On the Vector Broadcast Channel with Alternating CSIT: A Topological Perspective

Jinyuan Chen, Petros Elia and Syed Ali Jafar

Abstract

In many wireless networks, link strengths are affected by many topological factors such as different distances, shadowing and inter-cell interference, thus resulting in some links being generally stronger than other links. Accounting for such topological aspects has remained largely unexplored, despite strong indications that such aspects can crucially affect transceiver and feedback design, as well as the overall performance.

The work here takes a first step in exploring this interplay between topology, feedback and performance. This is done for the two user broadcast channel with random fading, in the presence of a simple two-state topological setting of statistically strong vs. weaker links, and in the presence of a practical ternary feedback setting of *alternating channel state information at the transmitter* (alternating CSIT) where for each channel realization, this CSIT can be perfect, delayed, or not available.

In this setting, the work derives generalized degrees-of-freedom bounds and exact expressions, that capture performance as a function of feedback statistics and topology statistics. The results are based on novel *topological signal management* (TSM) schemes that account for topology in order to fully utilize feedback. This is achieved for different classes of feedback mechanisms of practical importance, from which we identify specific feedback mechanisms that are best suited for different topologies. This approach offers further insight on how to split the effort - of channel learning and feeding back CSIT - for the strong versus for the weaker link. Further intuition is provided on the possible gains from topological spatio-temporal diversity, where topology changes in time and across users.

I. Introduction

A vector Gaussian broadcast channel, also known as the Gaussian MISO BC (multiple-input single-output broadcast channel) is comprised of a transmitter with multiple antennas that wishes to send independent messages to different receivers, each equipped with a single antenna. In addition to its direct relevance to cellular downlink communications, the MISO BC has attracted much attention for the critical role played in this setting by the feedback mechanism through which channel state information at the transmitter (CSIT) is typically acquired. Interesting insights into the dependence of the capacity limits of the MISO BC on the timeliness and quality of feedback have been found through degrees of freedom (DoF) characterizations under perfect CSIT [1], no CSIT [2]–[5], compound CSIT [6]–[8], delayed CSIT [9], CSIT comprised of channel coherence patterns [10], mixed CSIT [11]–[14], and alternating CSIT [15]. Other related work can be found in [16]–[28].

As highlighted recently in [29], while the insights obtained from DoF studies are quite profound, they are implicitly limited to settings where all users experience comparable signal strengths. This is due to the fundamental limitation of the DoF metric which treats each user with a non-zero channel coefficient, as capable of carrying exactly 1 DoF by itself, regardless of the statistical strength of the channel coefficients. Thus, the DoF metric ignores the diversity of link strengths, which is perhaps the most essential aspect of wireless communications from the perspective of interference management. Indeed, in wireless communication settings, the link strengths are affected by many topological factors, such as propagation path loss, shadow fading and inter-cell interference [30], which lead to statistically unequal channel gains, with some links being much weaker or stronger than others (See Figures 1, 2). Accounting

Jinyuan Chen is with Stanford University, Department of Electrical Engineering, CA 94305, USA (email: jinyuanc@stanford.edu). Petros Elia is with the Mobile Communications Department, EURECOM, Sophia Antipolis, 06410, France (email: elia@eurecom.fr). Syed Ali Jafar is with the Center for Pervasive Communications and Computing (CPCC) at the University of California Irvine, Irvine, CA 92697, USA (email: syed@uci.edu).

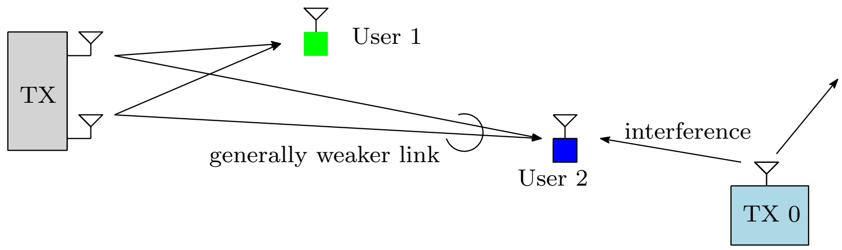

Fig. 1. Topology where link 2 is weaker due to distance and interference.

for these topological aspects, by going beyond the DoF framework into the *generalized* degrees of freedom (GDoF) framework, is the focus of the topological perspective that we seek here.

The work here combines considerations of topology with considerations of feedback timeliness and quality, and addresses questions on performance bounds, on encoding designs that account for topology and feedback, on feedback and channel learning mechanisms that adapt to topology, and on handling and even exploiting fluctuations in topology.

II. SYSTEM MODEL FOR THE TOPOLOGICAL BC

A. Channel, topology, and feedback models

We consider the broadcast channel, with a two-antenna transmitter sending information to two single-antenna receivers. The corresponding received signals at the first and second receiver at time t, can be modeled as

$$y_t = \sqrt{\rho}\boldsymbol{h}_t^{'\mathsf{T}}\boldsymbol{x}_t + u_t^{'} \tag{1}$$

$$z_t = \sqrt{\rho}\boldsymbol{g}_t^{'\mathsf{T}}\boldsymbol{x}_t + v_t^{'} \tag{2}$$

where ρ is defined by a power constraint, where $\boldsymbol{x}_t$ is the normalized transmitted vector at time t — normalized here to satisfy $||\boldsymbol{x}_t||^2 \leq 1$ — where $\boldsymbol{h}_t^{'}, \boldsymbol{g}_t^{'}$ represent the vector fading channels to the first and second receiver respectively, and where $u_t^{'}, v_t^{'}$ represent equivalent receiver noise.

1) Topological diversity: In the general topological broadcast channel setting, the variance of the above fading and equivalent noise, may be uneven across users, and may indeed fluctuate in time and frequency. These fluctuations may be a result of movement, but perhaps more importantly, topological changes in the time scales of interest, can be attributed to fluctuating inter-cell interference. Such fluctuations are in turn due to different allocations of carriers in different cells or — similarly — due to the fact that one carrier can experience more interference from adjacent cells than another.

The above considerations can be concisely captured by the following simple model

$$y_t = \rho^{A_{1,t}/2}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t + u_t \tag{3}$$

$$z_t = \rho^{A_{2,t}/2}\boldsymbol{g}_t^{\mathsf{T}}\boldsymbol{x}_t + v_t \tag{4}$$

where now $\boldsymbol{h}_t, \boldsymbol{g}_t$ and u_t, v_t are assumed to be spatially and temporally i.i.d[1] Gaussian with zero mean and *unit variance*. With $||\boldsymbol{x}_t||^2 \leq 1$, the parameter ρ and the *link power exponents* $A_{1,t}, A_{2,t}$ reflect — for each link, at time t — an *average* received signal-to-noise ratio (SNR)

$$\mathbb{E}_{\boldsymbol{h}_t,\boldsymbol{x}_t}[\rho^{A_{1,t}/2}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t]^2 = \rho^{A_{1,t}} \tag{5}$$

$$\mathbb{E}_{\boldsymbol{g}_t,\boldsymbol{x}_t}[\rho^{A_{2,t}/2}\boldsymbol{g}_t^{\mathsf{T}}\boldsymbol{x}_t]^2 = \rho^{A_{2,t}}. \tag{6}$$

In this setting we adopt a simple two-state topological model where the link exponents can each take, at a given time t, one of two values

[1]This suggests the simplifying formulation of unit coherence time.

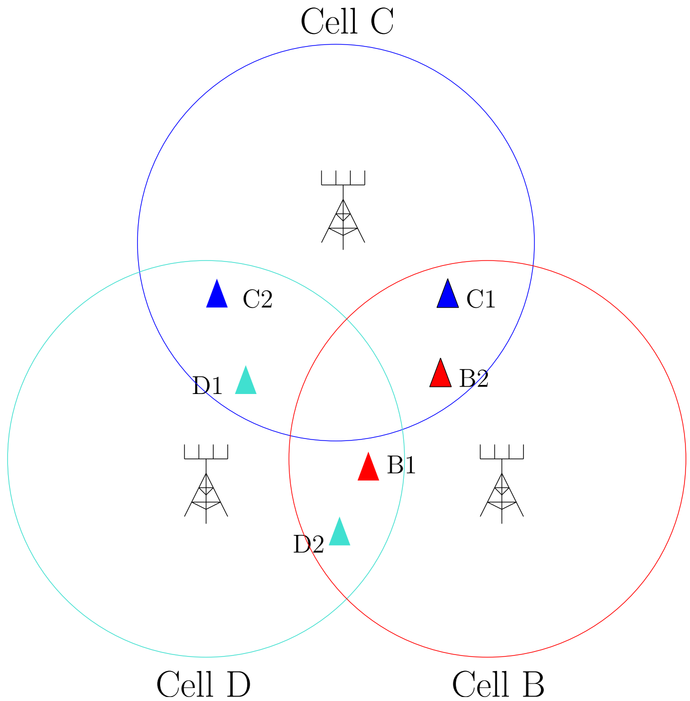

Fig. 2. Cell edge users experience fluctuating interference due to changing frequency allocation in the multi-cell system.

$$A_{k,t} \in \{1, \alpha\} \quad \text{for} \quad 0 \le \alpha \le 1, \quad k = 1, 2$$

reflecting the possibility of either a strong link ($A_{k,t} = 1$), or a weaker link ($A_{k,t} = \alpha$). The adopted small number of topological states, as opposed to a continuous range of $A_{k,t}$ values, is motivated by static multi-carrier settings with adjacent cell interference, where the number of topological states can be proportional to the number of carriers.

Remark 1: We clarify that the rate of change of the topology — despite the use of a common time index for $A_{k,t}$ and $\boldsymbol{h}_t, \boldsymbol{g}_t$ — need not match in any way, the rate of change of fading. We also clarify that our use of the term 'link' carries a statistical connotation, so for example when we say that at time t the first link is stronger than the second link, we refer to a statistical comparison where $A_{1,t} > A_{2,t}$.

2) Alternating CSIT formulation: Finally in terms of feedback, we draw from the alternating CSIT formulation by Tandon et al. [15], which can nicely capture simple feedback policies. In this setting, the CSIT for each channel realization can be immediately available and perfect (P), or it can be delayed (D), or not available (N). In our notation, $I_{k,t} \in \{P, D, N\}$ will represent the CSIT about the fading channel of user k at time t.

B. Problem statement: generalized degrees-of-freedom, feedback and topology statistics

1) Generalized Degrees-of-Freedom: In a setting where (R_1, R_2) denotes an achievable rate pair for the first and second user respectively, we focus on the high-SNR regime and seek to characterize sum GDoF

$$d_\Sigma = \lim_{\rho \to \infty} \max_{(R_1, R_2)} \frac{R_1 + R_2}{\log \rho}$$

performance bounds.

It is easy to see that in the current two-state topological setting, a strong link by itself has capacity that scales as $\log \rho + o(\log \rho)$, while[2] a weak link has a capacity that scales as $\alpha \log \rho + o(\log \rho)$. Setting

[2] $o(\bullet)$ comes from the standard Landau notation, where $f(x) = o(g(x))$ implies $\lim_{x \to \infty} f(x)/g(x) = 0$. Logarithms are of base 2.

$\alpha = 1$ removes topology considerations, while setting $\alpha = 0$ almost entirely removes the weak link, as its capacity does not scale with SNR.

Example 1: One can see that, in the current setting of the two-user MISO BC, having always perfect feedback (P) for both users' channels, and having a static topology where the first link is stronger ($A_{1,t} = 1, \forall t$) than the second throughout the communication process ($A_{2,t} = \alpha, \forall t$), the sum GDoF is $d_\Sigma = 1 + \alpha$, and it is achieved by zero forcing.

Example 2: Furthermore a quick back-of-the-envelope calculation (see the appendix in Section VIII-F), can show that in the same fixed topology $A_{1,t} = 1, A_{2,t} = \alpha, \forall t$, the original MAT scheme — originally designed in [9] without topology considerations for the $\alpha = 1$ case — after a small modification that regulates the rate of the private information to the weaker user, achieves a sum GDoF of $d_\Sigma = \frac{2}{3}(1 + \alpha)$. This performance will be surpassed by a more involved topological signal management (TSM) scheme, to be described later on.

2) Feedback and topology statistics: Naturally performance is a function of the feedback and topology statistics. Towards capturing these statistics, we draw from the formulation in [15] and consider

$$\lambda_{I_1,I_2}$$

to denote the fraction of the time during which the CSIT state is described by a pair $(I_1, I_2) \in (P, D, N) \times (P, D, N)$, as well as consider

$$\lambda_{A_1,A_2}$$

to denote the fraction of the time during which the gain exponents of the two links are some pair $(A_1, A_2) \in (1, \alpha) \times (1, \alpha)$, where naturally $\lambda_{1,\alpha} + \lambda_{\alpha,1} + \lambda_{1,1} + \lambda_{\alpha,\alpha} = 1$.

Example 3: $\lambda_{P,P} = 1$ (resp. $\lambda_{D,D} = 1$, $\lambda_{N,N} = 1$) implies perfect CSIT (resp. delayed CSIT, no CSIT) for both users' channels, throughout the communication process. Similarly $\lambda_{P,N} + \lambda_{N,P} = 1$ restricts to a family of feedback schemes where only one user sends CSIT at a time (more precisely, per channel realization), and does so perfectly. From this family, $\lambda_{P,N} = \lambda_{N,P} = 1/2$ is the symmetric option. Similarly, in terms of topology, $\lambda_{1,\alpha} = 1,\ \alpha < 1$ implies that the first link is stronger than the second throughout the communication process, while $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$ implies that half of the time, the first user is statistically stronger, and vice versa.

In addition to the feedback and topology statistics, the formulation here can allow for description of feedback mechanisms. Towards this we use

$$\lambda_{I_1,I_2}^{A_1,A_2}$$

to denote the fraction of the time during which the CSIT state is (I_1, I_2) and the topology state is (A_1, A_2).

Example 4: Having $\lambda_{P,D}^{1,\alpha} + \lambda_{D,P}^{\alpha,1} = 1$ implies a mechanism that asks — for any channel realization — the statistically stronger user to send perfect feedback, and the statistically weaker user to send delayed feedback.

C. Conventions and structure

In terms of notation, $(\bullet)^{\mathsf{T}}$ and $(\bullet)^{\mathsf{H}}$ denote the transpose and conjugate transpose operations respectively, while $|| \bullet ||$ denotes the Euclidean norm, and $| \bullet |$ denotes either the magnitude of a scalar or the cardinality of a set. We also use $\doteq$ to denote *exponential equality*, i.e., we write $f(\rho) \doteq \rho^B$ to denote $\lim\limits_{\rho \to \infty} \log f(\rho)/\log \rho = B$. Similarly $\dot{\geq}$ and $\dot{\leq}$ denote exponential inequalities. $\boldsymbol{e}^\perp$ denotes a unit-norm vector orthogonal to vector $\boldsymbol{e}$.

Throughout this paper, we adhere to the common convention and assume perfect and global knowledge of channel state information at the receivers (perfect and global CSIR).

We proceed with the main results. We first present sum GDoF outer bounds as a function of the CSIT and topology statistics, and then proceed to derive achievable and often optimal sum GDoF expressions for pertinent cases of practical significance.

III. OUTER BOUNDS

We first proceed with a simpler version of the outer bound, which encompasses all cases of alternating CSIT, and all *fixed* topologies ($\lambda_{1,\alpha}=1$, or $\lambda_{\alpha,1}=1,\ \alpha\in[0,1]$).

Lemma 1: The sum GDoF of the two-user MISO BC with alternating CSIT and a fixed topology, is upper bounded as $d_\Sigma \leq \min\{d_\Sigma^{(1)}, d_\Sigma^{(2)}\}$, where

$$\begin{aligned}
d_\Sigma^{(1)} \triangleq &(1+\alpha)\lambda_{P,P} + \frac{3+2\alpha}{3}(\lambda_{P,D}+\lambda_{D,P}+\lambda_{P,N}+\lambda_{N,P}) \\
&+ \frac{3+\alpha}{3}(\lambda_{D,D}+\lambda_{D,N}+\lambda_{N,D}+\lambda_{N,N}) \\
d_\Sigma^{(2)} \triangleq &(1+\alpha)(\lambda_{P,P}+\lambda_{P,D}+\lambda_{D,P}+\lambda_{D,D}) \\
&+ \frac{2+\alpha}{2}(\lambda_{P,N}+\lambda_{N,P}+\lambda_{D,N}+\lambda_{N,D}) + \lambda_{N,N}.
\end{aligned}$$

The proof of the above lemma, can be found as part of the proof of the following more general lemma, in the appendix of Section VII.

We now proceed with the general outer bound, for any alternating CSIT mechanism, and any topology, i.e., for any $\lambda_{I_1,I_2}^{A_1,A_2}$. In order to achieve a concise description of the bound, we provide the following notation.

$$\begin{aligned}
\lambda_{P\leftrightarrow N}^{A_1,A_2} &\triangleq \lambda_{P,N}^{A_1,A_2} + \lambda_{N,P}^{A_1,A_2} \\
\lambda_{D\leftrightarrow N}^{A_1,A_2} &\triangleq \lambda_{D,N}^{A_1,A_2} + \lambda_{N,D}^{A_1,A_2} \\
\lambda_{P\leftrightarrow D}^{A_1,A_2} &\triangleq \lambda_{P,D}^{A_1,A_2} + \lambda_{D,P}^{A_1,A_2}.
\end{aligned}$$

As a clarifying example, $\lambda_{P\leftrightarrow D}^{1,\alpha}$ simply denotes the fraction of the communication time during which the first link is stronger than the second, and during which, *any one* of the users feeds back perfect CSIT while the other feeds back delayed CSIT.

Lemma 2: The sum GDoF of the topological two-user MISO BC with alternating CSIT, is upper bounded as

$$d_\Sigma \leq \min\{d_\Sigma^{(3)}, d_\Sigma^{(4)}\} \tag{7}$$

where

$$\begin{aligned}
d_\Sigma^{(3)} \triangleq &(1+\alpha)(\lambda_{P,P}^{\alpha,1}+\lambda_{P,P}^{1,\alpha}) + \frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow D}^{\alpha,1}+\lambda_{P\leftrightarrow D}^{1,\alpha}) + \frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow N}^{\alpha,1}+\lambda_{P\leftrightarrow N}^{1,\alpha}) \\
&+ \frac{3+\alpha}{3}(\lambda_{D,D}^{\alpha,1}+\lambda_{D,D}^{1,\alpha}) + \frac{3+\alpha}{3}(\lambda_{D\leftrightarrow N}^{\alpha,1}+\lambda_{D\leftrightarrow N}^{1,\alpha}) + \frac{3+\alpha}{3}(\lambda_{N,N}^{\alpha,1}+\lambda_{N,N}^{1,\alpha}) \\
&+ 2\lambda_{P,P}^{1,1} + \frac{5}{3}\lambda_{P\leftrightarrow D}^{1,1} + \frac{5}{3}\lambda_{P\leftrightarrow N}^{1,1} + \frac{4}{3}\lambda_{D,D}^{1,1} + \frac{4}{3}\lambda_{D\leftrightarrow N}^{1,1} + \frac{4}{3}\lambda_{N,N}^{1,1} \\
&+ 2\alpha\lambda_{P,P}^{\alpha,\alpha} + \frac{5\alpha}{3}\lambda_{P\leftrightarrow D}^{\alpha,\alpha} + \frac{5\alpha}{3}\lambda_{P\leftrightarrow N}^{\alpha,\alpha} + \frac{4\alpha}{3}\lambda_{D,D}^{\alpha,\alpha} + \frac{4\alpha}{3}\lambda_{D\leftrightarrow N}^{\alpha,\alpha} + \frac{4\alpha}{3}\lambda_{N,N}^{\alpha,\alpha}
\end{aligned} \tag{8}$$

$$\begin{aligned}
d_\Sigma^{(4)} \triangleq &(1+\alpha)(\lambda_{P,P}^{1,\alpha}+\lambda_{P,P}^{\alpha,1}) + (1+\alpha)(\lambda_{P\leftrightarrow D}^{1,\alpha}+\lambda_{P\leftrightarrow D}^{\alpha,1}) + (1+\alpha)(\lambda_{D,D}^{1,\alpha}+\lambda_{D,D}^{\alpha,1}) \\
&+ \frac{2+\alpha}{2}(\lambda_{P\leftrightarrow N}^{1,\alpha}+\lambda_{P\leftrightarrow N}^{\alpha,1}) + \frac{2+\alpha}{2}(\lambda_{D\leftrightarrow N}^{1,\alpha}+\lambda_{D\leftrightarrow N}^{\alpha,1}) + \lambda_{N,N}^{1,\alpha} + \lambda_{N,N}^{\alpha,1} \\
&+ 2\lambda_{P,P}^{1,1} + 2\alpha\lambda_{P,P}^{\alpha,\alpha} + 2\lambda_{P\leftrightarrow D}^{1,1} + 2\alpha\lambda_{P\leftrightarrow D}^{\alpha,\alpha} + 2\lambda_{D,D}^{1,1} + 2\alpha\lambda_{D,D}^{\alpha,\alpha} \\
&+ \frac{3}{2}\lambda_{P\leftrightarrow N}^{1,1} + \frac{3\alpha}{2}\lambda_{P\leftrightarrow N}^{\alpha,\alpha} + \frac{3}{2}\lambda_{D\leftrightarrow N}^{1,1} + \frac{3\alpha}{2}\lambda_{D\leftrightarrow N}^{\alpha,\alpha} + \lambda_{N,N}^{1,1} + \alpha\lambda_{N,N}^{\alpha,\alpha}.
\end{aligned} \tag{9}$$

The above bounds will be used to establish the optimality of different encoding schemes and practical feedback mechanisms.

IV. Practical feedback schemes over a fixed topology

We proceed to derive different results for the case of any fixed topology. Here, without loss of generality, we will consider the case where $\lambda_{1,\alpha} = 1$, while the case of $\lambda_{\alpha,1} = 1$ is handled simply by interchanging the role of the two users. In the presence of a fixed topology, we initially focus on different practical feedback schemes for which we derive the exact sum GDoF expressions, and then proceed to explore the delayed CSIT case for which we derive a bound.

With emphasis on practicality, we first focus on three families of simple mechanisms which can be implemented so that, per coherence interval, only one user sends feedback[3].

Proposition 1: For the two-user MISO BC with a fixed topology and a feedback constraint $\lambda_{P,N} + \lambda_{N,P} = 1$ or $\lambda_{P,N} + \lambda_{N,P} = \lambda_{N,D} + \lambda_{D,N} = 1/2$ or $\lambda_{P,D} + \lambda_{D,P} = \lambda_{N,N} = 1/2$, the optimal sum GDoF is

$$d_\Sigma = 1 + \frac{\alpha}{2}. \tag{10}$$

In the first case, this is achieved by the symmetric mechanism $\lambda_{P,N} = \lambda_{N,P} = 1/2$, in the second case it is achieved by the symmetric mechanism $\lambda_{P,N} = \lambda_{N,D} = 1/2$ which associates delayed feedback with the weak user, and in the third case it is achieved by the mechanism $\lambda_{P,D} = \lambda_{N,N} = 1/2$, which again associates delayed feedback with the weak user.

Proof: All GDoF expressions are optimal as they meet the outer bound in Lemma 1. For the first case ($\lambda_{P,N} + \lambda_{N,P} = 1$), the GDoF optimal scheme can be found in Section VIII-E1. For the case where $\lambda_{P,N} + \lambda_{N,P} = \lambda_{N,D} + \lambda_{D,N} = 1/2$, the optimal scheme can be found in Section VIII-A, while for the last case of $\lambda_{P,D} + \lambda_{D,P} = \lambda_{N,N} = 1/2$ the optimal scheme is in Section VIII-B. ■

Remark 2: The optimality of $\lambda_{P,N} = \lambda_{N,D} = 1/2$ (resp. $\lambda_{P,D} = \lambda_{N,N} = 1/2$) among all possible mechanisms $\lambda_{P,N} + \lambda_{N,P} = \lambda_{D,N} + \lambda_{N,D} = 1/2$ (resp. $\lambda_{P,D} + \lambda_{D,P} = \lambda_{N,N} = 1/2$), is due to the fact that delayed CSIT is associated to the weak link, which in turn allows for the unintended interference — resulting from communicating without current CSIT — to be naturally reduced in the direction of the weak link.

Remark 3: It is easy to see that the family $\lambda_{P,D} + \lambda_{D,P} = \lambda_{N,N} = 1/2$ is again a 'one-user-per-channel' family of feedback policies since it can be implemented by having half of the channel states not fed back, while having the other half fed back by any one user with no delay, and by the other user with delay.

A. Delayed CSIT and fixed topology

For the same setting of fixed topologies ($\lambda_{1,\alpha} = 1$ or $\lambda_{\alpha,1} = 1$, $\alpha \in [0, 1]$), we lower bound the sum GDoF performance for the well known delayed CSIT setting of Maddah-Ali and Tse [9], where feedback is always delayed ($\lambda_{D,D} = 1$). A brief description of the corresponding new encoding scheme will appear immediately afterwards.

Proposition 2: For the two-user MISO BC with a fixed topology and delayed CSIT ($\lambda_{D,D} = 1$), the sum GDoF is lower bounded as

$$d_\Sigma \geq 1 + \frac{\alpha^2}{2 + \alpha}. \tag{11}$$

Proof: The scheme that achieves the lower bound can be found in Section VIII-C. The scheme is optimal as it meets the lower bound in Lemma 1. ■

It is worth noting that the above sum GDoF surpasses the aforementioned performance of the original — and slightly modified MAT scheme [9] — over the same topology, which was mentioned in example 2 to be $d_\Sigma = \frac{2}{3}(1 + \alpha)$.

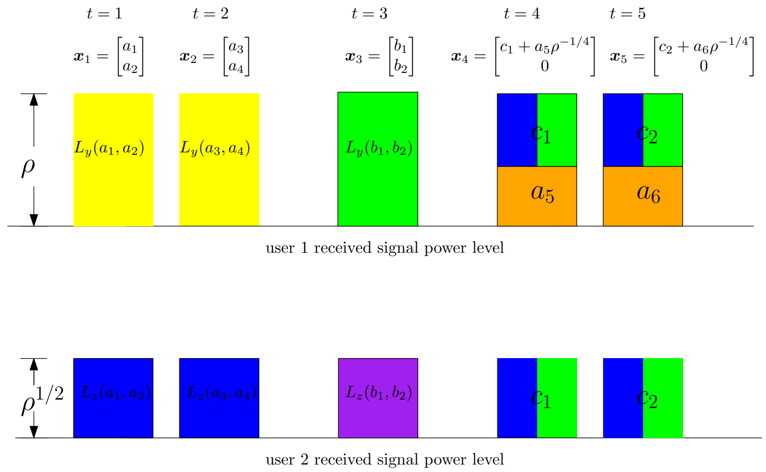

Fig. 3. Received signal power level illustration for the proposed TSM scheme: The case with $\lambda_{D,D}^{1,\alpha} = 1$ and $\alpha = 1/2$.

1) Topological signal management scheme: a sketch for the $\lambda_{D,D}^{1,\alpha} = 1$ case where $\alpha = 1/2$:

We now briefly sketch the description of the encoding scheme that achieves, in the presence of delayed CSIT, the sum GDoF $d_\Sigma = 1 + \frac{\alpha^2}{2+\alpha}$. For brevity we consider the simple setting where $\alpha = 1/2$, in which case the scheme has three phases with respective phase durations $T_1 = 2,\ T_2 = 1,\ T_3 = 2$ which, as we will see later on, are so chosen in order to balance the amount of side information that accumulates at the two users.

• During phase 1 ($t = 1, 2$) the transmitter sends $\boldsymbol{x}_1 = \begin{bmatrix} a_1 & a_2 \end{bmatrix}^\intercal$ and $\boldsymbol{x}_2 = \begin{bmatrix} a_3 & a_4 \end{bmatrix}^\intercal$ intended for user 1, where we recall that $\boldsymbol{x}_1, \boldsymbol{x}_2$ are normalized to have an average unit-power constraint. These are received by user 2 as interference, in the form of two linear combinations which we denote as $L_z(a_1, a_2)$ and $L_z(a_3, a_4)$. a_1 and a_3 each carry $\log \rho$ bits, while a_2 and a_4 each carry $\frac{1}{2} \log \rho$ bits.
• During phase 2 ($t = 3$), the transmitter sends — after normalization — $\boldsymbol{x}_3 = \begin{bmatrix} b_1 & b_2 \end{bmatrix}^\intercal$ intended for user 2, where again $\boldsymbol{x}_3$ is normalized to have an average unit-power constraint. This is received by user 1 as interference, in the form of a linear combination $L_y(b_1, b_2)$. b_1 carries $\log \rho$ bits, while b_2 carries $\frac{1}{2} \log \rho$ bits of information.
• Given delayed CSIT, at the beginning of the third phase ($t = 4, 5$), the transmitter can faithfully *reconstruct* the interference terms $L_z(a_1, a_2), L_z(a_3, a_4), L_y(b_1, b_2)$. As a result of the topology, $L_z(a_1, a_2), L_z(a_3, a_4)$ have power $\rho^{1/2}$, and can thus be reconstructed with $\frac{1}{2} \log \rho$ quantization bits each, with quantization error that is sufficiently small to not affect the DoF performance [31]. Similarly $L_y(b_1, b_2)$, which arrives with power ρ, is faithfully quantized with a total of $\log \rho$ quantization bits, which matches the number of quantization bits from the previous phase. Then these bits are mapped into common information symbols $\{c_1, c_2\}$ that are though represented by $\log \rho + o(\log \rho)$ bits, after the bits from the two phases are additively combined (vector XOR). Once this common information is eventually decoded, one user will be able to learn the other user's side information sufficiently well, by additively combining these bits with its own side information.
As a result, during phase 3 ($t = 4, 5$), the transmitter sends — after normalization — $\boldsymbol{x}_4 = \begin{bmatrix} c_1 + a_5\rho^{-1/4} & 0 \end{bmatrix}^\intercal$ and $\boldsymbol{x}_5 = \begin{bmatrix} c_2 + a_6\rho^{-1/4} & 0 \end{bmatrix}^\intercal$, which means that it sends high-power common symbols $\{c_1, c_2\}$ to both users, and low-power private symbols $\{a_5, a_6\}$ for user 1, where this power is sufficiently lowered to account for the topology. Each c_1, c_2, a_5, a_6 carries $\frac{1}{2} \log \rho$ bits. As a result, summing up the bits, we have a total of $\frac{11}{2} \log \rho$ information bits, over 5 channel uses, which gives a sum GDoF of $\frac{11}{10}$, and which in turn matches the expression of the proposition for $\alpha = 1/2$.

[3]In our formulation, which uses the simplifying assumption of having a unit coherence period, this simply refers to the case where only one user sends feedback at a time.

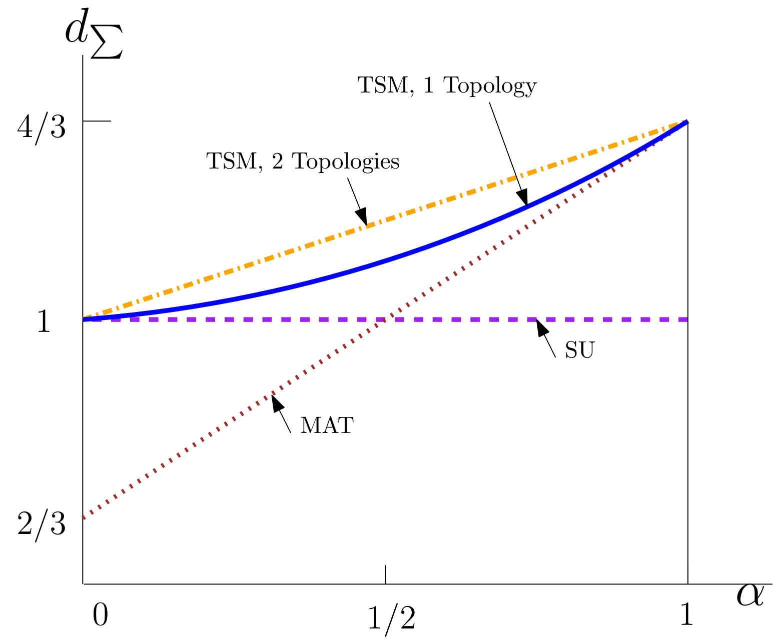

Fig. 4. Sum GDoF performance for the slightly modified MAT scheme, the single user (SU) case, and the topological signal management (TSM) scheme, all for the setting $\lambda_{D,D}^{1,\alpha} = 1$. Additionally the plot shows the sum GDoF of the proper TSM scheme for the setting $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$.

V. Optimal sum GDoF of practical feedback schemes for the BC with topological diversity

We here explore a class of alternating topologies and reveal a gain — in certain instances — that is associated to topologies that vary in time and across users.

We first proceed, and for the delayed CSIT setting $\lambda_{D,D} = 1$, derive the optimal sum GDoF in the presence of the symmetrically *alternating topology* where $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$.

Proposition 3: For the two-user MISO BC with delayed CSIT $\lambda_{D,D} = 1$ and topological spatio-temporal diversity such that $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$, the optimal sum GDoF is

$$d_\Sigma = 1 + \frac{\alpha}{3} \qquad (12)$$

which exceeds the optimal sum GDoF $d'_\Sigma = \frac{2}{3}(1+\alpha)$ of the same feedback scheme, over an equivalent[4] but spatially non-diverse topology $\lambda_{1,1} = \lambda_{\alpha,\alpha} = 1/2$.

Proof: The GDoF is optimal as it meets the general outer bound in Lemma 2. The optimal TSM scheme is described in Section VIII-D. ■

We also briefly note that for the same feedback policy $\lambda_{D,D} = 1$, the optimal sum GDoF $d_\Sigma = 1 + \frac{\alpha}{3}$ corresponding to the topologically diverse setting $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$, also exceeds the sum GDoF performance in Proposition 2 of the TSM scheme in the presence of any static topology (e.g. $\lambda_{1,\alpha} = 1$).

A similar observation to that of the above proposition, is derived below, now for the feedback mechanism $\lambda_{P,N} = \lambda_{N,P} = 1/2$.

Proposition 4: For the two-user MISO BC with $\lambda_{P,N} = \lambda_{N,P} = 1/2$ and topological diversity such that $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$, the optimal sum GDoF is

$$d_\Sigma = 1 + \frac{\alpha}{2} \qquad (13)$$

[4] The compared topologies are considered equivalent in the sense that the overall duration of weak links, is the same for the two topologies.

which exceeds the optimal sum GDoF $d'_\Sigma = \frac{3}{4}(1+\alpha)$ of the same feedback mechanism over the equivalent but spatially non-diverse topology $\lambda_{1,1} = \lambda_{\alpha,\alpha} = 1/2$.

Proof: The sum GDoF is optimal as it achieves the general outer in Lemma 2. The optimal scheme is described in Section VIII-E. ■

Regarding this same feedback policy $\lambda_{P,N} = \lambda_{N,P} = 1/2$, it is worth to now note this policy's very broad applicability. This is shown in the following proposition.

Proposition 5: For the two-user MISO BC with any strictly uneven topology $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$ and a feedback constraint $\lambda_{P,N} + \lambda_{N,P} = 1$, the optimal sum GDoF is

$$d_\Sigma = 1 + \frac{\alpha}{2} \tag{14}$$

and it is achieved by the symmetric feedback policy $\lambda_{P,N} = \lambda_{N,P} = 1/2$.

Proof: The sum GDoF is optimal as it achieves the general outer bound in Lemma 1. The optimal scheme is described in Section VIII-E. ■

Remark 4: This broad applicability of mechanism $\lambda_{P,N} = \lambda_{N,P} = 1/2$, implies a simpler process of learning the channel and generating CSIT, which now need not consider the specific topology as long as this is strictly uneven ($\lambda_{1,1} = \lambda_{\alpha,\alpha} = 0$).

VI. CONCLUSIONS

The work explored the interplay between topology, feedback and performance, for the specific setting of the two-user MISO broadcast channel. Adopting a generalized degrees of freedom framework, and addressing feedback and topology jointly, the work revealed new aspects on encoding design that accounts for topology and feedback, as well as new aspects on how to handle and even exploit topologically diverse settings where the topology varies across users and across time.

In addition to the bounds and encoding schemes, the work offers insight on how to feedback — and naturally how to learn the channel — in the presence of uneven and possibly fluctuating topologies. This insight came in the form of simple feedback mechanisms that achieve optimality — under specific constraints — often without knowledge of topology and its fluctuations.

VII. APPENDIX - PROOF OF OUTER BOUND (LEMMA 2)

We here provide the proof of the general outer bound in Lemma 2. Let W_1, W_2 respectively denote the messages of user 1 and user 2, let R_1, R_2 denote the two users' rates, and let Ω^n be the global channel state information about all the channel states of the BC. The communication duration is n channel uses, where n is large. We use $y^n_{I_1,I_2}$ (respectively $z^n_{I_1,I_2}$) to denote the received signals of user 1 (respectively of user 2) when the CSIT state was some fixed I_1, I_2. The entirety of the received signals are then

$$y^n = \left(y^n_{P,P}, y^n_{P,D}, y^n_{D,P}, y^n_{P,N}, y^n_{N,P}, y^n_{D,D}, y^n_{D,N}, y^n_{N,D}, y^n_{N,N}\right)$$

for the first user, and

$$z^n = \left(z^n_{P,P}, z^n_{P,D}, z^n_{D,P}, z^n_{P,N}, z^n_{N,P}, z^n_{D,D}, z^n_{D,N}, z^n_{N,D}, z^n_{N,N}\right)$$

for the second user.

We proceed with the proof of the outer bound, starting with the proof of (7).

A. Proof for $d_1 + d_2 \le d_{\Sigma}^{(3)}$

We first enhance the BC by offering user 2, complete knowledge of y^n and of W_1. Having now constructed a degraded BC, we proceed to remove all delayed feedback. This removal, which is equivalent to substituting the CSIT state $I_k = D$ with $I_k = N$, is shown in [32] to not affect capacity.

We then proceed to construct a degraded compound BC by adding an additional user, denoted as user $\tilde{1}$, sharing the same desired message W_1 as user 1. The received signal of user $\tilde{1}$ is set as

$$\tilde{y}^n = \left(y^n_{P,P}, y^n_{P,D}, \tilde{y}^n_{D,P}, y^n_{P,N}, \tilde{y}^n_{N,P}, \tilde{y}^n_{D,D}, \tilde{y}^n_{D,N}, \tilde{y}^n_{N,D}, \tilde{y}^n_{N,N}\right)$$

which means that when the CSIT states are $(I_1, I_2) = (P, P), (P, D), (P, N)$ (i.e., whenever the first user sends perfect CSIT) then the received signal of user $\tilde{1}$ is identical to that of user 1, whereas when the CSIT states are $(I_1, I_2) = (D, P), (D, D), (D, N), (N, P), (N, D), (N, N)$ (i.e., when the first user does not send perfect feedback), the received signals of user $\tilde{1}$ are just *identically distributed* to those of user 1. We note the common requirement that, throughout the communication process, user $\tilde{1}$ and user 1 experience the same channel gain exponent $A_{1,t}$ for all t (cf. (3)).

In this degraded compound BC, we proceed and further give the channel output $\tilde{y}^n$ of user $\tilde{1}$, to user 2. Since user 1 and user $\tilde{1}$ have the same decodability, the capacity of this degraded compound BC cannot be worse than that of the original degraded BC.

Finally, we enhance the degraded compound BC, by giving user 2 complete knowledge of

$$s_0^n \triangleq \{s^n_{D,P}, s^n_{N,P}, s^n_{D,N}, s^n_{N,D}, s^n_{D,D}, s^n_{N,N}\}$$

where, as described in (15), $\{s^n_{D,P}, s^n_{N,P}, s^n_{D,N}, s^n_{N,D}, s^n_{D,D}, s^n_{N,N}\}$ are the auxiliary random variables such as, for a given time-slot t during which the transmitter has no CSIT about channel $\tilde{\boldsymbol{h}}_t$ (for user $\tilde{1}$) and channel $\boldsymbol{h}_t$ (for user 1), we let

$$\rho^{\frac{A_{2,t}-A_{1,t}}{2}} \begin{bmatrix} \boldsymbol{h}_t^\mathsf{T} \\ \boldsymbol{g}_t^\mathsf{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_t^\mathsf{T} \\ \tilde{\boldsymbol{h}}_t^\mathsf{T} \end{bmatrix}^{-1} \begin{bmatrix} y_t \\ \tilde{y}_t \end{bmatrix} = \underbrace{\rho^{\frac{A_{2,t}}{2}} \begin{bmatrix} \boldsymbol{h}_t^\mathsf{T} \\ \boldsymbol{g}_t^\mathsf{T} \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} 0 \\ v_t \end{bmatrix}}_{= \begin{bmatrix} \star \\ z_t \end{bmatrix}} + \underbrace{\begin{bmatrix} 0 \\ -v_t \end{bmatrix} + \rho^{\frac{A_{2,t}-A_{1,t}}{2}} \begin{bmatrix} \boldsymbol{h}_t^\mathsf{T} \\ \boldsymbol{g}_t^\mathsf{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_t^\mathsf{T} \\ \tilde{\boldsymbol{h}}_t^\mathsf{T} \end{bmatrix}^{-1} \begin{bmatrix} u_t \\ \tilde{u}_t \end{bmatrix}}_{\triangleq \begin{bmatrix} \star \\ s_t \end{bmatrix}} \tag{15}$$

where s_t is a random valuable with average power

$$\mathbb{E}|s_t|^2 \doteq \rho^{(A_{2,t}-A_{1,t})^+}$$

and where $\tilde{\boldsymbol{h}}_t$ is independently and identically distributed to $\boldsymbol{h}_t$, and $\tilde{u}_t$ is independently and identically distributed to u_t. What this means is that knowledge of $\{s_t, y_t, \tilde{y}_t, \Omega^n\}$, implies the knowledge of z_t, again whenever $(I_1, I_2) = (D, P), (D, D), (D, N), (N, P), (N, D), (N, N)$. In the above, we used $s^n_{I_1,I_2}$ to denote the set of $\{s_t\}_t$, for any t during which the CSIT states are some fixed I_1, I_2.

We now see that

$$\begin{aligned} & nR_1 - n\epsilon_n \\ & = H(W_1) - n\epsilon_n \\ & = H(W_1|\Omega^n) - n\epsilon_n \\ & \le I(W_1; y^n|\Omega^n) && (16) \\ & = h(y^n|\Omega^n) - h(y^n|W_1, \Omega^n) && (17) \end{aligned}$$

where (16) results from Fano's inequality.

Similarly, for virtual user $\tilde{1}$, we have

$$\begin{aligned} & nR_1 - n\epsilon_n \\ & = h(\tilde{y}^n|\Omega^n) - h(\tilde{y}^n|W_1, \Omega^n). && (18) \end{aligned}$$

As a result, adding (17) and (18) gives

$$\begin{aligned}&2nR_1 - 2n\epsilon_n \\ &\le h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) - h(y^n|W_1,\Omega^n) - h(\tilde{y}^n|W_1,\Omega^n) \\ &\le h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) - h(y^n,\tilde{y}^n|W_1,\Omega^n) \qquad (19)\end{aligned}$$

where (19) uses the fact that conditioning reduces the entropy.

Now recalling that user 2 has knowledge of $\{W_1, z^n, y^n, \tilde{y}^n, s_0^n\}$, gives

$$\begin{aligned}&nR_2 - n\epsilon_n \\ &= H(W_2) - n\epsilon_n \\ &= H(W_2|\Omega^n) - n\epsilon_n \\ &\le I(W_2; W_1, z^n, y^n, \tilde{y}^n, s_0^n|\Omega^n) && (20) \\ &= I(W_2; z^n, y^n, \tilde{y}^n, s_0^n|W_1,\Omega^n) && (21) \\ &= h(z^n, y^n, \tilde{y}^n, s_0^n|W_1,\Omega^n) - \underbrace{h(z^n, y^n, \tilde{y}^n, s_0^n|W_1,W_2,\Omega^n)}_{\ge no(\log\rho)} && (22) \\ &\le h(z^n, y^n, \tilde{y}^n, s_0^n|W_1,\Omega^n) - no(\log\rho) && (23) \\ &= h(y^n,\tilde{y}^n|W_1,\Omega^n) + \underbrace{h(s_0^n|y^n,\tilde{y}^n,W_1,\Omega^n)}_{\le h(s_0^n)} + h(z^n|y^n,\tilde{y}^n,s_0^n,W_1,\Omega^n) - no(\log\rho) && (24) \\ &\le h(y^n,\tilde{y}^n|W_1,\Omega^n) + h(s_0^n) + h(z^n|y^n,\tilde{y}^n,s_0^n,W_1,\Omega^n) - no(\log\rho) && (25) \\ &\le h(y^n,\tilde{y}^n|W_1,\Omega^n) + h(s_0^n) + h(z^n_{P,P}, z^n_{P,D}, z^n_{P,N}) \\ &\quad + \underbrace{h(z^n_{D,P}, z^n_{N,P}, z^n_{D,N}, z^n_{N,D}, z^n_{D,D}, z^n_{N,N}|y^n,\tilde{y}^n,s_0^n,W_1,\Omega^n)}_{\le no(\log\rho)} - no(\log\rho) && (26) \\ &\le h(y^n,\tilde{y}^n|W_1,\Omega^n) + h(s_0^n) + h(z^n_{P,P}, z^n_{P,D}, z^n_{P,N}) + no(\log\rho) && (27)\end{aligned}$$

where (20) comes from Fano's inequality, where (23) follows from $h(z^n, y^n, \tilde{y}^n, s_0^n|W_1,W_2,\Omega^n) = h(z^n, y^n, \tilde{y}^n|W_1,W_2,\Omega^n) + \underbrace{h(s_0^n|z^n, y^n, \tilde{y}^n, W_1,W_2,\Omega^n)}_{=0} = h(z^n, y^n, \tilde{y}^n|W_1,W_2,\Omega^n) \ge no(\log\rho)$ by using the fact that the knowledge of $\{z^n, y^n, \tilde{y}^n, \Omega^n\}$ allows for the reconstruction of s_0^n (cf. (15)) and the fact that the knowledge of $\{W_1, W_2, \Omega^n\}$ allows for reconstructing $\{z^n, y^n, \tilde{y}^n\}$ up to noise level, where (24) is from the entropy chain rule, where the transitions to (25) and to (26) use the fact that conditioning reduces entropy, and where (27) is from the fact that the knowledge of $\{y^n, \tilde{y}^n, s_0^n, \Omega^n\}$ allows for the reconstruction of $\{z^n_{D,P}, z^n_{N,P}, z^n_{D,N}, z^n_{N,D}, z^n_{D,D}, z^n_{N,N}\}$; for example the knowledge of $\{y^n_{D,P}, \tilde{y}^n_{D,P}, s^n_{D,P}, \Omega^n\}$ allows for the reconstruction of $\{z^n_{D,P}\}$.

By adding (19) and (27), we have

$$\begin{aligned}&2nR_1 + nR_2 - 3n\epsilon_n \\ &\le h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) + h(s_0^n) + h(z^n_{P,P}, z^n_{P,D}, z^n_{P,N}) + no(\log\rho) && (28) \\ &\le 2\Big(\sum_{\forall I_1 I_2}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_1\lambda^{A_1,A_2}_{I_1,I_2}\Big)\log\rho \\ &\quad + \sum_{I_1 I_2\in\{DP,NP,DN,ND,DD,NN\}} (1-\alpha)\lambda^{\alpha,1}_{I_1,I_2}\log\rho \\ &\quad + \sum_{I_1 I_2\in\{PP,PD,PN\}}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_2\lambda^{A_1,A_2}_{I_1,I_2}\log\rho + no(\log\rho) && (29)\end{aligned}$$

and consequently have

$$2d_1+d_2 \le 2\Big(\sum_{\forall I_1 I_2}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_1\lambda_{I_1,I_2}^{A_1,A_2}\Big) + \sum_{I_1I_2\in\{DP,NP,DN,ND,DD,NN\}}(1-\alpha)\lambda_{I_1,I_2}^{\alpha,1} + \sum_{I_1I_2\in\{PP,PD,PN\}}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_2\lambda_{I_1,I_2}^{A_1,A_2}. \tag{30}$$

Similarly, exchanging the roles of user 1 and user 2, gives

$$2d_2+d_1 \le 2\Big(\sum_{\forall I_1 I_2}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_2\lambda_{I_1,I_2}^{A_1,A_2}\Big) + \sum_{I_1I_2\in\{PD,PN,ND,DN,DD,NN\}}(1-\alpha)\lambda_{I_1,I_2}^{1,\alpha} + \sum_{I_1I_2\in\{PP,DP,NP\}}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_1\lambda_{I_1,I_2}^{A_1,A_2}. \tag{31}$$

Consequently, summing up the two bounds in (30) and (31) gives the following sum GDoF bound

$$\begin{aligned} d_1+d_2 \le \frac{1}{3}\Big[2\Big(\sum_{\forall I_1 I_2}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} (A_1+A_2)\lambda_{I_1,I_2}^{A_1,A_2}\Big) \\ + \sum_{I_1I_2\in\{DP,NP,DN,ND,DD,NN\}}(1-\alpha)\lambda_{I_1,I_2}^{\alpha,1} + \sum_{I_1I_2\in\{PD,PN,ND,DN,DD,NN\}}(1-\alpha)\lambda_{I_1,I_2}^{1,\alpha} \\ + \sum_{I_1I_2\in\{PP,PD,PN\}}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_2\lambda_{I_1,I_2}^{A_1,A_2} + \sum_{I_1I_2\in\{PP,DP,NP\}}\sum_{A_2\in\{1,\alpha\}}\sum_{A_1\in\{1,\alpha\}} A_1\lambda_{I_1,I_2}^{A_1,A_2}\Big] \end{aligned} \tag{32}$$

which, after some manipulation gives

$$\begin{aligned} d_1+d_2 &\le (1+\alpha)(\lambda_{P,P}^{\alpha,1}+\lambda_{P,P}^{1,\alpha}) + \frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow D}^{\alpha,1}+\lambda_{P\leftrightarrow D}^{1,\alpha}) + \frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow N}^{\alpha,1}+\lambda_{P\leftrightarrow N}^{1,\alpha}) \\ &+ \frac{3+\alpha}{3}(\lambda_{D,D}^{\alpha,1}+\lambda_{D,D}^{1,\alpha}) + \frac{3+\alpha}{3}(\lambda_{D\leftrightarrow N}^{\alpha,1}+\lambda_{D\leftrightarrow N}^{1,\alpha}) + \frac{3+\alpha}{3}(\lambda_{N,N}^{\alpha,1}+\lambda_{N,N}^{1,\alpha}) \\ &+ 2\lambda_{P,P}^{1,1} + 2\alpha\lambda_{P,P}^{\alpha,\alpha} + \frac{5}{3}\lambda_{P\leftrightarrow D}^{1,1} + \frac{5\alpha}{3}\lambda_{P\leftrightarrow D}^{\alpha,\alpha} + \frac{5}{3}\lambda_{P\leftrightarrow N}^{1,1} + \frac{5\alpha}{3}\lambda_{P\leftrightarrow N}^{\alpha,\alpha} \\ &+ \frac{4}{3}\lambda_{D,D}^{1,1} + \frac{4\alpha}{3}\lambda_{D,D}^{\alpha,\alpha} + \frac{4}{3}\lambda_{D\leftrightarrow N}^{1,1} + \frac{4\alpha}{3}\lambda_{D\leftrightarrow N}^{\alpha,\alpha} + \frac{4}{3}\lambda_{N,N}^{1,1} + \frac{4\alpha}{3}\lambda_{N,N}^{\alpha,\alpha}. \end{aligned} \tag{33}$$

B. *Proof for* $d_1+d_2 \le d_{\sum}^{(4)}$

We continue with the proof of the general outer bound, and provide the proof of (7).

We first enhance the BC, by substituting delayed CSIT with perfect CSIT, i.e., by treating CSIT state $I_k = D$ as if it corresponded to $I_k = P$. We then transition to the compound BC by introducing the additional user $\tilde{1}$, as well as an extra additional user, denoted as user $\tilde{2}$. By definition, user 1 and user $\tilde{1}$ share the same desired message W_1. The received signal of user $\tilde{1}$ is set as

$$\tilde{y}^n = \big(y_0^n, y_{P,N}^n, \tilde{y}_{N,P}^n, y_{D,N}^n, \tilde{y}_{N,D}^n, \tilde{y}_{N,N}^n\big), \quad \text{for} \quad y_0^n \triangleq \big(y_{P,P}^n, y_{P,D}^n, y_{D,P}^n, y_{D,D}^n\big)$$

which means that user 1 and user $\tilde{1}$ share the same received signal whenever the CSIT states are $(I_1,I_2) = (P,P),(P,D),(D,P),(D,D),(P,N),(D,N)$ (i.e., whenever the channel of user 1 is treated with either perfect or delayed feedback), while when $(I_1,I_2) = (N,P),(N,D),(N,N)$ (i.e., whenever the channel of user 1 is not fed back), user 1 and user $\tilde{1}$ have received signals that are statistically identical, but not necessarily the same.

Similarly, user 2 and user $\tilde{2}$ share the same desired message W_2. The received signals for user $\tilde{2}$ take the form

$$\tilde{z}^n = \left(z_0^n, z_{N,P}^n, \tilde{z}_{P,N}^n, z_{N,D}^n, \tilde{z}_{D,N}^n, \tilde{z}_{N,N}^n\right), \quad \text{for} \quad z_0^n \triangleq \left(z_{P,P}^n, z_{D,P}^n, z_{P,D}^n, z_{D,D}^n\right)$$

which again means that user 2 and user $\tilde{2}$ share the same received signal whenever the CSIT states are $(I_1, I_2) = (P, P), (D, P), (P, D), (D, D), (N, P), (N, D)$, else $((I_1, I_2) = (P, N), (D, N), (N, N))$ user 2 and user $\tilde{2}$ have received signals that are statistically identical. In this compound BC, the capacity remains the same as in the original enhanced BC, since user 1 and user $\tilde{1}$ have the same decodability, and because the same holds for user 2 and user $\tilde{2}$.

Furthermore, since the capacity depends only on the marginals, whenever $(I_1, I_2) = (N, N)$ we can assume without an effect to the result, that the channel vectors $\boldsymbol{g}_t, \tilde{\boldsymbol{g}}_t, \tilde{\boldsymbol{h}}_t, \boldsymbol{h}_t$ are the same for all four users, i.e., $\boldsymbol{g}_t = \tilde{\boldsymbol{g}}_t = \tilde{\boldsymbol{h}}_t = \boldsymbol{h}_t$, ($\tilde{\boldsymbol{g}}_t$ and $\tilde{\boldsymbol{h}}_t$ for user $\tilde{2}$ and user $\tilde{1}$ respectively).

Additionally for any t during which $(I_1, I_2) = (N, N)$, we define

$$\bar{y}_t = \sqrt{\rho^{\min\{A_{1,t}, A_{2,t}\}}} \boldsymbol{h}_t^{\mathsf{T}} \boldsymbol{x}_t + \bar{u}_t \tag{34}$$

where $\bar{u}_t$ is a unit-power AWGN random variable, where

$$\sqrt{\rho^{A_{1,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{y}_t = \underbrace{\sqrt{\rho^{A_{1,t}}}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t + u_t}_{=y_t} + \underbrace{\sqrt{\rho^{A_{1,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{u}_t - u_t}_{\triangleq \omega_t} \tag{35}$$

$$\sqrt{\rho^{A_{2,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{y}_t = \underbrace{\sqrt{\rho^{A_{2,t}}}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t + v_t}_{=z_t} + \underbrace{\sqrt{\rho^{A_{2,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{u}_t - v_t}_{\triangleq \psi_t} \tag{36}$$

and where the two new random variables ω_t, ψ_t have power

$$\mathbb{E}|\omega_t|^2 \doteq \rho^{(A_{1,t}-A_{2,t})^+}$$

and

$$\mathbb{E}|\psi_t|^2 \doteq \rho^{(A_{2,t}-A_{1,t})^+}.$$

The collection of all $\{\bar{y}_t\}_t$ for all t such that $(I_1, I_2) = (N, N)$, is denoted by $\bar{y}_{N,N}^n$, and similarly $\omega_{N,N}^n$ and $\psi_{N,N}^n$ respectively denote the set of $\{\omega_t\}_t$ and $\{\psi_t\}_t$ for all t such that $(I_1, I_2) = (N, N)$.

Finally we give each user the observation $\bar{y}_{NN}^n$ (enhanced compound BC).

At this point we have

$$\begin{aligned}
&nR_1 - n\epsilon_n \\
&= H(W_1) - n\epsilon_n \\
&= H(W_1|\Omega^n) - n\epsilon_n \\
&\le I(W_1; y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, y_{N,N}^n, \bar{y}_{N,N}^n|\Omega^n) && (37)\\
&= I(W_1; y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n|\Omega^n) + I(W_1; y_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) && (38)\\
&= I(W_1; y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n|\Omega^n) \\
&\quad + \underbrace{h(y_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\le h(y_{N,N}^n|\bar{y}_{N,N}^n, \Omega^n)} - \underbrace{h(y_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n, W_1, \Omega^n)}_{\ge h(y_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n, W_1, W_2, \Omega^n) \ge no(\log\rho)} && (39)\\
&\le I(W_1; y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n|\Omega^n) + \underbrace{h(y_{N,N}^n|\bar{y}_{N,N}^n, \Omega^n)}_{= h(\omega_{N,N}^n|\bar{y}_{N,N}^n, \Omega^n) \le h(\omega_{N,N}^n)} + no(\log\rho) && (40)\\
&\le I(W_1; y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, \bar{y}_{N,N}^n|\Omega^n) + h(\omega_{N,N}^n) + no(\log\rho) && (41)
\end{aligned}$$

$$= h(\omega_{N,N}^n) + no(\log\rho) + \underbrace{I(W_1; y_0^n | y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\leq h(y_0^n) + no(\log\rho)}$$
$$+ I(W_1; y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n | \Omega^n) \quad (42)$$
$$\leq h(\omega_{N,N}^n) + h(y_0^n) + no(\log\rho) + I(W_1; y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n | \Omega^n) \quad (43)$$
$$= h(\omega_{N,N}^n) + h(y_0^n) + no(\log\rho) + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) + I(W_1; y_{N,P}^n, y_{N,D}^n | y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) \quad (44)$$
$$= h(\omega_{N,N}^n) + \underbrace{h(y_0^n)}_{\leq n\Phi_{10} + no(\log\rho)} + no(\log\rho) + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n)$$
$$+ \underbrace{I(W_1, W_2; y_{N,P}^n, y_{N,D}^n | y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\leq h(y_{N,P}^n, y_{N,D}^n) + no(\log\rho) \leq n\Phi_{11} + no(\log\rho)} - I(W_2; y_{N,P}^n, y_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) \quad (45)$$
$$\leq h(\omega_{N,N}^n) + n\Phi_{10} + n\Phi_{11} + no(\log\rho)$$
$$+ I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; y_{N,P}^n, y_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) \quad (46)$$

where

$$\Phi_{10} \triangleq \Big(\sum_{I_1 I_2 \in \{PP, PD, DP, DD\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda_{I_1, I_2}^{A_1, A_2} \Big) \log\rho$$

and

$$\Phi_{11} \triangleq \Big(\sum_{I_1 I_2 \in \{NP, ND\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda_{I_1, I_2}^{A_1, A_2} \Big) \log\rho$$

where (37) results from Fano's inequality, where the transition to (40) uses the fact that conditioning reduces entropy and the fact that $y_{N,N}^n$ can be reconstructed with errors up to noise level by using the knowledge of $\{W_1, W_2, \Omega^n\}$, where (41) follows from the definition in (35) and from the fact that conditioning reduces entropy, and where (43) - (46) are derived using basic entropy rules.

Similarly for user $\tilde{1}$, we have

$$nR_1 - n\epsilon_n$$
$$\leq h(\omega_{N,N}^n) + n\Phi_{10} + n\Phi_{11} + no(\log\rho)$$
$$+ I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n). \quad (47)$$

Adding (46) and (47), gives

$$2nR_1 - 2\Phi_{10} - 2\Phi_{11} - no(\log\rho) - 2n\epsilon_n$$
$$\leq 2h(\omega_{N,N}^n) + 2I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; y_{N,P}^n, y_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)$$
$$- I(W_2; \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) \quad (48)$$
$$= 2h(\omega_{N,N}^n) + 2I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n)$$
$$\underbrace{- h(y_{N,P}^n, y_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n) - h(\tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\leq -h(y_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}$$
$$+ \underbrace{h(y_{N,P}^n, y_{N,D}^n | W_2, W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\leq no(\log\rho)} + \underbrace{h(\tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_2, W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\leq no(\log\rho)}$$
$$\leq 2h(\omega_{N,N}^n) + 2I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - h(W_2; y_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)$$
$$+ no(\log\rho) \quad (49)$$
$$= 2h(\omega_{N,N}^n) + 2I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; y_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)$$
$$- \underbrace{h(y_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n | W_2, W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\geq no(\log\rho)} + no(\log\rho) \quad (50)$$

$$
\begin{aligned}
&\leq 2h(\omega_{N,N}^n) + 2I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) \\
&\quad + I(W_2; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) + no(\log\rho) && (51) \\
&= 2h(\omega_{N,N}^n) + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) \\
&\quad + \underbrace{I(W_1, W_2; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n)}_{\leq h(y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n) + no(\log\rho)} + no(\log\rho) && (52) \\
&\leq h(\omega_{N,N}^n) + \underbrace{h(\omega_{N,N}^n) + h(y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n)}_{\leq n\Phi_{12} + no(\log\rho)} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) \\
&\quad - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) + no(\log\rho) && (53) \\
&\leq h(\omega_{N,N}^n) + n\Phi_{12} + no(\log\rho) + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) \\
&\quad - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (54) \\
&= h(\omega_{N,N}^n) + n\Phi_{12} + no(\log\rho) + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) \\
&\quad - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, s_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, s_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) \\
&\quad + \underbrace{I(W_2; s_{N,P}^n, s_{N,D}^n | y_{N,P}^n, \tilde{y}_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, W_1, \Omega^n)}_{\leq h(s_{N,P}^n, s_{N,D}^n) + no(\log\rho)} && (55) \\
&\leq h(\omega_{N,N}^n) + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) \\
&\quad - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, s_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, s_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) + h(s_{N,P}^n, s_{N,D}^n) + no(\log\rho) && (56) \\
&= h(\omega_{N,N}^n) + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) + h(s_{N,P}^n, s_{N,D}^n) + no(\log\rho) \\
&\quad - I(W_2; y_{N,P}^n, \tilde{y}_{N,P}^n, s_{N,P}^n, z_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, s_{N,D}^n, z_{N,D}^n, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (57) \\
&\leq h(\omega_{N,N}^n) + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) + h(s_{N,P}^n, s_{N,D}^n) + no(\log\rho) \\
&\quad - I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (58) \\
&\leq h(\omega_{N,N}^n) + n\Phi_{12} + I(W_1; W_2, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | \Omega^n) + h(s_{N,P}^n, s_{N,D}^n) + no(\log\rho) \\
&\quad - I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (59) \\
&= h(\omega_{N,N}^n) + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_2, \Omega^n) + \underbrace{h(s_{N,P}^n, s_{N,D}^n)}_{\leq n\Phi_{13} + no(\log\rho)} + no(\log\rho) \\
&\quad - I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (60) \\
&\leq \underbrace{h(\omega_{N,N}^n)}_{\leq n\Phi_{14} + no(\log\rho)} + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_2, \Omega^n) + n\Phi_{13} + no(\log\rho) \\
&\quad - I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (61) \\
&\leq n\Phi_{14} + n\Phi_{12} + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_2, \Omega^n) + n\Phi_{13} + no(\log\rho) \\
&\quad - I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) && (62)
\end{aligned}
$$

where

$$\Phi_{12} \triangleq \Big(\sum_{I_1 I_2 \in \{PN, DN, NN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \log\rho$$

$$\Phi_{13} \triangleq \Big(\sum_{I_1 I_2 \in \{NP, ND\}} (1-\alpha) \lambda_{I_1,I_2}^{\alpha,1} \Big) \log\rho$$

$$\Phi_{14} \triangleq (1-\alpha) \lambda_{N,N}^{1,\alpha} \log\rho$$

where $s^n_{N,P}$ and $z^n_{N,D}$ (see (55)) are defined in (15). Furthermore (57) is from the fact that the knowledge of $\{y^n_{N,P}, \tilde{y}^n_{N,P}, s^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, s^n_{N,D}, \Omega^n\}$ implies the knowledge of $z^n_{N,P}$ and $z^n_{N,D}$ (cf. (15)). In the above, most of the steps use basic entropy rules.

Similarly, considering user 2 and user $\tilde{2}$, we have

$$\begin{aligned} & 2nR_2 - 2\Phi_{20} - 2\Phi_{21} - no(\log\rho) - 2n\epsilon_n \\ & \leq n\Phi_{24} + n\Phi_{22} + I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N} | W_1, \Omega^n) + n\Phi_{23} + no(\log\rho) \\ & \quad - I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N} | W_2, \Omega^n) \end{aligned} \tag{63}$$

where

$$\Phi_{20} \triangleq \Big(\sum_{I_1 I_2 \in \{PP,PD,DP,DD\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_2 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \log\rho$$

$$\Phi_{21} \triangleq \Big(\sum_{I_1 I_2 \in \{PN,DN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \log\rho$$

$$\Phi_{22} \triangleq \Big(\sum_{I_1 I_2 \in \{NP,ND,NN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \log\rho$$

$$\Phi_{23} \triangleq \Big(\sum_{I_1 I_2 \in \{PN,DN\}} (1-\alpha) \lambda^{1,\alpha}_{I_1,I_2} \Big) \log\rho$$

$$\Phi_{24} \triangleq (1-\alpha) \lambda^{\alpha,1}_{N,N} \log\rho.$$

Finally, combining (62) and (63), gives

$$\begin{aligned} & d_1 + d_2 \\ & \leq \frac{1}{2\log\rho} \Big[2\Phi_{10} + 2\Phi_{11} + \Phi_{12} + \Phi_{13} + \Phi_{14} + 2\Phi_{20} + 2\Phi_{21} + \Phi_{22} + \Phi_{23} + \Phi_{24} \Big] \\ & = \frac{1}{2} \Big[2\Big(\sum_{I_1 I_2 \in \{PP,PD,DP,DD\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \\ & \quad + 2\Big(\sum_{I_1 I_2 \in \{NP,ND\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \\ & \quad + \sum_{I_1 I_2 \in \{PN,DN,NN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_1 \lambda^{A_1,A_2}_{I_1,I_2} \\ & \quad + \sum_{I_1 I_2 \in \{NP,ND\}} (1-\alpha) \lambda^{\alpha,1}_{I_1,I_2} + (1-\alpha) \lambda^{1,\alpha}_{N,N} \\ & \quad + 2\Big(\sum_{I_1 I_2 \in \{PP,PD,DP,DD\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_2 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \\ & \quad + 2\Big(\sum_{I_1 I_2 \in \{PN,DN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \Big) \\ & \quad + \sum_{I_1 I_2 \in \{NP,ND,NN\}} \sum_{A_2 \in \{1,\alpha\}} \sum_{A_1 \in \{1,\alpha\}} A_2 \lambda^{A_1,A_2}_{I_1,I_2} \\ & \quad + \sum_{I_1 I_2 \in \{PN,DN\}} (1-\alpha) \lambda^{1,\alpha}_{I_1,I_2} + (1-\alpha) \lambda^{\alpha,1}_{N,N} \Big] \\ & = \sum_{I_1 I_2 \in \{PP,PD,DP,DD\}} (1+\alpha) \big(\lambda^{1,\alpha}_{I_1,I_2} + \lambda^{\alpha,1}_{I_1,I_2} \big) \end{aligned}$$

TABLE I

SUMMARY OF SCHEMES

Scheme #	Section #	CSIT, topology	achieved $d_{\sum}$	for Proposition #
1	VIII-A	$\lambda_{1,\alpha}=1$ $\lambda_{N,D}=\lambda_{P,N}=1/2$	$1+\frac{\alpha}{2}$ optimal	Proposition 1
2	VIII-B	$\lambda_{1,\alpha}=1$ $\lambda_{P,D}=\lambda_{N,N}=1/2$	$1+\frac{\alpha}{2}$ optimal	Proposition 1
3	VIII-C	$\lambda_{1,\alpha}=1$ $\lambda_{D,D}=1$	$1+\frac{\alpha^2}{2+\alpha}$	Proposition 2
4	VIII-D	$\lambda_{1,\alpha}=\lambda_{\alpha,1}=1/2$ $\lambda_{D,D}=1$	$1+\frac{\alpha}{3}$ optimal	Proposition 3
5	VIII-E	any $\lambda_{1,\alpha}+\lambda_{\alpha,1}=1$ $\lambda_{P,N}=\lambda_{N,P}=1/2$	$1+\frac{\alpha}{2}$ optimal	Propositions 1, 4, 5
MAT	VIII-F	$\lambda^{1,\alpha}_{D,D}=1$	$\frac{2(1+\alpha)}{3}$ sub-optimal	-

$$
\begin{aligned}
&+\sum_{I_1I_2\in\{NP,PN,ND,DN\}} \frac{2+\alpha}{2}\big(\lambda^{1,\alpha}_{I_1,I_2}+\lambda^{\alpha,1}_{I_1,I_2}\big)+\big(\lambda^{1,\alpha}_{N,N}+\lambda^{\alpha,1}_{N,N}\big)\\
&+\sum_{I_1I_2\in\{PP,PD,DP,DD\}} \big(2\lambda^{1,1}_{I_1,I_2}+2\alpha\lambda^{\alpha,\alpha}_{I_1,I_2}\big)\\
&+\sum_{I_1I_2\in\{NP,PN,ND,DN\}} \big(\frac{3}{2}\lambda^{1,1}_{I_1,I_2}+\frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{I_1,I_2}\big)+\big(\lambda^{1,1}_{N,N}+\alpha\lambda^{\alpha,\alpha}_{N,N}\big)\\
=&\big(1+\alpha\big)\big(\lambda^{1,\alpha}_{P,P}+\lambda^{\alpha,1}_{P,P}\big)+\big(1+\alpha\big)\big(\lambda^{1,\alpha}_{P\leftrightarrow D}+\lambda^{\alpha,1}_{P\leftrightarrow D}\big)+\big(1+\alpha\big)\big(\lambda^{1,\alpha}_{D,D}+\lambda^{\alpha,1}_{D,D}\big)\\
&+\frac{2+\alpha}{2}\big(\lambda^{1,\alpha}_{P\leftrightarrow N}+\lambda^{\alpha,1}_{P\leftrightarrow N}\big)+\frac{2+\alpha}{2}\big(\lambda^{1,\alpha}_{D\leftrightarrow N}+\lambda^{\alpha,1}_{D\leftrightarrow N}\big)+\big(\lambda^{1,\alpha}_{N,N}+\lambda^{\alpha,1}_{N,N}\big)\\
&+\big(2\lambda^{1,1}_{P,P}+2\alpha\lambda^{\alpha,\alpha}_{P,P}\big)+\big(2\lambda^{1,1}_{P\leftrightarrow D}+2\alpha\lambda^{\alpha,\alpha}_{P\leftrightarrow D}\big)+\big(2\lambda^{1,1}_{D,D}+2\alpha\lambda^{\alpha,\alpha}_{D,D}\big)\\
&+\big(\frac{3}{2}\lambda^{1,1}_{P\leftrightarrow N}+\frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{P\leftrightarrow N}\big)+\big(\frac{3}{2}\lambda^{1,1}_{D\leftrightarrow N}+\frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{D\leftrightarrow N}\big)+\big(\lambda^{1,1}_{N,N}+\alpha\lambda^{\alpha,\alpha}_{N,N}\big)
\end{aligned}
\tag{64}
$$

which completes the proof.

VIII. APPENDIX - SCHEMES

We here design the topological signal management schemes for the different topology and feedback scenarios (see Table I for a summary). In what follows, we will generally associate the use of symbol a to denote the private symbol for user 1, while we will associate symbol b to denote the private symbol for user 2, and symbol c to denote common symbol for both users. We will also use $P^{(q)}\triangleq\mathbb{E}|q|^2$ to denote the average power of some symbol q, and will use $r^{(q)}$ to denote the prelog factor of the number of bits $[r^{(q)}\log\rho-o(\log\rho)]$ carried by symbol q. In the interest of brevity, we will on occasion neglect the additive noise terms, without an effect on the GDoF analysis.

A. *TSM scheme for* $\lambda^{1,\alpha}_{N,D}=\lambda^{1,\alpha}_{P,N}=1/2$ *achieving the optimal sum GDoF* $1+\alpha/2$

For the setting of $\lambda^{1,\alpha}_{N,D}=\lambda^{1,\alpha}_{P,N}=1/2$, the proposed scheme consists of two transmissions. Without loss of generality, we will assume that during the first channel use, $t=1$, the feedback-and-topology state

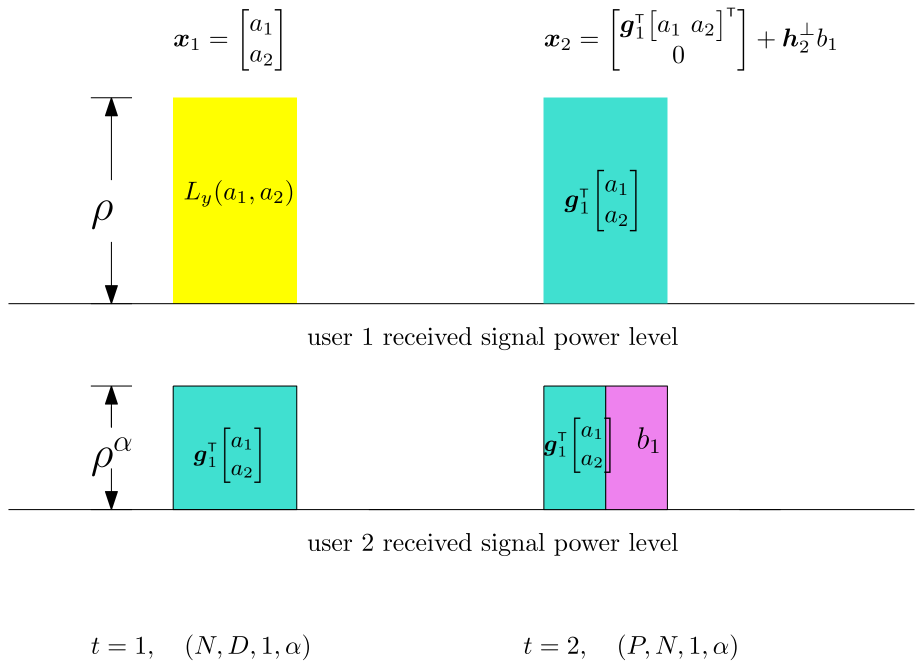

Fig. 5. Illustration of received signal power level for the TSM scheme for $\lambda_{N,D}^{1,\alpha} = \lambda_{P,N}^{1,\alpha} = 1/2$.

is $(I_1, I_2, A_1, A_2) = (N, D, 1, \alpha)$, while during the second channel use, $t = 2$, the feedback-and-topology state is $(I_1, I_2, A_1, A_2) = (P, N, 1, \alpha)$.

At time $t = 1$ there is no CSIT, and the transmitter sends (see Figure 5)

$$\boldsymbol{x}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{65}$$

where a_1, a_2 are symbols meant for user 1, where

$$\begin{aligned} P^{(a_1)} \doteq 1, \quad r^{(a_1)} = 1 \\ P^{(a_2)} \doteq 1, \quad r^{(a_2)} = 1 \end{aligned} \tag{66}$$

resulting in received signals of the form

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho} + \underbrace{u_t}_{\rho^0} \tag{67}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho^{\alpha}} + \underbrace{v_t}_{\rho^0} \tag{68}$$

where under each term we noted the order of the summand's average power. One can briefly note that the unintended interference is naturally attenuated due to the weak link.

At time $t = 2$, the transmitter has knowledge of $\boldsymbol{g}_1$ (delayed feedback) and of $\boldsymbol{h}_2$ (current feedback). As a result, the transmitter reconstructs $\boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 & a_2 \end{bmatrix}^{\intercal}$ and sends

$$\boldsymbol{x}_2 = \begin{bmatrix} \boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 & a_2 \end{bmatrix}^{\intercal} \\ 0 \end{bmatrix} + \boldsymbol{h}_2^{\perp} b_1 \tag{69}$$

where b_1 is meant for user 2, and where

$$P^{(b_1)} \doteq 1, \quad r^{(b_1)} = \alpha. \tag{70}$$

Then the processed (normalized) received signals take the form

$$y_2/h_{2,1} = \underbrace{\sqrt{\rho}\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}}_{\rho} + \underbrace{\frac{u_t}{h_{2,1}}}_{\rho^0} \tag{71}$$

$$z_2/g_{2,1} = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{\alpha}}\frac{\boldsymbol{g}_2^{\mathsf{T}}\boldsymbol{h}_2^{\perp}}{g_{2,1}}b_1}_{\rho^{\alpha}} + \underbrace{\frac{v_t}{g_{2,1}}}_{\rho^0} \tag{72}$$

where $h_{t,1} \triangleq \boldsymbol{h}_t^{\mathsf{T}}\begin{bmatrix}1 & 0\end{bmatrix}^{\mathsf{T}}$, $g_{t,1} \triangleq \boldsymbol{g}_t^{\mathsf{T}}\begin{bmatrix}1 & 0\end{bmatrix}^{\mathsf{T}}$, and where the normalized noise power (of $\frac{u_t}{h_{2,1}}$ and $\frac{v_t}{g_{2,1}}$) is noted to be *typically* bounded, since $Pr(|h_{2,1}|^2 \leq \rho^{-\epsilon}) \doteq Pr(|h_{2,1}|^2 \leq \rho^{-\epsilon}) \doteq \rho^{-\epsilon}$ for arbitrarily small positive ϵ.

At this point, it is easy to see that user 1 can recover a_1, a_2 at the declared rates, by MIMO decoding based on (67), (71), while user 2 can recover b_1 by employing interference cancelation based on (68), (72). This provides for the optimal sum GDoF $d_{\sum} = 1 + \alpha/2$.

B. *TSM scheme for* $\lambda_{P,D}^{1,\alpha} = \lambda_{N,N}^{1,\alpha} = 1/2$*, achieving the optimal sum GDoF* $1 + \alpha/2$

For the setting where $\lambda_{P,D}^{1,\alpha} = \lambda_{N,N}^{1,\alpha} = 1/2$, the proposed scheme has two channel uses. Again without loss of generality, we assume that during $t = 1$ the state is $(I_1, I_2, A_1, A_2) = (P, D, 1, \alpha)$, while during $t = 2$, the state is $(I_1, I_2, A_1, A_2) = (N, N, 1, \alpha)$.

At $t = 1$ the transmitted knows $\boldsymbol{h}_1$ (current CSIT) and sends (see Figure 6)

$$\boldsymbol{x}_1 = \begin{bmatrix}a_1\\a_2\end{bmatrix} + \boldsymbol{h}_1^{\perp}b_1 \tag{73}$$

where a_1, a_2 are unit-power symbols meant for user 1, and b_1 is unit-power symbol meant for user 2, and where

$$r^{(a_1)} = 1, \quad r^{(a_2)} = 1, \quad r^{(b_1)} = \alpha. \tag{74}$$

Then the received signals (in their noiseless form) are

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}}_{\rho} \tag{75}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1^{\perp}b_1}_{\rho^{\alpha}} \tag{76}$$

where in the above we omitted the noise without an effect to the derived DoF expressions.

At $t = 2$ $((I_1, I_2, A_1, A_2) = (N, N, 1, \alpha))$, the transmitter knows $\boldsymbol{g}_1$ (delayed CSIT), reconstructs $\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}$, and sends

$$\boldsymbol{x}_2 = \begin{bmatrix}\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1 & a_2\end{bmatrix}^{\mathsf{T}}\\ 0\end{bmatrix}. \tag{77}$$

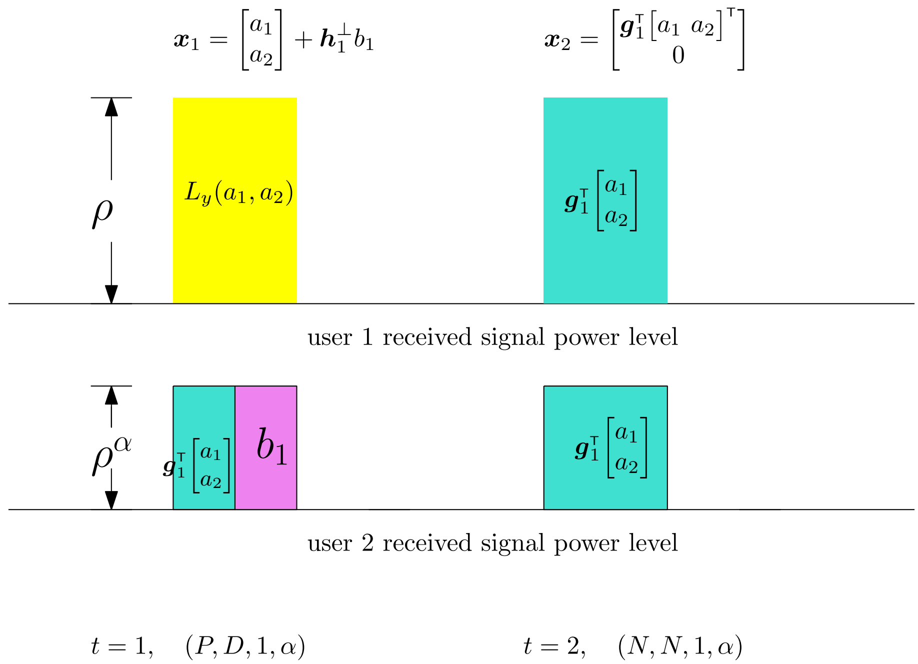

Fig. 6. Illustration of received signal power level for the TSM scheme: The setting with $\lambda_{P,D}^{1,\alpha} = \lambda_{N,N}^{1,\alpha} = 1/2$.

After normalization, the received signals (in their noiseless form) are

$$y_2/h_{2,1} = \underbrace{\sqrt{\rho}\boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho} \tag{78}$$

$$z_2/g_{2,1} = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho^{\alpha}}. \tag{79}$$

One can now easily see that, user 1 can MIMO decode a_1, a_2 based on (75) and (78), while user 2 can recover b_1 by employing interference cancelation based on (76) and (79) (see also Figure 6). This achieves the optimal sum GDoF $d_{\sum} = 1 + \alpha/2$.

C. TSM scheme for the case with $\lambda_{D,D}^{1,\alpha} = 1$

The proposed scheme has three phases, of respective durations T_1, T_2, T_3 channel uses[5],[6]

$$T_2/\alpha = T_1 = T_3 \tag{80}$$

which - as we will see later on - are chosen so that the amount of side information, at user 1 and user 2, are properly balanced.

1) Phase 1: When $t = 1, 2, \cdots, T_1$, the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} a_{t,1} \\ a_{t,2} \end{bmatrix} \tag{81}$$

where $a_{t,1}$ and $a_{t,2}$ are unit-power symbols meant for user 1, and where

$$r^{(a_{t,1})} = 1, \quad r^{(a_{t,2})} = \alpha. \tag{82}$$

[5]Here we assume that α is rational, such that T_1, T_2, T_3 can be integer. The case of irrational α, can be handled with minor modifications to the scheme.

[6]As a clarifying example, when $\alpha = 1/2$, the phase durations are $T_1 = 2,\ T_2 = 1,\ T_3 = 2$.

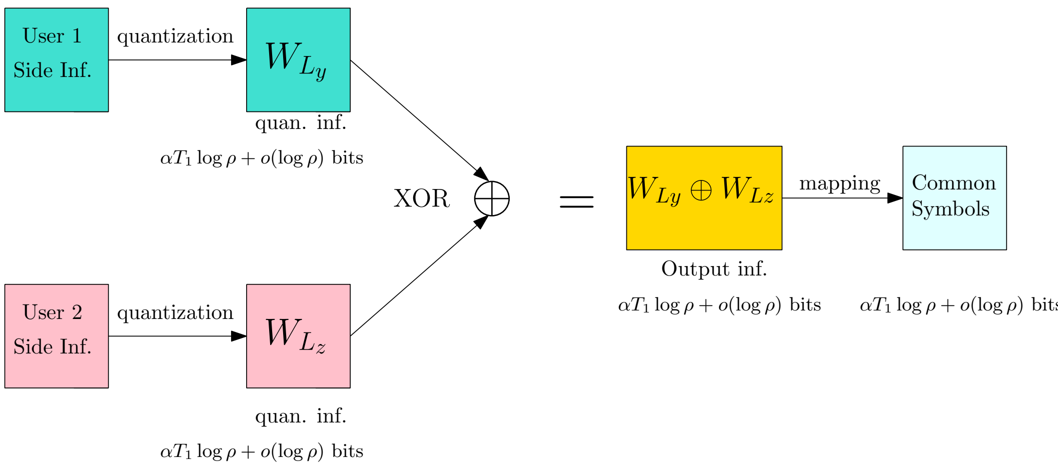

Fig. 7. Illustration for side information reconstruction and quantization, bitwise XOR operation, and symbol mapping.

The received signals then take the form

$$y_t = \underbrace{\sqrt{\rho}\boldsymbol{h}_t^\mathsf{T} \begin{bmatrix} a_{t,1} \\ a_{t,2} \end{bmatrix}}_{\rho} + \underbrace{u_t}_{\rho^0} \tag{83}$$

$$z_t = \underbrace{\sqrt{\rho^\alpha}\boldsymbol{g}_t^\mathsf{T} \begin{bmatrix} a_{t,1} \\ a_{t,2} \end{bmatrix}}_{\rho^\alpha} + \underbrace{v_t}_{\rho^0} = \sqrt{\rho^\alpha} L_z(a_{t,1}, a_{t,2}) + v_t \tag{84}$$

where $L_z(a_{t,1}, a_{t,2}) \triangleq \boldsymbol{g}_t^\mathsf{T} \begin{bmatrix} a_{t,1} \\ a_{t,2} \end{bmatrix}$ represents interference at the second receiver.

2) Phase 2: When $t = T_1 + 1, \cdots, T_1 + T_2$, the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} b_{t,1} \\ b_{t,2} \end{bmatrix} \tag{85}$$

where $b_{t,1}, b_{t,2}$ are unit-power symbols meant for user 2, and where

$$r^{(b_{t,1})} = 1, \quad r^{(b_{t,2})} = \alpha \tag{86}$$

resulting in received signals of the form

$$y_t = \underbrace{\sqrt{\rho}\boldsymbol{h}_t^\mathsf{T} \begin{bmatrix} b_{t,1} \\ b_{t,2} \end{bmatrix}}_{\rho} + \underbrace{u_t}_{\rho^0} = \sqrt{\rho} L_y(b_{t,1}, b_{t,2}) + u_t \tag{87}$$

$$z_t = \underbrace{\sqrt{\rho^\alpha}\boldsymbol{g}_t^\mathsf{T} \begin{bmatrix} b_{t,1} \\ b_{t,2} \end{bmatrix}}_{\rho^\alpha} + \underbrace{v_t}_{\rho^0} \tag{88}$$

where $L_y(b_{t,1}, b_{t,2}) \triangleq \boldsymbol{h}_t^\mathsf{T} \begin{bmatrix} b_{t,1} \\ b_{t,2} \end{bmatrix}$ represents interference at the first receiver.

3) Phase 3: At the end of the second phase, user 1 knows $\{y_t = \sqrt{\rho} L_y(b_{t,1}, b_{t,2}) + u_t\}_{t=T_1+1}^{T_1+T_2}$, while user 2 knows $\{z_t = \sqrt{\rho^\alpha} L_z(a_{t,1}, a_{t,2}) + v_t\}_{t=1}^{T_1}$. At the same time, with the help of delayed CSIT, the transmitter *reconstructs* and *quantizes* the above side information, up to noise level (see Figure 7). Specifically, the transmitter *reconstructs*

$$\left[\sqrt{\rho^\alpha} L_z(a_{1,1}, a_{1,2}) \quad \sqrt{\rho^\alpha} L_z(a_{2,1}, a_{2,2}) \ \cdots \ \sqrt{\rho^\alpha} L_z(a_{T_1,1}, a_{T_1,2})\right] \tag{89}$$

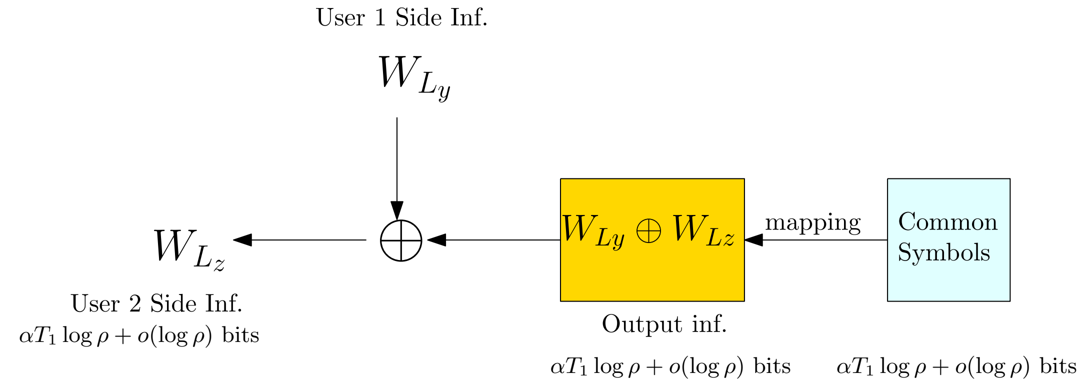

Fig. 8. Illustration for side information decoding at user 1: Learning user 2's side information from the common information and its side information.

and *quantizes* the vector using

$$\alpha T_1 \log\rho + o(\log\rho) \tag{90}$$

quantization bits, allowing for bounded quantization error because $\mathbb{E}|\sqrt{\rho^{\alpha}} L_z(a_{t,1}, a_{t,2})|^2 \doteq \rho^{\alpha},\ t = 1, 2, \cdots, T_1$ (cf. [31]). Similarly the transmitter reconstructs

$$\left[\sqrt{\rho}L_y(b_{T_1+1,1}, b_{T_1+1,2}) \quad \sqrt{\rho}L_y(b_{T_1+2,1}, b_{T_1+2,2}) \ \cdots \ \sqrt{\rho}L_y(b_{T_1+T_2,1}, b_{T_1+T_2,2})\right] \tag{91}$$

and *quantizes* it using

$$T_2 \log\rho + o(\log\rho) \tag{92}$$

quantization bits, which allows for bounded quantization error since $\mathbb{E}|\sqrt{\rho}L_y(b_{t,1}, b_{t,2})|^2 \doteq \rho,\ t = T_1 + 1, \cdots, T_1 + T_2$.

Next the transmitter performs the bitwise exclusive-or (XOR) operation on the two sets of quantization bits, i.e., proceeds to bitwise XOR W_{L_z} and W_{L_y} (see Figure 7), where W_{L_z} denotes the vector of $\alpha T_1 \log\rho + o(\log\rho)$ quantization bits corresponding to (89), and where W_{L_y} denotes the vector of (again[7] $\alpha T_1 \log\rho + o(\log\rho)$) quantization bits corresponding to (91).

Then the $\alpha T_1 \log\rho + o(\log\rho)$ bits in XOR (W_{L_z}, W_{L_y}) are mapped into the common symbols $\{c_t\}$ that will be transmitted in the next phase, in order to eventually allow for recovering the other user's side information (see Figure 8).

As a result, for $t = T_1 + T_2 + 1, \cdots, T_1 + T_2 + T_3$, the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} c_t + a_{t,3}\rho^{-\alpha/2} \\ 0 \end{bmatrix} \tag{93}$$

where c_t is a common symbol meant for both users, where $a_{t,3}$ is meant for user 1, where

$$\begin{aligned} P^{(c_t)} \doteq 1, &\quad r^{(c_t)} = \alpha \\ P^{(a_{t,3})} \doteq 1, &\quad r^{(a_{t,3})} = 1 - \alpha \end{aligned} \tag{94}$$

and where the normalized received signals take the form

$$y_t/h_{t,1} = \underbrace{\sqrt{\rho}c_t}_{\rho} + \underbrace{\sqrt{\rho^{1-\alpha}}a_{t,3}}_{\rho^{1-\alpha}} + \underbrace{u_t/h_{t,1}}_{\rho^0} \tag{95}$$

$$z_t/g_{t,1} = \underbrace{\sqrt{\rho^{\alpha}}c_t}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{0}}a_{t,3}}_{\rho^0} + \underbrace{v_t/g_{t,1}}_{\rho^0}. \tag{96}$$

[7]With phase durations designed such that $T_2 = T_1\alpha = T_3\alpha$ (cf. (80)), the number of quantization bits in (90) and (92) match, and are both equal to $\lceil \alpha T_1 \log\rho + o(\log\rho) \rceil$.

At this point, for $t = T_1+T_2+1, \cdots, T_1+T_2+T_3$, user 1 can successively decode the common symbol c_t and the private symbol $a_{t,3}$ from y_t (cf. (95)), while user 2 can decode the common symbol c_t from z_t by treating the other signals as noise (cf. (96)).

Upon decoding $\{c_t\}_{t=T_1+T_2+1}^{T_1+T_2+T_3}$, user 1 can recover XOR (W_{L_z}, W_{L_y}), and can thus sufficiently-well recover W_{L_z} using its own side information $\{y_t\}_{t=T_1+1}^{T_1+T_2}$, thus recovering $\{\sqrt{\rho^\alpha} L_z(a_{t,1}, a_{t,2})\}_{t=1}^{T_1}$ up to noise level. This in turn allows user 1 to obtain the following 'MIMO observations' for $t = 1, 2, \cdots, T_1$

$$\begin{bmatrix} y_t \\ \sqrt{\rho^\alpha} L_z(a_{t,1}, a_{t,2}) + \tilde{\iota}_{z,t} \end{bmatrix} = \begin{bmatrix} \sqrt{\rho}\boldsymbol{h}_t^\intercal \\ \sqrt{\rho^\alpha}\boldsymbol{g}_t^\intercal \end{bmatrix} \begin{bmatrix} a_{t,1} \\ a_{t,2} \end{bmatrix} + \begin{bmatrix} u_t \\ \tilde{\iota}_{z,t} \end{bmatrix} \tag{97}$$

and to MIMO decode $a_{t,1}, a_{t,2}$ at the declared rates (cf. (82)). In the above we used $\tilde{\iota}_{z,t}$ to denote the aforementioned quantization and reconstruction noise, which - given the number of quantization bits - can be seen to have bounded power.

Similarly, upon decoding $\{c_t\}_{t=T_1+T_2+1}^{T_1+T_2+T_3}$, user 2 uses $\{z_t\}_{t=1}^{T_1}$ to recover $\{\sqrt{\rho} L_y(b_{t,1}, b_{t,2})\}_{t=T_1+1}^{T_1+T_2}$ sufficiently well, and to allow for a MIMO observation

$$\begin{bmatrix} z_t \\ \sqrt{\rho} L_y(b_{t,1}, b_{t,2}) + \tilde{\iota}_{y,t} \end{bmatrix} = \begin{bmatrix} \sqrt{\rho^\alpha}\boldsymbol{g}_t^\intercal \\ \sqrt{\rho}\boldsymbol{h}_t^\intercal \end{bmatrix} \begin{bmatrix} b_{t,1} \\ b_{t,2} \end{bmatrix} + \begin{bmatrix} v_t \\ \tilde{\iota}_{y,t} \end{bmatrix} \tag{98}$$

which results in the subsequent decoding of $b_{t,1}, b_{t,2}$ ($t = T_1+1, \cdots, T_1+T_2$) at the declared rates (86). In the above, we used $\tilde{\iota}_{y,t}$ to denote the previous quantization and reconstruction noise, which can be shown to have bounded power.

As a result, summing up the number of information bits, allows us to conclude that the proposed scheme achieves a sum GDoF

$$\begin{aligned} d_\Sigma &= \frac{T_1(1+\alpha) + T_2(1+\alpha) + T_3(1-\alpha)}{T_1+T_2+T_3} \\ &= \frac{2+\alpha+\alpha^2}{2+\alpha} \\ &= 1 + \frac{\alpha^2}{2+\alpha}. \end{aligned}$$

D. TSM scheme for $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$, achieving the optimal sum GDoF $(1+\alpha/3)$

We now transition to an alternating topology.

The scheme can be described as having three channel uses, $t = 1, 2, 3$. We will first, without loss of generality, describe the scheme for the setting where, for $t = 1, 3$, the feedback-and-topology state is $(I_1, I_2, A_1, A_2) = (D, D, 1, \alpha)$, and for $t = 2$ the state is $(I_1, I_2, A_1, A_2) = (D, D, \alpha, 1)$. The scheme can be slightly modified for the case where $(I_1, I_2, A_1, A_2) = \underbrace{(D, D, 1, \alpha)}_{t=1}, \underbrace{(D, D, \alpha, 1)}_{t=2}, \underbrace{(D, D, \alpha, 1)}_{t=3}$. In both cases, the scheme can achieve the optimal sum GDoF $(1+\alpha/3)$.

1) Phase 1: At $t = 1$ ($(I_1, I_2, A_1, A_2) = (D, D, 1, \alpha)$, link 1 is strong) the transmitter sends (see Figure 9)

$$\boldsymbol{x}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{99}$$

where a_1 and a_2 are unit-power symbols meant for user 1, where

$$r^{(a_1)} = 1, \quad r^{(a_2)} = \alpha. \tag{100}$$

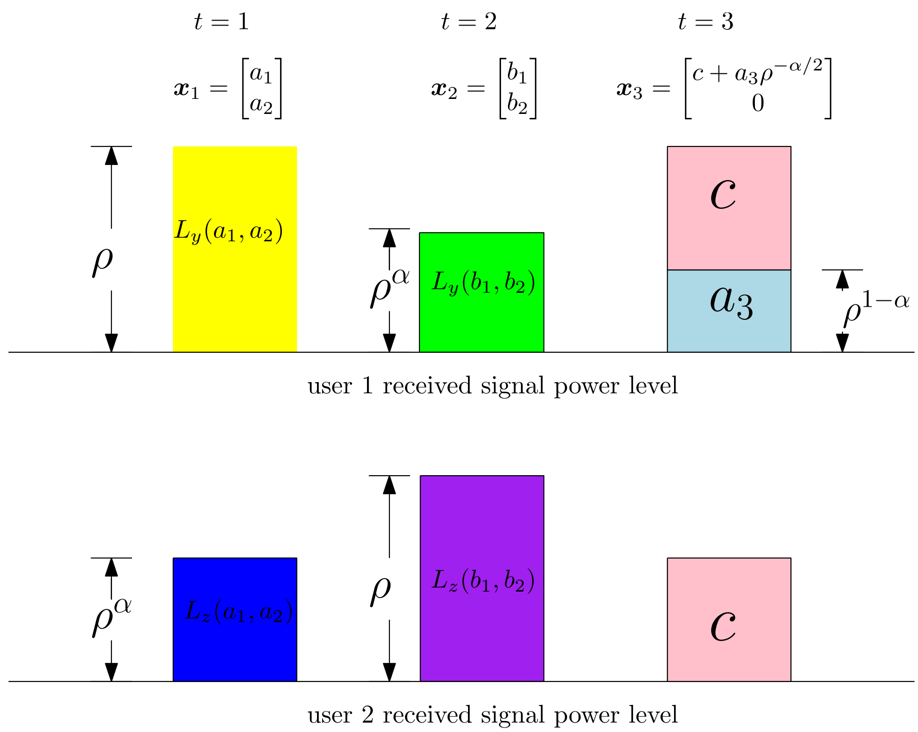

Fig. 9. Received signal power level illustration for the TSM scheme, for the setting where $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$.

resulting in received signals of the form

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^\intercal \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho} + u_1 \tag{101}$$

$$z_1 = \underbrace{\sqrt{\rho^\alpha}\boldsymbol{g}_1^\intercal \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho^\alpha} + v_1 \tag{102}$$

where we note that the unintended interfering signal is attenuated due to the weak link.

2) Phase 2: At time $t = 2$ ($(I_1, I_2, A_1, A_2) = (D, D, \alpha, 1)$, link 1 is weak) the transmitter sends

$$\boldsymbol{x}_2 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{103}$$

where b_1, b_2 are unit-power symbols meant for user 2, where

$$r^{(b_1)} = 1, \quad r^{(b_2)} = \alpha \tag{104}$$

resulting in received signals of the form

$$y_2 = \underbrace{\sqrt{\rho^\alpha}\boldsymbol{h}_2^\intercal \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{\rho^\alpha} + u_2 \tag{105}$$

$$z_2 = \underbrace{\sqrt{\rho}\boldsymbol{g}_2^\intercal \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{\rho} + v_2 \tag{106}$$

where again the unintended interfering signal is attenuated due to the weak link.

3) Phase 3: At this point the transmitter - using delayed CSIT - knows $\boldsymbol{g}_1$ and $\boldsymbol{h}_2$. It then proceeds to reconstruct $(z_1 - v_1)$ and $(y_2 - u_2)$, and to quantize the sum

$$\iota \triangleq (z_1 - v_1) + (y_2 - u_2) \tag{107}$$

using $\alpha \log \rho + o(\log \rho)$ quantization bits, in order to get the quantized version $\bar{\iota}$. Given the number of quantization bits, and given that $\mathbb{E}|\iota|^2 \doteq \rho^{\alpha}$, the quantization error

$$\tilde{\iota} = \iota - \bar{\iota}$$

is bounded and does not scale with ρ (cf. [31]). The above quantized information is then mapped into a *common* symbol c.

At time $t = 3$, with state $(I_1, I_2, A_1, A_2) = (D, D, 1, \alpha)$ (link 2 is weak), the transmitter sends

$$\boldsymbol{x}_3 = \begin{bmatrix} c + a_3\rho^{-\alpha/2} \\ 0 \end{bmatrix} \tag{108}$$

where c is the aforementioned common symbol meant for both users, where a_3 is a symbol meant for user 1, where

$$\begin{aligned} P^{(c)} &\doteq 1, & r^{(c)} &= \alpha \\ P^{(a_3)} &\doteq 1, & r^{(a_3)} &= 1 - \alpha \end{aligned} \tag{109}$$

and where the (normalized) received signals (in their noiseless form) are

$$y_3/h_{3,1} = \sqrt{\rho}c + \sqrt{\rho^{1-\alpha}}a_3 \tag{110}$$

$$z_3/g_{3,1} = \sqrt{\rho^{\alpha}}c + \sqrt{\rho^{0}}a_3. \tag{111}$$

Now we see from (110) (111) that c can be decoded by both users. Similarly we can readily see that a_3 can be decoded by user 1.

At this point, knowing c allows both users to recover $\bar{\iota}$ (cf. (107)), and to then decode the private symbols. Specifically, user 1 obtains a MIMO observation

$$\begin{bmatrix} y_1 \\ \bar{\iota} - y_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}} \\ \sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ -u_2 - \tilde{\iota} \end{bmatrix} \tag{112}$$

which allows for decoding of a_1, a_2 at the declared rates (cf. (100)). Similarly, user 2 obtains another MIMO observation

$$\begin{bmatrix} z_2 \\ \bar{\iota} - z_1 \end{bmatrix} = \begin{bmatrix} \sqrt{\rho}\boldsymbol{g}_2^{\mathsf{T}} \\ \sqrt{\rho^{\alpha}}\boldsymbol{h}_2^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} v_2 \\ -v_1 - \tilde{\iota} \end{bmatrix} \tag{113}$$

and can decode b_1, b_2 at the declared rates (cf. (104)), Summing up the information bits concludes that the scheme achieves the optimal sum GDoF $d_{\sum} = \frac{1+\alpha+1+\alpha+(1-\alpha)}{3} = 1 + \frac{\alpha}{3}$ (see Figure 9).

Remark 5: As stated above, when $(I_1, I_2, A_1, A_2) = (D, D, 1, \alpha), (D, D, \alpha, 1), (D, D, \alpha, 1)$ for $t = 1, 2, 3$ respectively, we can slightly modify the scheme such that at $t = 3$, instead of sending the private symbol a_3 for the first user (see (108)), to instead send a private symbol b_3 for the second user (i.e., again to the stronger user). Following the same steps, one can easily show that the sum GDoF $d_{\sum} = 1 + \alpha/3$ is again achievable.

Remark 6: It is interesting to note that the proposed scheme needs delayed CSIT for only a fraction of the channels (the channels with weak channel gain in phase 1 and phase 2), and in essence only needs $\lambda_{N,D}^{1,\alpha} = \lambda_{D,N}^{\alpha,1} = \lambda_{N,N}^{1,\alpha} = 1/3$, or $\lambda_{N,D}^{1,\alpha} = \lambda_{D,N}^{\alpha,1} = \lambda_{N,N}^{\alpha,1} = 1/3$, or $\lambda_{N,D}^{1,\alpha} = \lambda_{D,N}^{\alpha,1} = \frac{1}{2}\lambda_{N,N}^{1,\alpha} = \frac{1}{2}\lambda_{N,N}^{\alpha,1} = 1/3$, to achieve the same optimal sum GDoF.

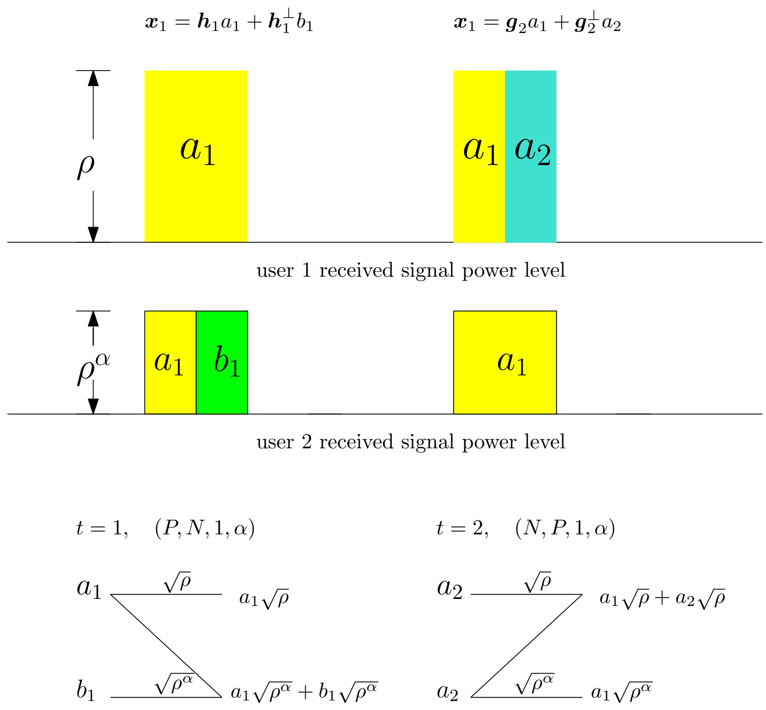

Fig. 10. Illustration of TSM coding and of received signal power levels, for $\lambda^{1,\alpha}_{P,N} = \lambda^{1,\alpha}_{N,P} = 1/2$.

E. TSM schemes for $\lambda_{P,N} = \lambda_{N,P} = 1/2$ and for any $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$; achieving the optimal sum GDoF $1 + \frac{\alpha}{2}$

We will now show that the optimal sum GDoF $(1 + \frac{\alpha}{2})$ is achievable for any topology $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$ using $\lambda_{P,N} = \lambda_{N,P} = 1/2$ and a sequence of TSM schemes proposed for the different settings of

$$\lambda^{1,\alpha}_{P,N} = \lambda^{1,\alpha}_{N,P} = 1/2; \quad \lambda^{\alpha,1}_{P,N} = \lambda^{\alpha,1}_{N,P} = 1/2; \quad \lambda^{1,\alpha}_{P,N} = \lambda^{\alpha,1}_{N,P} = 1/2; \quad \lambda^{\alpha,1}_{P,N} = \lambda^{1,\alpha}_{N,P} = 1/2$$

respectively. Each scheme achieves the optimal sum GDoF $(1 + \frac{\alpha}{2})$, and each scheme is designed with only two channel uses during which the two users take turn to feed back current CSIT (only one user feeds back at a time). The general result is proven by properly concatenating the proposed schemes for the different cases.

1) TSM scheme for $\lambda^{1,\alpha}_{P,N} = \lambda^{1,\alpha}_{N,P} = 1/2$: Without loss of generality, we focus on the specific sub-case where $(I_1, I_2, A_1, A_2) = (P, N, 1, \alpha)$ for $t = 1$, and $(I_1, I_2, A_1, A_2) = (N, P, 1, \alpha)$ for $t = 2$.

At $t = 1$ the transmitter knows $\boldsymbol{h}_1$ (current CSIT), and sends (see Figure 10)

$$\boldsymbol{x}_1 = \boldsymbol{h}_1 a_1 + \boldsymbol{h}_1^{\perp} b_1 \tag{114}$$

where a_1 and b_1 are intended for user 1 and user 2 respectively, and where

$$\begin{aligned} P^{(a_1)} \doteq 1, \quad r^{(a_1)} = 1 \\ P^{(b_1)} \doteq 1, \quad r^{(b_1)} = \alpha. \end{aligned} \tag{115}$$

Then the received signals (in their noiseless form) are

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho} \tag{116}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1^{\perp} b_1}_{\rho^{\alpha}}. \tag{117}$$

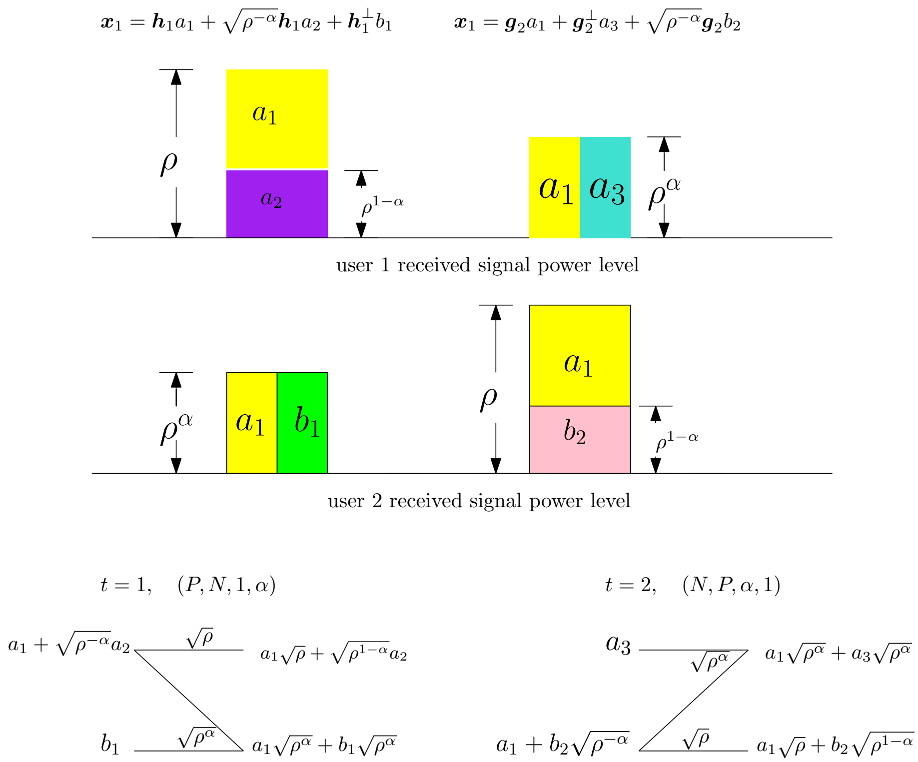

Fig. 11. Illustration of coding and received signal power levels for $\lambda_{P,N}^{1,\alpha} = \lambda_{N,P}^{\alpha,1} = 1/2$.

At $t = 2$ ($(I_1, I_2, A_1, A_2) = (N, P, 1, \alpha)$), the transmitter knows $\boldsymbol{g}_2$ (current CSIT) and sends

$$\boldsymbol{x}_2 = \boldsymbol{g}_2 a_1 + \boldsymbol{g}_2^{\perp} a_2 \tag{118}$$

where a_2 is intended for user 1, and where

$$P^{(a_2)} \doteq 1, \quad r^{(a_2)} = 1. \tag{119}$$

Then the received signals (in their noiseless form) are as follows

$$y_2 = \underbrace{\sqrt{\rho}\boldsymbol{h}_2^{\mathsf{T}}\boldsymbol{g}_2 a_1}_{\rho} + \underbrace{\sqrt{\rho}\boldsymbol{h}_2^{\mathsf{T}}\boldsymbol{g}_2^{\perp} a_2}_{\rho} \tag{120}$$

$$z_2 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_2^{\mathsf{T}}\boldsymbol{g}_2 a_1}_{\rho^{\alpha}}. \tag{121}$$

At this point, we can see that user 1 can MIMO decode a_1, a_2 based on (116), (120), while user 2 recover b_1, b_2 by employing interference cancelation based on (117), (121). This gives a sum DoF of $1 + \alpha/2$.

Remark 7: We can now easily see that for the setting where $(I_1, I_2, A_1, A_2) = \overbrace{(N, P, 1, \alpha)}^{t=1}, \overbrace{(P, N, 1, \alpha)}^{t=2}$, we can easily modify the above scheme to achieve the same performance, just by reordering the transmissions such that $\boldsymbol{x}_1 = \boldsymbol{g}_1 a_1 + \boldsymbol{g}_1^{\perp} a_2$ and $\boldsymbol{x}_2 = \boldsymbol{h}_2 a_1 + \boldsymbol{h}_2^{\perp} b_1$.

Similarly when $\lambda_{P,N}^{\alpha,1} = \lambda_{N,P}^{\alpha,1} = 1/2$, we can simply take the above scheme (of Section VIII-E1), and simply interchange the roles of the users, to again achieve the optimal sum GDoF $1 + \alpha/2$.

2) TSM scheme for $\lambda_{P,N}^{1,\alpha} = \lambda_{N,P}^{\alpha,1} = 1/2$: We focus on the case where we first have $(I_1, I_2, A_1, A_2) = (P, N, 1, \alpha)$ (at $t = 1$), followed by $(I_1, I_2, A_1, A_2) = (N, P, \alpha, 1)$ ($t = 2$).

At $t = 1$, the transmitter knows $\boldsymbol{h}_1$, and sends (see Figure 11)

$$\boldsymbol{x}_1 = \boldsymbol{h}_1 a_1 + \sqrt{\rho^{-\alpha}}\boldsymbol{h}_1 a_2 + \boldsymbol{h}_1^{\perp} b_1 \tag{122}$$

where a_1, a_2 are the unit-power symbols intended for user 1, b_1 is the unit-power symbol intended for user 2, where

$$r^{(a_1)} = \alpha, \quad r^{(a_2)} = 1 - \alpha, \quad r^{(b_1)} = \alpha \tag{123}$$

and where the received signals, in their noiseless form, are

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\intercal}\boldsymbol{h}_1 a_1}_{\rho} + \underbrace{\sqrt{\rho^{1-\alpha}}\boldsymbol{h}_1^{\intercal}\boldsymbol{h}_1 a_2}_{\rho^{1-\alpha}} \tag{124}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\intercal}\boldsymbol{h}_1 a_1}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{0}}\boldsymbol{g}_1^{\intercal}\boldsymbol{h}_1 a_2}_{\rho^{0}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\intercal}\boldsymbol{h}_1^{\perp} b_1}_{\rho^{\alpha}}. \tag{125}$$

At $t = 2$ ($(I_1, I_2, A_1, A_2) = (N, P, \alpha, 1)$) the transmitter knows $\boldsymbol{g}_2$ (user 1 is weak), and sends

$$\boldsymbol{x}_2 = \boldsymbol{g}_2 a_1 + \boldsymbol{g}_2^{\perp} a_3 + \sqrt{\rho^{-\alpha}}\boldsymbol{g}_2 b_2 \tag{126}$$

where a_3, b_2 are the unit-power symbols intended for user 1 and user 2 respectively, where

$$r^{(a_3)} = \alpha, \quad r^{(b_2)} = 1 - \alpha \tag{127}$$

and where the received signals, in their noiseless form, are

$$y_2 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2 a_1}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2^{\perp} a_3}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{0}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2^{\perp} b_2}_{\rho^{0}} \tag{128}$$

$$z_2 = \underbrace{\sqrt{\rho}\boldsymbol{g}_2^{\intercal}\boldsymbol{g}_2 a_1}_{\rho} + \underbrace{\sqrt{\rho^{1-\alpha}}\boldsymbol{g}_2^{\intercal}\boldsymbol{g}_2 b_2}_{\rho^{1-\alpha}}. \tag{129}$$

At this point, it is easy to see that user 1 can recover a_1, a_2, a_3 by MIMO decoding based on (124) and (128), while user 2 can recover b_1 by employing interference cancelation based on (125) and (129) (see also Figure 11). This provides for $d_{\sum} = 1 + \alpha/2$.

a) Modifying the scheme for the setting where (I_1, I_2, A_1, A_2) is $(N, P, \alpha, 1)$ or $(P, N, 1, \alpha)$: Similarly for the setting where (I_1, I_2, A_1, A_2) is $(N, P, \alpha, 1)$ or $(P, N, 1, \alpha)$, we can modify the previous scheme — to achieve the same optimal sum DoF — by interchanging the transmissions of the first and second channel uses, i.e., of $t = 1, 2$.

b) Modifying the scheme for the setting where $\lambda_{P,N}^{\alpha,1} = \lambda_{N,P}^{1,\alpha} = 1/2$: Furthermore when $\lambda_{P,N}^{\alpha,1} = \lambda_{N,P}^{1,\alpha} = 1/2$, we can simply interchange the roles of users in the previous scheme, to again achieve the same optimal sum GDoF.

c) Spanning the entire setting $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$, $\lambda_{P,N} = \lambda_{N,P}$: Finally, by using $\lambda_{P,N} = \lambda_{N,P}$ and by properly concatenating the above scheme variants, gives the optimal performance $d_{\sum} = 1 + \alpha/2$, for the entire range $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$.

F. Original MAT scheme in the fixed topological setting ($\lambda_{1,\alpha} = 1$)

We recall that the original MAT scheme in [9] consists of three phases, each of duration one. At time $t = 1, 2$, the transmitter sends

$$\boldsymbol{x}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \qquad \boldsymbol{x}_2 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

where a_1, a_2 are for user 1, b_1, b_2 for user 2, and where the received signals, in their noiseless form, are

$$y_1 = \sqrt{\rho}\boldsymbol{h}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad z_1 = \sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\intercal} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \triangleq \sqrt{\rho^{\alpha}} L_z(a_1, a_2) \tag{130}$$

$$y_2 = \sqrt{\rho}\boldsymbol{h}_2^{\intercal} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \triangleq \sqrt{\rho} L_y(b_1, b_2) \qquad z_2 = \sqrt{\rho^{\alpha}}\boldsymbol{g}_2^{\intercal} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{131}$$

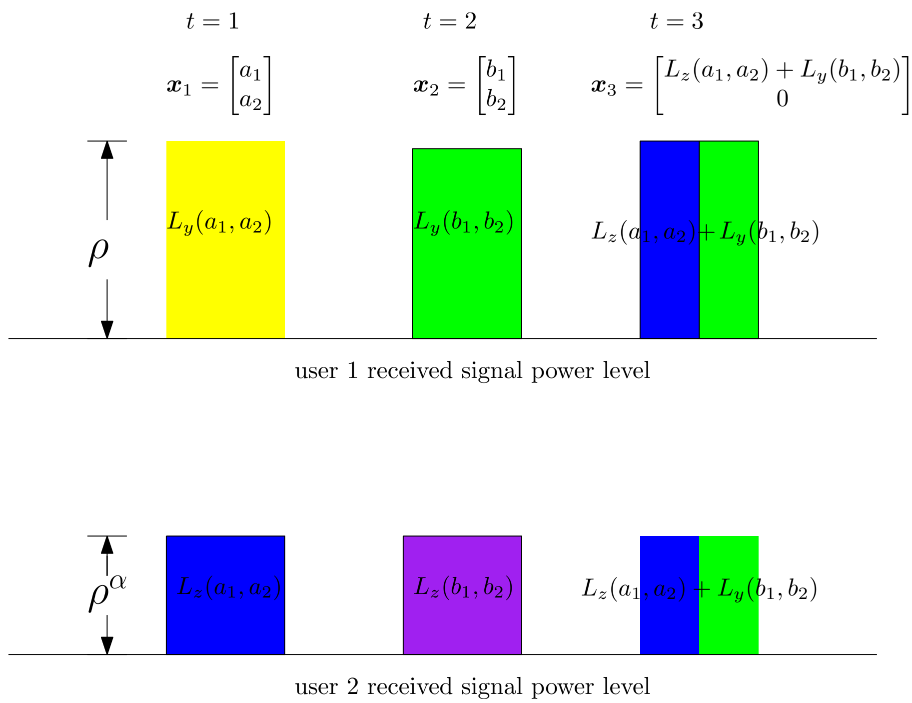

Fig. 12. Illustration of received power level for the original MAT scheme in the fixed topology setting $\lambda_{1,\alpha}=1$.

At $t=3$, the transmitter knows $\boldsymbol{g}_1$ and $\boldsymbol{h}_2$ (delayed CSIT), reconstructs $L_z(a_1,a_2), L_y(b_1,b_2)$ (cf. (130), (131)), and sends

$$\boldsymbol{x}_3 = \begin{bmatrix} L_z(a_1,a_2)+L_y(b_1,b_2) \\ 0 \end{bmatrix}$$

with normalized/processed received signals which, in their noiseless form, are

$$y_3/h_{3,1} = \sqrt{\rho}L_z(a_1,a_2)+\sqrt{\rho}L_y(b_1,b_2) \tag{132}$$

$$z_3/g_{3,1} = \sqrt{\rho^\alpha}L_z(a_1,a_2)+\sqrt{\rho^\alpha}L_y(b_1,b_2). \tag{133}$$

At this point, we recall that user 1 combines the above with y_1, y_2, y_3, to design a MIMO system

$$\begin{bmatrix} y_1 \\ y_3/h_{3,1}-y_2 \end{bmatrix} = \sqrt{\rho}\begin{bmatrix} \boldsymbol{h}_1^\intercal \\ \boldsymbol{g}_1^\intercal \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_3/h_{3,1}-u_2 \end{bmatrix} \tag{134}$$

and to MIMO decode a_1, a_2, which carry a total of $[2\log\rho + o(\log\rho)]$ bits. Similarly, user 2 is presented with another MIMO system

$$\begin{bmatrix} z_2 \\ z_3/g_{3,1}-z_1 \end{bmatrix} = \sqrt{\rho^\alpha}\begin{bmatrix} \boldsymbol{g}_2^\intercal \\ \boldsymbol{h}_2^\intercal \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} v_2 \\ v_3/g_{3,1}-v_1 \end{bmatrix} \tag{135}$$

of less power, from which it can MIMO decode b_1, b_2, which though now carry a total of $2\alpha\log\rho + o(\log\rho)$ bits. As a result, the original MAT scheme achieves a sum GDoF $d_{\sum} = \frac{2(1+\alpha)}{3}$.

REFERENCES

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691 – 1706, Jul. 2003.
[2] S. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848 – 857, Mar. 2005.
[3] 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.
[4] A. Lapidoth, S. Shamai, and M. A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proc. Allerton Conf. Communication, Control and Computing*, Sep. 2005.

[5] C. Vaze and M. Varanasi, “The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT,” *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5254 – 5374, Aug. 2012.
[6] H. Weingarten, S. Shamai, and G. Kramer, “On the compound MIMO broadcast channel,” in *Proc. Inf. Theory and App. Workshop (ITA)*, Jan. 2007.
[7] T. Gou, S. A. Jafar, and C. Wang, “On the degrees of freedom of finite state compound wireless networks,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3286 – 3308, Jun. 2011.
[8] M. A. Maddah-Ali, “On the degrees of freedom of the compound MIMO broadcast channels with finite states,” Oct. 2009, available on arXiv:0909.5006v3.
[9] M. A. Maddah-Ali and D. N. C. Tse, “Completely stale transmitter channel state information is still very useful,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418 – 4431, Jul. 2012.
[10] S. A. Jafar, “Blind interference alignment,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 216 – 227, Jun. 2012.
[11] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315 – 328, Jan. 2013.
[12] T. Gou and S. Jafar, “Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT,” *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084 – 1087, Jul. 2012.
[13] J. Chen and P. Elia, “Degrees-of-freedom region of the MISO broadcast channel with general mixed-CSIT,” in *Proc. Inf. Theory and App. Workshop (ITA)*, Feb. 2013, (also available on arXiv:1205.3474, May 2012).
[14] ——, “Toward the performance vs. feedback tradeoff for the two-user MISO broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8336 – 8356, Dec. 2013.
[15] R. Tandon, S. A. Jafar, S. Shamai, and H. V. Poor, “On the synergistic benefits of alternating CSIT for the MISO broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106 – 4128, Jul. 2013.
[16] 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 (ISIT)*, Jul. 2011.
[17] J. Xu, J. G. Andrews, and S. A. Jafar, “Broadcast channels with delayed finite-rate feedback: Predict or observe?” *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1456 – 1467, Apr. 2012.
[18] Y. Lejosne, D. Slock, and Y. Yuan-Wu, “Degrees of freedom in the MISO BC with delayed-CSIT and finite coherence time: A simple optimal scheme,” in *Proc. IEEE Int. Conf. on Signal Processing, Communications and Control (ICSPCC)*, Aug. 2012.
[19] J. Chen and P. Elia, “MISO broadcast channel with delayed and evolving CSIT,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013.
[20] ——, “MIMO BC with imperfect and delayed channel state information at the transmitter and receivers,” in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2013.
[21] N. Lee and R. W. Heath Jr., “Not too delayed CSIT achieves the optimal degrees of freedom,” in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2012.
[22] C. Hao and B. Clerckx, “Imperfect and unmatched CSIT is still useful for the frequency correlated MISO broadcast channel,” in *Proc. IEEE Int. Conf. Communications (ICC)*, Jun. 2013.
[23] J. Chen, S. Yang, and P. Elia, “On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013.
[24] J. Chen and P. Elia, “Symmetric two-user MIMO BC and IC with evolving feedback,” Jun. 2013, available on arXiv:1306.3710.
[25] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, “The degrees of freedom region of temporally-correlated MIMO networks with delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 494 – 514, Jan. 2014.
[26] C. S. Vaze and M. K. Varanasi, “The degrees of freedom region of two-user and certain three-user MIMO broadcast channel with delayed CSI,” Dec. 2011, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1101.0306.
[27] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Capacity results for binary fading interference channels with delayed CSIT,” Jan. 2013, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1301.5309.
[28] G. Caire, N. Jindal, and S. Shamai, “On the required accuracy of transmitter channel state information in multiple antenna broadcast channels,” in *Proc. Allerton Conf. Communication, Control and Computing*, Nov. 2007.
[29] S. A. Jafar, “Topological interference management through index coding,” *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529 – 568, Jan. 2014.
[30] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
[31] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.
[32] A. E. Gamal, “The feedback capacity of degraded broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 379–381, Apr. 1978.