user case. The goal is to explain the main ingredients of the proposed scheme in an accessible and less formal way, whereas a rigorous and detailed description will be provided for the K-user case in the next section. Hereafter, we also refer to our scheme as the JSC scheme.

A. Scheme description

The JSC scheme consists in two phases with n_1 and n_2 being the length of phase 1 and 2, respectively. The total transmission length is $n = n_1 + n_2$. In the first phase, the original messages M_1 and M_2, for receiver 1 and receiver 2, respectively, are encoded and transmitted. At the end of phase 1, the transmitter obtains the state feedback and thus some side information about the

received signals at both users during phase 1. In the second phase, the transmitter compresses the side information that are useful to both users into $\hat{Y}$, and transmits the compression index M_{12} with a channel code. Each receiver decodes the compression index first, but with the observation from *both* phases. With the compression index, the side information $\hat{Y}$ is recovered and combined with the observation from phase 1 by each receiver $k \in \{1, 2\}$ to finally decode the message M_k. The main information-theoretic tools that we use in this scheme are the following ones:

- Coded time-sharing for the transmission in phase 1;
- Joint source-channel coding in phase 2;
- Joint source-channel decoding with side information on the source.

Phase 1

At the beginning, we randomly generate a sequence of time-sharing variables $\boldsymbol{q} \triangleq (q_1, \ldots, q_{n_1})$ according to $\prod_{t=1}^{n_1} p(q_t)$. For each user $k \in \{1, 2\}$, we generate a random codebook of 2^{nR_k} independent codewords, $\boldsymbol{v}_k(m_k)$, $m_k \in [1 : 2^{nR_k}]$, each according to $\prod_{t=1}^{n_1} p(v_{k,t})$. Both the time-sharing sequence and the codebooks are revealed to the transmitter and all the receivers.

To send messages m_1 and m_2 to user 1 and 2, respectively, we use coded time-sharing in phase 1. Specifically, at time t, the transmitter sends $x_t^{(1)} = v_{1,t}$ if $q = 1$ and $x_t^{(1)} = v_{2,t}$ if $q = 2$. It is similar to a TDMA scheme controlled by the time-sharing variables $\{q_t\}$. Here the superscript $\cdot^{(1)}$ stands for phase 1.

At the end of phase 1, each receiver k observes $y_{k,t}$ that depends on x_t and the channel state s_t, for $t = 1, \ldots, n_1$. The transmitter obtains through feedback the sequence $s_1, \ldots, s_{n_1}$. At this point, the transmitter knows the following i.i.d. triples

$$(v_{1,1}, v_{2,1}, s_1), \ldots, (v_{1,n_1}, v_{2,n_1}, s_{n_1}), \tag{5}$$

which can be regarded as a source sequence of length n_1. Alternatively, it can be represented by $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{s}^{(1)})$.

Phase 2

The transmitter creates a source codebook and a channel codebook, both with the same size $2^{n_1 R_{12}}$. Specifically, the source codebook contains $2^{n_1 R_{12}}$ i.i.d. sequences $\hat{\boldsymbol{y}}(m_{12})$, each generated according to $\prod_{t=1}^{n_1} p(\hat{y}_t \,|\, s_t, q_t)$, whereas the channel codebook contains $2^{n_1 R_{12}}$ sequences $\boldsymbol{v}_{12}(m_{12})$, each generated according to $\prod_{t=1}^{n_2} p(v_{12,t})$. Note that while the codebook size is the

same, the codeword lengths are different for the source and channel codebooks. This is because the source codebook is used to describe the source sequence (5) from phase 1 while the channel codebook is used to send the index m_{12} in phase 2.

First, the transmitter finds a sequence $\hat{\boldsymbol{y}}(m_{12})$ from the source codebook that is jointly typical with the source sequence (5). This can be done successfully provided that

$$R_{12} \geq I(\hat{Y}; V_1, V_2 \,|\, S^{(1)}, Q). \tag{6}$$

Then, the source sequence is associated with the channel codeword $\boldsymbol{v}_{12}(m_{12})$ through the index m_{12}. The transmission in phase 2 is simply specified by $x_t^{(2)} = v_{12,t}$, $t = 1, \ldots, n_2$. The above procedure can be seen as a joint source-channel coding.

Decoding

We focus on the decoding at receiver k without loss of generality. First, the receiver tries to find out m_{12} with the observations from the two phases: $\boldsymbol{y}_k^{(1)}$ and $\boldsymbol{y}_k^{(2)}$. Intuitively, $\boldsymbol{y}^{(1)}$ is correlated with the source $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{s}^{(1)})$ and thus the source codeword $\hat{\boldsymbol{y}}(m_{12})$, whereas $\boldsymbol{y}_k^{(2)}$ is correlated with the channel codeword $\boldsymbol{v}_{12}(m_{12})$. Hence, both observations can help find the *same* index m_{12}. Specifically, decoder k looks for $\hat{m}_{12}$ such that $\left(\boldsymbol{v}_{12}(\hat{m}_{12}), \boldsymbol{y}_k^{(2)}, \boldsymbol{s}^{(2)}\right)$ are jointly typical and that $\left(\hat{\boldsymbol{y}}(\hat{m}_{12}), \boldsymbol{y}_k^{(1)}, \boldsymbol{s}^{(1)}, \boldsymbol{q}\right)$ are jointly typical. It turns out that one can recover m_{12} correctly as long as the rate satisfies

$$n_1 R_{12} \leq n_1 I(\hat{Y}; Y_k^{(1)} \,|\, S^{(1)}, Q) + n_2 I(V_{12}; Y_k^{(2)} \,|\, S^{(2)}), \tag{7}$$

where we see clearly the contribution of the observations from both phases. This is essentially Tuncel's scheme [14] of separate source-channel encoding but joint source-channel decoding.

Then the receiver uses $\hat{\boldsymbol{y}}(\hat{m}_{12})$ jointly with the observation from phase 1, $\boldsymbol{y}^{(1)}$, to decode the original message. Specifically, it looks for the unique $\hat{m}_k$ such that $(\boldsymbol{v}_k(\hat{m}_k), \boldsymbol{y}_k^{(1)}, \hat{\boldsymbol{y}}(\hat{m}_{12}), \boldsymbol{q}, \boldsymbol{s}^{(1)})$ is jointly typical. The original message can be decoded correctly if the message rate satisfies

$$nR_k \leq n_1 I(V_k; Y_k^{(1)}, \hat{Y} \,|\, S^{(1)}, Q). \tag{8}$$

From (6)-(8), we see that for any fixed distribution $p(v_1)p(v_2)p(v_{12})p(\hat{y} \,|\, v_1, v_2, s^{(1)}, q)$ the rate pair (R_1, R_2) is achievable if, for $k \in \{1, 2\}$,

$$R_k \leq \alpha_1 I(V_k; Y_k^{(1)}, \hat{Y} \,|\, S^{(1)}, Q), \tag{9}$$

$$I(\hat{Y}; V_1, V_2 \,|\, S^{(1)}, Q) \leq I(\hat{Y}; Y_k^{(1)} \,|\, S^{(1)}, Q) + \frac{\alpha_2}{\alpha_1} I(V_{12}; Y_k^{(2)} \,|\, S^{(2)}), \tag{10}$$

where $\alpha_1 \triangleq \frac{n_1}{n}$ and $\alpha_2 \triangleq \frac{n_2}{n}$ with $\alpha_1 + \alpha_2 = 1$ can be optimized.

Remark 1 (Comparison to other information-theoretic schemes)**.** *Although in the two-user case our result is closely connected to the works [10] and [12], unfortunately it is hard to make a fair comparison in the Gaussian case. First, the fact that the achievable regions depend on different sets of pmf's prohibits the analytical comparison. Then, since we cannot find the exact optimal solution for any of the regions (e.g. Gaussian input is not even proved to be optimal in general), any numerical comparison must be based on a particular choice of distribution, which cannot be conclusive. Indeed, the underlying transmission schemes are conceptually different. Both schemes in [10] and [12] use Marton coding and block-Markov coding. While a separate source channel coding was used to compress the side information in [10], the authors in [12] adopted a joint source-channel coding approach. In our scheme, we do not use Marton coding nor block-Markov coding, but, as in [12], we use a joint source-channel coding for the transmission of side information. It is worth noting that our scheme is based on Tuncel's scheme [14] and is different from the one used in [12].*

From the complexity perspective, our scheme is conceptually simpler since neither block-Markov coding nor binning is required. Furthermore, due to the relative simplicity, we manage to derive a rate region for the K-user case with a reasonable number of parameters, as will be shown in the upcoming sections. Such an advantage allows us to easily obtain numerical results for $K \geq 3$, whereas the counterpart of the existing two-user schemes is still missing in the literature due to their complexity for extension.

B. Application to the fading GBC

Let us consider a two-user MISO fading GBC with $Y_k^{(j)} = \mathbf{H}_k^{(j)H}\mathbf{X}^{(j)} + Z_k^{(j)}$, for $k \in \{1, 2\}$ and $j \in \{1, 2\}$. Here $\mathbf{H}_k^{(j)H}$ is the channel $\mathbf{H}_k^{(j)}$ in the vector case. We consider the symmetric case and let $Q \in \{1, 2\}$ be uniform with probability $\frac{1}{2}$ for each user. In phase 1, we have $\mathbf{X}^{(1)} = \mathbf{V}_1$ if $Q = 1$ and $\mathbf{X}^{(1)} = \mathbf{V}_2$ if $Q = 2$, with $\mathbf{V}_1, \mathbf{V}_2 \sim \mathcal{CN}(0, \frac{P}{2}\boldsymbol{I}_2)$ being independent. At the end of phase 1, the transmitter sets the side information as follows

$$\hat{Y} = \begin{cases} \mathbf{H}_2^{(1)H}\mathbf{X}^{(1)} + \hat{Z}_2 = \mathbf{H}_2^{(1)H}\mathbf{V}_1 + \hat{Z}_2, & \text{if } Q = 1, \\ \mathbf{H}_1^{(1)H}\mathbf{X}^{(1)} + \hat{Z}_1 = \mathbf{H}_1^{(1)H}\mathbf{V}_2 + \hat{Z}_1, & \text{if } Q = 2. \end{cases} \tag{11}$$

Intuitively, $\hat{Y}$ is the compression of the *overheard* signal at the *unintended* receiver, with $\hat{Z}_1, \hat{Z}_2 \sim \mathcal{CN}(0, \hat{\sigma}^2)$ being the independent compression noises. The idea is not to retransmit everything

about $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{H})$ as side information, since this would be too costly. Instead, sending compressed version of a function of these information, namely, $(\mathbf{H}_2^{(1)H}\mathbf{V}_1, \mathbf{H}_1^{(1)H}\mathbf{V}_2)$, would be helpful. The compression noise can balance the precision of side information and the cost for the transmission, as can be observed in (10). To get more insight on the choice of $\hat{Y}$, let us rewrite (10) as

$$I(\hat{Y}; V_1, V_2 \,|\, Y_k^{(1)}, S^{(1)}, Q) \leq \frac{\alpha_2}{\alpha_1} I(V_{12}; Y_k^{(2)} \,|\, S^{(2)}), \tag{12}$$

using the Markovity $\hat{Y} \leftrightarrow (V_1, V_2, S^{(1)}, Q) \leftrightarrow Y_k^{(1)}$ and the chain rule. The left-hand side of (12) is the average amount of side information remaining in $\hat{Y}$ after observing $Y_k^{(1)}$, whereas the right hand side is the achievable transmission rate to user k in phase 2. With the choice (11), on the one hand, we make sure that during half of the time, $I(\hat{Y}; V_1, V_2 \,|\, Y_k^{(1)}, S^{(1)}, Q = q)$ is small since $Y_k^{(1)}$ already almost contains the information in $\hat{Y}$. This makes sure that the constraint (12) can be met. On the other hand, from (11) and (9), we notice that $\hat{Y}$ provides to receiver k an extra observation approximative to the other receiver's signal. Such observation helps create a virtual MIMO system for each user. From (9), it readily follows that the symmetric rate is

$$R_{\text{sym}}^{\text{JSC}} = \frac{\alpha_1}{2} \mathbb{E}\left[\log\det\left(\boldsymbol{I} + \left[\begin{smallmatrix}\sigma^{-2} & 0\\ 0 & \hat{\sigma}^{-2}\end{smallmatrix}\right] \frac{P}{2}\mathbf{H}\mathbf{H}^H\right)\right], \tag{13}$$

where α_1 should be chosen to satisfy (12) with equality,[2] that is,

$$\frac{1}{2}\sum_{l=1}^{2} \mathbb{E}\left[\log\left(1 + \frac{P}{2\hat{\sigma}^2}\mathbf{H}_l\left(1 + \frac{P}{2\sigma^2}\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_l^H\right)\right] = \frac{\alpha_2}{\alpha_1}\mathbb{E}\left[\log\left(1 + \frac{P}{2\sigma^2}\|\mathbf{H}_k\|^2\right)\right]. \tag{14}$$

Combining the above equation with $\alpha_1 + \alpha_2 = 1$, we obtain the solution

$$\alpha_1 = \frac{\mathbb{E}\left[\log\left(1 + \frac{P}{2\sigma^2}\|\mathbf{H}_k\|^2\right)\right]}{\mathbb{E}\left[\log\left(1 + \frac{P}{2\sigma^2}\|\mathbf{H}_k\|^2\right)\right] + \frac{1}{2}\sum_{l=1}^{2}\mathbb{E}\left[\log\left(1 + \frac{P}{2\hat{\sigma}^2}\mathbf{H}_l\left(1 + \frac{P}{2\sigma^2}\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_l^H\right)\right]}. \tag{15}$$

Comparison to the MAT scheme

The proposed scheme can be compared to the original two-user MAT scheme which also works in two phases. There are three slots in total: two slots in phase 1 and one slot in phase 2. First, $\mathbf{V}_1$ and $\mathbf{V}_2$ are sent in slot 1 and 2, respectively, in a TDMA fashion. At the end of phase 1, the transmitter receives the CSI feedback and linearly combines the *overheard* signal from phase 1 as $L \triangleq \mathbf{H}_2^H\mathbf{V}_1 + \mathbf{H}_1^H\mathbf{V}_2$. In phase 2, the symbol L is scaled, e.g., to $\frac{1}{\sqrt{2}}L$ if the average transmit power constraint is imposed and if the channel coefficients are normalized.

[2]If the inequality (12) is strict, then one can always increase α_1 to achieves equality. This is without loss of optimality since increasing α_1 only increases the symmetric rate.

Then, the scaled signal is transmitted in one slot using one antenna. At the end, user 1 receives the noisy versions $\mathbf{H}_1^H\mathbf{V}_1 + Z_{11}$, $\mathbf{H}_1^H\mathbf{V}_2 + Z_{12}$, and $\frac{H'_{11}}{\sqrt{2}}(\mathbf{H}_2^H\mathbf{V}_1 + \mathbf{H}_1^H\mathbf{V}_2) + Z_{13}$ in three slots, where H'_{11} is the channel from the first antenna to user 1 at slot 3. From the three observations, receiver 1 gets the following virtual MIMO output

$$Y'_{11} = \mathbf{H}_1^H\mathbf{V}_1 + Z_{11} \tag{16}$$

$$Y'_{12} = \frac{1}{\sqrt{2}}H'_{11}\mathbf{H}_2^H\mathbf{V}_1 - \frac{1}{\sqrt{2}}H'_{11}Z_{12} + Z_{13}, \tag{17}$$

Due to the symmetry, receiver 2 has the similar form on $\mathbf{V}_2$. Finally, we conclude that the symmetric MAT rate is

$$R_{\text{sym}}^{\text{MAT}} = \frac{1}{3}\mathbb{E}\left[\log\det\left(\boldsymbol{I} + \begin{bmatrix}\sigma^{-2} & 0\\ 0 & \tilde{\sigma}_H^{-2}\end{bmatrix}\frac{P}{2}\mathbf{H}\mathbf{H}^H\right)\right], \tag{18}$$

where $\tilde{\sigma}_H^2 \triangleq \sigma^2(1 + \frac{2}{|H'_{11}|^2})$. Comparing (13) and (18), we notice the similarity of the rate expressions. Indeed, from (15), we see that if we set $\hat{\sigma}^2 = \sigma^2$, then $\alpha_1 \xrightarrow{P\to\infty} \frac{2}{3}$ which implies that both the MAT and the JSC schemes have the same prelog factor and thus the same DoF. However, the noise covariance inside the determinant is different in (13) and (18) since $\tilde{\sigma}_H^2$ is almost triple of $\hat{\sigma}^2$ if we approximate $|H'_{11}|^2$ by 1. Although the above comparison is not precise, it provides an idea that the power gain of the JSC scheme over the MAT scheme at high SNR is mainly due to the fact that the linear operations in MAT cause cumulation of noises from different phases. Intuitively, it is analogous to the advantage of compress-forward like schemes over amplify-forward like schemes in relay channels. At finite SNR, the proposed JSC scheme also provides the flexibility of choosing an appropriate compression noise variance $\hat{\sigma}^2$ as a function of P. Obviously, this flexibility requires that α_1 in (15) can be changed accordingly. In other words, with the JSC scheme one can adjust the length of the phases to achieve better performance, which is essential at finite SNR. Such flexibility is not possible with the MAT scheme since the length of each phase is fixed. Therefore, although the MAT scheme is DoF optimal, it may suffer from rate loss at finite SNR. More comments on the differences between the JSC scheme and the MAT-like schemes shall be made in Section VI.

IV. The General Case with K Users

In this section, we describe the general JSC scheme for the K-user BC with state feedback. The rate region is given in the following main result of this paper.

Theorem 1. *A rate tuple $(R_1, \ldots, R_K)$ is achievable in the K-user BC with causal state feedback if*

$$R_k \leq \alpha_1 I(V_k; Y_k^{(1)}, \{\hat{Y}_{1\to\mathcal{U}}\}_{\mathcal{U}\ni k} \,|\, S^{(1)}, Q^{(1)}), \tag{19}$$

$$0 \leq \min_{\substack{i,j,k,\mathcal{J}:\\ i<j, k\in\mathcal{J}}} \Big\{ \alpha_j I(V_{i\to\mathcal{J}}; Y_k^{(j)}, \{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}} \,|\, S^{(j)}, Q^{(j)}) - \alpha_i I(\{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}}; \hat{Y}_{i\to\mathcal{J}} \,|\, Y_k^{(i)}, S^{(i)}, Q^{(i)}) \Big\}. \tag{20}$$

for some K-tuple $(\alpha_1, \ldots, \alpha_K) \in \mathbb{R}_+^K$ with $\sum_k \alpha_k = 1$, and some pmf [3]

$$\left(\prod_{j=1}^{K} p(x^{(j)} \,|\, \{v_{\mathcal{J}}\}_{\mathcal{J}}, q^{(j)}) \right) \prod_{k=1}^{K} p(v_k) \prod_{j=2}^{K} \prod_{\mathcal{J}} \prod_{i=1}^{j-1} p(v_{i\to\mathcal{J}}) p(\hat{y}_{i\to\mathcal{J}} \,|\, \{v_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}}, s^{(i)}, q^{(i)}), \tag{21}$$

If we let $K = 2$ and $Q^{(2)}$ be deterministic, and identify $(V_{1\to 12}, \hat{Y}_{1\to 12})$ with $(V_{12}, \hat{Y})$, then we recover the results (9) and (12) in the two-user case. In order to have a general scheme for the K-user case, however, we need a much heavier notation as shown in the above theorem. As it will become clearer later, the complexity originates from the need to introduce K phases in each one of which different types of information are created and sent. In particular, as compared to the two-user case, there are in general a coded time-sharing random variable (RV) $Q^{(j)}$ in each phase $j \in \mathcal{K}$. The subscript "$i \to \mathcal{J}$" can be understood as related to the side information created in phase i and intended for the users in group $\mathcal{J}$ of j users. We recall that in the two-user case, there is only one type of side information that is intended for both users ($|\mathcal{J}| = 2$).

In the rest of the section, we present the general JSC scheme in detail and prove the achievability of the rate region given by Theorem 1. We divide the n-slot transmission into K phases, each phase $j \in \mathcal{K}$ having length n_j such that $n = n_1 + \cdots + n_K$. We define the normalized length of phase j as $\alpha_j \triangleq \frac{n_j}{n}$ with $\sum_{j=1}^{K} \alpha_j = 1$.

The K-user scheme works in a similar manner as the two-user scheme. In each phase $j \in \mathcal{K}$,

- the input, output, and state are denoted by $X^{(j)}$, $Y^{(j)}$, and $S^{(j)}$, respectively;
- if $j = 1$, the original messages $\{M_k \in \mathcal{M}_k \triangleq [1 : 2^{nR_k}] :\ k \in \mathcal{K}\}$ are sent, otherwise a set of side information messages $\{M_{i\to\mathcal{J}} \in \mathcal{M}_{i\to\mathcal{J}} \triangleq [1 : 2^{n_i R_{i\to\mathcal{J}}}] :\ |\mathcal{J}| = j,\ i < j\}$ are sent;
- each message $M_{i\to\mathcal{J}}$ is related to the source RV $\hat{Y}_{i\to\mathcal{J}}$ created in a previous phase i, and is carried by $V_{i\to\mathcal{J}}$ for transmission;
- the transmission is controlled by a coded time-sharing RV $Q^{(j)}$.

[3] We define $v_{\mathcal{U}} \triangleq \{v_{k\to\mathcal{U}} :\ k < |\mathcal{U}|\}$ and $V_{\mathcal{U}} \triangleq \{V_{k\to\mathcal{U}} :\ k < |\mathcal{U}|\}$ for brevity. We also recall that $\mathcal{I}$ and $\mathcal{J}$ are subsets of size i and j, respectively.

Codebook Generation: Fix the pmf as described in (21).

1) Before the beginning of phase $j \in \mathcal{K}$, randomly generate the time-sharing sequence according to $\prod_{t=1}^{n_j} p(q_t^{(j)})$.
2) At the beginning of phase 1, for each user $k \in \mathcal{K}$, randomly generate 2^{nR_k} independent sequences $\boldsymbol{v}_k(m_k)$, $m_k \in [1 : 2^{nR_k}]$, each according to $\prod_{t=1}^{n_1} p(v_{k,t})$.
3) At the end of phase $i \in \{1, \ldots, K-1\}$, for each $j > i$ and each $\mathcal{J}$ with $|\mathcal{J}| = j$, randomly generate $2^{n_i R_{i\to\mathcal{J}}}$ independent sequences $\hat{\boldsymbol{y}}_{i\to\mathcal{J}}(m_{i\to\mathcal{J}})$ and $2^{n_i R_{i\to\mathcal{J}}}$ independent sequences $\boldsymbol{v}_{i\to\mathcal{J}}(m_{i\to\mathcal{J}})$, $m_{i\to\mathcal{J}} \in [1 : 2^{n_i R_{i\to\mathcal{J}}}]$, each according to $\prod_{t=1}^{n_i} p(\hat{y}_{i\to\mathcal{J},t} \,|\, s_t^{(i)}, q_t^{(i)})$ and $\prod_{t=1}^{n_j} p(v_{i\to\mathcal{J},t})$, respectively.

Encoding:

1) In phase 1, to send the original messages $(M_1, \ldots, M_K)$, a sequence $\boldsymbol{x}^{(1)}$ is first generated based on $(\boldsymbol{v}_1(M_1), \cdots, \boldsymbol{v}_K(M_K), \boldsymbol{q}^{(1)})$ according to $\prod_{t=1}^{n_1} p(x^{(1)} \,|\, v_{1,t}, \ldots, v_{K,t}, q_t^{(1)})$ and then transmitted.
2) At the end of phase i, $i = 1, \ldots, K-1$, and for each $j > i$ and each $\mathcal{J}$ with $|\mathcal{J}| = j$, given the state feedback of all the previous phases, the source searches for an index $M_{i\to\mathcal{J}}$ such that $(\hat{\boldsymbol{y}}_{i\to\mathcal{J}}(M_{i\to\mathcal{J}}), \{\boldsymbol{v}_{l\to\mathcal{I}}\}_{l<i,\mathcal{I}\subset\mathcal{J}}, \boldsymbol{s}^{(i)}, \boldsymbol{q}^{(i)}) \in \mathcal{T}_{\epsilon_{n_i}}^{n_i}(\hat{Y}_{i\to\mathcal{J}}, \{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}}, S, Q)$[4]. According to the covering lemma [16], this is feasible with probability going to 1 when $n_i \to \infty$, if

$$n_i R_{i\to\mathcal{J}} \geq n_i I(\hat{Y}_{i\to\mathcal{J}}; \{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}} \,|\, S^{(i)}, Q^{(i)}) + n_i \epsilon_{n_i}. \tag{22}$$

Decoding: We focus on the decoding procedure of a particular receiver $k \in \mathcal{K}$ without loss of generality. At the end of phase K, a backward decoding is performed. Specifically, for phase $j = K, K-1, \ldots, 1$, the set of messages, $\{M_{\mathcal{J}}\}_{\mathcal{J}\ni k}$, intended for user k is decode as follows.

1) For phase j, $j = K, K-1, \ldots, 2$, by construction the "future" message set $\{\hat{M}_{j\to\mathcal{U}} : \mathcal{U} \ni k, |\mathcal{U}| > j\}$ has been decoded previously. The goal is to decode the "current" messages $\hat{M}_{i\to\mathcal{J}}$ for each $\mathcal{J} \ni k$ and each $i < j$. To that end, the decoder looks for a *unique* index

[4]It is worth clarifying that the *weak* typicality compatible with both discrete and continuous RVs is used. As in [15], weak typicality is defined as $\mathcal{T}_\epsilon^n(x^n) \triangleq \{x^n : |-\frac{1}{n}\log p(x^n) - H(X)| \leq \epsilon\}$ for discrete RVs and $\mathcal{T}_\epsilon^n(x^n) \triangleq \{x^n : |-\frac{1}{n}\log p(x^n) - h(X)| \leq \epsilon\}$ for continuous RVs, where with a bit abuse of notation we denote pmf and probability density function with the same notation $p(x^n)$ for discrete and continuous cases, respectively.

$\hat{M}_{i\to\mathcal{J}}$ such that the following joint typicalities are satisfied *simultaneously*

$$\begin{aligned}\left(\boldsymbol{v}_{i\to\mathcal{J}}(\hat{M}_{i\to\mathcal{J}}), \boldsymbol{y}_k^{(j)}, \{\hat{\boldsymbol{y}}_{j\to\mathcal{U}}(\hat{M}_{j\to\mathcal{U}})\}_{\mathcal{U}\supset\mathcal{J}}, \boldsymbol{s}^{(j)}, \boldsymbol{q}^{(j)}\right) &\in \mathcal{T}_{\epsilon_{n_j}}^{n_j}(V_{i\to\mathcal{J}}, Y_k, \{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}}, S^{(j)}, Q^{(j)})\\ \left(\hat{\boldsymbol{y}}_{i\to\mathcal{J}}(\hat{M}_{i\to\mathcal{J}}), \boldsymbol{y}_k^{(i)}, \boldsymbol{s}^{(i)}, \boldsymbol{q}^{(i)}\right) &\in \mathcal{T}_{\epsilon_{n_i}}^{n_i}(\hat{Y}_{i\to\mathcal{J}}, Y_k, S^{(i)}, Q^{(i)}). \end{aligned} \tag{23}$$

The probability that such an index cannot be found or is not correct ($\hat{M}_{i\to\mathcal{J}} \neq M_{i\to\mathcal{J}}$) vanishes when $n_i \to \infty$ provided that

$$n_i R_{i\to\mathcal{J}} \leq n_j I(V_{i\to\mathcal{J}}; Y_k^{(j)}, \{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}} \,|\, S^{(j)}, Q^{(j)}) + n_i I(\hat{Y}_{i\to\mathcal{J}}; Y_k^{(i)} \,|\, S^{(i)}, Q^{(i)}) - n_j\epsilon_{n_j} - n_i\epsilon_{n_i}. \tag{24}$$

The error event analysis that leads to the above rate follows the exact same steps as the one in [14, Sec. IV, p.1476], and is omitted here due to the space limitation.

2) Finally, for phase 1, the decoder searches for a unique $\hat{M}_k$ such that

$$\left(\boldsymbol{v}_k(\hat{M}_k), \boldsymbol{y}_k^{(1)}, \{\hat{\boldsymbol{y}}_{1\to\mathcal{U}}(\hat{M}_{1\to\mathcal{U}})\}_{\mathcal{U}\ni k}, \boldsymbol{s}^{(1)}, \boldsymbol{q}^{(1)}\right) \in \mathcal{T}_{\epsilon_{n_1}}^{n_1}(V_k, Y_k, \{\hat{Y}_{1\to\mathcal{U}}\}_{\mathcal{U}\ni k}, S^{(1)}, Q^{(1)})$$

According to the packing lemma, the probability that such an index cannot be found or is not correct ($\hat{M}_k \neq M_k$) vanishes when $n_1 \to \infty$ provided that

$$nR_k \leq n_1 I(V_k; Y_k^{(1)}, \{\hat{Y}_{1\to\mathcal{U}}\}_{\mathcal{U}\ni k} \,|\, S^{(1)}, Q^{(1)}) - n_1\epsilon_{n_1}. \tag{25}$$

To summarize, $(R_1, \ldots, R_K)$ is achievable if for each k, (25) is satisfied subject to the existence of $\{R_{i\to\mathcal{J}} :\ i < j, |\mathcal{J}| = j\}$ that verify (22) and (24). Thus let $n_1, \ldots, n_K$ go to infinity by keeping the same ratio $\alpha_1, \ldots, \alpha_n$, we obtain the rate (20) if, for each $(k, i, j, \mathcal{J})$ with $i < j$ and $|\mathcal{J}| = j$,

$$I(\hat{Y}_{i\to\mathcal{J}}; \{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}} \,|\, S^{(i)}, Q^{(i)}) \leq \frac{\alpha_j}{\alpha_i} I(V_{i\to\mathcal{J}}; Y_k^{(j)}, \{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}} \,|\, S^{(j)}, Q^{(j)}) + I(\hat{Y}_{i\to\mathcal{J}}; Y_k^{(i)} \,|\, S^{(i)}, Q^{(i)}).$$

Using the Markovity $\hat{Y}_{i\to\mathcal{J}} \leftrightarrow (\{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}}, S^{(i)}, Q^{(i)}) \leftrightarrow Y_k^{(i)}$ and the chain rule, we have

$$I(\hat{Y}_{i\to\mathcal{J}}; \{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}} \,|\, S^{(i)}, Q^{(i)}) - I(\hat{Y}_{i\to\mathcal{J}}; Y_k^{(i)} \,|\, S^{(i)}, Q^{(i)}) = I(\hat{Y}_{i\to\mathcal{J}}; \{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}} \,|\, Y_k^{(i)}, S^{(i)}, Q^{(i)}) \tag{26}$$

which leads to the constraint (20). This completes of proof of Theorem 1.

V. Applications to the GBC and EBC

In this section, we apply the general result in Theorem 1 to the fading Gaussian BC and the erasure BC. The key is to fix the distribution (21) of the RVs involved in the rate region appropriately for each channel.

In particular, the coded time-sharing random variable $Q^{(j)}$, $j \in \mathcal{K}$, is used to indicate which of the messages $M_{i\to\mathcal{J}}$ is to be sent.[5] Thus, it is natural to define $Q^{(j)}$ as $(Q_1^{(j)}, Q_2^{(j)})$, where $Q_1^{(j)} \in \mathcal{Q}_1^{(j)}$ and $Q_2^{(j)} \in \mathcal{Q}_2^{(j)}$ with

$$\mathcal{Q}_1^{(j)} \triangleq \begin{cases} \{0\}, & j = 1, \\ \{1, \ldots, j-1\}, & j > 1, \end{cases} \quad \text{and} \quad \mathcal{Q}_2^{(j)} \triangleq \{\mathcal{J} : \ |\mathcal{J}| = j\}. \tag{27}$$

We also let the channel input X be a deterministic function of V and Q. Specifically, when $Q^{(j)} = (i, \mathcal{J})$, the message $M_{i\to\mathcal{J}}$ is carried by $V_{i\to\mathcal{J}}$. Hence, we set

$$X^{(j)} = V_{Q_1^{(j)} \to Q_2^{(j)}}. \tag{28}$$

A. Fading Gaussian BC

For the fading GBC, we focus on the symmetric channel and the corresponding symmetric rate for simplicity. To that end, we make the following choices on the RVs:

- Time-sharing RVs. We let $Q_1^{(j)}$ be deterministic, namely, $Q_1^{(j)} = j - 1$, and let $Q_2^{(j)}$ be uniformly distributed over $\mathcal{Q}_2^{(j)}$, namely

  $$\mathrm{P}\left(Q_2^{(j)} = \mathcal{J}\right) = \binom{K}{j}^{-1}, \quad \forall\, \mathcal{J} \in \mathcal{Q}_2^{(j)}. \tag{29}$$

  Intuitively, the above choice means that in phase j, we only transmit side information created in the previous phase $j-1$. This is similar to the general MAT scheme [4]. We can write $Q^{(j)} = (j-1, Q_2^{(j)})$. The uniformity is simply due to the symmetry of the setting.
- Gaussian distributed V's. The RVs V's with different subscripts are independent and identically distributed (i.i.d.) according to $\mathcal{CN}(\mathbf{0}, \frac{P}{n_{\mathrm{t}}}\boldsymbol{I}_{n_{\mathrm{t}}})$. It means that all the transmit antennas are used in each phase with isotropic signaling, which can be justified by the lack of instantaneous CSIT.
- Side information $\hat{Y}$ as compression of the overheard signal. In phase i, when $Q^{(i)} = (i-1, \mathcal{I})$ for some $\mathcal{I}$, we set

  $$\hat{\mathbf{Y}}_{i\to\mathcal{J}} = \begin{cases} \mathsf{H}_{\mathcal{J}\setminus\mathcal{I}}\mathbf{X}^{(i)} + \hat{\mathbf{Z}}_{\mathcal{J}\setminus\mathcal{I}}, & \text{if } \mathcal{J} \supset \mathcal{I} \text{ and } |\mathcal{J}| = i+1, \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

  where $\hat{\mathbf{Z}}_{\mathcal{J}\setminus\mathcal{I}} \sim \mathcal{CN}(\mathbf{0}, \beta_i\sigma^2\boldsymbol{I})$ is the compression noise with $\beta_i > 0$ being a parameter to be fixed later. The intuition behind (30) is the following. When $Q^{(i)} = (i-1, \mathcal{I})$, the information

[5]Slightly abusing the subscript notation, we sometimes write $(\cdot)_k$ as $(\cdot)_{0\to\{k\}}$, e.g., the original message M_k is also $M_{0\to\{k\}}$

intended for the users in the set $\mathcal{I}$ is being sent and is overheard by some user $k \notin \mathcal{I}$. Let $\mathcal{J} = \{k\} \cup \mathcal{I}$ be the new group. Then, the overheard signal $\mathsf{H}_{\mathcal{J}\setminus\mathcal{I}}\mathbf{X}^{(i)}$ is indeed interested by the users in group $\mathcal{I}$ since it provides an extra observation[6]. Furthermore, thanks to the joint source-channel coding, $\mathsf{H}_{\mathcal{J}\setminus\mathcal{I}}\mathbf{X}^{(i)}$ as a side information does not cost receiver k much to decode since it already has some noisy version of the information.

Applying the above RVs to the general region in Theorem 1, we obtain the following corollary. Some intermediate steps are rather technical and deferred to Appendix A.

Corollary 1. *For the K-user symmetric $n_t \times n_r$ fading GBC, the symmetric rate:*

$$R_{sym}^{GBC} = \max_{(\beta_1,\ldots,\beta_{K-1})\in\mathbb{R}_+^{K-1}} \left(K + \sum_{j=2}^{K} \binom{K}{j} \prod_{t=2}^{j} \frac{\sum_{l\le t} b_{l,t}}{a_t} \right)^{-1} a_1, \tag{31}$$

is achievable, where, for $t = 1, \ldots, K$,

$$a_t \triangleq \mathbb{E}\left[\log\det\left(\boldsymbol{I} + \mathsf{snr}\,\mathsf{H}_{\mathcal{T}}^H \boldsymbol{\Lambda}_t \mathsf{H}_{\mathcal{T}}\right)\right], \tag{32}$$

$$b_{l,t} \triangleq \mathbb{E}\left[\log\det\left(\boldsymbol{I} + \frac{\mathsf{snr}}{\beta_{t-1}}\mathsf{H}_l(\boldsymbol{I} + \mathsf{snr}\,\mathsf{H}_1^H\mathsf{H}_1)^{-1}\mathsf{H}_l^H\right)\right], \tag{33}$$

with $\mathcal{T} \triangleq \{1\} \cup \{t+1, \ldots, K\}$ and $\boldsymbol{\Lambda}_t \triangleq \mathrm{diag}\{\boldsymbol{I}_{n_r}, \beta_t^{-1}\boldsymbol{I}_{(K-t)n_r}\}$.

Although the maximization in (31) is not convex in general, it can be done numerically. We shall comment more on this aspect in the next section with some examples.

Now let us take a look at the high SNR regime. We consider the MISO case with $n_{\mathrm{t}} = K$. We shall show from the above rate (31) that the optimal symmetric DoF can be achieved. To that end, we let the compression noise variance be $\beta_i = 1$, $i = 1, \ldots, K-1$. From (32) and (33), one can verify that, at high SNR,

$$a_t = |\mathcal{T}|\log\mathsf{snr} + O(1) = (K - t + 1)\log\mathsf{snr} + O(1), \tag{34}$$

$$b_{l,t} = \begin{cases} O(1), & l = 1, \\ \log\mathsf{snr} + O(1), & l \neq 1. \end{cases} \tag{35}$$

[6]When $n_{\mathrm{r},1} + \cdots + n_{\mathrm{r},K} \le n_{\mathrm{t}}$, such observation is linearly independent of what each user in $\mathcal{I}$ already has.

Since the DoF is defined as $\mathsf{DoF}_{\text{sym}} \triangleq \lim_{\mathsf{snr}\to\infty} \frac{R_{\text{sym}}}{\log \mathsf{snr}}$, it follows from (31) that

$$\mathsf{DoF}_{\text{sym}} = \left(K + \sum_{j=2}^{K} \binom{K}{j} \prod_{t=2}^{j} \frac{t-1}{K-t+1} \right)^{-1} K \tag{36}$$

$$= \left(K + \sum_{j=2}^{K} \binom{K}{j} \binom{K-1}{j-1}^{-1} \right)^{-1} K \tag{37}$$

$$= \left(\sum_{j=1}^{K} \frac{1}{j} \right)^{-1}, \tag{38}$$

which coincides with the optimal symmetric DoF derived in [4] for the same channel. Note that the DoF achievability holds for all $\{\beta_i > 0\}_i$ that do not scale with the SNR, while at finite SNR the exact values of the β's actually matter for the rate performance.

B. Erasure BC

Next let us consider the erasure BC. We make the following choices on the RVs:

- Time-sharing RVs. Let us recall that $Q^{(j)} = (Q_1^{(j)}, Q_2^{(j)})$. Here we let $Q_1^{(j)}$ and $Q_2^{(j)}$ be independent for each $j \in \mathcal{K}$, i.e.,

$$p(q^{(j)}) = p(q_1^{(j)}) p(q_2^{(j)}), \quad \forall\, q^{(j)} \in \mathcal{Q}^{(j)}. \tag{39}$$

  However, we do not specify the distribution of $(Q_1^{(j)}, Q_2^{(j)})$.

- Uniformly distributed V's. The RVs V's with different subscripts are i.i.d. over the input alphabet $\mathcal{X}$ according to a uniform distribution. Specifically, the distribution of $V_{i\to\mathcal{J}}$, for each $i < j$ and $|\mathcal{J}| = j$, is

$$p(v_{i\to\mathcal{J}}) = \frac{1}{|\mathcal{X}|}, \quad \forall\, v_{i\to\mathcal{J}} \in \mathcal{X}. \tag{40}$$

  This choice guarantees the maximum entropy of the V's, with $H(V_{i\to\mathcal{J}}) = \log|\mathcal{X}|$.

- Side information $\hat{Y}$ as the overheard signal. In phase i, when $Q^{(i)} = (l, \mathcal{I})$ for some $l < i$ and $|\mathcal{I}| = i$, we set

$$\hat{\mathbf{Y}}_{i\to\mathcal{J}} = \begin{cases} X^{(i)}, & \text{if } \mathcal{J} \supset \mathcal{I},\ S_{\mathcal{I}} \neq \mathbf{1},\ S_{\mathcal{J}\setminus\mathcal{I}} = \mathbf{1}, \text{and } S_{\bar{\mathcal{J}}} = \mathbf{0}, \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

  The intuition behind (41) is the following. When $Q^{(i)} = (i-1, \mathcal{I})$, the information intended for the users in the set $\mathcal{I}$ is being sent. If this information is not received by some of the users in $\mathcal{I}$ (i.e. $S_{\mathcal{I}} \neq \mathbf{1}$), and meanwhile received by some unintended users defined by $\mathcal{U}$

with $\mathcal{U}\cap\mathcal{I}=\emptyset$, then we define a new group $\mathcal{J}=\mathcal{U}\cup\mathcal{I}$. We have the conditions $S_{\mathcal{J}\setminus\mathcal{I}}=\mathbf{1}$ and $S_{\bar{\mathcal{J}}}=\mathbf{0}$. Thanks to the joint source-channel coding, such signal does not cost receivers in $\mathcal{J}\setminus\mathcal{I}$ anything to decode since they already have the information.

Applying the above RVs to the general region in Theorem 1, we obtain the following corollary. As in the Gaussian case, the technical intermediate steps are deferred to Appendix B.

Corollary 2. *The rate tuple $(R_1,\ldots,R_K)\in\mathbb{R}_+^K$ is achievable in the EBC with state feedback if*

$$R_k\le\alpha_1\mathrm{P}(Q_2^{(1)}=k)(1-\delta_{\mathcal{K}})\log|\mathcal{X}|, \tag{42}$$

$$0\le\min_{\substack{j,k,\mathcal{J}:\\ k\in\mathcal{J}}}\Big\{\alpha_j\mathrm{P}(Q_2^{(j)}=\mathcal{J})(1-\delta_{\mathcal{K}\setminus\mathcal{J}\cup\{k\}})-\sum_{i=1}^{j-1}\alpha_i\sum_{\mathcal{I}\subset\mathcal{J},\mathcal{I}\ni k}\mathrm{P}(Q_2^{(i)}=\mathcal{I})\phi_{\mathcal{K}\setminus\mathcal{J}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}\Big\}, \tag{43}$$

for some K-tuple $(\alpha_1,\ldots,\alpha_K)\in\mathbb{R}_+^K$ with $\sum_k\alpha_k=1$, and some distribution of $\{Q_2^{(j)}\}_{j\in\mathcal{K}}$.

For the symmetric EBC, the above region coincides with the capacity region, as will be shown in Corollary 3.

Definition 1. *An EBC is said to be symmetric if the erasure probability $\delta_{\mathcal{U}}$ only depends on the cardinality of the set $\mathcal{U}$, that is, $\delta_{\mathcal{U}}=\delta_{\mathcal{U}'}$ if $|\mathcal{U}|=|\mathcal{U}'|$.*

Corollary 3. *The JSC scheme achieves the following capacity region of the symmetric EBC.*

$$\mathcal{C}_{sym}^{EBC}=\bigcap_{\boldsymbol{\pi}}\left\{\begin{array}{c}(R_{\pi(1)},\ldots,R_{\pi(K)})\in\mathbb{R}_+^K:\\ \sum_{k=1}^K\dfrac{R_{\pi(k)}}{1-\delta_{\{\pi(1),\cdots,\pi(k)\}}}\le\log|\mathcal{X}|\end{array}\right\}, \tag{44}$$

where the intersection is over all permutations $\boldsymbol{\pi}\triangleq(\pi(1),\cdots,\pi(K))$ of $(1,\ldots,K)$.

Proof. The converse can be found in [2], [3] with the standard outer bound techniques by creating an artificial degraded BC. A detailed proof on the achievability is provided in Appendix C. ■

Remark 2. *The capacity region for the general EBC with state feedback is still unknown. In [2], [3], the authors designed a special scheme that can achieve the capacity region for the general EBC with three users. In their capacity-achieving scheme, the transmitted signal can depend simultaneously on messages from different phases, e.g., M_1 and $M_{2\to\{2,3\}}$. Such a result suggests that coded time-sharing may not be enough to achieve the capacity with JSC scheme in general. We believe that it is possible to set the RVs in our region in a similar way as the scheme in [2], [3] to achieve the three-user capacity region. However, it is out of the scope of*

the current paper and is not considered here. Nevertheless, the capacity region beyond three users remains unknown.

VI. Numerical Examples

In this section, we consider the Gaussian MISO channel with i.i.d. Rayleigh fading for $K = 2$ and 3 users. We let $n_{\mathrm{t}} = K$ and evaluate the symmetric rate (31) of the JSC scheme. The maximization (31) over $\{\beta_1, \ldots, \beta_{K-1}\}$ is done numerically. Since K is small in our examples, we simply sample each β_i uniformly within a given region of β_i with a small step size and then find out the maximum value of (31).[7] For larger values of K, however, more sophisticated numerical methods may be needed.

Our scheme is compared to the following baseline schemes:

1) The TDMA scheme. It is optimal for the no CSIT case, and achieves the following symmetric rate

$$R_{\mathrm{sym}}^{\mathrm{TDMA}} \triangleq \frac{1}{K} I(X; Y_1 \,|\, S) \tag{45}$$

$$= \frac{1}{K} \mathbb{E}\left[\log(1 + \mathsf{snr}\|\mathbf{H}_1\|^2)\right]. \tag{46}$$

2) The original MAT scheme from [4]. Here we consider the MAT scheme as described in [4], except for adding a proper linear scaling factor to meet the transmit power constraint.
3) The generalized MAT (GMAT) scheme from [5]. With GMAT, a precoder is designed to balance the alignment of interference and the enhancement of each signal. The GMAT scheme includes the MAT scheme as a special case by letting the precoder be the respective channel matrix $\boldsymbol{H}$ to reconstruct the overheard observations.
4) The quantized MAT (QMAT) scheme from [6]. Instead of sending the analogy linear combinations as the MAT scheme does, the QMAT transmits a *quantized* version of each linear combination. In phase j, it turns out that $(j-1)(j+2)$ is the minimum quantization noise variance such that the message for group $\mathcal{J}$ can be recovered at each user through a $(K-j+1)\times 1$ MISO channel. The achievable symmetric rate is

$$R_{\mathrm{sym}}^{\mathrm{QMAT}} = \frac{1}{K} \mathsf{DoF}_{\mathrm{sym}}\, \mathbb{E}\left[\log\det\left(\boldsymbol{I} + \mathsf{snr}\, \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^H \tilde{\boldsymbol{N}}^{-1}\right)\right], \tag{47}$$

where $\tilde{\boldsymbol{N}} \triangleq \mathrm{diag}\Big(\{1+(j-1)(j+2)\}_{j=1:K}\Big)$ and $\mathsf{DoF}_{\mathrm{sym}}$ is given by (38). The fundamental differences between our JSC scheme and the QMAT are: 1) we use joint source-channel

[7] Although the numerical maximum value with such a method may not be optimal, but it still represents an achievable rate.

coding while QMAT uses separate coding, 2) our source codebook is generated by $\hat{Y}$ that indicates what the users need while QMAT explicitly generates linear combinations and the quantization of each combinations, and 3) the JSC scheme uses all the transmit antennas all the time while the QMAT uses only a subset of $K-j+1$ transmit antennas in each phase $j \in \mathcal{K}$.

5) The genie-aided upper bound. For $k \in \mathcal{K}$, a genie provides the output Y_k to users $l \in \{k+1, \cdots, K\}$. The new channel can only have a larger capacity region than the original one, and it is a *physically degraded* BC whose capacity region cannot be enlarged with feedback. The single-letter characterization of the capacity region of such degraded BC is, for some pmf $p(x \,|\, u_{K-1}) \prod_{k=2}^{K-1} p(u_k \,|\, u_{k-1}) p(u_1)$, is

$$R_k \le I(U_k; Y_1^k \,|\, U_{k-1}, S), \forall k \in \mathcal{K}, \tag{48}$$

where we define $U_K = X$ and $U_0 = 0$ for convenience [16]. Thus, the symmetric capacity of the original channel must satisfy (48), which yields the following upper bound on the weighted sum

$$\sum_{k=1}^{K} \frac{C_{\text{sym}}}{k} \le I(U_1; Y_1 \,|\, S) + \cdots + \frac{1}{K} I(U_K; Y_1^K \,|\, U_{K-1}, S) \tag{49}$$

$$\le h(Y_1 \,|\, S) + \sum_{k=1}^{K-1} \left(\frac{h(Y_1^{k+1} \,|\, U_k, S)}{k+1} - \frac{h(Y_1^k \,|\, U_k, S)}{k} \right) - \frac{1}{K} h(Y_1^K \,|\, X, S) \tag{50}$$

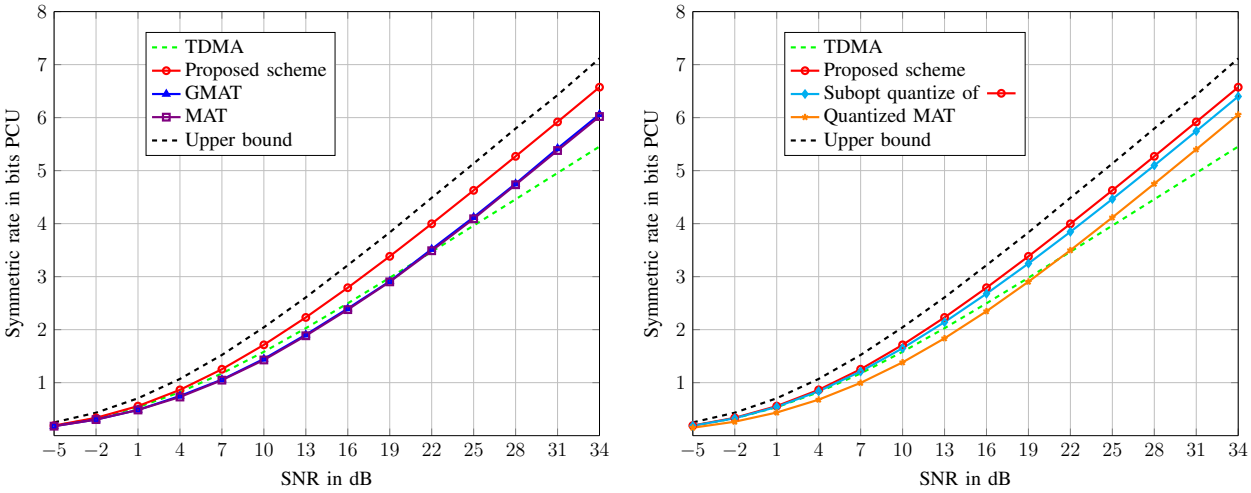
$$\le h(Y_1 \,|\, S) - \frac{1}{K} h(Y_1^K \,|\, X, S) \tag{51}$$

$$= I(X; Y_1 \,|\, S) = K R_{\text{sym}}^{\text{TDMA}}, \tag{52}$$

where the second inequality is from the Markovity $h(Y_1^k \,|\, U_k, U_{k-1}, S) = h(Y_1^k \,|\, U_k, S)$ by the construction of the pmf; the third inequality follows from the symmetry of the channel output in a symmetric fading channel [17], i.e., when $|\mathcal{J}| \ge |\mathcal{I}|$, $\frac{h(Y_{\mathcal{J}} \,|\, U, S)}{|\mathcal{J}|} \le \frac{h(Y_{\mathcal{I}} \,|\, U, S)}{|\mathcal{I}|}$; and the last equality holds since $h(Y_1^K \,|\, X, S) = h(Z_1^K \,|\, S) = K\, h(Y_1 \,|\, X, S)$. Hence, we have the following upper bound on the symmetric capacity

$$C_{\text{sym}} \le \mathsf{DoF}_{\text{sym}} R_{\text{sym}}^{\text{TDMA}}. \tag{53}$$

The two-user and three-user cases are evaluated separately in Fig. 2 and Fig. 3, respectively. In both Fig. 2b and Fig. 3b, the curve ⟷ denotes a variant of our proposed scheme where the quantization noises β_j's are not optimized. Instead, we apply the same equivalent compression



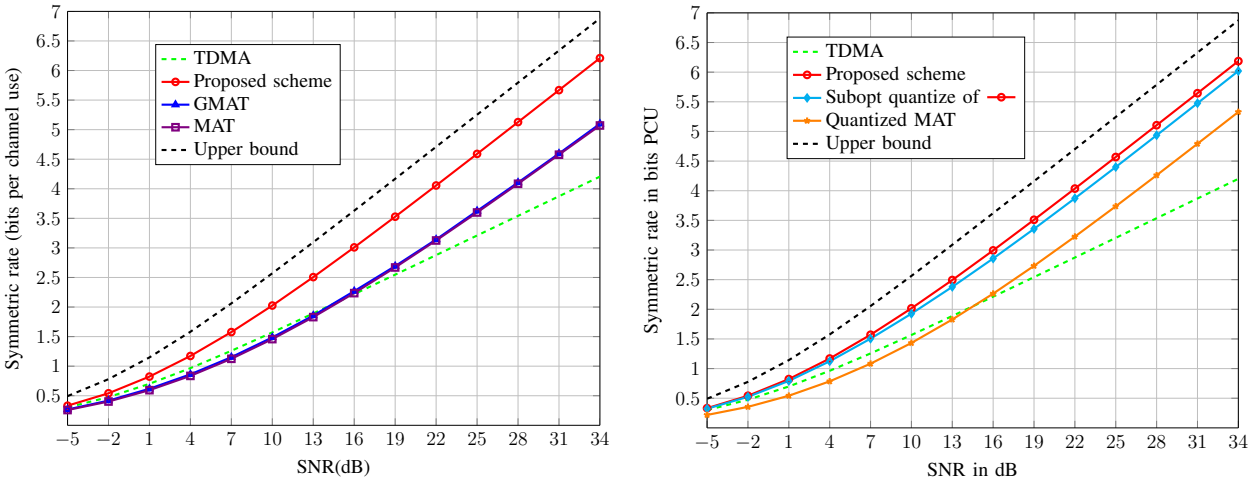
(a) Proposed scheme versus MAT, GMAT.

(b) Proposed scheme versus Quantized MAT.

Figure 2: The proposed scheme versus the baseline schemes: two-user BC.

noise variance used in the QMAT scheme, that is, $\beta_{j-1} = 1 + (j-1)(j+2)$. We have the following comments on the results.

- From the plots, we see that the curves of the GMAT scheme proposed in [5] and the MAT curves almost overlap in all SNR regime for $K = 2, 3$. It shows that the performance improvement brought by carefully designing the linear combinations (referred to as pre-coder) is marginal in the i.i.d. isotropic fading case. Another generalization direction within the MAT framework is the quantization of linear combinations. Although the MAT and QMAT are not compared directly in the same plot, we can still observe that the QMAT does outperform the MAT scheme especially when $K = 3$ in medium-to-high SNR regime. However, appreciable gain appears only at high SNR.
- In the low-to-medium SNR regime, the MAT/GMAT/QMAT schemes are outperformed by the TDMA. This result is somewhat surprising since, unlike the other schemes, TDMA does not exploit the state feedback. Indeed, the MAT-like schemes use the state feedback to perform interference alignment which is known to be optimal at high SNR but is usually less good when the SNR is not high. In such regime, the channel is not interference limited and sending linear equations may be too costly for the marginal interference mitigation effect. In the high SNR regime, the MAT-like schemes dominates the TDMA scheme eventually thanks to a larger DoF gain, which is reflected by the slopes of the curves. The optimal



(a) Proposed scheme versus MAT, GMAT.

(b) Proposed scheme versus Quantized MAT.

Figure 3: The proposed scheme versus the baseline schemes: three-user BC.

DoF of the MAT-like schemes is also confirmed by the fact that the corresponding curves are almost parallel to the upper bound curve.

- In all SNR regimes, the proposed JSC scheme outperforms all four baseline schemes (TDMA, MAT, GMAT, QMAT) and has a non-negligible power gain over the MAT-like schemes. This gain becomes more appealing in the medium-to-low SNR regime in which the MAT-like schemes are not even better than the simple TDMA scheme. Our scheme can still take advantage of the state feedback to achieve a better performance. This is mainly thanks to the flexibility over the duration of each phase (time-slot) and the compression parameters as a function of the SNR, which is not possible with the MAT/GMAT schemes. The comparison to the QMAT scheme is even more interesting, since both the JSC scheme and QMAT are based on compression. We see that the performance gain over QMAT is almost 3 dB for $K = 2$ and is up to 6 dB for $K = 3$. To analyze the causes of such a significant gain, we fix the $\beta_{j-1} = 1 + (j-1)(j+2)$ which corresponds to the same setting in the QMAT. As can be seen from the plots in Fig. 2b and Fig. 3b, the JSC scheme still dominates the QMAT with a slight performance degradation from the case with optimized β's. Such an observation suggests that the main performance gain of our scheme over the QMAT comes from the joint source-channel coding.

VII. Conclusion

In this paper, we proposed a novel scheme for the general state-dependent K-user broadcast channel with state feedback. The proposed scheme is based on joint source-channel coding and coded time-sharing. Thanks to the systematic and scalable structure of this scheme, we managed to derive the corresponding achievable region in terms of a reasonable number of parameters. Such region was then evaluated for two special cases, namely, the erasure BC and fading Gaussian BC. In particular, we showed that our results covered the previously known capacity region for the erasure BC. In addition, for the fading Gaussian BC, we demonstrated through numerical evaulation a non-negligible power gain of our scheme over the existing ones in the literature. We argued that such a substantial performance gain comes from the use of joint source-channel coding which is still highly theoretical. Practical implementation of such schemes would be an interesting direction to explore in the future.

Appendix A
Proof of Corollary 1

To study the symmetric rate, we let $R_1 = \cdots = R_K = R_{\text{sym}}$. Due to the symmetry of the channel, it is without loss of generality to consider receiver 1. In particular, we apply the RVs choice given in Section V-A, and compute the quantities in (19) and (20). We define $\tilde{\mathcal{J}} \triangleq \{1\} \cup \{j+1, \cdots, K\}$, $\mathbf{W}_j \triangleq \left[\mathbf{Z}_1^T, \hat{\mathbf{Z}}^T_{\mathcal{U}_{j+1}\setminus\mathcal{J}}, \cdots, \hat{\mathbf{Z}}^T_{\mathcal{U}_K\setminus\mathcal{J}}\right]^T$. Then, we have

$$
\begin{aligned}
&I(V_1; Y_1^{(1)}, \{\hat{Y}_{1\to\mathcal{U}}\}_{\mathcal{U}\ni k} \,|\, S^{(1)}, Q^{(1)}) \\
&\quad = \mathrm{P}(Q^{(1)} = (0, \{1\})) I(V_1; Y_1^{(1)}, \hat{Y}_{1\to\{1,2\}}, \ldots, \hat{Y}_{1\to\{1,K\}} \,|\, S^{(1)}, Q^{(1)} = (0, \{1\})) && (54)\\
&\quad = K^{-1} I(\mathbf{V}_1; \mathsf{H}_{\mathcal{K}}\mathbf{V}_1 + \mathbf{W}_1 \,|\, \mathsf{H}_{\mathcal{K}}) && (55)\\
&\quad = K^{-1}\mathbb{E}\left[\log\det\left(\boldsymbol{I} + \mathsf{snr}\,\mathsf{H}_{\mathcal{K}}^H \boldsymbol{\Lambda}_1 \mathsf{H}_{\mathcal{K}}\right)\right] = K^{-1} a_1. && (56)
\end{aligned}
$$

Similarly, assuming $i=j-1$ and $|\mathcal{U}|=|\mathcal{J}|+1=j+1$, we obtain

$$\begin{aligned}
I(V_{i\to\mathcal{J}};Y_1^{(j)},\{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}}\,|\,S^{(j)},Q^{(j)}) &= \binom{K}{j}^{-1} I(\mathbf{V}_{i\to\mathcal{J}};\mathbf{H}_{\tilde{\mathcal{J}}}\mathbf{V}_{i\to\mathcal{J}}+\mathbf{W}_j\,|\,\mathbf{H}_{\tilde{\mathcal{J}}}) \\
&= \binom{K}{j}^{-1}\mathbb{E}\left[\log\det\left(\boldsymbol{I}+\mathsf{snr}\,\mathbf{H}_{\tilde{\mathcal{J}}}^H\boldsymbol{\Lambda}_j\mathbf{H}_{\tilde{\mathcal{J}}}\right)\right] = \binom{K}{j}^{-1} a_j, \\
I(\{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}};\hat{Y}_{i\to\mathcal{J}}\,|\,Y_1^{(i)},S^{(i)},Q^{(i)}) &= \sum_{l=1}^{j}\binom{K}{i}^{-1} I(\mathbf{V}_{\mathcal{I}};\mathbf{H}_l\mathbf{V}_{\mathcal{I}}+\hat{\mathbf{Z}}_{\mathcal{J}\setminus\mathcal{I}}\,|\,\mathbf{H}_1\mathbf{V}_{\mathcal{I}}+\mathbf{Z}_1,\mathbf{H}_1,\mathbf{H}_l) \\
&= \sum_{l=1}^{j}\binom{K}{i}^{-1}\mathbb{E}\left[\log\det\left(\boldsymbol{I}+\frac{\mathsf{snr}}{\beta_i}\mathbf{H}_l(\boldsymbol{I}+\mathsf{snr}\,\mathbf{H}_1^H\mathbf{H}_1)^{-1}\mathbf{H}_l^H\right)\right] \\
&= \sum_{l=1}^{j}\binom{K}{i}^{-1} b_{l,j}. \qquad (57)
\end{aligned}$$

Thus, we can rewrite the Gaussian rate region as below.

$$R_{\text{sym}} \le \alpha_1 K^{-1} a_1 \qquad (58)$$

$$0 \le \alpha_j\binom{K}{j}^{-1} a_j - \alpha_{j-1}\binom{K}{j-1}^{-1}\sum_{l=1}^{j} b_{l,j}. \qquad (59)$$

For a given set of $\{\beta_j\}$ and for a given SNR, $\{a_j\}$ and $\{b_{l,j}\}$ are fixed. Hence, the maximum achievable rate of R_{sym} can be reached when α_1 is maximized. However, the selection of $\{\alpha_i\}$ is subject to the constraint $\sum_{j=1}^{K}\alpha_j=1$. Applying (59) $K-1$ times, for $j=K,K-1,\ldots,2$, we obtain

$$\alpha_1 \le c_1\alpha_2 \le \cdots \le c_{K-1}\alpha_K, \qquad (60)$$

where c_k's are nonnegative and can be found from (59). We argue that from (60) it is without loss of optimality to assume that (59) should hold with equality for all j. To see this, let $\{\alpha_j, j\in\mathcal{K}\}$ be such that $\sum_{j=1}^{K}\alpha_j=1$ and assume that some of the inequalities in (60) are strict. Then, we can always reduce some of $\alpha_2,\ldots,\alpha_K$ and make sure that all the equalities hold, which would in turn lower the value of summation of α, i.e., $\sum_{j=1}^{K}\alpha_j=c<1$. In this case, we can make the scaling $\alpha_j^*=\frac{\alpha_j}{c}$ which increases α_1 and also the objective function. With this reasoning, we conclude that the optimal value of α_1 should be such that (59) holds with equality for all j, which leads to

$$\alpha_1^* = \left(1+\sum_{j=2}^{K}\frac{\binom{K}{j}}{K}\prod_{t=2}^{j}\frac{\sum_{l\le t}b_{l,t}}{a_t}\right)^{-1}. \qquad (61)$$

Plugging α_1^* back to (58), we obtain the optimal symmetric rate (31) in the Gaussian case.

APPENDIX B
PROOF OF COROLLARY 2

In the following, we first apply the RVs selected in Section V-B and evaluate the quantities in (19) and (20), that is, for user $k \in \mathcal{K}$,

$$I(V_k; Y_k^{(1)}, \{\hat{Y}_{1\to\mathcal{U}}\}_{\mathcal{U}\ni k} \,|\, S^{(1)}, Q^{(1)}) = \mathrm{P}(Q_2^{(1)} = k)(1-\delta_{\mathcal{K}}) \log|\mathcal{X}|, \tag{62}$$

$$I(V_{i\to\mathcal{J}}; Y_k^{(j)}, \{\hat{Y}_{j\to\mathcal{U}}\}_{\mathcal{U}\supset\mathcal{J}} \,|\, S^{(j)}, Q^{(j)}) = \mathrm{P}(Q^{(j)} = (i,\mathcal{J}))(1-\delta_{\bar{\mathcal{J}}\cup\{k\}}) \log|\mathcal{X}|, \tag{63}$$

$$= \mathrm{P}(Q_1^{(j)} = i)\mathrm{P}(Q_2^{(j)} = \mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}}) \log|\mathcal{X}| \tag{64}$$

$$I(\{V_{\mathcal{I}}\}_{\mathcal{I}\subset\mathcal{J}}; \hat{Y}_{i\to\mathcal{J}} \,|\, Y_k^{(i)}, S^{(i)}, Q^{(i)}) = \sum_{\mathcal{I}\subset\mathcal{J}} \mathrm{P}(Q_2^{(i)} = \mathcal{I}) I(V_{\mathcal{I}}; \hat{Y}_{i\to\mathcal{J}} \,|\, Y_k^{(i)}, S^{(i)}, Q_1^{(i)}, Q_2^{(i)} = \mathcal{I})$$

$$= \sum_{\mathcal{I}\subset\mathcal{J}, \mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)} = \mathcal{I}, \hat{Y}_{i\to\mathcal{J}} \neq 0, Y_k^{(i)} = ?) H(V_{\mathcal{I}}) \tag{65}$$

$$= \sum_{\mathcal{I}\subset\mathcal{J}, \mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)} = \mathcal{I}) \phi_{\bar{\mathcal{J}}\cup\{k\}, \mathcal{J}\setminus\mathcal{I}} \log|\mathcal{X}|, \tag{66}$$

where (62) can be interpreted as: receiver k can recover the intended signal on M_k unless all the receivers are in erasure; (63) and (66) are obtained with the same reasoning on the choice of the side information $\hat{Y}_{i\to\mathcal{J}}$ as defined in (41). From (19) and (62), we obtain (42).

Applying (20), we have, for all $i, j, k, \mathcal{J}$ with $i < j$ and $k \in \mathcal{I}$,

$$0 \le \alpha_j \mathrm{P}(Q_1^{(j)} = i)\mathrm{P}(Q_2^{(j)} = \mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}}) \log|\mathcal{X}| - \alpha_i \sum_{\mathcal{I}\subset\mathcal{J}, \mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)} = \mathcal{I}) \phi_{\bar{\mathcal{J}}\cup\{k\}, \mathcal{J}\setminus\mathcal{I}} \log|\mathcal{X}|.$$

We assume that the probabilities and α's are bounded away from zero or one so that the following inequality holds ($\forall i, j, k, \mathcal{I}$ with $i < j$ and $k \in \mathcal{J}$).

$$\frac{\alpha_i \sum_{\mathcal{I}\subset\mathcal{J}, \mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)} = \mathcal{I}) \phi_{\bar{\mathcal{J}}\cup\{k\}, \mathcal{J}\setminus\mathcal{I}}}{\alpha_j \mathrm{P}(Q_2^{(j)} = \mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}})} \le \mathrm{P}(Q_1^{(j)} = i). \tag{67}$$

There are $j-1$ such inequalities for each given set of $(j, k, \mathcal{J})$. In addition, $\mathrm{P}(Q_1^{(j)} = i)$ should also fulfil $0 \le \mathrm{P}(Q_1^{(j)} = i) \le 1$ and $\sum_{i=1}^{j-1} \mathrm{P}(Q_1^{(j)} = i) = 1$. Then, we can eliminate the set $\{\mathrm{P}(Q_1^{(j)} = i)\}_{i=1\ldots j-1}$ with the Fourier–Motzkin elimination (FME) to obtain $K-1$ constraints on the α's. Let us take $i = 1$ as an example, as show below.

$$\frac{\alpha_1 \sum_{\mathcal{I}\subset\mathcal{J}, \mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)} = \mathcal{I}) \phi_{\bar{\mathcal{J}}\cup\{k\}, \mathcal{J}\setminus\mathcal{I}}}{\alpha_j \mathrm{P}(Q_2^{(j)} = \mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}})} \le \mathrm{P}(Q_1^{(j)} = 1), \tag{68}$$

$$0 \le \mathrm{P}(Q_1^{(j)} = 1), \tag{69}$$

$$\mathrm{P}(Q_1^{(j)} = 1) \le 1 - \sum_{i=2}^{j-1} \mathrm{P}(Q_1^{(j)} = i). \tag{70}$$

We obtain

$$\frac{\alpha_1 \sum_{\mathcal{I}\subset\mathcal{J},\mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)}=\mathcal{I})\phi_{\bar{\mathcal{J}}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}}{\alpha_j \mathrm{P}(Q_2^{(j)}=\mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}})} \leq 1-\sum_{i=2}^{j-1}\mathrm{P}(Q_1^{(j)}=i). \tag{71}$$

Therefore, we can apply $j-1$ times the same type FME and the rate constraints after these FME are

$$0 \leq \alpha_j \mathrm{P}(Q_2^{(j)}=\mathcal{J})(1-\delta_{\bar{\mathcal{J}}\cup\{k\}}) - \sum_{i=1}^{j-1}\alpha_i \sum_{\mathcal{I}\subset\mathcal{J},\mathcal{I}\ni k} \mathrm{P}(Q_2^{(i)}=\mathcal{I})\phi_{\bar{\mathcal{J}}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}, \tag{72}$$

for $j=2,\ldots,K$. This completes the proof.

APPENDIX C
PROOF OF PROPOSITION 3

We define $\mu_{\mathcal{J}} \triangleq \alpha_j P(Q_2^{(j)}=\mathcal{J})$ as the normalized length such that $\sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{K}}\mu_{\mathcal{J}}=1$. The rate region in Corollary 2 can be rewritten as

$$R_k \leq \mu_{\{k\}}(1-\delta_{\mathcal{K}})\log|\mathcal{X}|, \tag{73}$$

$$\mu_{\mathcal{J}} \geq \max_{k:\, k\in\mathcal{J}} \left\{ \sum_{\mathcal{I}:\, k\in\mathcal{I}\subset\mathcal{J}} \frac{\phi_{\bar{\mathcal{J}}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}}{1-\delta_{\bar{\mathcal{J}}\cup\{k\}}}\mu_{\mathcal{I}} \right\}. \tag{74}$$

First, we show that (74) should be satisfied with equality for all $\mathcal{J}\subseteq\mathcal{K}$. It follows the similar steps as those in Appendix A. We assume that there exist $\{\mu_{\mathcal{J}}\}_{\mathcal{J}\subseteq\mathcal{K}}$ such that $\sum_{\mathcal{J}\subseteq\mathcal{K}}\mu_{\mathcal{J}}=1$ holds, and that the inequality (74) is strict for some $\mathcal{J}'$. In this case, one can always reduce the value of $\mu_{\mathcal{J}'}$ to achieve equality in (74), which leads to a smaller sum $\sum\mu_{\mathcal{J}}=c<1$. Then, we can scale the whole set $\{\mu_{\mathcal{J}}\}_{\mathcal{J}\subseteq\mathcal{K}}$ by c to make sure that $\sum\mu_{\mathcal{J}}=1$ holds again. This will increase the values of $\{\mu_{\{k\}}\}_{k\in\mathcal{K}}$ by a factor $\frac{1}{c}$, and will increase simultaneously the rate in (73). Therefore, it is without loss of optimality to assume that (74) is satisfied with equality.

Then, we focus on the symmetric EBC, for which the optimal normalized lengths are characterized by the following lemma.

Lemma 2. *Let us define* $k^*_{\mathcal{J}} \triangleq \min_{k\in\mathcal{J}} k$ *and*

$$\mu_{k,\mathcal{J}} \triangleq \sum_{\mathcal{I}:\, k\in\mathcal{I}\subset\mathcal{J}} \frac{\phi_{\bar{\mathcal{J}}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}}{1-\delta_{\bar{\mathcal{J}}\cup\{k\}}}\mu_{\mathcal{I}}. \tag{75}$$

For a symmetric EBC, the optimal $\mu_{\mathcal{J}}$, $\mathcal{J}\subseteq\mathcal{K}$, *is*

$$\mu_{\mathcal{J}} = \mu_{k^*_{\mathcal{J}},\mathcal{J}}. \tag{76}$$

Proof. We prove the lemma by induction on j. Note that by definition $\mu_{k^*_{\mathcal{I}},\mathcal{I}} = \mu_{k^*_{\mathcal{J}},\mathcal{I}}$ for $\mathcal{I} \subset \mathcal{J}$. We define two sets $\mathcal{J}_1$ and $\mathcal{J}_2$ that verify $|\mathcal{J}_1| = |\mathcal{J}_2| = j = |\mathcal{J}| \geq 2$ and $\mathcal{J}_1 \setminus \{k^*_{\mathcal{J}_1}\} = \mathcal{J}_2 \setminus \{k^*_{\mathcal{J}_2}\}$. If $k^*_{\mathcal{J}_1} \leq k^*_{\mathcal{J}_2}$, it can be proved with induction that $\mu_{\mathcal{J}_1} \geq \mu_{\mathcal{J}_2}$. To initiate the induction, we assume that the maximal $\mu_{\mathcal{I}}$ is obtained with $k^*_{\mathcal{I}}$ and $\mu_{\mathcal{I}_1} \geq \mu_{\mathcal{I}_2}$ is correct with the analogously defined $\mathcal{I}_1, \mathcal{I}_2$. Due to the channel's symmetry, we assume without loss of generality that $\mu_{\{1\}} \geq \mu_{\{2\}} \geq \cdots \geq \mu_{\{K\}}$ and we use abusively the following notations in this appendix $\delta_{\bar{\mathcal{J}} \cup \{k^*_{\mathcal{J}}\}} = \delta_{K-j+1}$ and $\phi_{\bar{\mathcal{J}} \cup \{k^*_{\mathcal{J}}\}, \mathcal{J} \setminus \mathcal{I}} = \phi_{K-j+1,j-i}$. Hence, (75) and (76) reduce to $\mu_{k^*_{\mathcal{J}},\mathcal{J}} = \frac{1}{1-\delta_{K-j+1}} \sum_{k^*_{\mathcal{J}} \in \mathcal{I} \subset \mathcal{J}} \mu_{k^*_{\mathcal{I}},\mathcal{I}} \phi_{K-j+1,j-i}$.

As Lemma 2 focuses on $j \geq 2$ case, we start by verify the case with $j = 2$. We assume that $\mu_{\mathcal{J}_1} = \mu_{\{t_1,t_3\}}$ and $\mu_{\mathcal{J}_2} = \mu_{\{t_2,t_3\}}$ where $t_1 \leq t_2 \leq t_3$. Then, we notice that $\mu_{\mathcal{J}_m} = \mu_{\{t_m,t_3\}} = \frac{\phi_{K-1,1}}{1-\delta_{K-1}} \max_{k \in \{t_m,t_3\}} \mu_{\{k\}} = \frac{\phi_{K-1,1}}{1-\delta_{K-1}} \mu_{\{t_m\}}$, $m = 1, 2$ and $\mu_{\mathcal{J}_1} = \frac{\phi_{K-1,1}}{1-\delta_{K-1}} \mu_{\{t_1\}} \geq \frac{\phi_{K-1,1}}{1-\delta_{K-1}} \mu_{\{t_2\}} = \mu_{\mathcal{J}_2}$.

Let us assume that (76) and $\mu_{\mathcal{J}_1} \geq \mu_{\mathcal{J}_2}$ hold for any $\mathcal{J}_m \subset \mathcal{K}$ with $|\mathcal{J}_m| = j = l-1$ ($3 \leq l \leq K$), $k^*_{\mathcal{J}_1} \leq k^*_{\mathcal{J}_2}$, and $\mathcal{J}_1 \setminus \{k^*_{\mathcal{J}_1}\} = \mathcal{J}_2 \setminus \{k^*_{\mathcal{J}_2}\}$, for $m = 1, 2$. We show that (76) and $\mu_{\mathcal{L}_1} \geq \mu_{\mathcal{L}_2}$ hold for any $\mathcal{J}_m \subset \mathcal{L}_m \subseteq \mathcal{K}$ with $|\mathcal{L}_m| = l$, $k^*_{\mathcal{L}_1} \leq k^*_{\mathcal{L}_2}$, and $\mathcal{L}_1 \setminus \{k^*_{\mathcal{L}_1}\} = \mathcal{L}_2 \setminus \{k^*_{\mathcal{L}_2}\}$, for $m = 1, 2$. Let us take $\mu_{\mathcal{L}_1}$ as an example. The $\mu_{\mathcal{L}_1}$ writes as

$$\mu_{\mathcal{L}_1} = \frac{\phi_{K-l+1,l-j}}{1-\delta_{K-l+1}} \max_{k \in \mathcal{L}_1} \sum_{k \in \mathcal{J}_1 \subset \mathcal{L}_1} \mu_{\mathcal{J}_1} \tag{77}$$

$$= \frac{\phi_{K-l+1,l-j}}{1-\delta_{K-l+1}} \max \left\{ \sum_{k^*_{\mathcal{L}_1} \in \mathcal{J}_1 \subset \mathcal{L}_1} \mu_{\mathcal{J}_1}, \left\{ \sum_{k' \in \mathcal{J}' \subset \mathcal{L}_1} \mu_{\mathcal{J}'} \right\}_{k' \neq k^*_{\mathcal{L}_1}} \right\} \tag{78}$$

$$= \frac{\phi_{K-l+1,l-j}}{1-\delta_{K-l+1}} \max_{\substack{k' \neq k^*_{\mathcal{L}_1}, \\ k' \in \mathcal{L}_1}} \left\{ \max \left\{ \sum_{k^*_{\mathcal{L}_1} \in \mathcal{J}_1 \subset \mathcal{L}_1} \mu_{\mathcal{J}_1}, \sum_{k' \in \mathcal{J}' \subset \mathcal{L}_1} \mu_{\mathcal{J}'} \right\} \right\} \tag{79}$$

$$= \frac{\phi_{K-l+1,l-j}}{1-\delta_{K-l+1}} \sum_{k^*_{\mathcal{L}_1} \in \mathcal{J}_1 \subset \mathcal{L}_1} \mu_{\mathcal{J}_1}. \tag{80}$$

To prove (80), we consider four types of subsets of $\mathcal{L}_1$ depending on whether $k^*_{\mathcal{L}_1}$ and k' are included in the subset. In particular, a subset including both $k^*_{\mathcal{L}_1}$ and k' appears in both terms inside the inner maximization of (79) which yields $k^*_{\mathcal{J}'} = k^*_{\mathcal{J}_1} = k^*_{\mathcal{L}_1}$, while a subset containing neither $k^*_{\mathcal{L}_1}$ nor k' does not appear inside the inner maximization of (79). Note that the other subsets have either $k^*_{\mathcal{L}_1}$ or k' such that $k' \notin \mathcal{J}_1$ and $k^*_{\mathcal{L}_1} \notin \mathcal{J}'$. There always exists a mapping that projects a subset $\mathcal{J}_1$ including $k^*_{\mathcal{L}_1}$ into another subset $\mathcal{J}'$ including k' by substituting $k^*_{\mathcal{L}_1}$ for k', i.e., $\mathcal{J}' = \mathcal{J}_1 \setminus \{k^*_{\mathcal{L}_1}\} \cup \{k'\}$. Thus, $\sum_{k' \in \mathcal{J}' \subset \mathcal{L}_1} \mu_{\mathcal{J}'} \leq \sum_{k^*_{\mathcal{L}_1} \in \mathcal{J}_1 \subset \mathcal{L}_1} \mu_{\mathcal{J}_1}$ holds for any $k' \neq k^*_{\mathcal{L}_1}$

given that $k^*_{\mathcal{L}_1} \leq k'$ and the property $\mu_{\mathcal{J}_1} \geq \mu_{\mathcal{J}_2}$ is true for $k^*_{\mathcal{J}_1} \leq k^*_{\mathcal{J}_2}$. Therefore, (80) holds. The proof completes by

$$\sum_{\mathcal{I}\subset\mathcal{L}_1\setminus\{k^*_{\mathcal{L}_1}\}} \mu_{\mathcal{I}\cup\{k^*_{\mathcal{L}_1}\}} \geq \sum_{\mathcal{I}\subset\mathcal{L}_2\setminus\{k^*_{\mathcal{L}_2}\}} \mu_{\mathcal{I}\cup\{k^*_{\mathcal{L}_2}\}} \Longleftrightarrow \mu_{\mathcal{L}_1} \geq \mu_{\mathcal{L}_2}, \tag{81}$$

where the inequality follows by identifying $\mathcal{J}_1 = \mathcal{I}\cup\{k^*_{\mathcal{L}_1}\}$, $\mathcal{J}_2 = \mathcal{I}\cup\{k^*_{\mathcal{L}_2}\}$ and the property $\mu_{\mathcal{J}_1} \geq \mu_{\mathcal{J}_2}$ is true when $k^*_{\mathcal{J}_1} = k^*_{\mathcal{L}_1} \leq k^*_{\mathcal{L}_2} = k^*_{\mathcal{J}_2}$ and $\mathcal{J}_1 \setminus \{k^*_{\mathcal{J}_1}\} = \mathcal{I} = \mathcal{J}_2 \setminus \{k^*_{\mathcal{J}_2}\}$. ■

In the following, we show that the μ's in Lemma 2 lead to the capacity region (3). From Lemma 10 in [3], we know that, for any disjoint sets $\mathcal{F}, \mathcal{T} \subseteq \mathcal{K}$,

$$\phi_{\mathcal{F},\mathcal{T}} = \sum_{\mathcal{U}\subseteq\mathcal{T}} (-1)^{|\mathcal{U}|}\delta_{\mathcal{F}\cup\mathcal{U}} = \sum_{\mathcal{U}\subseteq\mathcal{T}} (-1)^{|\mathcal{U}|+1}(1-\delta_{\mathcal{F}\cup\mathcal{U}}). \tag{82}$$

Then, we extend $\phi_{\bar{\mathcal{J}}\cup\{k\},\mathcal{J}\setminus\mathcal{I}}$ analogously such that (76) writes as

$$\begin{aligned}
\mu_{\mathcal{J}} &= (1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}})^{-1} \sum_{\mathcal{I}:\, k^*_{\mathcal{J}}\in\mathcal{I}\subset\mathcal{J}} \mu_{\mathcal{I}} \sum_{\mathcal{U}:\, \mathcal{U}\subseteq\mathcal{J}\setminus\mathcal{I}} (-1)^{|\mathcal{U}|+1}(1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}\cup\mathcal{U}}) \\
&= (1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}})^{-1} \sum_{k^*_{\mathcal{J}}\in\mathcal{I}\subset\mathcal{J}} \mu_{\mathcal{I}} \Bigg(-(1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}}) + \sum_{\phi\neq\mathcal{U}\subseteq\mathcal{J}\setminus\mathcal{I}} (-1)^{|\mathcal{U}|+1}(1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}\cup\mathcal{U}}) \Bigg) \\
&= \frac{\sum_{\phi\neq\mathcal{U}\subseteq\mathcal{J}\setminus\{k^*_{\mathcal{J}}\}} \sum_{k^*_{\mathcal{J}}\in\mathcal{I}\subseteq\mathcal{J}\setminus\mathcal{U}} \mu_{\mathcal{I}} (-1)^{|\mathcal{U}|+1}(1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}\cup\mathcal{U}})}{1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}}} - \sum_{k^*_{\mathcal{J}}\in\mathcal{I}\subset\mathcal{J}} \mu_{\mathcal{I}},
\end{aligned} \tag{83}$$

where we change the summation order over $\mathcal{I}$ and $\mathcal{U}$ to obtain (83). We simplify (83) by adding $\sum_{k^*_{\mathcal{J}}\in\mathcal{I}\subset\mathcal{J}} \mu_{\mathcal{I}}$ to both sides of (83), as shown below.

$$\sum_{\mathcal{I}:\, k^*_{\mathcal{J}}\in\mathcal{I}\subseteq\mathcal{J}} \mu_{\mathcal{I}} = \frac{\sum_{\mathcal{U}:\, \phi\neq\mathcal{U}\subseteq\mathcal{J}\setminus\{k^*_{\mathcal{J}}\}} (-1)^{|\mathcal{U}|+1}(1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}\cup\mathcal{U}}) \sum_{\mathcal{I}:\, k^*_{\mathcal{J}}\in\mathcal{I}\subseteq\mathcal{J}\setminus\mathcal{U}} \mu_{\mathcal{I}}}{1-\delta_{\bar{\mathcal{J}}\cup\{k^*_{\mathcal{J}}\}}}. \tag{84}$$

Next, we show that the μ's satisfying the recursive relation, i.e., (84), also verify Lemma 3.

Lemma 3. *For a given k and for any $\mathcal{W}_k$ such that $k \in \mathcal{W}_k \subseteq \{k, k+1, \cdots, K\}$, we have*

$$\sum_{\mathcal{I}:\, k\in\mathcal{I}\subseteq\mathcal{W}_k} \mu_{\mathcal{I}} = \frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}}}. \tag{85}$$

Proof. The proof is done by induction on the cardinality of $\mathcal{W}_k$. For arbitrary k and $|\mathcal{W}_k| = 1$ (i.e., $\mathcal{W}_k = \{k\}$), one can easily verify (85) is true. We now assume that (85) holds for all $\mathcal{W}_k$ with $|\mathcal{W}_k| \leq w$ and show that it also holds for all $\mathcal{W}_k$ with $|\mathcal{W}_k| = w+1$. Note that $\mathcal{J}$ is a

set whose minimal element is $k^*_{\mathcal{J}}$ and $\mathcal{W}_k$ is a set with its minimum being k. Since (84) is true for any $\mathcal{J}$, we can substitute k and $\mathcal{W}_k$ for $k^*_{\mathcal{J}}$ and $\mathcal{J}$, respectively, in (84) and have

$$\begin{aligned}\sum_{k\in\mathcal{I}\subseteq\mathcal{W}_k}\mu_{\mathcal{I}} &= (1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}})^{-1}\sum_{\phi\neq\mathcal{U}\subseteq\mathcal{W}_k\setminus\{k\}}(-1)^{|\mathcal{U}|+1}(1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}\cup\mathcal{U}})\sum_{k\in\mathcal{I}\subseteq\mathcal{W}_k\setminus\mathcal{U}}\mu_{\mathcal{I}}\\ &= (1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}})^{-1}\sum_{\phi\neq\mathcal{U}\subseteq\mathcal{W}_k\setminus\{k\}}(-1)^{|\mathcal{U}|+1}(1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}\cup\mathcal{U}})\frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{(1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}\cup\mathcal{U}})}\\ &= \frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}}}\sum_{\phi\neq\mathcal{U}\subseteq\mathcal{W}_k\setminus\{k\}}(-1)^{|\mathcal{U}|+1}\\ &= \frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{1-\delta_{\mathcal{K}\setminus\mathcal{W}_k\cup\{k\}\cup\mathcal{U}}},\end{aligned} \tag{86}$$

where second equality holds because of the assumption and (86) follows the binomial theorem. ■

We sum up the $\mu_{\mathcal{J}}$'s ($\sum_{\mathcal{J}\subseteq\mathcal{K}}\mu_{\mathcal{J}}=1$) to obtain $1=\sum_{\mathcal{J}\subseteq\mathcal{K}}\mu_{\mathcal{J}}=\sum_{k=1}^{K}\sum_{k\in\mathcal{I}\subseteq\{k,k+1,\cdots,K\}}\mu_{\mathcal{I}}=\sum_{k=1}^{K}\frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{1-\delta_{\mathcal{K}\setminus\{k,k+1,\cdots,K\}\cup\{k\}}}=\sum_{k=1}^{K}\frac{(1-\delta_{\mathcal{K}})\mu_{\{k\}}}{1-\delta_{\{1,2,\cdots,k\}}}$, where the third equality comes from Lemma 3. Additionally, we rewrite (73) as $\frac{R_k}{(1-\delta_{\mathcal{K}})\log|\mathcal{X}|}\leq\mu_{\{k\}}$, and apply this inequality to the above sum, which yields $\log|\mathcal{X}|\geq\sum_{k=1}^{K}\frac{R_k}{1-\delta_{\{1,2,\cdots,k\}}}$. The above proof also holds if we swap the roles of the users according to the permutation π. This completes the proof.

REFERENCES

[1] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transaction on Information Theory*, vol. 56, no. 6, pp. 2845–2866, May. 2010.

[2] C.-C. Wang, "On the capacity of 1-to-K broadcast packet erasure channels with channel output feedback," *IEEE Transaction on Information Theory*, vol. 58, no. 2, pp. 931–956, Feb. 2012.

[3] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback-capacity and algorithms," *IEEE Transaction on Information Theory*, vol. 59, no. 9, pp. 5779–5804, May. 2013.

[4] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Transaction on Information Theory*, vol. 58, no. 7, pp. 4418–4431, April. 2012.

[5] X. Yi and D. Gesbert, "Precoding methods for the MISO broadcast channel with delayed CSIT," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1–11, May. 2013.

[6] M. A. Maddah-Ali and A. S. Avestimehr, "Approximate capacity region of the MISO broadcast channels with delayed CSIT," *IEEE Transaction on Communications*, vol. 64, no. 7, pp. 2913 – 2924, June. 2016.

[7] J. Wang, M. Matthaiou, S. Jin, and X. Gao, "Precoder design for multiuser MISO systems exploiting statistical and outdated CSIT," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4551–4564, Sept. 2013.

[8] B. Clerckx and D. Gesbert, "Space-Time encoded MISO broadcast channel with outdated CSIT: An error rate and diversity performance analysis," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1661–1675, Mar. 2015.

[9] C. He, S. Yang, and P. Piantanida, "On the capacity of the fading broadcast channel with state feedback," in *IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP)*, May. 2014.

[10] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *IEEE Transaction on Information Theory*, vol. 59, no. 3, pp. 1329–1345, Mar. 2013.

[11] H. Kim, Y.-K. Chia, and A. El Gamal, "A note on the broadcast channel with stale state information at the transmitter," *IEEE Transaction on Information Theory*, vol. 61, no. 7, pp. 3622–3631, May. 2015.

[12] R. Venkataramanan and S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Transaction on Information Theory*, vol. 59, no. 10, pp. 6175–6191, June. 2013.

[13] Y. Wu and M. Wigger, "Coding schemes with rate-limited feedback that improve over the no feedback capacity for a large class of broadcast channels," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2009–2033, Feb. 2016.

[14] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Transaction on Information Theory*, vol. 52, no. 4, pp. 1469–1482, April. 2006.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[16] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[17] C. S. Vaze and M. K. Varanasi, "The Degree-of-Freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5354–5374, May. 2012.