On the Two-User MISO 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. From an information theoretic point of view, accounting for such topological aspects is still a novel approach, that has been recently fueled by strong indications that such aspects can crucially affect transceiver and feedback design, as well as the overall performance.

The work here takes a 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

The Gaussian multiple-input single-output broadcast channel (MISO BC) 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]–[30].

As highlighted recently in [31], 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 [32], which lead to statistically unequal channel gains, with some links being much weaker or stronger than others (See Figures 1, 2). Accounting for these topological aspects, by going beyond the DoF framework into the *generalized* degrees of freedom (GDoF) framework (cf. [33]–[39]), 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

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

Jinyuan Chen is with Stanford University, Department of Electrical Engineering, CA 94305, USA (email: jinyuanc@stanford.edu). Petros Elia is with EURECOM, Sophia Antipolis, 06410, France (email: elia@eurecom.fr). Syed Ali Jafar is with University of California, Irvine, CA 92697, USA (email: syed@uci.edu). The work of Jinyuan Chen was supported by the NSF Center for Science of Information: NSF-CCF-0939370. The work of Petros Elia was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) / grant agreement no.318306 (NEWCOM#), from the FP7 CELTIC SPECTRA project, and from the ANR project IMAGENET. The work of Syed Ali Jafar was supported by NSF grants CCF 1319104 and CCF 1317351.

A smaller version of this paper was presented at ISIT 2014.

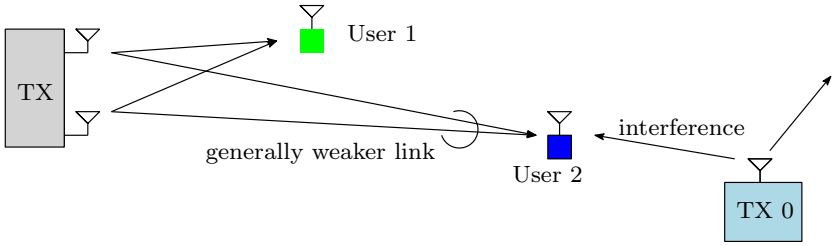

Fig. 1. Topology where link 2 is weaker due to distance and interference.

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

where ρ is defined by a power constraint, $\boldsymbol{x}_t$ is the normalized transmitted vector at time t — normalized here to satisfy $||\boldsymbol{x}_t||^2 \leq 1$ — $\boldsymbol{h}_t^{'}, \boldsymbol{g}_t^{'}$ represent the vector fading channels to the first and second receiver respectively, and $u_t^{'}, v_t^{'}$ 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 simplified model, the difference in link strengths (in a statistical sense) reflects the differences due to the propagation setting or due to inter-cell interference. While more motivation for this simplified multiplicative model will be given later on in the context of generalized degrees-of-freedom, we hasten to note that the multiplicative dependency of received power to input power, is meant to capture the possibility of a substantial difference in the high-SNR capacities of any two links.

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

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

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: 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 characterize 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 this work we focus on the generalized degrees-of-freedom (GDoF) performance of the system. This approach goes back to Etkin and Tse in [33] which studied the Gaussian interference channel (IC), and which was followed by many GDoF related works such as that by Mohapatra and Murthy in [40] which analyzed the GDoF of the K-user symmetric IC, as well as the work by Karmakar and Varanasi in [37] which analyzed the GDoF of the

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

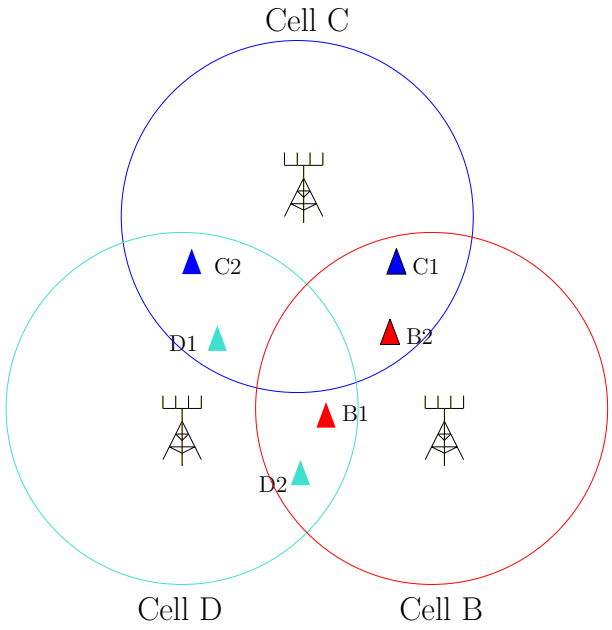

Fig. 2. Cell edge users experience fluctuating interference due to changing frequency allocation in the multi-cell system.

multiple-input multiple-output (MIMO) IC. Combining topology and feedback considerations, Vaze et al. in [34] employed the GDoF measure in the MIMO IC setting without CSIT under statistically weak interference links, while Karmakar and Varanasi in [36] analyzed the GDoF of the MIMO IC with limited feedback. Further interesting works include the work by Gherekhloo et al. in [38] which considered interference management issues in the presence of an alternating connectivity ($\alpha = 0$).

In this setting, for an achievable rate pair (R_1, R_2) for the first and second user respectively, the corresponding GDoF pair (d_1, d_2) is given by

$$d_k = \lim_{\rho\to\infty} \frac{R_k}{\log \rho}, \quad k = 1, 2. \tag{7}$$

The corresponding GDoF region $\mathcal{D}$ is then the set of all achievable DoF pairs (d_1, d_2), and the sum GDoF is

$$d_\Sigma = \sup\{d_1 + d_2 : (d_1, d_2) \in \mathcal{D}\}. \tag{8}$$

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 $\alpha = 1$ removes topology considerations, while setting $\alpha = 0$ almost entirely removes the weak link, as its capacity does not scale with SNR. Needless to say that setting the stronger link to correspond to a unit link-power exponent, is a result of normalization, and thus imposes no loss in generality.

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 than the second throughout the communication process ($A_{1,t} = 1, A_{2,t} = \alpha,\ \forall t$), 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 Section IV-G), can show that in the same static topology $A_{1,t} = 1, A_{2,t} = \alpha, \forall t$, the original Maddah-Ali and Tse (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) Motivation of the GDoF setting: Often, taking a strict interpretation of the limiting nature of GDoF, leads to confusion because, strictly speaking, any reasonable channel model would force a limiting α to be 1, since all powers would go to infinity the same way. Towards convincing the skeptical reader of the usefulness of our approach, we offer the following thoughts which can help clarify any misconceptions.

Our GDoF approach here is based on two crucial premises.

i) Network links generally have different capacities, and in the perfectly conceivable case where a link has a capacity that is a fraction α of another link's capacity, a good approximation is that the weaker link has average power that is close to the α^{th}

[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.

power of the aforementioned power of the strong link.

ii) Even though, strictly speaking, GDoF results are by definition associated to the infinite SNR limit (cf. (8)) where the *limiting* behavior of random variables allows for more analytical tractability, it is crucial to note that this tractable interpretation applies and offers insight in operational *moderate-to-large* SNR regimes. The crucial element that binds infinite-SNR mathematical analysis to engineering insight over operational SNR values, can be found in the above observation regarding the ratios of link capacities. This says that our analysis would apply in a broadcast channel setting, where the two links independently have sufficiently high capacity — which would in turn imply a moderate-to-large SNR regime — and where the ratio of these capacities is close to a certain value α. Once this α is picked and fixed, the derived high-SNR approximations will yield (capacity) expressions which, as SNR increases, are expected to offer an increasingly faithful representation of the actual behavior of the system, i.e., are expected to offer an increasingly better qualitative estimate of the overall system behavior. Avoiding a strict and literal interpretation of asymptotics, while still mathematically rigorous, the GDoF approach allows for consideration of topological settings that are motivated by reasonable scenarios that include distance variations and interference fluctuations. In other words, while the mathematics use scaling laws and limits as tools for tractability of randomness, the GDoF approach does not require the actual real-life nature of the problem to scale with SNR, as this would related to awkward scenarios where variable geometries have distances that scale in different specific ways.

With the above premises in mind, one can now better appreciate the utility of the simple multiplicative model in (5) which — employing a multiplicative dependency of the received power to the input power — manages to concisely capture substantial differences in the high-SNR capacities of any two links, and thus fits well with the GDoF setting. While other, more refined models could certainly be conceived that could potentially better map the intricacies of what causes topological diversity in networks, we have yet to see such models that allow for analysis that offers insight. Additionally, we believe that such complex and involved models would be more susceptible to losing some of their refinement in the high SNR regime of GDoF asymptotics. No such loss of model information is suffered — in the transition to the asymptotic setting — by the chosen multiplicative model, exactly because of this model's inherent simplicity and its direct association to the GDoF measure.

3) Feedback and topology statistics: Naturally performance is a function of the feedback and topology statistics. In terms of feedback 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 $(I_1, I_2) \in (P, D, N) \times (P, D, N)$.

We similarly 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 $\lambda_{1,\alpha} + \lambda_{\alpha,1} + \lambda_{1,1} + \lambda_{\alpha,\alpha} = 1$. Finally 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 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 a *static (or fixed) topology* where the first link is stronger than the second throughout the communication process, $\lambda_{1,1} = \lambda_{\alpha,\alpha} = 1/2$ implies a topology where half of the time both links are strong and then both are weak, while $\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$ implies an *alternating topology* where half of the time, the first user is statistically stronger, and vice versa.

Finally having $\lambda_{P,D}^{1,\alpha} + \lambda_{D,P}^{\alpha,1} = 1$ does not impose any restriction on the topology statistics, but it implies a feedback 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, $(\bullet)^{\mathsf{T}}$, $(\bullet)^{\mathsf{H}}$, $(\bullet)^{-1}$, and $\mathrm{tr}(\bullet)$ denote the transpose, conjugate transpose, inverse and the trace of a matrix respectively, while $(\bullet)^*$ denotes the complex conjugate, $||\bullet||$ 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 $\lim_{\rho\to\infty} \log f(\rho)/\log\rho = B$. Similarly $\dot{\geq}$ and $\dot{\leq}$ denote exponential inequalities. $\boldsymbol{e}^{\perp}$ denotes a unit-norm vector orthogonal to vector $\boldsymbol{e}$. We define that $(\bullet)^+ = \max\{\bullet, 0\}$. Throughout this work, we adhere to the common convention and assume perfect and global knowledge of channel state information at the receivers (perfect and global CSIR). We also make the soft assumption that the transmitter is aware of the feedback statistics and the topology statistics. Furthermore, for some cases, we will consider the broad 'symmetric' alternating CSIT setting, corresponding to the symmetry assumption that

$$\text{i.e.,}\ \lambda_{P,N} = \lambda_{N,P},\ \lambda_{D,N} = \lambda_{N,D},\ \lambda_{P,D} = \lambda_{D,P}.$$

For this *symmetric* CSIT setting we will often use the following notations

$$\lambda_P \triangleq \sum_{(I_1,I_2):I_1=P} \lambda_{I_1,I_2} \triangleq \sum_{(I_1,I_2):I_2=P} \lambda_{I_1,I_2}, \qquad \lambda_D \triangleq \sum_{(I_1,I_2):I_1=D} \lambda_{I_1,I_2} \triangleq \sum_{(I_1,I_2):I_2=D} \lambda_{I_1,I_2}, \qquad \lambda_N \triangleq \sum_{(I_1,I_2):I_1=N} \lambda_{I_1,I_2} \triangleq \sum_{(I_1,I_2):I_2=N} \lambda_{I_1,I_2}.$$

In terms of the feedback statistics, we will here adopt a commonly used soft assumption that the long term feedback statistics defining λ_{I_1,I_2}, $(I_1,I_2) \in (P,D,N) \times (P,D,N)$, still hold for reasonably large but finite durations. While there are some specific cases of non-homogeneous feedback statistics for which this assumption does not hold, the assumption in general can be achieved, up to a certain point, by interchanging of the time index, as well as fits well to feedback mechanisms that are periodic in time.

In Section III we present the GDoF bounds for the topological BC with alternating CSIT. Specifically in Section III-A we present the general GDoF outer bounds, in Section III-B we present a unified GDoF inner bound for the BC with symmetrically alternating CSIT and a static topology, while in Section III-C we present the optimal sum GDoF for different practical CSIT schemes, for general *fluctuating* (non-static) topology settings. In Section IV we present a general topological signal management scheme for the entire spectrum of static topologies and alternating CSIT settings (this serves as a proof for Theorem 1), as well as provide two illustrative examples, where the general scheme is distilled down to specific simpler instances that can help the reader better understand the idea behind these schemes. Then in Section V we describe sum-GDoF optimal schemes for the fluctuating topology setting. In Section VI we offer some conclusions, while in the appendix of Section VII we have the proof of the general outer bound of Lemma 2.

We proceed with the main results, starting with the GDoF region outer bounds, and then proceeding with achievable and often optimal GDoF expressions for pertinent cases of practical significance.

III. GDoF bounds for the topological BC with alternating CSIT

A. GDoF outer bounds for the topological BC with alternating CSIT

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

Lemma 1: For the two-user MISO BC with alternating CSIT and a static topology ($\lambda_{1,\alpha}=1$), the GDoF region is outer bounded as

$$\begin{aligned} d_1 &\le 1, \qquad d_2 \le \alpha, \\ d_1+\frac{d_2}{2} &\le 1+\sum_{(I_1,I_2):I_1=P}\frac{\alpha}{2}\lambda_{I_1,I_2}, \\ d_2+\frac{d_1}{2} &\le \alpha+\sum_{(I_1,I_2):I_2=P}\frac{1}{2}\lambda_{I_1,I_2}+\sum_{(I_1,I_2):I_2\neq P}\frac{1-\alpha}{2}\lambda_{I_1,I_2}, \\ d_1+d_2 &\le d_\Sigma^{(2)}, \end{aligned}$$

and the sum GDoF is upper bounded as $d_\Sigma \le \min\{d_\Sigma^{(1)}, d_\Sigma^{(2)}\}$, 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 $\lambda_{I_1,I_2}^{A_1,A_2}$. For conciseness we use

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

so for 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, the CSIT for the channel of *any one* of the users, is being fed back in a perfect and instantaneous manner, while the CSIT for the channel of the other user, is fed back later in a delayed manner.

Lemma 2: For the topological two-user MISO BC with alternating CSIT, the GDoF region is outer bounded as

$$d_1 \leq \sum_{\forall(A_1,A_2)} A_1\lambda_{A_1,A_2}, \tag{9}$$

$$d_2 \leq \sum_{\forall(A_1,A_2)} A_2\lambda_{A_1,A_2}, \tag{10}$$

$$d_1+\frac{d_2}{2} \leq \Big(\sum_{\forall(I_1,I_2)}\ \sum_{\forall(A_1,A_2)} A_1\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\Big(\sum_{(I_1,I_2):I_1=P}\ \sum_{\forall(A_1,A_2)} \frac{A_2}{2}\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\Big(\sum_{(I_1,I_2):I_1\neq P} \frac{1-\alpha}{2}\lambda_{I_1,I_2}^{\alpha,1}\Big), \tag{11}$$

$$d_2+\frac{d_1}{2} \leq \Big(\sum_{\forall(I_1,I_2)}\ \sum_{\forall(A_1,A_2)} A_2\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\Big(\sum_{(I_1,I_2):I_2=P}\ \sum_{\forall(A_1,A_2)} \frac{A_1}{2}\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\Big(\sum_{(I_1,I_2):I_2\neq P} \frac{1-\alpha}{2}\lambda_{I_1,I_2}^{1,\alpha}\Big), \tag{12}$$

$$d_1+d_2 \leq d_\Sigma^{(4)}, \tag{13}$$

and the sum GDoF is upper bounded as $d_\Sigma \leq \min\{d_\Sigma^{(3)}, d_\Sigma^{(4)}\}$, where

$$\begin{aligned} d_\Sigma^{(3)} \triangleq&(1+\alpha)(\lambda_{P,P}^{\alpha,1}+\lambda_{P,P}^{1,\alpha})+\frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow D}^{\alpha,1}+\lambda_{P\leftrightarrow D}^{1,\alpha})+\frac{3+2\alpha}{3}(\lambda_{P\leftrightarrow N}^{\alpha,1}+\lambda_{P\leftrightarrow N}^{1,\alpha})\\ &+\frac{3+\alpha}{3}(\lambda_{D,D}^{\alpha,1}+\lambda_{D,D}^{1,\alpha})+\frac{3+\alpha}{3}(\lambda_{D\leftrightarrow N}^{\alpha,1}+\lambda_{D\leftrightarrow N}^{1,\alpha})+\frac{3+\alpha}{3}(\lambda_{N,N}^{\alpha,1}+\lambda_{N,N}^{1,\alpha})\\ &+2\lambda_{P,P}^{1,1}+\frac{5}{3}\lambda_{P\leftrightarrow D}^{1,1}+\frac{5}{3}\lambda_{P\leftrightarrow N}^{1,1}+\frac{4}{3}\lambda_{D,D}^{1,1}+\frac{4}{3}\lambda_{D\leftrightarrow N}^{1,1}+\frac{4}{3}\lambda_{N,N}^{1,1}\\ &+2\alpha\lambda_{P,P}^{\alpha,\alpha}+\frac{5\alpha}{3}\lambda_{P\leftrightarrow D}^{\alpha,\alpha}+\frac{5\alpha}{3}\lambda_{P\leftrightarrow N}^{\alpha,\alpha}+\frac{4\alpha}{3}\lambda_{D,D}^{\alpha,\alpha}+\frac{4\alpha}{3}\lambda_{D\leftrightarrow N}^{\alpha,\alpha}+\frac{4\alpha}{3}\lambda_{N,N}^{\alpha,\alpha} \end{aligned} \tag{14}$$

$$\begin{aligned} d_\Sigma^{(4)} \triangleq&(1+\alpha)(\lambda_{P,P}^{1,\alpha}+\lambda_{P,P}^{\alpha,1})+(1+\alpha)(\lambda_{P\leftrightarrow D}^{1,\alpha}+\lambda_{P\leftrightarrow D}^{\alpha,1})+(1+\alpha)(\lambda_{D,D}^{1,\alpha}+\lambda_{D,D}^{\alpha,1})\\ &+\frac{2+\alpha}{2}(\lambda_{P\leftrightarrow N}^{1,\alpha}+\lambda_{P\leftrightarrow N}^{\alpha,1})+\frac{2+\alpha}{2}(\lambda_{D\leftrightarrow N}^{1,\alpha}+\lambda_{D\leftrightarrow N}^{\alpha,1})+\lambda_{N,N}^{1,\alpha}+\lambda_{N,N}^{\alpha,1}\\ &+2\lambda_{P,P}^{1,1}+2\alpha\lambda_{P,P}^{\alpha,\alpha}+2\lambda_{P\leftrightarrow D}^{1,1}+2\alpha\lambda_{P\leftrightarrow D}^{\alpha,\alpha}+2\lambda_{D,D}^{1,1}+2\alpha\lambda_{D,D}^{\alpha,\alpha}\\ &+\frac{3}{2}\lambda_{P\leftrightarrow N}^{1,1}+\frac{3\alpha}{2}\lambda_{P\leftrightarrow N}^{\alpha,\alpha}+\frac{3}{2}\lambda_{D\leftrightarrow N}^{1,1}+\frac{3\alpha}{2}\lambda_{D\leftrightarrow N}^{\alpha,\alpha}+\lambda_{N,N}^{1,1}+\alpha\lambda_{N,N}^{\alpha,\alpha}. \end{aligned} \tag{15}$$

Note that bound $d_\Sigma^{(3)}$ results from the combination of bound (11) and bound (12).

The above bounds will be used to establish, particularly in the fluctuating topology setting, the optimality of different encoding schemes and practical feedback mechanisms.

Remark 2: The derived outer bound here expands on the classical compound BC techniques, to account for uneven link strengths. The original idea of the compound BC technique is that, two statistically equivalent observations may allow for approximate reconstruction of another observation (assuming two transmit-antennas). However, in this setting, two statistically equivalent observations may not allow approximate reconstruction of another observation, due to the uneven nature of the links. Towards this, we introduced a different auxiliary random variable structure such that, together with the two statistically equivalent observations — that are common in these type of bounds — can allow for approximate reconstruction of another observation.

B. Unified GDoF inner bound for the BC with symmetrically alternating CSIT and a static topology

We first proceed to bound the GDoF region for the entire *symmetric* alternating CSIT setting with a *static* topology ($\lambda_{1,\alpha}=1$, or $\lambda_{\alpha,1}=1,\ \alpha\in[0,1]$).

Theorem 1: The GDoF region of the two-user MISO BC with *symmetric* alternating CSIT and a *static* topology ($\lambda_{1,\alpha}=1$) is inner bounded by the region described as

$$\begin{aligned} d_1\leq 1,\quad d_2\leq\alpha,\\ d_1+\frac{d_2}{1+\alpha}\leq 1+\frac{\alpha}{1+\alpha}\lambda_P,\\ d_2+\frac{d_1}{2}\leq\frac{1+\alpha}{2}+\frac{\alpha}{2}\lambda_P,\\ d_1+d_2\leq 1+\alpha\lambda_P+\alpha\lambda_D. \end{aligned}$$

Proof: The achievability of the bound is described in Section IV. ■

The GDoF bound in Theorem 1 is depicted in Fig. 3. Note that for $\alpha=1$, our result covers the previous result in [15]. From Theorem 1 we directly have the following corollaries for the setting with delayed CSIT and a static topology ($\lambda_{D,D}=1,\lambda_{1,\alpha}=1$).

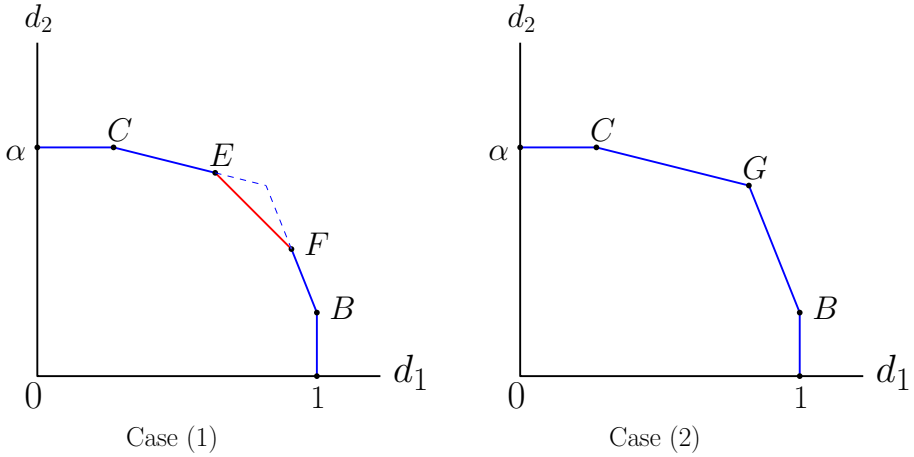

Fig. 3. GDoF inner bound for the two-user MISO BC with symmetric alternating CSIT and a static topology ($\lambda_{1,\alpha}=1$), for case (1) of $\lambda_D < \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$, and for case (2) of $\lambda_D \geq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$, respectively. Corner points take the values: $B=(1,\ \alpha\lambda_P)$, $C=(1-\alpha+\alpha\lambda_P,\ \alpha)$, $E=\big(1-\alpha+2\alpha\lambda_D+\alpha\lambda_P,\ \alpha-\alpha\lambda_D\big)$, $F=\big(1-\lambda_D,\ \alpha\lambda_P+(1+\alpha)\lambda_D\big)$ and $G=\big(\frac{1+\alpha}{1+2\alpha}+\frac{\alpha}{1+2\alpha}\lambda_P,\ \frac{\alpha(1+\alpha)}{1+2\alpha}+\frac{\alpha^2}{1+2\alpha}\lambda_P\big)$.

Corollary 1a: The GDoF region of the two-user MISO BC with delayed CSIT and a static topology $(\lambda_{D,D}=1, \lambda_{1,\alpha}=1)$ is inner bounded by the region characterized as

$$\begin{aligned} d_1 \leq 1, \quad d_2 \leq \alpha, \\ d_1 + \frac{d_2}{1+\alpha} \leq 1, \\ d_2 + \frac{d_1}{2} \leq \frac{1+\alpha}{2}, \end{aligned}$$

i.e., is inner bounded by the region with GDoF corner points $(0,0)$, $(1,0)$, $(\frac{1+\alpha}{1+2\alpha}, \frac{\alpha(1+\alpha)}{1+2\alpha})$, $(1-\alpha,\alpha)$ and $(0,\alpha)$.

Corollary 1b: The sum GDoF of the two-user MISO BC with delayed CSIT and a static topology $(\lambda_{D,D}=1, \lambda_{1,\alpha}=1)$ is lower bounded as

$$d_\Sigma \geq \frac{(1+\alpha)^2}{1+2\alpha}.$$

C. Optimal sum GDoF for the topological BC with practical CSIT schemes: fluctuating topology

We here explore a class of dynamically fluctuating topologies and reveal a certain *topological diversity gain* — in specific instances — that is associated to topologies that vary in time and across users. Emphasis is mainly given to statistically symmetric topologies, as well as to a certain class of practical feedback schemes.

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

Proposition 1: 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

$$d_\Sigma = 1+\frac{\alpha}{3}. \tag{16}$$

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

Remark 3: We see that the above result corresponding to an alternating topology ($\lambda_{1,\alpha}=\lambda_{\alpha,1}=1/2$) exhibiting a certain spatio-temporal topological diversity, exceeds the achievable sum GDoF $d=\frac{(1+\alpha)^2}{1+2\alpha}$ of the corresponding setting with a static topology ($\lambda_{1,\alpha}=1$ or $\lambda_{\alpha,1}=1$), as well as exceeds the optimal sum GDoF $d=\frac{2}{3}(1+\alpha)$ for the equivalent delayed-CSIT setting over a topology ($\lambda_{1,1}=\lambda_{\alpha,\alpha}=1/2$) that lacks the alternating and spatial-diversity elements that we find in the first topology ($\lambda_{1,\alpha}=\lambda_{\alpha,1}=1/2$).

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

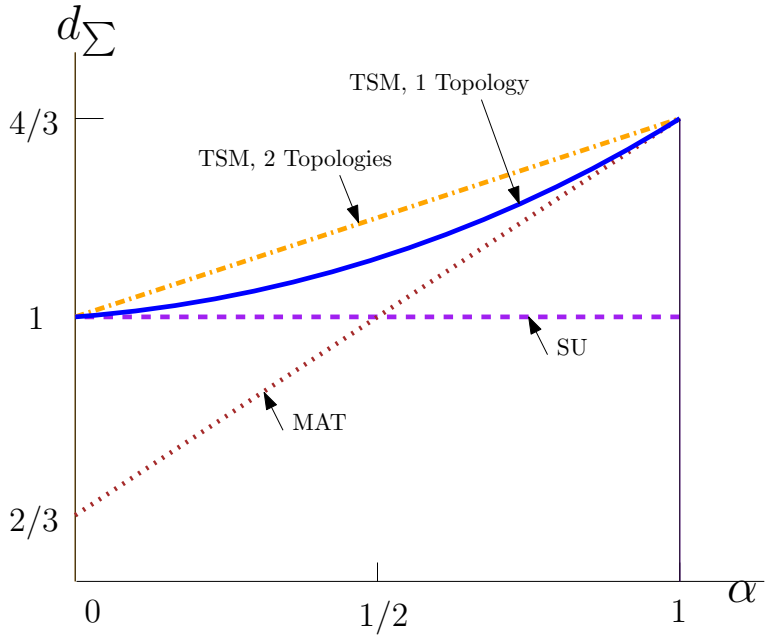

Fig. 4. Sum GDoF performance for the naively modified Maddah-Ali and Tse scheme (MAT), the single user case (SU), and the topological signal management scheme (TSM 1), all for the setting $\lambda_{D,D}^{1,\alpha} = 1$. Additionally the plot (TSM 2) describes the optimal sum GDoF for the fluctuating topology setting where $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$.

Proposition 2: 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

$$d_\Sigma = 1 + \frac{\alpha}{2} \tag{17}$$

which can be seen to exceed 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 bound in Lemma 2. The optimal scheme is described in Section V-B. ■

Regarding this same feedback policy $\lambda_{P,N} = \lambda_{N,P} = 1/2$, it is worth noting this policy's optimality, in the following broad context.

Proposition 3: 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

$$d_\Sigma = 1 + \frac{\alpha}{2} \tag{18}$$

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 2. The optimal scheme is described in Section V-B. ■

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 ($\lambda_{1,1} = \lambda_{\alpha,\alpha} = 0$). In essence, what the last two propositions say is that the design of the CSIT feedback protocol that indicates which user offers feedback at any given time, does not have to depend on the knowledge of the topology, and only needs to know that $\lambda_{P,N} = \lambda_{N,P} = 1/2$. Such CSIT feedback design can hence be agreed upon before the communication process.

IV. Topological signal management schemes for static topologies and for symmetric alternating CSIT (Proof of Theorem 1)

We proceed to derive a broad scheme for the general static topology, i.e., for the case of $\lambda_{1,\alpha} = 1$, which will constructively support the result in Theorem 1. This will entail achieving GDoF corner points (see Figure 3) $B = (1,\ \alpha\lambda_P)$, $C = (1-\alpha+\alpha\lambda_P,\ \alpha)$, $(0,\alpha)$,$(1,0)$, GDoF corner points

$$E = \big(1-\alpha+2\alpha\lambda_D+\alpha\lambda_P,\ \alpha-\alpha\lambda_D\big), \quad F = \big(1-\lambda_D,\ \alpha\lambda_P+(1+\alpha)\lambda_D\big)$$

for

$$\lambda_D < \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P \tag{19}$$

and point

$$G = \big(\frac{1+\alpha}{1+2\alpha} + \frac{\alpha}{1+2\alpha}\lambda_P,\ \frac{\alpha(1+\alpha)}{1+2\alpha} + \frac{\alpha^2}{1+2\alpha}\lambda_P\big)$$

for $\lambda_D \geq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$. Proper time sharing allows for the entire GDoF region in Theorem 1.

a) Intuition behind schemes: In a nutshell, the schemes will alternate between the actions of *overloading* and of *multicasting*, where *overloading* refers to having the transmitter send at a rate that is larger than what can be supported by the MISO BC, while *multicasting* refers to having the transmitter compensating for this excess by transmitting additional information that eventually assists both users in decoding. This interplay will naturally be a function of the topology. Such overload-multicast strategy was explored in different settings, including in [41] for the heterogeneous parallel channel with delayed CSIT. It is worth noting that one of the main differences between the new schemes, and the older schemes by Tandon et al. [15] as well as the schemes from the general CSIT setting in [14] — and by extension, the difference between the new schemes here and other block-Markov related schemes [42]–[46] (see also [47], [48]) — relates the new schemes' ability to properly capitalize on the inherent weakness of a link in order to (often optimally) reduce interference in at least one direction.

b) General notation used in schemes: In describing any scheme, we will generally associate the use of symbol a to denote a private symbol for user 1, while we will associate symbol b to denote a private symbol for user 2, and symbol c to denote a common symbol meant 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 pre-log 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.

We first describe the encoding, interference quantization and mapping, and the backward decoding for the scheme. Upon achieving points E and F for when $\lambda_D < \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$, we will do the same for point G for the case of $\lambda_D \geq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$ by slightly modifying the scheme such that it uses delayed CSIT for a lesser fraction of the time $\lambda_D^{'} \triangleq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P \leq \lambda_D$. Similar modifications will allow for the other corner points[3].

The general scheme will consist of L communication blocks, with T consecutive channel uses in each block, where T is finite while L can grow as large as we need it to be. We recall our soft assumption that, in every T consecutive channel uses — without loss of generality, in every time period $t = T(\ell-1)+1, T(\ell-1)+2, \cdots, T\ell$ for $\ell = 1, 2, 3, \cdots$, — the fraction of time associated with CSIT state (I_1, I_2) converges to the long-term statistic λ_{I_1,I_2}, for any $(I_1, I_2) \in (P, D, N) \times (P, D, N)$. This is commonly used in the setting of alternating CSIT.

A. Encoding

We now describe the encoding in block $\ell, \ell \in [1, L-1]$, which takes place over time $t = T(\ell-1)+1, T(\ell-1)+2, \cdots, T\ell$. During block ℓ, the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} c_t + \sqrt{\rho^{-\alpha}}a_t^{'''} \\ 0 \end{bmatrix} + \phi_t^{P2}\boldsymbol{g}_t^{\perp}a_t + \phi_t^{P1}\boldsymbol{h}_t^{\perp}b_t + \phi_t^{D2}\begin{bmatrix} a_t^{'} \\ a_t^{''} \end{bmatrix} + \phi_t^{D1}\begin{bmatrix} b_t^{'} \\ b_t^{''} \end{bmatrix} \tag{20}$$

where $a_t, a_t^{'}, a_t^{''}, a_t^{'''}$ are the private symbols meant for user 1, $b_t, b_t^{'}, b_t^{''}$ for user 2, where c_t is a common symbol, where the average power of each of those eight symbols is $1/8$ (the effective average power of $\sqrt{\rho^{-\alpha}}a_t^{'''}$ is $\rho^{-\alpha}$), where $\boldsymbol{e}^{\perp}$ denotes a unit-norm vector orthogonal to $\boldsymbol{e}$, and where

$$\phi_t^{P2} \triangleq \begin{cases} 1 & \text{if } I_2 = P \text{ at time } t \\ 0 & \text{else} \end{cases}, \qquad \phi_t^{P1} \triangleq \begin{cases} 1 & \text{if } I_1 = P \text{ at time } t \\ 0 & \text{else} \end{cases}, \tag{21}$$

$$\phi_t^{D2} \triangleq \begin{cases} 1 & \text{if } I_2 = D \text{ at time } t \\ 0 & \text{else} \end{cases}, \qquad \phi_t^{D1} \triangleq \begin{cases} 1 & \text{if } I_1 = D \text{ at time } t \\ 0 & \text{else} \end{cases}, \tag{22}$$

where we note that

$$\frac{1}{T}\sum_{t=T(\ell-1)+1}^{T\ell}\phi_t^{P1} = \frac{1}{T}\sum_{t=T(\ell-1)+1}^{T\ell}\phi_t^{P2} = \lambda_P, \qquad \frac{1}{T}\sum_{t=T(\ell-1)+1}^{T\ell}\phi_t^{D1} = \frac{1}{T}\sum_{t=T(\ell-1)+1}^{T\ell}\phi_t^{D2} = \lambda_D \tag{23}$$

after recalling the symmetric alternating CSIT assumption, and the assumption that the long term feedback statistics defining λ_{I_1,I_2}, $(I_1, I_2) \in (P, D, N) \times (P, D, N)$, still hold for finite durations.

[3]Section IV-F introduces a special example of this scheme for a specific setting.

After transmission during each $t = T(\ell-1)+1, \cdots, t = T\ell$, the received signals take the form

$$y_t = h_{t,1}(\sqrt{\rho}c_t + \sqrt{\rho^{1-\alpha}}a_t^{'''}) + \phi_t^{P2}\sqrt{\rho}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{g}_t^{\perp}a_t + \phi_t^{D2}\sqrt{\rho}\boldsymbol{h}_t^{\mathsf{T}}\begin{bmatrix}a_t^{'}\\ a_t^{''}\end{bmatrix} + \underbrace{\phi_t^{D1}\sqrt{\rho}\boldsymbol{h}_t^{\mathsf{T}}\begin{bmatrix}b_t^{'}\\ b_t^{''}\end{bmatrix}}_{s_{1,t}} + u_t \tag{24}$$

$$z_t = \sqrt{\rho^{\alpha}}g_{t,1}c_t + \phi_t^{P1}\sqrt{\rho^{\alpha}}\boldsymbol{g}_t^{\mathsf{T}}\boldsymbol{h}_t^{\perp}b_t + \phi_t^{D1}\sqrt{\rho^{\alpha}}\boldsymbol{g}_t^{\mathsf{T}}\begin{bmatrix}b_t^{'}\\ b_t^{''}\end{bmatrix} + \underbrace{\phi_t^{D2}\sqrt{\rho^{\alpha}}\boldsymbol{g}_t^{\mathsf{T}}\begin{bmatrix}a_t^{'}\\ a_t^{''}\end{bmatrix}}_{s_{2,t}} + \sqrt{\rho^{0}}g_{t,1}a_t^{'''} + v_t \tag{25}$$

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

$$s_{1,t} \triangleq \phi_t^{D1}\sqrt{\rho}\boldsymbol{h}_t^{\mathsf{T}}\begin{bmatrix}b_t^{'}\\ b_t^{''}\end{bmatrix}, \quad s_{2,t} \triangleq \phi_t^{D2}\sqrt{\rho^{\alpha}}\boldsymbol{g}_t^{\mathsf{T}}\begin{bmatrix}a_t^{'}\\ a_t^{''}\end{bmatrix} \tag{26}$$

denote the interference signals at user 1 and user 2 respectively.

B. Interference quantization and mapping

At the end of block ℓ, $\ell \in [1, L-1]$, the transmitter *reconstructs* $s_{1,t}$ and $s_{2,t}$ using delayed CSIT, and then *quantizes* these into $\bar{s}_{1,t}$ and $\bar{s}_{2,t}$ with $\phi_t^{D1}\log\rho - o(\log\rho)$ quantization bits and $\phi_t^{D2}\alpha\log\rho - o(\log\rho)$ quantization bits, respectively, allowing for bounded quantization errors $\tilde{s}_{1,t} \triangleq s_{1,t} - \bar{s}_{1,t}$ and $\tilde{s}_{2,t} \triangleq s_{2,t} - \bar{s}_{2,t}$ because $\mathbb{E}|s_{1,t}|^2 \doteq \phi_t^{D1}\rho$ and $\mathbb{E}|s_{2,t}|^2 \doteq \phi_t^{D2}\rho^{\alpha}$ (cf. [49]). The total of

$$\sum_{t=T(\ell-1)+1}^{T\ell}(\phi_t^{D1} + \phi_t^{D2}\alpha)\log\rho - To(\log\rho) = T\lambda_D(1+\alpha)\log\rho - To(\log\rho) \tag{27}$$

quantization bits for block ℓ (cf. (23)) is then mapped into common information symbols $\{c_t\}_{t=T\ell+1}^{T(\ell+1)}$ that will be transmitted in the next block, together with *new* information bits. In the last block (block L), the transmitter simply sends the common information symbols $\{c_t\}_{t=T(L-1)+1}^{TL}$ carrying a total of $T\lambda_D(1+\alpha)\log\rho$ information bits to both users, which can be done in T channel uses.

C. Backward decoding

We proceed to describe the decoding for each block. The decoding starts from the last block and moves backward. Specifically after decoding the common information in the last block, each user reconstructs $\{\bar{s}_{1,t}\}_{t=T(L-2)+1}^{T(L-1)}$ and $\{\bar{s}_{2,t}\}_{t=T(L-2)+1}^{T(L-1)}$ (corresponding to the quantized interference of block $L-1$) and uses them to decode its private symbols and common information symbols of block $L-1$; naturally the common information of block $L-1$ can accommodate decoding of the previous block (block $L-2$), and so on. Specifically after decoding the common information $\{c_t\}_{t=T\ell+1}^{T(\ell+1)}$ in block $\ell+1$, user 1 reconstructs $\{\bar{s}_{1,t}\}_{t=T(\ell-1)+1}^{T\ell}$, $\{\bar{s}_{2,t}\}_{t=T(\ell-1)+1}^{T\ell}$ and forms a MIMO observation for block ℓ, $\ell \in [1, L-1]$, which takes the form

$$\begin{bmatrix} y_{T\ell} - \bar{s}_{1,T\ell} \\ \bar{s}_{2,T\ell} \\ \vdots \\ y_{T(\ell-1)+1} - \bar{s}_{1,T(\ell-1)+1} \\ \bar{s}_{2,T(\ell-1)+1}\end{bmatrix} = \begin{bmatrix} \sqrt{\rho}h_{T\ell,1}c_{T\ell} \\ 0 \\ \vdots \\ \sqrt{\rho}h_{T(\ell-1)+1,1}c_{T(\ell-1)+1} \\ 0\end{bmatrix} + \begin{bmatrix} \sqrt{\rho^{1-\alpha}}h_{T\ell,1}a_{T\ell}^{'''} \\ 0 \\ \vdots \\ \sqrt{\rho^{1-\alpha}}h_{T(\ell-1)+1,1}a_{T(\ell-1)+1}^{'''} \\ 0\end{bmatrix}$$

$$+ \begin{bmatrix} \phi_{T\ell}^{P2}\sqrt{\rho}\boldsymbol{h}_{T\ell}^{\mathsf{T}}\boldsymbol{g}_{T\ell}^{\perp}a_{T\ell} \\ 0 \\ \vdots \\ \phi_{T(\ell-1)+1}^{P2}\sqrt{\rho}\boldsymbol{h}_{T(\ell-1)+1}^{\mathsf{T}}\boldsymbol{g}_{T(\ell-1)+1}^{\perp}a_{T(\ell-1)+1} \\ 0\end{bmatrix} + \begin{bmatrix} \phi_{T\ell}^{D2}\begin{bmatrix}\sqrt{\rho}\boldsymbol{h}_{T\ell}^{\mathsf{T}} \\ \sqrt{\rho^{\alpha}}\boldsymbol{g}_{T\ell}^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}a_{T\ell}^{'} \\ a_{T\ell}^{''}\end{bmatrix} \\ \vdots \\ \phi_{T(\ell-1)+1}^{D2}\begin{bmatrix}\sqrt{\rho}\boldsymbol{h}_{T(\ell-1)+1}^{\mathsf{T}} \\ \sqrt{\rho^{\alpha}}\boldsymbol{g}_{T(\ell-1)+1}^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}a_{T(\ell-1)+1}^{'} \\ a_{T(\ell-1)+1}^{''}\end{bmatrix}\end{bmatrix} + \underbrace{\begin{bmatrix} u_{T\ell} + \tilde{s}_{1,T\ell} \\ -\tilde{s}_{2,T\ell} \\ \vdots \\ u_{T(\ell-1)+1} + \tilde{s}_{1,T(\ell-1)+1} \\ -\tilde{s}_{2,T(\ell-1)+1}\end{bmatrix}}_{\text{power } \rho^0}.$$

One can easily show that, with successive decoding on this MIMO, user 1 can *jointly* decode the common symbols $\{c_t\}_{t=T(\ell-1)+1}^{T\ell}$ by treating other signals as noise, allowing for decoding a total of

$$T\alpha(1-\lambda_P-\lambda_D)\log\rho + To(\log\rho) \tag{28}$$

information bits. After removal of the common symbols from the received signals, the decoder can decode the private symbols $\{a_t^{'}, a_t^{''}\}_{t=T(\ell-1)+1}^{T\ell}$ by treating other signals as noise, thus allowing for decoding of up to

$$2T\alpha\lambda_D \log\rho + To(\log\rho) \tag{29}$$

further information bits. Again, after removal of these symbols, the decoder can now decode $\{a_t\}_{t=T(\ell-1)+1}^{T\ell}$ containing a total of

$$T\alpha\lambda_P \log\rho + To(\log\rho) \tag{30}$$

information bits, and finally after removing these last decoded symbols, the decoder can decode $\{a_t^{'''}\}_{t=T(\ell-1)+1}^{T\ell}$ containing a total of

$$T(1-\alpha) \log\rho + To(\log\rho) \tag{31}$$

information bits. Once the common information symbols $\{c_t\}_{t=T(\ell-1)+1}^{T\ell}$ (with a total of $T\alpha(1-\lambda_P-\lambda_D)\log\rho + To(\log\rho)$ information bits) are decoded at user 1, the $T\lambda_D(1+\alpha)\log\rho - To(\log\rho)$ (cf. (27)) side-information bits of these common symbols, can be used to recover the quantized interference $\{\bar{s}_{1,t}\}_{t=T(\ell-2)+1}^{T(\ell-1)}$ and $\{\bar{s}_{2,t}\}_{t=T(\ell-2)+1}^{T(\ell-1)}$ of block $\ell-1$, which in turn allows for completing decoding of block $\ell-1$. Backward decoding naturally stops at block 1.

Similarly, user 2 reconstructs $\{\bar{s}_{1,t}\}_{t=T(\ell-1)+1}^{T\ell}$ and $\{\bar{s}_{2,t}\}_{t=T(\ell-1)+1}^{T\ell}$ with the knowledge of common information $\{c_t\}_{t=T\ell+1}^{T(\ell+1)}$, and forms a MIMO observation for block ℓ, $\ell \in [1, L-1]$, which takes the form

$$\begin{bmatrix} z_{T\ell} - \bar{s}_{2,T\ell} \\ \bar{s}_{1,T\ell} \\ \vdots \\ z_{T(\ell-1)+1} - \bar{s}_{2,T(\ell-1)+1} \\ \bar{s}_{1,T(\ell-1)+1} \end{bmatrix} = \begin{bmatrix} \sqrt{\rho^\alpha} g_{T\ell,1} c_{T\ell} \\ 0 \\ \vdots \\ \sqrt{\rho^\alpha} g_{T(\ell-1)+1,1} c_{T(\ell-1)+1} \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_{T\ell}^{P1} \sqrt{\rho^\alpha} \boldsymbol{g}_{T\ell}^{\mathsf{T}} \boldsymbol{h}_{T\ell}^{\perp} b_{T\ell} \\ 0 \\ \vdots \\ \phi_{T(\ell-1)+1}^{P1} \sqrt{\rho^\alpha} \boldsymbol{g}_{T(\ell-1)+1}^{\mathsf{T}} \boldsymbol{h}_{T(\ell-1)+1}^{\perp} b_{T(\ell-1)+1} \\ 0 \end{bmatrix}$$
$$+ \begin{bmatrix} \phi_{T\ell}^{D1} \begin{bmatrix} \sqrt{\rho^\alpha} \boldsymbol{g}_{T\ell}^{\mathsf{T}} \\ \sqrt{\rho} \boldsymbol{h}_{T\ell}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} b_{T\ell}^{'} \\ b_{T\ell}^{''} \end{bmatrix} \\ \vdots \\ \phi_{T(\ell-1)+1}^{D1} \begin{bmatrix} \sqrt{\rho^\alpha} \boldsymbol{g}_{T(\ell-1)+1}^{\mathsf{T}} \\ \sqrt{\rho} \boldsymbol{h}_{T(\ell-1)+1}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} a_{T(\ell-1)+1}^{'} \\ a_{T(\ell-1)+1}^{''} \end{bmatrix} \end{bmatrix} + \underbrace{\begin{bmatrix} v_{T\ell} + \tilde{s}_{2,T\ell} + \sqrt{\rho^0} g_{T\ell,1} a_{T\ell}^{'''} \\ -\tilde{s}_{1,T\ell} \\ \vdots \\ v_{T(\ell-1)+1} + \tilde{s}_{2,T(\ell-1)+1} + \sqrt{\rho^0} g_{T(\ell-1)+1,1} a_{T(\ell-1)+1}^{'''} \\ -\tilde{s}_{1,T(\ell-1)+1} \end{bmatrix}}_{\text{power } \rho^0}$$

from which one can easily show that — again by using successive decoding — the common symbols $\{c_t\}_{t=T(\ell-1)+1}^{T\ell}$ can be *jointly* decoded with a total of

$$T\alpha(1-\lambda_P-\lambda_D)\log\rho + To(\log\rho) \tag{32}$$

information bits, while the private symbols $\{b_t^{'}, b_t^{''}\}_{t=T(\ell-1)+1}^{T\ell}$ can be decoded with a total of

$$T(1+\alpha)\lambda_D \log\rho + To(\log\rho) \tag{33}$$

information bits, whereas the private symbols $\{b_t\}_{t=T(\ell-1)+1}^{T\ell}$ can be decoded with a total of

$$T\alpha\lambda_P \log\rho + To(\log\rho) \tag{34}$$

information bits. As with the first user case, once the common information symbols $\{c_t\}_{t=T(\ell-1)+1}^{T\ell}$ are decoded by user 2, the side information bits can be used to recover the quantized interference of block $\ell-1$, which allows for completion of decoding for block $\ell-1$. This continues until we reach block 1.

D. Achieving the GDoF corner points

We proceed to calculate the GDoF performance of the designed scheme. We here consider a large L, in order to be able to neglect the necessary inefficiency of the last block.

1) Achieving GDoF points E and F for the case of $\lambda_D < \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$: In calculating the total number of *information bits*, we start by recalling that the common symbols $\{c_t\}_{t=T(\ell-1)+1}^{T\ell}$ of block ℓ, $\ell \in [1, L-1]$, carry a total of $T\alpha(1-\lambda_P-\lambda_D)\log\rho + To(\log\rho)$ bits (cf. (28), (32)), out of which $T\lambda_D(1+\alpha)\log\rho - To(\log\rho)$ bits (cf. (27)) are used as side information to convey the information of quantized interference $\{\bar{s}_{1,t}\}_{t=T(\ell-2)+1}^{T(\ell-1)}$ and $\{\bar{s}_{2,t}\}_{t=T(\ell-2)+1}^{T(\ell-1)}$ of block $\ell-1$. This leaves

$$\Delta_{\text{com}} \triangleq T\alpha(1-\lambda_P-\lambda_D)\log\rho - T\lambda_D(1+\alpha)\log\rho + To(\log\rho) = T\big(\alpha(1-\lambda_P)-\lambda_D(1+2\alpha)\big)\log\rho + To(\log\rho) \quad (35)$$

remaining information bits in these common symbols (this number is non-negative when $\lambda_D < \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$ (cf. (19))). Assigning all Δ_{com} information bits to user 1, achieves the GDoF point F, i.e., allows for

$$d_1 = \underbrace{d_{\Delta_{\text{com}}}}_{\text{cf. (35)}} + \underbrace{2\alpha\lambda_D}_{\text{cf. (29)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (30)}} + \underbrace{(1-\alpha)}_{\text{cf. (31)}} = \big(\alpha(1-\lambda_P)-\lambda_D(1+2\alpha)\big) + 2\alpha\lambda_D + \alpha\lambda_P + (1-\alpha) = 1-\lambda_D$$
$$d_2 = \underbrace{(1+\alpha)\lambda_D}_{\text{cf. (33)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (34)}}$$

where $d_{\Delta_{\text{com}}} \triangleq \lim_{\rho\to\infty}\frac{\Delta_{\text{com}}}{\log\rho}$ (cf. (35)). On the other hand, assigning all these Δ_{com} information bits to user 2, allows for GDoF point E, i.e., allows for

$$d_1 = \underbrace{2\alpha\lambda_D}_{\text{cf. (29)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (30)}} + \underbrace{(1-\alpha)}_{\text{cf. (31)}}$$
$$d_2 = \underbrace{d_{\Delta_{\text{com}}}}_{\text{cf. (35)}} + \underbrace{(1+\alpha)\lambda_D}_{\text{cf. (33)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (34)}} = \big(\alpha(1-\lambda_P)-\lambda_D(1+2\alpha)\big) + (1+\alpha)\lambda_D + \alpha\lambda_P = \alpha - \alpha\lambda_D.$$

2) Achieving GDoF point G for the case of $\lambda_D \geq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$: To achieve GDoF point G associated to the case where $\lambda_D \geq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P$, we simply apply the proposed scheme, except that now, instead of using delayed CSIT for the allowable λ_D fraction of the time (fraction of the block), we only use delayed CSIT for $\lambda_D^{'}$ fraction of the time, where

$$\lambda_D^{'} \triangleq \frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P.$$

Simple calculations show that this allows for a total of $\Delta_{\text{com}} = T\big(\alpha(1-\lambda_P)-\lambda_D^{'}(1+2\alpha)\big)\log\rho + To(\log\rho) = To(\log\rho)$ information bits (cf. (35)), and for the GDoF point G corresponding to

$$d_1 = \underbrace{2\alpha\lambda_D^{'}}_{\text{cf. (29)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (30)}} + \underbrace{(1-\alpha)}_{\text{cf. (31)}} = 2\alpha\Big(\frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P\Big) + \alpha\lambda_P + (1-\alpha) = \frac{1+\alpha}{1+2\alpha} + \frac{\alpha}{1+2\alpha}\lambda_P$$
$$d_2 = \underbrace{(1+\alpha)\lambda_D^{'}}_{\text{cf. (33)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (34)}} = (1+\alpha)\Big(\frac{\alpha}{1+2\alpha} - \frac{\alpha}{1+2\alpha}\lambda_P\Big) + \alpha\lambda_P = \frac{\alpha(1+\alpha)}{1+2\alpha} + \frac{\alpha^2}{1+2\alpha}\lambda_P.$$

3) Achieving GDoF points B, C, $(0,\alpha)$ and $(1,0)$: It is easy to show that the two GDoF points $(0,\alpha)$ and $(1,0)$ are easily achievable with simple time division between the two users. For achieving GDoF point $C = (1-\alpha+\alpha\lambda_P,\ \alpha)$, we repeat the same relegation of λ_D as before, except that now this λ_D is relegated all the way down to $\lambda_D^{''} = 0$, which simply means that we disregard entirely delayed CSIT. Proceeding as above, allocating Δ_{com} information bits to user 2 gives GDoF point C, corresponding to

$$d_1 = \underbrace{2\alpha\lambda_D^{''}}_{\text{cf. (29)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (30)}} + \underbrace{(1-\alpha)}_{\text{cf. (31)}} = \alpha\lambda_P + (1-\alpha)$$
$$d_2 = \underbrace{d_{\Delta_{\text{com}}}}_{\text{cf. (35)}} + \underbrace{(1+\alpha)\lambda_D^{''}}_{\text{cf. (33)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (34)}} = \big(\alpha(1-\lambda_P)-\lambda_D^{''}(1+2\alpha)\big) + (1+\alpha)\lambda_D^{''} + \alpha\lambda_P = \alpha$$

while allocating the Δ_{com} information bits to user 1, gives GDoF point B, corresponding to

$$d_1 = \underbrace{d_{\Delta_{\text{com}}}}_{\text{cf. (35)}} + \underbrace{2\alpha\lambda_D^{''}}_{\text{cf. (29)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (30)}} + \underbrace{(1-\alpha)}_{\text{cf. (31)}} = \big(\alpha(1-\lambda_P)-\lambda_D^{''}(1+2\alpha)\big) + \alpha\lambda_P + (1-\alpha) = 1$$
$$d_2 = \underbrace{(1+\alpha)\lambda_D^{''}}_{\text{cf. (33)}} + \underbrace{\alpha\lambda_P}_{\text{cf. (34)}} = (1+\alpha)\lambda_D^{''} + \alpha\lambda_P = \alpha\lambda_P.$$

Having completed the description of the general TSM design, we proceed to provide two illustrative examples, where the general scheme described above, is distilled down to specific instances. In the first example, the overloading and multicasting

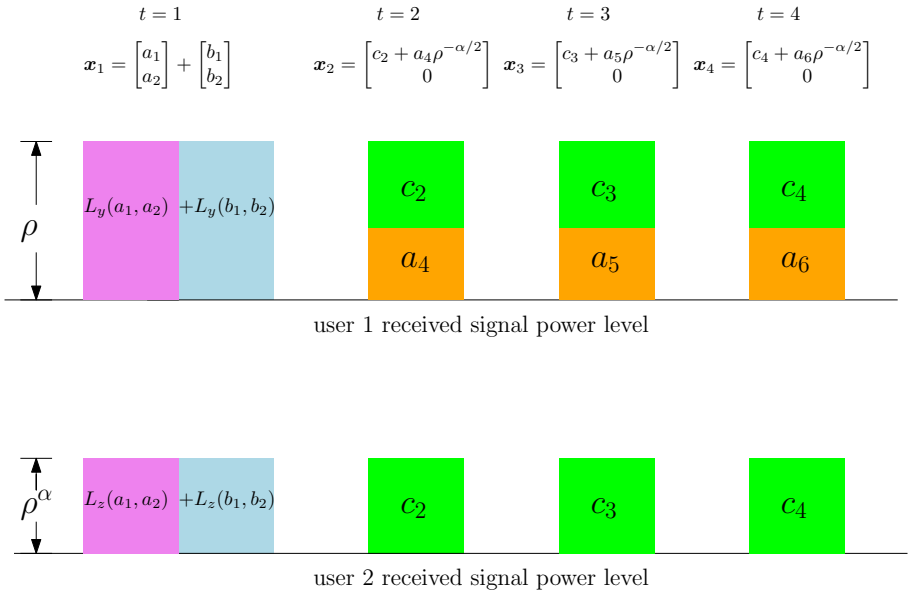

Fig. 5. Illustration of received power level for the proposed scheme, on the setting with delayed CSIT and one topology ($\lambda_{1,\alpha} = 1$ and $\lambda_{D,D} = 1,\ \alpha = 1/2$), where $L_y(\bullet)$ and $L_z(\bullet)$ denote the linear function of the argument at user 1 and user 2, respectively.

phases are operated in a consecutive manner, and the entire scheme has a finite and small duration. In the second example — where CSIT has a periodic structure — the two phases are jointly performed in the same communication block, and this block is repeated many times, in a block Markov manner where the multicasting phase in one block is designed to aid for the overloading phase from the previous block. This sequence follows closely from the scheme in [14] that considered a similar setting without though any topology considerations ($\alpha = 1$).

E. Illustrative Example: Fixed topology, delayed CSIT ($\lambda_{1,\alpha} = 1,\ \alpha = 1/2$ *and* $\lambda_{D,D} = 1$)

For the setting with constantly available delayed CSIT ($\lambda_{D,D} = 1$), and a specific static topology $\lambda_{1,\alpha} = 1$ with $\alpha = 1/2$, the scheme *overloads* for one channel use, and *multicasts* in three other channel uses, to achieve GDoF ($d_1 = \frac{1+\alpha}{1+2\alpha} = 3/4,\ d_2 = \frac{\alpha(1+\alpha)}{1+2\alpha} = 3/8$).

1) Overloading phase: During the overloading phase, taking place at $t = 1$, the transmitter sends (as illustrated in Fig. 5)

$$\boldsymbol{x}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

where a_1 and a_2 are the private symbols for user 1, where b_1, b_2 are the private symbols for user 2, and where the power of each symbol is $1/4$. The received signals then take the form

$$y_1 = \sqrt{\rho}\boldsymbol{h}_1^\intercal \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \underbrace{\sqrt{\rho}\boldsymbol{h}_1^\intercal \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{s_1} + u_1 \qquad z_1 = \sqrt{\rho^\alpha}\boldsymbol{g}_1^\intercal \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \underbrace{\sqrt{\rho^\alpha}\boldsymbol{g}_1^\intercal \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{s_2} + v_1 \tag{36}$$

where $s_1 \triangleq \sqrt{\rho}\boldsymbol{h}_1^\intercal \begin{bmatrix} b_1 & b_2 \end{bmatrix}^\intercal$ and $s_2 \triangleq \sqrt{\rho^\alpha}\boldsymbol{g}_1^\intercal \begin{bmatrix} a_1 & a_2 \end{bmatrix}^\intercal$ correspond to the interference signals at user 1 and user 2 respectively. Note that for user 1, knowing s_1 allows for removal of interference from y_1, while knowing s_2 allows for an extra observation that can assist in decoding a_1 and a_2. Similarly user 2 can use possible knowledge of s_1 and s_2 towards decoding b_1 and b_2. This knowledge will be provided in the next phase, where the transmitter will multicast the information about s_1 and s_2 to both users.

2) Multicasting phase: After time $t = 1$, and after having access to delayed CSIT of channels $\boldsymbol{g}_1$ and $\boldsymbol{h}_2$, the transmitter *reconstructs* s_1 and s_2, and then *quantizes* them into $\bar{s}_1$ and $\bar{s}_2$ with approximately $\log\rho$ quantization bits[4] and $\alpha\log\rho$ quantization bits respectively, allowing for bounded quantization errors $\tilde{s}_1 \triangleq s_1 - \bar{s}_1$ and $\tilde{s}_2 \triangleq s_2 - \bar{s}_2$ since $\mathbb{E}|s_1|^2 \doteq \rho$ and $\mathbb{E}|s_2|^2 \doteq \rho^\alpha$ (cf. [49]). All $(1+\alpha)\log\rho$ quantization bits are then mapped into the common information symbols c_2, c_3, c_4 that will be transmitted to both users in this multicasting phase, during $t = 2, 3, 4$. Specifically, at each time $t = 2, 3, 4$, the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} c_t + \sqrt{\rho^{-\alpha}}a_{t+2} \\ 0 \end{bmatrix}$$

[4]The use of the term 'approximately', refers to the fact that we are using $\log\rho - o(\log\rho)$ (rather than $\log\rho$) quantization bits.

where a_{t+2} is the private symbol for user 1, and where the average power of each c_t and a_{t+2} is $1/2$, i.e., the effective average power of $\sqrt{\rho^{-\alpha}}a_{t+2}$ is $\rho^{-\alpha}$. Then the *processed* received signals during $t=2,3,4$, are of the form

$$y_t/h_{t,1}=\sqrt{\rho}c_t+\sqrt{\rho^{1-\alpha}}a_{t+2}+u_t/h_{t,1}\qquad z_t/g_{t,1}=\sqrt{\rho^{\alpha}}c_t+\sqrt{\rho^{0}}a_{t+2}+v_t/g_{t,1} \tag{37}$$

where $h_{t,1}\triangleq \boldsymbol{h}_t^{\mathsf{T}}\left[1\ 0\right]^{\mathsf{T}}$, $g_{t,1}\triangleq \boldsymbol{g}_t^{\mathsf{T}}\left[1\ 0\right]^{\mathsf{T}}$. One can see that both users can decode the common symbol c_t with $\alpha\log\rho$ information bits, and additionally that user 1 can decode the private symbol a_{t+2} with approximately $(1-\alpha)\log\rho$ information bits, for each $t=2,3,4$.

After decoding the common information symbols c_2,c_3,c_4, corresponding to a total of $3\alpha\log\rho=\frac{3}{2}\log\rho$ bits, both users can reconstruct $\bar{s}_1$ and $\bar{s}_2$ — represented by a total of $(1+\alpha)\log\rho=\frac{3}{2}\log\rho$ information bits — in order to decode the private symbols a_1,a_2 at user 1 and b_1,b_2 at user 2. Specifically user 1 and user 2 each form their 2×2 MIMO observations, respectively taking the form

$$\begin{bmatrix}y_1-\bar{s}_1\\ \bar{s}_2\end{bmatrix}=\begin{bmatrix}\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\\ \sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}a_1\\ a_2\end{bmatrix}+\underbrace{\begin{bmatrix}u_1+\tilde{s}_1\\ -\tilde{s}_2\end{bmatrix}}_{\text{power }\rho^0},\qquad \begin{bmatrix}y_2-\bar{s}_2\\ \bar{s}_1\end{bmatrix}=\begin{bmatrix}\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\\ \sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix}+\underbrace{\begin{bmatrix}z_2+\tilde{s}_2\\ -\tilde{s}_1\end{bmatrix}}_{\text{power }\rho^0}.$$

One can easily show that the private symbols a_1,a_2 can be decoded by user 1 with a total of approximately $(1+\alpha)\log\rho$ bits, while the private symbols b_1,b_2 can be decoded by user 2 with a total of approximately $(1+\alpha)\log\rho$ bits. Finally a simple calculation can show that the GDoF $(d_1=\frac{1+\alpha+3(1-\alpha)}{4}=\frac{3}{4},\ d_2=\frac{1+\alpha}{4}=\frac{3}{8})$ is achievable.

Remark 5: Note that in this scheme, during the four channel uses, we only use delayed CSIT on $\boldsymbol{h}_1$ and $\boldsymbol{g}_1$, i.e., only for the first channels, which implies that the scheme and result still hold when $(I_1,I_2)=\underbrace{(D,D)}_{t=1},\underbrace{(N,N)}_{t=2},\underbrace{(N,N)}_{t=3},\underbrace{(N,N)}_{t=4}$.

F. Illustrative Example: Fixed topology, partially available and periodic delayed CSIT

We now consider a specific static topology ($\lambda_{1,\alpha}=1$, $\alpha=1/2$) and delayed CSIT that is *periodic, but only partially available*. Specifically we consider a setting where

$$(I_1,I_2)=\underbrace{(D,N)}_{t=1},\underbrace{(N,D)}_{t=2},\underbrace{(N,N)}_{t=3},\underbrace{(N,N)}_{t=4},\Big|\underbrace{(D,N)}_{t=5},\underbrace{(N,D)}_{t=6},\underbrace{(N,N)}_{t=7},\underbrace{(N,N)}_{t=8},\Big|\underbrace{(D,N)}_{t=9},\underbrace{(N,D)}_{t=10},\underbrace{(N,N)}_{t=11},\underbrace{(N,N)}_{t=12},\Big|\cdots$$

corresponding to having $\lambda_{D,N}=\lambda_{N,D}=\frac{1}{2}\lambda_{N,N}=1/4$. This scheme consists of L communication blocks, where each block ℓ $(\ell=1,2,\cdots,L)$ has duration of 4 channel uses $t=4\ell-3,4\ell-2,4\ell-1,4\ell$. In the end, the scheme will achieve GDoF $(d_1=\frac{1+\alpha}{1+2\alpha}=3/4,\ d_2=\frac{\alpha(1+\alpha)}{1+2\alpha}=3/8)$.

1) Encoding: We proceed to describe the encoding during each block $\ell,\ell\in[1,L-1]$. The last block will be omitted without a GDoF effect, given that L will be chosen to be large.

In the first channel use of block ℓ ($t=4\ell-3$) we have $(I_1,I_2)=(D,N)$, and the transmitter sends

$$\boldsymbol{x}_{4\ell-3}=\begin{bmatrix}c_{4\ell-3}+\sqrt{\rho^{-\alpha}}a_{4\ell-3}\\ 0\end{bmatrix}+\begin{bmatrix}b_{4\ell-3}\\ b'_{4\ell-3}\end{bmatrix} \tag{38}$$

where $a_{4\ell-3}$ is the private symbol for user 1, $c_{4\ell-3}$ is a common symbol for both users, where $b_{4\ell-3},b'_{4\ell-3}$ are the private symbols meant for user 2[5], and where the power of each of these four symbols is $1/4$. The corresponding received signals take the form

$$y_{4\ell-3}=h_{4\ell-3,1}(\sqrt{\rho}c_{4\ell-3}+\sqrt{\rho^{1-\alpha}}a_{4\ell-3})+\underbrace{\sqrt{\rho}\boldsymbol{h}_{4\ell-3}^{\mathsf{T}}\begin{bmatrix}b_{4\ell-3}\\ b'_{4\ell-3}\end{bmatrix}}_{s_{1,\ell}}+u_{4\ell-3} \tag{39}$$

$$z_{4\ell-3}=\sqrt{\rho^{\alpha}}g_{4\ell-3,1}c_{4\ell-3}+\sqrt{\rho^{\alpha}}\boldsymbol{g}_{4\ell-3}^{\mathsf{T}}\begin{bmatrix}b_{4\ell-3}\\ b'_{4\ell-3}\end{bmatrix}+\sqrt{\rho^{0}}g_{4\ell-3,1}a_{4\ell-3}+v_{4\ell-3} \tag{40}$$

where $s_{1,\ell}\triangleq\sqrt{\rho}\boldsymbol{h}_{4\ell-3}^{\mathsf{T}}\begin{bmatrix}b_{4\ell-3} & b'_{4\ell-3}\end{bmatrix}^{\mathsf{T}}$ corresponds to the interference signal at user 1.

In the second channel use of block ℓ ($t=4\ell-2$), we have $(I_1,I_2)=(N,D)$, and the transmitter sends

$$\boldsymbol{x}_{4\ell-2}=\begin{bmatrix}c_{4\ell-2}+\sqrt{\rho^{-\alpha}}a_{4\ell-2}\\ 0\end{bmatrix}+\begin{bmatrix}a'_{4\ell-2}\\ a''_{4\ell-2}\end{bmatrix} \tag{41}$$

[5]These symbols can be considered as 'overloaded', in the sense that user 2 would not have been able to decode them, even if there was no interference.

where $a_{4\ell-2}, a_{4\ell-2}^{'}, a_{4\ell-2}^{''}$ are the private symbols meant for user 1 (now $a_{4\ell-2}, a_{4\ell-2}^{'}, a_{4\ell-2}^{''}$ are the overloaded symbols), where $c_{4\ell-2}$ is a common symbol, and where the power of each symbol is $1/4$. The received signals then take the form

$$y_{4\ell-2} = h_{4\ell-2,1}(\sqrt{\rho}c_{4\ell-2} + \sqrt{\rho^{1-\alpha}}a_{4\ell-2}) + \sqrt{\rho}\boldsymbol{h}_{4\ell-2}^{\mathsf{T}} \begin{bmatrix} a_{4\ell-2}^{'} \\ a_{4\ell-2}^{''} \end{bmatrix} + u_{4\ell-2} \tag{42}$$

$$z_{4\ell-2} = \sqrt{\rho^{\alpha}}g_{4\ell-2,1}c_{4\ell-2} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_{4\ell-2}^{\mathsf{T}} \begin{bmatrix} a_{4\ell-2}^{'} \\ a_{4\ell-2}^{''} \end{bmatrix}}_{s_{2,\ell}} + \sqrt{\rho^{0}}g_{4\ell-2,1}a_{4\ell-2} + v_{4\ell-2}, \tag{43}$$

where $s_{2,\ell} \triangleq \sqrt{\rho^{\alpha}}\boldsymbol{g}_{4\ell-2}^{\mathsf{T}} \begin{bmatrix} a_{4\ell-2}^{'} & a_{4\ell-2}^{''} \end{bmatrix}^{\mathsf{T}}$ corresponds to the interference signal at user 2.

In the last two channel uses of block ℓ ($t = 4\ell - 1, 4\ell$), we have $(I_1, I_2) = (N, N)$, and the transmitter sends

$$\boldsymbol{x}_t = \begin{bmatrix} c_t + \sqrt{\rho^{-\alpha}}a_t \\ 0 \end{bmatrix} \tag{44}$$

where again a_t is the private symbol for user 1, c_t is a common symbol, and where both symbols have power $1/2$. This results in received signals of the following form

$$y_t = \sqrt{\rho}h_{t,1}c_t + \sqrt{\rho^{1-\alpha}}h_{t,1}a_t + + u_t \tag{45}$$

$$z_t = \sqrt{\rho^{\alpha}}g_{t,1}c_t + \sqrt{\rho^{0}}g_{t,1}a_t + v_t. \tag{46}$$

2) Interference quantization and mapping: At the end of each block ℓ, $\ell \in [1, L-1]$, the transmitter *reconstructs* $s_{1,\ell}$ and $s_{2,\ell}$ using delayed CSIT, and *quantizes* these into $\bar{s}_{1,\ell}$ and $\bar{s}_{2,\ell}$ using respectively $\log\rho$ and $\alpha\log\rho$ quantization bits, thus allowing for bounded quantization errors $\tilde{s}_{1,\ell} \triangleq s_{1,\ell} - \bar{s}_{1,\ell}$ and $\tilde{s}_{2,\ell} \triangleq s_{2,\ell} - \bar{s}_{2,\ell}$ since $\mathbb{E}|s_{1,\ell}|^2 \doteq \rho$ and $\mathbb{E}|s_{2,\ell}|^2 \doteq \rho^{\alpha}$. The total of $(1+\alpha)\log\rho = \frac{3}{2}\log\rho$ quantization bits is then mapped into the common symbols $\{c_t\}_{t=4\ell+1}^{4(\ell+1)}$ that will be transmitted to both users in the next block. Note that in the last block — block L, again of length 4 — the transmitter simply sends to both users the common information symbols $\{c_t\}_{t=4L-3}^{4L}$ containing a total of $\frac{3}{2}\log\rho$ bits.

3) Backward decoding: As with all other schemes here, decoding starts from the last block and moves backward. Specifically after decoding the common symbols of the last block, each user reconstructs $\bar{s}_{1,L-1}$ and $\bar{s}_{2,L-1}$, recovers the quantized interference of block $L-1$, and uses this to decode its private and common symbols of block $L-1$. This last common information of block $L-1$, can now be used for decoding of block $L-2$, and so on. In general, after decoding the common information $\{c_t\}_{t=4\ell+1}^{4(\ell+1)}$ of block $\ell+1$, user 1 reconstructs $\bar{s}_{1,\ell}$ and $\bar{s}_{2,\ell}$, to form a MIMO observation

$$\begin{bmatrix} y_{4\ell-3} - \bar{s}_{1,\ell} \\ y_{4\ell-2} \\ \bar{s}_{2,\ell} \\ y_{4\ell-1} \\ y_{4\ell} \end{bmatrix} = \begin{bmatrix} \sqrt{\rho}h_{4\ell-3,1}c_{4\ell-3} \\ \sqrt{\rho}h_{4\ell-2,1}c_{4\ell-2} \\ 0 \\ \sqrt{\rho}h_{4\ell-1,1}c_{4\ell-1} \\ \sqrt{\rho}h_{4\ell,1}c_{4\ell} \end{bmatrix} + \begin{bmatrix} \sqrt{\rho^{1-\alpha}}h_{4\ell-3,1}a_{4\ell-3} \\ \sqrt{\rho^{1-\alpha}}h_{4\ell-2,1}a_{4\ell-2} \\ 0 \\ \sqrt{\rho^{1-\alpha}}h_{4\ell-1,1}a_{4\ell-1} \\ \sqrt{\rho^{1-\alpha}}h_{4\ell,1}a_{4\ell} \end{bmatrix} + \begin{bmatrix} 0 \\ \begin{bmatrix} \sqrt{\rho}\boldsymbol{h}_{4\ell-2}^{\mathsf{T}} \\ \sqrt{\rho^{\alpha}}\boldsymbol{g}_{4\ell-2}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} a_{4\ell-2}^{'} \\ a_{4\ell-2}^{''} \end{bmatrix} \\ 0 \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} u_{4\ell-3} + \tilde{s}_{1,\ell} \\ u_{4\ell-2} \\ -\tilde{s}_{2,\ell} \\ u_{4\ell-1} \\ u_{4\ell} \end{bmatrix}}_{\text{power } \rho^0}.$$

In a similar manner to the previous schemes, one can show that successive decoding on the above MIMO setting, allows user 1 to *jointly* decode the common symbols $\{c_t\}_{t=4\ell-3}^{4\ell}$ by treating the other signals as noise, decoding a total of $3\alpha\log\rho$ information bits. After removing the common symbols, the user can decode the private symbols $a_{4\ell-2}^{'}$ and $a_{4\ell-2}^{''}$ by treating the other signals as noise, thus decoding a total of $2\alpha\log\rho$ information bits. Similarly, after removing these last decoded symbols, user 1 can decode the private symbols $\{a_t\}_{t=4\ell-3}^{4\ell}$ carrying a total of $4(1-\alpha)\log\rho$ information bits. Furthermore, having already decoded the common information in symbols $\{c_t\}_{t=4\ell-3}^{4\ell}$, user 1 can — as we have seen for block ℓ — complete decoding for the previous block (block $\ell-1$). Such backward decoding stops at block 1. A similar procedure is followed for user 2. Consequently, for large L, the achievable GDoF can easily be calculated to be

$$d_1 = \frac{2\alpha + 4(1-\alpha)}{4} = \frac{3}{4}, \quad d_2 = \frac{1+\alpha}{4} = \frac{3}{8}.$$

G. Example: Naive topological modifications to the original MAT scheme ($\lambda_{D,D} = 1$) for the setting $\lambda_{1,\alpha} = 1$

The following — which is meant to accentuate the need for proper TSM design — describes a naive variant of the original MAT scheme, which fails to properly account for topology and thus under-performs compared to the corresponding TSM in the same $\lambda_{1,\alpha} = 1, \lambda_{D,D} = 1$ setting.

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

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

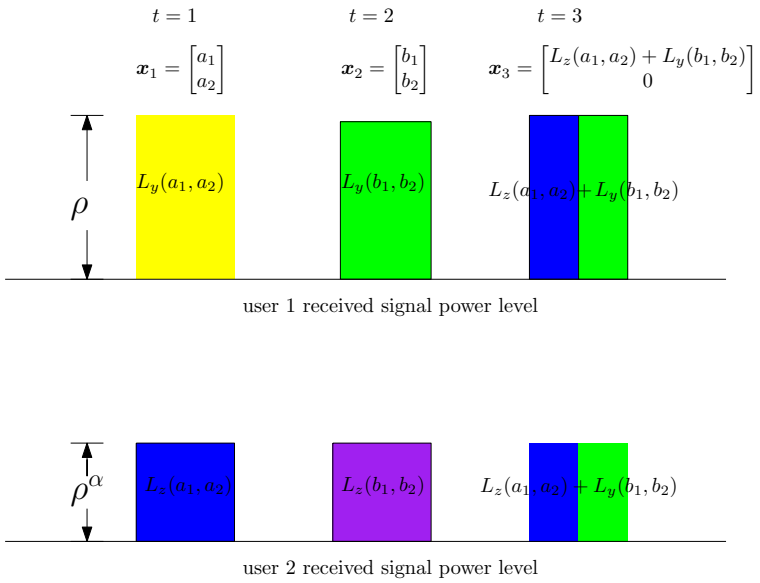

Fig. 6. Illustration of the received power level for the naively-modified MAT scheme in the static topology setting $\lambda_{1,\alpha}=1$.

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 now (in the current, topologically sensitive setting)

$$y_1=\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix} \qquad z_1=\sqrt{\rho^\alpha}\boldsymbol{g}_1^{\mathsf{T}}\begin{bmatrix}a_1\\a_2\end{bmatrix}\triangleq\sqrt{\rho^\alpha}L_z(a_1,a_2) \tag{47}$$

$$y_2=\sqrt{\rho}\boldsymbol{h}_2^{\mathsf{T}}\begin{bmatrix}b_1\\b_2\end{bmatrix}\triangleq\sqrt{\rho}L_y(b_1,b_2) \qquad z_2=\sqrt{\rho^\alpha}\boldsymbol{g}_2^{\mathsf{T}}\begin{bmatrix}b_1\\b_2\end{bmatrix}. \tag{48}$$

At time $t=3$, the transmitter knows $\boldsymbol{g}_1$ and $\boldsymbol{h}_2$ by using delayed CSIT, reconstructs $L_z(a_1,a_2),L_y(b_1,b_2)$ (cf. (47), (48)), and sends

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

The normalized/processed received signals, in their noiseless form, are

$$y_3/h_{3,1}=\sqrt{\rho}L_z(a_1,a_2)+\sqrt{\rho}L_y(b_1,b_2) \tag{49}$$

$$z_3/g_{3,1}=\sqrt{\rho^\alpha}L_z(a_1,a_2)+\sqrt{\rho^\alpha}L_y(b_1,b_2). \tag{50}$$

At this point, we recall from [9] that user 1 combines the above with y_1,y_2,y_3, to design a MIMO system

$$\begin{bmatrix}y_1\\y_3/h_{3,1}-y_2\end{bmatrix}=\sqrt{\rho}\begin{bmatrix}\boldsymbol{h}_1^{\mathsf{T}}\\\boldsymbol{g}_1^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}a_1\\a_2\end{bmatrix}+\begin{bmatrix}u_1\\u_3/h_{3,1}-u_2\end{bmatrix} \tag{51}$$

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

$$\begin{bmatrix}z_2\\z_3/g_{3,1}-z_1\end{bmatrix}=\sqrt{\rho^\alpha}\begin{bmatrix}\boldsymbol{g}_2^{\mathsf{T}}\\\boldsymbol{h}_2^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}b_1\\b_2\end{bmatrix}+\begin{bmatrix}v_2\\v_3/g_{3,1}-v_1\end{bmatrix} \tag{52}$$

over a weaker link, 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 $d_{\sum}=\frac{2(1+\alpha)}{3}$.

V. Sum-GDoF optimal topological signal management schemes for the fluctuating topology setting

We proceed to build on the topological signal management schemes in Section IV and to design schemes for the alternating topology settings in Section III-C. These schemes will be sum-GDoF optimal.

TABLE I
SUMMARY OF SCHEMES

Scheme #	Section #	CSIT, topology	achieved $d_{\sum}$	for Proposition #
1	V-A	$\lambda_{1,\alpha} = \lambda_{\alpha,1} = 1/2$ $\lambda_{D,D} = 1$	$1 + \frac{\alpha}{3}$ optimal	Proposition 1
2	V-B	any $\lambda_{1,\alpha} + \lambda_{\alpha,1} = 1$ $\lambda_{P,N} = \lambda_{N,P} = 1/2$	$1 + \frac{\alpha}{2}$ optimal	Propositions 2, 3

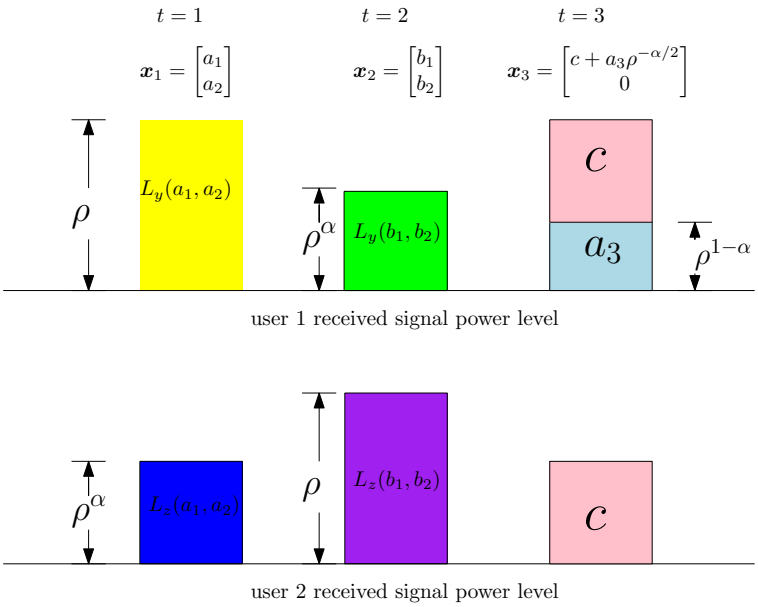

Fig. 7. Received signal power level illustration for the TSM scheme, for the setting where $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$.

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

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 $d_{\sum} = (1 + \alpha/3)$. By averaging over the two schemes, we can get the optimal sum GDoF $d_{\sum} = (1 + \alpha/3)$ with $\lambda_{D,D}^{1,\alpha} = \lambda_{D,D}^{\alpha,1} = 1/2$.

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 7)

$$\boldsymbol{x}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{53}$$

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

$$r^{(a_1)} = 1, \quad r^{(a_2)} = \alpha \tag{54}$$

resulting in received signals of the form

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho} + u_1 \tag{55}$$

$$z_1 = \underbrace{\sqrt{\rho^\alpha}\boldsymbol{g}_1^{\mathsf{T}} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\rho^\alpha} + v_1 \tag{56}$$

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

$$\boldsymbol{x}_2=\begin{bmatrix}b_1\\ b_2\end{bmatrix} \tag{57}$$

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

$$r^{(b_1)}=1,\quad r^{(b_2)}=\alpha \tag{58}$$

resulting in received signals of the form

$$y_2=\underbrace{\sqrt{\rho^\alpha}\boldsymbol{h}_2^{\mathsf{T}}\begin{bmatrix}b_1\\ b_2\end{bmatrix}}_{\rho^\alpha}+u_2 \tag{59}$$

$$z_2=\underbrace{\sqrt{\rho}\boldsymbol{g}_2^{\mathsf{T}}\begin{bmatrix}b_1\\ b_2\end{bmatrix}}_{\rho}+v_2 \tag{60}$$

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

$$\iota\triangleq(z_1-v_1)+(y_2-u_2) \tag{61}$$

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. [49]). 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

$$\boldsymbol{x}_3=\begin{bmatrix}c+a_3\rho^{-\alpha/2}\\ 0\end{bmatrix} \tag{62}$$

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

$$\begin{aligned}P^{(c)}&\doteq 1, & r^{(c)}&=\alpha\\ P^{(a_3)}&\doteq 1, & r^{(a_3)}&=1-\alpha\end{aligned} \tag{63}$$

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

$$y_3/h_{3,1}=\sqrt{\rho}c+\sqrt{\rho^{1-\alpha}}a_3 \tag{64}$$

$$z_3/g_{3,1}=\sqrt{\rho^\alpha}c+\sqrt{\rho^0}a_3. \tag{65}$$

Now we see from (64),(65) 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. (61)), and to then decode the private symbols. Specifically, user 1 obtains a MIMO observation

$$\begin{bmatrix}y_1\\ \bar{\iota}-y_2\end{bmatrix}=\begin{bmatrix}\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\\ \sqrt{\rho^\alpha}\boldsymbol{g}_1^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}a_1\\ a_2\end{bmatrix}+\begin{bmatrix}u_1\\ -u_2-\tilde{\iota}\end{bmatrix} \tag{66}$$

which allows for decoding of a_1,a_2 at the declared rates (cf. (54)). Similarly, user 2 obtains another MIMO observation

$$\begin{bmatrix}z_2\\ \bar{\iota}-z_1\end{bmatrix}=\begin{bmatrix}\sqrt{\rho}\boldsymbol{g}_2^{\mathsf{T}}\\ \sqrt{\rho^\alpha}\boldsymbol{h}_2^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix}+\begin{bmatrix}v_2\\ -v_1-\tilde{\iota}\end{bmatrix} \tag{67}$$

and can decode b_1,b_2 at the declared rates (cf. (58)). Summing up the information bits concludes that the scheme achieves the optimal sum GDoF $d_{\sum}=\frac{1+\alpha+1+\alpha+(1-\alpha)}{3}=1+\frac{\alpha}{3}$ (also see Figure 7).

Remark 6: 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 (62)), 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 7: 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 $\lambda_{N,D}^{1,\alpha}=\lambda_{D,N}^{\alpha,1}=2\lambda_{N,N}^{1,\alpha}=2\lambda_{N,N}^{\alpha,1}=1/3$, to achieve the same optimal sum GDoF.

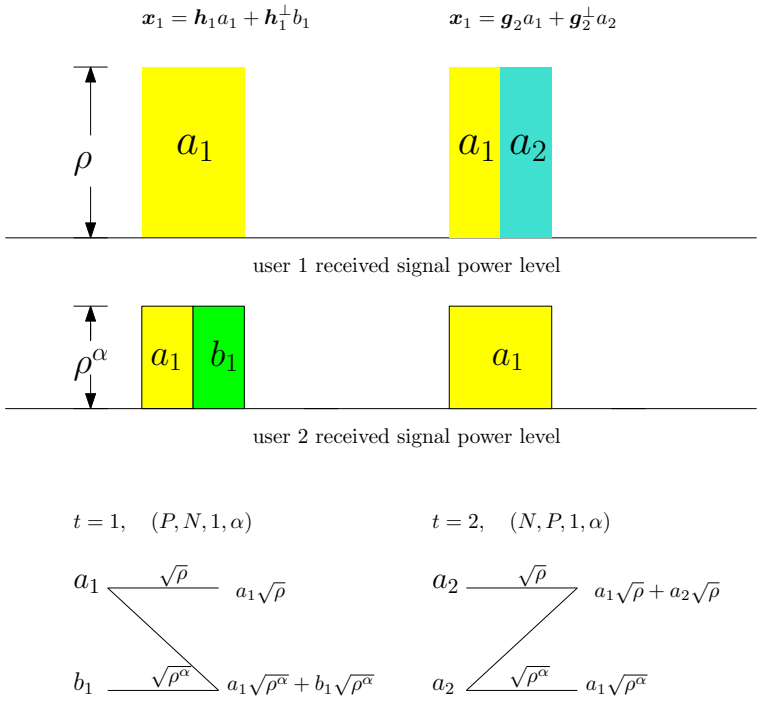

Fig. 8. Illustration of TSM coding and of received signal power levels, for $\lambda_{P,N}^{1,\alpha} = \lambda_{N,P}^{1,\alpha} = 1/2$.

B. 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

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

respectively. Each scheme achieves the optimal sum GDoF $(1 + \frac{\alpha}{2})$, and each scheme is designed to have 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 $\lambda_{P,N}^{1,\alpha} = \lambda_{N,P}^{1,\alpha} = 1/2$: 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 8)

$$\boldsymbol{x}_1 = \boldsymbol{h}_1 a_1 + \boldsymbol{h}_1^{\perp} b_1 \tag{68}$$

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

$$\begin{aligned} P^{(a_1)} &\doteq 1, \quad r^{(a_1)} = 1 \\ P^{(b_1)} &\doteq 1, \quad r^{(b_1)} = \alpha. \end{aligned} \tag{69}$$

Then the received signals (in their noiseless form) are

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho} \tag{70}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1^{\perp} b_1}_{\rho^{\alpha}}. \tag{71}$$

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

$$\boldsymbol{x}_2 = \boldsymbol{g}_2 a_1 + \boldsymbol{g}_2^{\perp} a_2 \tag{72}$$

where a_2 is intended for user 1, and where

$$P^{(a_2)} \doteq 1, \quad r^{(a_2)} = 1. \tag{73}$$

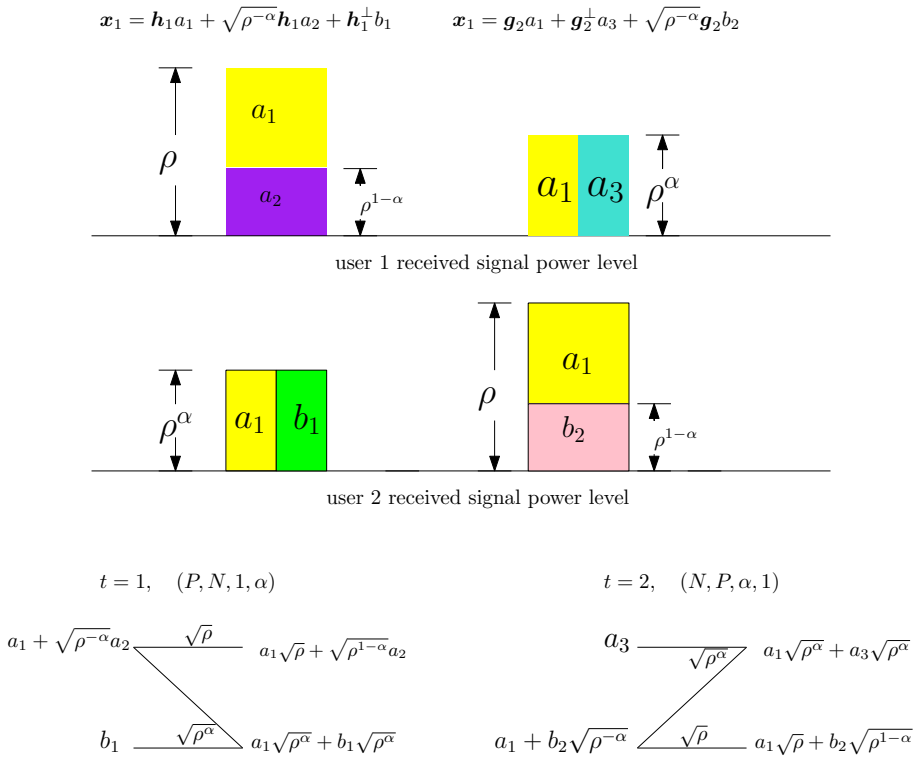

Fig. 9. Illustration of coding and received signal power levels for $\lambda_{P,N}^{1,\alpha} = \lambda_{N,P}^{\alpha,1} = 1/2$.

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

$$y_2 = \underbrace{\sqrt{\rho}\boldsymbol{h}_2^{\mathsf{T}}\boldsymbol{g}_2 a_1}_{\rho} + \underbrace{\sqrt{\rho}\boldsymbol{h}_2^{\mathsf{T}}\boldsymbol{g}_2^{\perp} a_2}_{\rho} \tag{74}$$

$$z_2 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_2^{\mathsf{T}}\boldsymbol{g}_2 a_1}_{\rho^{\alpha}}. \tag{75}$$

At this point, we can see that user 1 can MIMO decode a_1, a_2 based on (70), (74), while user 2 can recover b_1 by employing interference cancelation based on (71), (75). This gives a sum DoF of $1 + \alpha/2$.

Remark 8: We can now readily 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 take the above scheme (of Section V-B1), 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 9)

$$\boldsymbol{x}_1 = \boldsymbol{h}_1 a_1 + \sqrt{\rho^{-\alpha}}\boldsymbol{h}_1 a_2 + \boldsymbol{h}_1^{\perp} b_1 \tag{76}$$

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

$$r^{(a_1)} = \alpha, \quad r^{(a_2)} = 1 - \alpha, \quad r^{(b_1)} = \alpha \tag{77}$$

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

$$y_1 = \underbrace{\sqrt{\rho}\boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho} + \underbrace{\sqrt{\rho^{1-\alpha}}\boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{h}_1 a_2}_{\rho^{1-\alpha}} \tag{78}$$

$$z_1 = \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1 a_1}_{\rho^{\alpha}} + \underbrace{\sqrt{\rho^{0}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1 a_2}_{\rho^{0}} + \underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{h}_1^{\perp} b_1}_{\rho^{\alpha}}. \tag{79}$$

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

$$\boldsymbol{x}_2=\boldsymbol{g}_2a_1+\boldsymbol{g}_2^{\perp}a_3+\sqrt{\rho^{-\alpha}}\boldsymbol{g}_2b_2 \tag{80}$$

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

$$r^{(a_3)}=\alpha,\quad r^{(b_2)}=1-\alpha \tag{81}$$

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

$$y_2=\underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2a_1}_{\rho^{\alpha}}+\underbrace{\sqrt{\rho^{\alpha}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2^{\perp}a_3}_{\rho^{\alpha}}+\underbrace{\sqrt{\rho^{0}}\boldsymbol{h}_2^{\intercal}\boldsymbol{g}_2^{\perp}b_2}_{\rho^{0}} \tag{82}$$

$$z_2=\underbrace{\sqrt{\rho}\boldsymbol{g}_2^{\intercal}\boldsymbol{g}_2a_1}_{\rho}+\underbrace{\sqrt{\rho^{1-\alpha}}\boldsymbol{g}_2^{\intercal}\boldsymbol{g}_2b_2}_{\rho^{1-\alpha}}. \tag{83}$$

At this point, it is easy to see that user 1 can recover a_1,a_2,a_3 by MIMO decoding based on (78) and (82), while user 2 can recover b_1,b_2 by employing interference cancelation based on (79) and (83) (see also Figure 9). 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 $\lambda_{1,\alpha}+\lambda_{\alpha,1}=1$.

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.

VII. Appendix - Proof of general 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 denote all channel states that appear in the BC. Let the communication duration be n channel uses, where n is large. We use

$$y^n_{I_1,I_2}=\{y_t\}_t,\qquad z^n_{I_1,I_2}=\{z_t\}_t\qquad \forall t: I_{1,t}=I_1,I_{2,t}=I_2$$

to denote the accumulated set of received signals at user 1 and user 2 respectively, accumulated throughout the time when the CSIT state was some fixed I_1,I_2. As a result, the entirety of the received signals, at each user, is the following union of the above sets

$$y^n=\bigcup_{I_1,I_2}y^n_{I_1,I_2},\qquad z^n=\bigcup_{I_1,I_2}z^n_{I_1,I_2}.$$

A. Proof of bound (9) and bound (10)

Towards proving the bound in (9), we note that

$$\begin{aligned}
&nR_1 - n\epsilon_n \\
&= H(W_1) - n\epsilon_n \\
&= H(W_1|\Omega^n) - n\epsilon_n \\
&= I(W_1; y^n|\Omega^n) + \underbrace{H(W_1|y^n, \Omega^n)}_{\leq n\epsilon_n} - n\epsilon_n \\
&\leq I(W_1; y^n|\Omega^n) \qquad (84)\\
&= h(y^n|\Omega^n) - h(y^n|W_1, \Omega^n) \\
&\leq n\Big(\sum_{\forall(A_1,A_2)} A_1\lambda_{A_1,A_2}\Big)\log\rho + no(\log\rho) - \underbrace{h(y^n|W_1,\Omega^n)}_{\geq no(\log\rho)} \qquad (85)\\
&\leq n\Big(\sum_{\forall(A_1,A_2)} A_1\lambda_{A_1,A_2}\Big)\log\rho + no(\log\rho) - no(\log\rho) \qquad (86)
\end{aligned}$$

where (84) results from Fano's inequality which bounds $H(W_1|y^n,\Omega^n)$, where (85) follows from the fact that $h(y^n|\Omega^n) = \sum_{t=1}^n h(y_t|y^{t-1}\Omega^n) \leq \sum_{t=1}^n \max_{\Psi:\mathrm{tr}(\Psi)\leq 1}\log(1+\rho^{A_{1,t}}\boldsymbol{h}_t^{\mathsf{H}}\Psi\boldsymbol{h}_t) = \sum_{t=1}^n\Big(A_{1,t}\log\rho + o(\log\rho)\Big)$, that Gaussian input maximizes the differential entropy, and that $\Psi \triangleq \mathbb{E}[\boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{H}}]$, where (86) is from the fact that $h(y^n|W_1,\Omega^n) \geq h(y^n|W_1,\Omega^n,\{\boldsymbol{x}_t\}_{t=1}^n) = h\big(\{u_t\}_{t=1}^n\big) = no(\log\rho)$ and that conditioning reduces differential entropy. Finally dividing (86) by $n\log\rho$ and leting $\rho\to\infty$, provides for the bound in (9). Similarly, exchanging the roles of user 1 and user 2, proves (10).

B. Proof of bound (11) and bound (12)

Towards proving (11), 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 $I_k = N$, does not affect capacity, as one can deduce from the work in [50].

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

$$\tilde{y}^n = \big(y_{P,P}^n, y_{P,D}^n, y_{P,N}^n, \tilde{y}_{D,P}^n, \tilde{y}_{N,P}^n, \tilde{y}_{D,D}^n, \tilde{y}_{D,N}^n, \tilde{y}_{N,D}^n, \tilde{y}_{N,N}^n\big)$$

where specifically when $I_1 = P$ (i.e., whenever the first user sends perfect CSIT) then the received signal of user $\tilde{1}$ is identical to that of user 1, else when $I_1 \neq P$, the received signal of user $\tilde{1}$ is only assumed to be *identically distributed* to the signal y_t of user 1. We also assume 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)). We further enhance by assuming that $\tilde{y}^n$ is known to user 2. We note that, 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.

As a next step, we introduce the auxiliary random variable s_t, and define $s_{I_1,I_2}^n = \{s_t\}_{t:I_{1,t}=I_1,I_{2,t}=I_2}$. At this point we enhance the degraded compound BC, by giving user 2 complete knowledge of

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

where, as described below in (87), $\{s_{D,P}^n, s_{N,P}^n, s_{D,N}^n, s_{N,D}^n, s_{D,D}^n, s_{N,N}^n\}$ is the collection of auxiliary random variables $s_t,\ t: I_{1,t}\neq P$ accumulated whenever there is no CSIT on channel $\boldsymbol{h}_t$ of user 1 and no CSIT on channel $\tilde{\boldsymbol{h}}_t$ of user $\tilde{1}$, where specifically

$$\rho^{\frac{A_{2,t}-A_{1,t}}{2}}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \boldsymbol{g}_t^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \tilde{\boldsymbol{h}}_t^{\mathsf{T}}\end{bmatrix}^{-1}\begin{bmatrix}y_t\\ \tilde{y}_t\end{bmatrix} = \underbrace{\rho^{\frac{A_{2,t}}{2}}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \boldsymbol{g}_t^{\mathsf{T}}\end{bmatrix}\boldsymbol{x}_t + \begin{bmatrix}0\\ v_t\end{bmatrix}}_{=\begin{bmatrix}\star\\ z_t\end{bmatrix}} + \underbrace{\begin{bmatrix}0\\ -v_t\end{bmatrix} + \rho^{\frac{A_{2,t}-A_{1,t}}{2}}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \boldsymbol{g}_t^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \tilde{\boldsymbol{h}}_t^{\mathsf{T}}\end{bmatrix}^{-1}\begin{bmatrix}u_t\\ \tilde{u}_t\end{bmatrix}}_{\triangleq\begin{bmatrix}\star\\ s_t\end{bmatrix}} \tag{87}$$

i.e., where specifically s_t is the second element of the vector $\begin{bmatrix}0\\ -v_t\end{bmatrix} + \rho^{\frac{A_{2,t}-A_{1,t}}{2}}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \boldsymbol{g}_t^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_t^{\mathsf{T}}\\ \tilde{\boldsymbol{h}}_t^{\mathsf{T}}\end{bmatrix}^{-1}\begin{bmatrix}u_t\\ \tilde{u}_t\end{bmatrix}$, and where we have set $\tilde{\boldsymbol{h}}_t$ to be independently and identically distributed to $\boldsymbol{h}_t$, and $\tilde{u}_t$ to be independently and identically distributed to u_t. What the above means is that s_t has average power

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

as well as that knowledge of $\{s_t, y_t, \tilde{y}_t, \Omega^n\}$, implies the knowledge of z_t, again whenever $I_1 \neq P$.

At this point we can see that

$$\begin{aligned} & nR_1 - n\epsilon_n \\ & \leq I(W_1; y^n|\Omega^n) && (88) \\ & = h(y^n|\Omega^n) - h(y^n|W_1, \Omega^n) && (89) \end{aligned}$$

where (88) results from Fano's inequality which bounds $H(W_1|y^n, \Omega^n)$.

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

$$nR_1 - n\epsilon_n \leq h(\tilde{y}^n|\Omega^n) - h(\tilde{y}^n|W_1, \Omega^n). \tag{90}$$

As a result, adding (89) and (90) gives

$$\begin{aligned} & 2nR_1 - 2n\epsilon_n \\ & \leq h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) - h(y^n|W_1, \Omega^n) - h(\tilde{y}^n|W_1, \Omega^n) \\ & \leq h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) - h(y^n, \tilde{y}^n|W_1, \Omega^n) && (91) \end{aligned}$$

where (91) uses a basic entropy inequality.

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

$$\begin{aligned} & nR_2 - n\epsilon_n \\ & = H(W_2) - n\epsilon_n \\ & = H(W_2|\Omega^n) - n\epsilon_n \\ & \leq I(W_2; W_1, z^n, y^n, \tilde{y}^n, s_0^n|\Omega^n) && (92) \\ & = I(W_2; z^n, y^n, \tilde{y}^n, s_0^n|W_1, \Omega^n) + \underbrace{I(W_2; W_1|\Omega^n)}_{=0} && (93) \\ & = I(W_2; z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, \Omega^n) \\ & \quad + \underbrace{I(W_2; z_{D,P}^n, z_{N,P}^n, z_{D,N}^n, z_{N,D}^n, z_{D,D}^n, z_{N,N}^n|z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n, W_1, \Omega^n)}_{=0} && (94) \\ & = I(W_2; z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, \Omega^n) && (95) \\ & = h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, \Omega^n) - \underbrace{h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, W_2, \Omega^n)}_{=no(\log\rho)} && (96) \\ & = h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, \Omega^n) - no(\log\rho) && (97) \\ & = h(y^n, \tilde{y}^n|W_1, \Omega^n) + \underbrace{h(s_0^n|y^n, \tilde{y}^n, W_1, \Omega^n)}_{\leq h(s_0^n)} + \underbrace{h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n|y^n, \tilde{y}^n, s_0^n, W_1, \Omega^n)}_{\leq h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n)} - no(\log\rho) && (98) \\ & \leq h(y^n, \tilde{y}^n|W_1, \Omega^n) + h(s_0^n) + h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n) - no(\log\rho), && (99) \end{aligned}$$

where (92) comes from Fano's inequality, where (93), (96), (98) use basic chain rule, where (94) stems from messages independence, where (95) follows from that the knowledge of $\{y^n, \tilde{y}^n, s_0^n, \Omega^n\}$ allows for the reconstruction of $\{z_{D,P}^n, z_{N,P}^n, z_{D,N}^n, z_{N,D}^n, z_{D,D}^n, z_{N,N}^n\}$ (for example, knowing $\{y_{D,P}^n, \tilde{y}_{D,P}^n, s_{D,P}^n, \Omega^n\}$ allows for reconstructing $\{z_{D,P}^n\}$, cf. (87)), i.e., $\{z_{D,P}^n, z_{N,P}^n, z_{D,N}^n, z_{N,D}^n, z_{D,D}^n, z_{N,N}^n\} \leftrightarrow \{y^n, \tilde{y}^n, s_0^n, \Omega^n\} \leftrightarrow W_2$ forms a Markov chain, where (97) is from $h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, s_0^n|W_1, W_2, \Omega^n) = \underbrace{h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n|W_1, W_2, \Omega^n)}_{=no(\log\rho)} + \underbrace{h(s_0^n|z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n, W_1, W_2, \Omega^n)}_{=no(\log\rho)} = no(\log\rho)$ by using the fact that the knowledge of $\{W_1, W_2, \Omega^n\}$ allows for reconstructing $\{z_{P,P}^n, z_{P,D}^n, z_{P,N}^n, y^n, \tilde{y}^n\}$ up to noise level and the knowledge of $\{W_1, W_2, \Omega^n, y^n, \tilde{y}^n\}$ allows for reconstructing s_0^n up to noise level, where (99) uses the fact that conditioning reduces entropy.

By adding (91) and (99), and dividing by n, we have

$$\begin{aligned} & 2R_1 + R_2 - 3\epsilon_n \\ & \leq \frac{1}{n}\Big(h(y^n|\Omega^n) + h(\tilde{y}^n|\Omega^n) + h(s_0^n) + h(z_{P,P}^n, z_{P,D}^n, z_{P,N}^n) + no(\log\rho)\Big) && (100) \\ & \leq 2\Big(\sum_{\forall(I_1,I_2)} \sum_{\forall(A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2}\Big)\log\rho + \sum_{(I_1,I_2):I_1\neq P} (1-\alpha)\lambda_{I_1,I_2}^{\alpha,1}\log\rho + \sum_{(I_1,I_2):I_1=P} \sum_{\forall(A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2}\log\rho - o(\log\rho), && (101) \end{aligned}$$

and consequently have

$$2d_1+d_2\leq 2\Big(\sum_{\forall(I_1,I_2)}\sum_{\forall(A_1,A_2)}A_1\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\sum_{(I_1,I_2):I_1\neq P}(1-\alpha)\lambda_{I_1,I_2}^{\alpha,1}+\sum_{(I_1,I_2):I_1=P}\sum_{\forall(A_1,A_2)}A_2\lambda_{I_1,I_2}^{A_1,A_2} \tag{102}$$

which gives bound (11).

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

$$2d_2+d_1\leq 2\Big(\sum_{\forall(I_1,I_2)}\sum_{\forall(A_1,A_2)}A_2\lambda_{I_1,I_2}^{A_1,A_2}\Big)+\sum_{(I_1,I_2):I_2\neq P}(1-\alpha)\lambda_{I_1,I_2}^{1,\alpha}+\sum_{(I_1,I_2):I_2=P}\sum_{\forall(A_1,A_2)}A_1\lambda_{I_1,I_2}^{A_1,A_2} \tag{103}$$

which gives bound (12).

C. Proof for bound (13)

We continue with the proof of bound (13). 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 a first imaginary user $\tilde{1}$, and a second imaginary user $\tilde{2}$.

User $\tilde{1}$, which shares the same desired message W_1 as user 1, is supplied with a received signal that takes the form

$$\tilde{y}^n=\big(y_{P,P}^n,y_{P,D}^n,y_{D,P}^n,y_{D,D}^n,y_{P,N}^n,y_{D,N}^n,\tilde{y}_{N,P}^n,\tilde{y}_{N,D}^n,\tilde{y}_{N,N}^n\big)$$

which means that user 1 and user $\tilde{1}$ share the exact same received signal whenever $I_1\neq N$, while otherwise we only assume that user $\tilde{1}$ has a received signal that is statistically identical to that of user 1, but not necessarily the same.

Similarly user $\tilde{2}$, which shares the same desired message W_2 as user 2, is supplied with a received signal that takes the form

$$\tilde{z}^n=\big(z_{P,P}^n,z_{D,P}^n,z_{P,D}^n,z_{D,D}^n,z_{N,P}^n,z_{N,D}^n,\tilde{z}_{P,N}^n,\tilde{z}_{D,N}^n,\tilde{z}_{N,N}^n\big)$$

which again means that user 2 and user $\tilde{2}$ share the same received signal whenever $I_2\neq N$, while otherwise we only assume that user $\tilde{2}$ has a received signal that is statistically identical to that of user 2, but not necessarily the same.

This latter stage does not further alter the capacity - compared to the previously *enhanced* BC - since user 1 and user $\tilde{1}$ have the same long-term decoding ability; similarly for user 2 and user $\tilde{2}$.

Furthermore, whenever $(I_1,I_2)=(N,N)$ we can assume without an effect on 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), since the capacity depends only on the marginals for the channels associated with $(I_1,I_2)=(N,N)$.

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

$$\bar{y}_t=\sqrt{\rho^{\min\{A_{1,t},A_{2,t}\}}}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t+\bar{u}_t \tag{104}$$

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

$$\sqrt{\rho^{A_{1,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{y}_t=\underbrace{\sqrt{\rho^{A_{1,t}}}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t+u_t}_{=y_t}+\underbrace{\sqrt{\rho^{A_{1,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{u}_t-u_t}_{\triangleq\,\omega_t} \tag{105}$$

$$\sqrt{\rho^{A_{2,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{y}_t=\underbrace{\sqrt{\rho^{A_{2,t}}}\boldsymbol{h}_t^{\mathsf{T}}\boldsymbol{x}_t+v_t}_{=z_t}+\underbrace{\sqrt{\rho^{A_{2,t}-\min\{A_{1,t},A_{2,t}\}}}\bar{u}_t-v_t}_{\triangleq\,\psi_t} \tag{106}$$

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 provide each user with the observation $\bar{y}_{NN}^n$, to reach an 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) && (107)\\
&= 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) + I(W_1; y_{N,N}^n|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) \\
&\le 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) + I(W_1; y_{N,N}^n, \omega_{N,N}^n|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) && (108)\\
&= 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) \\
&\quad + I(W_1; \omega_{N,N}^n|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) + \underbrace{I(W_1; y_{N,N}^n|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,N}^n, \Omega^n)}_{=0} \\
& && (109)\\
&= 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) + \underbrace{I(W_1; \omega_{N,N}^n|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)}_{\le h(\omega_{N,N}^n) - no(\log\rho)} && (110)\\
&\le 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) + h(\omega_{N,N}^n) - no(\log\rho) && (111)\\
&= \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)}_{\le 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) + h(\omega_{N,N}^n) - no(\log\rho) && (112)\\
&\le 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) && (113)\\
&= 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) && (114)\\
&= h(\omega_{N,N}^n) + \underbrace{h(y_0^n)}_{\le 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) \\
&\quad + \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)}_{\le h(y_{N,P}^n, y_{N,D}^n) - no(\log\rho) \le 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) && (115)\\
&\le h(\omega_{N,N}^n) + n\Phi_{10} + n\Phi_{11} + no(\log\rho) \\
&\quad + 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) && (116)
\end{aligned}
$$

where

$$
\begin{aligned}
y_0^n &\triangleq \big(y_{P,P}^n, y_{P,D}^n, y_{D,P}^n, y_{D,D}^n\big) \\
\Phi_{10} &\triangleq \Big(\sum_{(I_1,I_2): I_1 \ne N, I_2 \ne N} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2}\Big)\log\rho \\
\Phi_{11} &\triangleq \Big(\sum_{(I_1,I_2) \in \{(N,P),(N,D)\}} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2}\Big)\log\rho
\end{aligned}
$$

where (107) results from Fano's inequality, where (108) uses the fact that adding information does not reduce mutual information, where (109) results from the chain rule, where (110) follows from the fact that the knowledge of $\{\bar{y}_{N,N}^n, \omega_{N,N}^n, \Omega^n\}$ allows for the reconstruction of $y_{N,N}^n$, i.e., follows from the fact that $y_{N,N}^n \leftrightarrow \{\bar{y}_{N,N}^n, \omega_{N,N}^n, \Omega^n\} \leftrightarrow W_1$ forms a Markov chain, where (111) is from the fact that $I(W_1; \omega_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n, \Omega^n) = \underbrace{h(\omega_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n, \Omega^n)}_{\le h(\omega_{N,N}^n)} - \underbrace{h(\omega_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n, \Omega^n, W_1)}_{\ge h(\omega_{N,N}^n|y_0^n, y_{P,N}^n, y_{N,P}^n, y_{D,N}^n, y_{N,D}^n, \bar{y}_{N,N}^n, \Omega^n, W_1, W_2) = no(\log\rho)} \le h(\omega_{N,N}^n) - no(\log\rho)$ by using the fact that the knowledge of $\{\bar{y}_{N,N}^n, W_1, W_2, \Omega^n\}$ allows for reconstructing $\omega_{N,N}^n$ up to noise level (cf. (105), (104)), and where (112) - (116) are derived using basic entropy rules.

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

$$
\begin{aligned}
&nR_1 - n\epsilon_n \\
&\le h(\omega_{N,N}^n) + n\Phi_{10} + n\Phi_{11} + no(\log\rho) \\
&\quad + I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n|\Omega^n) - I(W_2; \tilde{y}_{N,P}^n, \tilde{y}_{N,D}^n|W_1, y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n, \Omega^n). && (117)
\end{aligned}
$$

Adding (116) and (117), gives

$$
\begin{aligned}
&2nR_1 - 2n\Phi_{10} - 2n\Phi_{11} - no(\log\rho) - 2n\epsilon_n \\
&\le 2h(\omega^n_{N,N}) + 2I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) - I(W_2; y^n_{N,P}, y^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n) \\
&\quad - I(W_2; \tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n) && (118)\\
&= 2h(\omega^n_{N,N}) + 2I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) \\
&\quad \underbrace{-h(y^n_{N,P}, y^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n) - h(\tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n)}_{\le -h(y^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n)} \\
&\quad + \underbrace{h(y^n_{N,P}, y^n_{N,D}|W_2, W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n)}_{=no(\log\rho)} + \underbrace{h(\tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_2, W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n)}_{=no(\log\rho)} \\
&\le 2h(\omega^n_{N,N}) + 2I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) - h(y^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n) \\
&\quad + no(\log\rho) && (119)\\
&= 2h(\omega^n_{N,N}) + 2I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) - I(W_2; y^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n) \\
&\quad \underbrace{- h(y^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,P}, \tilde{y}^n_{N,D}|W_2, W_1, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, \Omega^n)}_{=no(\log\rho)} + no(\log\rho) && (120)\\
&= 2h(\omega^n_{N,N}) + 2I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) - I(W_2; y^n_{N,P}, \tilde{y}^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) \\
&\quad + I(W_2; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) + no(\log\rho) && (121)\\
&= 2h(\omega^n_{N,N}) + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) - I(W_2; y^n_{N,P}, \tilde{y}^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) \\
&\quad \underbrace{+ I(W_1, W_2; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n)}_{\le h(y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}) - no(\log\rho)} + no(\log\rho) && (122)\\
&\le h(\omega^n_{N,N}) + \underbrace{h(\omega^n_{N,N}) + h(y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N})}_{\le n\Phi_{12} + no(\log\rho)} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) \\
&\quad - I(W_2; y^n_{N,P}, \tilde{y}^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) + no(\log\rho) && (123)\\
&\le h(\omega^n_{N,N}) + n\Phi_{12} + no(\log\rho) + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) \\
&\quad - I(W_2; y^n_{N,P}, \tilde{y}^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (124)\\
&= h(\omega^n_{N,N}) + n\Phi_{12} + no(\log\rho) + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) \\
&\quad - I(W_2; 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}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) \\
&\quad \underbrace{+ I(W_2; s^n_{N,P}, s^n_{N,D}|y^n_{N,P}, \tilde{y}^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}, W_1, \Omega^n)}_{\le h(s^n_{N,P}, s^n_{N,D}) - no(\log\rho)} && (125)\\
&\le h(\omega^n_{N,N}) + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) \\
&\quad - I(W_2; 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}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) + h(s^n_{N,P}, s^n_{N,D}) + no(\log\rho) && (126)\\
&= h(\omega^n_{N,N}) + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) + h(s^n_{N,P}, s^n_{N,D}) + no(\log\rho) \\
&\quad - I(W_2; y^n_{N,P}, \tilde{y}^n_{N,P}, s^n_{N,P}, z^n_{N,P}, y^n_{N,D}, \tilde{y}^n_{N,D}, s^n_{N,D}, z^n_{N,D}, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (127)\\
&\le h(\omega^n_{N,N}) + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) + h(s^n_{N,P}, s^n_{N,D}) + no(\log\rho) \\
&\quad - I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (128)\\
&\le h(\omega^n_{N,N}) + n\Phi_{12} + I(W_1; W_2, y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|\Omega^n) + h(s^n_{N,P}, s^n_{N,D}) + no(\log\rho) \\
&\quad - I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (129)\\
&= h(\omega^n_{N,N}) + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_2, \Omega^n) + \underbrace{h(s^n_{N,P}, s^n_{N,D})}_{\le n\Phi_{13} + no(\log\rho)} + no(\log\rho) \\
&\quad - I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (130)\\
&\le \underbrace{h(\omega^n_{N,N})}_{\le n\Phi_{14} + no(\log\rho)} + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_2, \Omega^n) + n\Phi_{13} + no(\log\rho) \\
&\quad - I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (131)\\
&\le n\Phi_{14} + n\Phi_{12} + I(W_1; y^n_{P,N}, y^n_{D,N}, \bar{y}^n_{N,N}|W_2, \Omega^n) + n\Phi_{13} + no(\log\rho) \\
&\quad - I(W_2; z^n_{N,P}, z^n_{N,D}, \bar{y}^n_{N,N}|W_1, \Omega^n) && (132)
\end{aligned}
$$

where

$$\Phi_{12} \triangleq \Big(\sum_{(I_1,I_2):I_2=N} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \log \rho$$
$$\Phi_{13} \triangleq \Big(\sum_{(I_1,I_2)\in\{(N,P),(N,D)\}} (1-\alpha) \lambda_{I_1,I_2}^{\alpha,1} \Big) \log \rho$$
$$\Phi_{14} \triangleq (1-\alpha) \lambda_{N,N}^{1,\alpha} \log \rho$$

where $s_{N,P}^n$ and $z_{N,D}^n$ (cf. (125)) are defined in (87). In the above, (127) is from the fact that the knowledge of $\{y_{N,P}^n, \tilde{y}_{N,P}^n, s_{N,P}^n, y_{N,D}^n, \tilde{y}_{N,D}^n, s_{N,D}^n, \Omega^n\}$ implies the knowledge of $z_{N,P}^n$ and $z_{N,D}^n$ (cf. (87)). Most of the above steps are based on basic entropy rules.

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

$$\begin{aligned} &2nR_2 - 2n\Phi_{20} - 2n\Phi_{21} - no(\log\rho) - 2n\epsilon_n \\ &\le n\Phi_{24} + n\Phi_{22} + I(W_2; z_{N,P}^n, z_{N,D}^n, \bar{y}_{N,N}^n | W_1, \Omega^n) + n\Phi_{23} + no(\log\rho) - I(W_1; y_{P,N}^n, y_{D,N}^n, \bar{y}_{N,N}^n | W_2, \Omega^n) \end{aligned} \tag{133}$$

where

$$\Phi_{20} \triangleq \Big(\sum_{(I_1,I_2):I_1\neq N, I_2\neq N} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \log \rho$$
$$\Phi_{21} \triangleq \Big(\sum_{(I_1,I_2)\in\{(P,N),(D,N)\}} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \log \rho$$
$$\Phi_{22} \triangleq \Big(\sum_{(I_1,I_2):I_1=N} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \log \rho$$
$$\Phi_{23} \triangleq \Big(\sum_{(I_1,I_2)\in\{(P,N),(D,N)\}} (1-\alpha) \lambda_{I_1,I_2}^{1,\alpha} \Big) \log \rho$$
$$\Phi_{24} \triangleq (1-\alpha) \lambda_{N,N}^{\alpha,1} \log \rho.$$

Finally, combining (132) and (133), gives

$$\begin{aligned} &d_1 + d_2 \\ &\le \frac{1}{2\log\rho} \Big[2\Phi_{10} + 2\Phi_{11} + \Phi_{12} + \Phi_{13} + \Phi_{14} + 2\Phi_{20} + 2\Phi_{21} + \Phi_{22} + \Phi_{23} + \Phi_{24} \Big] \\ &= \frac{1}{2} \Big[2 \Big(\sum_{(I_1,I_2):I_1\neq N, I_2\neq N} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \\ &\quad + 2 \Big(\sum_{(I_1,I_2)\in\{(N,P),(N,D)\}} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \\ &\quad + \sum_{(I_1,I_2):I_2=N} \sum_{\forall (A_1,A_2)} A_1 \lambda_{I_1,I_2}^{A_1,A_2} \\ &\quad + \sum_{(I_1,I_2)\in\{(N,P),(N,D)\}} (1-\alpha) \lambda_{I_1,I_2}^{\alpha,1} + (1-\alpha) \lambda_{N,N}^{1,\alpha} \\ &\quad + 2 \Big(\sum_{(I_1,I_2):I_1\neq N, I_2\neq N} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \\ &\quad + 2 \Big(\sum_{(I_1,I_2)\in\{(P,N),(D,N)\}} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \Big) \\ &\quad + \sum_{(I_1,I_2):I_1=N} \sum_{\forall (A_1,A_2)} A_2 \lambda_{I_1,I_2}^{A_1,A_2} \\ &\quad + \sum_{(I_1,I_2)\in\{(P,N),(D,N)\}} (1-\alpha) \lambda_{I_1,I_2}^{1,\alpha} + (1-\alpha) \lambda_{N,N}^{\alpha,1} \Big] \\ &= \sum_{(I_1,I_2):I_1\neq N, I_2\neq N} (1+\alpha) \big(\lambda_{I_1,I_2}^{1,\alpha} + \lambda_{I_1,I_2}^{\alpha,1} \big) \\ &\quad + \sum_{(I_1,I_2)\in\{(N,P),(P,N),(N,D),(D,N)\}} \frac{2+\alpha}{2} \big(\lambda_{I_1,I_2}^{1,\alpha} + \lambda_{I_1,I_2}^{\alpha,1} \big) + \big(\lambda_{N,N}^{1,\alpha} + \lambda_{N,N}^{\alpha,1} \big) \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{(I_1,I_2):I_1\neq N, I_2\neq N} \big(2\lambda^{1,1}_{I_1,I_2} + 2\alpha\lambda^{\alpha,\alpha}_{I_1,I_2}\big) \\
&+ \sum_{(I_1,I_2)\in\{(N,P),(P,N),(N,D),(D,N)\}} \big(\frac{3}{2}\lambda^{1,1}_{I_1,I_2} + \frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{I_1,I_2}\big) + \big(\lambda^{1,1}_{N,N} + \alpha\lambda^{\alpha,\alpha}_{N,N}\big) \\
&= \big(1+\alpha\big)\big(\lambda^{1,\alpha}_{P,P} + \lambda^{\alpha,1}_{P,P}\big) + \big(1+\alpha\big)\big(\lambda^{1,\alpha}_{P\leftrightarrow D} + \lambda^{\alpha,1}_{P\leftrightarrow D}\big) + \big(1+\alpha\big)\big(\lambda^{1,\alpha}_{D,D} + \lambda^{\alpha,1}_{D,D}\big) \\
&+ \frac{2+\alpha}{2}\big(\lambda^{1,\alpha}_{P\leftrightarrow N} + \lambda^{\alpha,1}_{P\leftrightarrow N}\big) + \frac{2+\alpha}{2}\big(\lambda^{1,\alpha}_{D\leftrightarrow N} + \lambda^{\alpha,1}_{D\leftrightarrow N}\big) + \big(\lambda^{1,\alpha}_{N,N} + \lambda^{\alpha,1}_{N,N}\big) \\
&+ \big(2\lambda^{1,1}_{P,P} + 2\alpha\lambda^{\alpha,\alpha}_{P,P}\big) + \big(2\lambda^{1,1}_{P\leftrightarrow D} + 2\alpha\lambda^{\alpha,\alpha}_{P\leftrightarrow D}\big) + \big(2\lambda^{1,1}_{D,D} + 2\alpha\lambda^{\alpha,\alpha}_{D,D}\big) \\
&+ \big(\frac{3}{2}\lambda^{1,1}_{P\leftrightarrow N} + \frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{P\leftrightarrow N}\big) + \big(\frac{3}{2}\lambda^{1,1}_{D\leftrightarrow N} + \frac{3\alpha}{2}\lambda^{\alpha,\alpha}_{D\leftrightarrow N}\big) + \big(\lambda^{1,1}_{N,N} + \alpha\lambda^{\alpha,\alpha}_{N,N}\big)
\end{aligned} \tag{134}
$$

which completes the proof.

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 and P. Elia, "Can imperfect delayed CSIT be as useful as perfect delayed CSIT? DoF analysis and constructions for the BC," in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2012.

[24] 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.

[25] J. Chen and P. Elia, "Symmetric two-user MIMO BC and IC with evolving feedback," Jun. 2013, available on arXiv:1306.3710.

[26] 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.

[27] 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.

[28] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, "Capacity results for binary fading interference channels with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6093 – 6130, Oct. 2014.

[29] 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.

[30] J. Chen, S. Yang, A. Özgür, and A. Goldsmith, "Outdated CSIT can achieve full DoF in heterogeneous parallel channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2014.

[31] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529 – 568, Jan. 2014.

[32] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[33] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534 – 5562, Dec. 2008.
[34] C. S. Vaze, S. Karmakar, and M. K. Varanasi, "On the generalized degrees of freedom region of the MIMO interference channel with no CSIT," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011.
[35] S. Karmakar and M. K. Varanasi, "The generalized degrees of freedom of the MIMO interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011.
[36] ——, "The generalized multiplexing gain region of the slow fading MIMO interference channel and its achievability with limited feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012.
[37] ——, "The generalized degrees of freedom region of the MIMO interference channel and its achievability," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7188 – 7203, Dec. 2012.
[38] S. Gherekhloo, A. Chaaban, and A. Sezgin, "Topological interference management with alternating connectivity: The wyner-type three user interference channel," Oct. 2013, available on arXiv:1310.2385.
[39] C. Huang, V. R. Cadambe, and S. A. Jafar, "Interference alignment and the generalized degrees of freedom of the X channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5130 – 5150, May 2012.
[40] P. Mohapatra and C. Murthy, "On the generalized degrees of freedom of the K-user symmetric MIMO Gaussian interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 2188 –2192.
[41] J. Chen, S. Yang, A. Özgür, and A. Goldsmith, "Achieveing full DoF in heterogeneous parallel channels," 2014, in preparation.
[42] G. Dueck, "Partial feedback for two-way and broadcast channels," *Information and Control*, vol. 46, pp. 1 – 15, 1980.
[43] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1329 – 1345, Mar. 2013.
[44] H. Kim, Y.-K. Chia, and A. E. Gamal, "A note on broadcast channels with stale state information at the transmitter," 2013, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1309.7437 [cs.IT].
[45] C.-C. Wang, "On the capacity of 1-to-k broadcast packet erasure channels with channel output feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931 – 956, Feb. 2012.
[46] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback-capacity and algorithms," 2009, in Proc. 5th Workshop on Network Coding, Theory and Applications, pages 54 - 61, Lausanne, Switzerland.
[47] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 4 – 21, Apr. 2003.
[48] R. Venkataramanan and S. S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6175 – 6191, Oct. 2013.
[49] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.
[50] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 379–381, Apr. 1978.