Toward the Performance vs. Feedback Tradeoff for the Two-User MISO Broadcast Channel

Jinyuan Chen and Petros Elia

***Abstract*—For the two-user MISO broadcast channel with imperfect and delayed channel state information at the transmitter (CSIT), the work explores the tradeoff between performance, and CSIT timeliness and quality. The work considers a broad setting where communication takes place in the presence of a random fading process, and in the presence of a feedback process that, at any point in time, provides CSIT estimates - of some arbitrary quality - for any past, current or future channel realization. Under standard assumptions, the work derives the degrees-of-freedom (DoF) region, which is optimal for a large regime of sufficiently good (but potentially imperfect) delayed CSIT. This region concisely captures the effect of channel correlations, the quality of predicted, current, and delayed-CSIT, as well as concisely captures the effect of the quality of CSIT offered at any time, about any channel.**

The bounds are met with novel schemes which - in the context of imperfect and delayed CSIT - introduce here for the first time, encoding and decoding with a phase-Markov structure. The results hold for a large class of block and non-block fading channel models, and they unify and extend many prior attempts to capture the effect of imperfect and delayed feedback. This generality also allows for consideration of novel pertinent settings, such as the new periodically evolving feedback setting, where a gradual accumulation of feedback bits progressively improves CSIT as time progresses across a finite coherence period.

I. Introduction

A. Channel model

We consider the multiple-input single-output broadcast channel (MISO BC) with an M-transmit antenna ($M \geq 2$) transmitter communicating to two receiving users with a single receive antenna each. Let $\boldsymbol{h}_t, \boldsymbol{g}_t$ denote the channel of the first and second user respectively at time t, and let $\boldsymbol{x}_t$ denote the transmitted vector at time t, satisfying a power constraint $\mathbb{E}[||\boldsymbol{x}_t||^2] \leq P$, for some power P which also here takes the role of the signal-to-noise ratio (SNR). Here $\boldsymbol{h}_t$ and $\boldsymbol{g}_t$ are drawn from a random distribution, such that each has zero mean and identity covariance (spatially uncorrelated), and such that $\boldsymbol{h}_t$ is linearly independent of $\boldsymbol{g}_t$ with probability 1.

In this setting, the corresponding received signals at the first and second user take the form

$$y_t^{(1)} = \boldsymbol{h}_t^{\mathsf{T}} \boldsymbol{x}_t + z_t^{(1)} \quad (1)$$

$$y_t^{(2)} = \boldsymbol{g}_t^{\mathsf{T}} \boldsymbol{x}_t + z_t^{(2)} \quad (2)$$

$(t = 1, 2, \cdots)$, where $z_t^{(1)}, z_t^{(2)}$ denote the unit power AWGN noise at the receivers.

In the high-SNR setting of interest, for an achievable rate pair (R_1, R_2) for the first and second user respectively, the corresponding degrees-of-freedom (DoF) pair (d_1, d_2) is given by

$$d_i = \lim_{P\to\infty} \frac{R_i}{\log P}, \ i = 1, 2$$

and the corresponding DoF region is then the set of all achievable DoF pairs.

B. Delay-and-quality effects of feedback

As in many multiuser wireless communications scenarios, the performance of the broadcast channel depends on the timeliness and quality of channel state information at the transmitter (CSIT). This timeliness and quality though may be reduced by limited-capacity feedback links, which may offer feedback with consistently low quality and high delays, i.e., feedback that offers an inaccurate representation of the true state of the channel, as well feedback that can only be used for an insufficient fraction of the communication duration. The corresponding performance degradation, as compared to the case of having perfect feedback without delay, forces the delay-and-quality question of how much feedback quality is necessary, and when, in order to achieve a certain performance.

C. Channel process and feedback process with predicted, current, and delayed CSIT

We here consider communication of an infinite duration n, a channel fading process $\{\boldsymbol{h}_t, \boldsymbol{g}_t\}_{t=1}^n$ drawn from a statistical distribution, and a feedback process that provides CSIT estimates $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t,t'=1}^n$ (of channel $\boldsymbol{h}_t, \boldsymbol{g}_t$) at any time t' - before, during, or after materialization of $\boldsymbol{h}_t, \boldsymbol{g}_t$ at time t - and does so with quality defined by the statistics of

$$\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n \quad (3)$$

where we consider these estimation errors to have zero-mean circularly-symmetric complex Gaussian entries.

1) Predicted, current, and delayed CSIT: For the channel $\boldsymbol{h}_t, \boldsymbol{g}_t$ at time t, the set of all estimates $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t'}$ form what can be described as the set of *delayed CSIT* comprising of estimates that are not available at time t, the set of *current estimates* $\hat{\boldsymbol{h}}_{t,t}, \hat{\boldsymbol{g}}_{t,t}$ at time t, and the set of *predicted estimates* $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t'<t}$. Predicted CSIT may potentially allow for reduction of the effect of future interference, current CSIT may be used to orthogonalize the channels of the users, while

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 257616 (CONECT), from the FP7 CELTIC SPECTRA project, and from Agence Nationale de la Recherche project ANR-IMAGENET.

J. Chen and P. Elia are with the Mobile Communications Department, EURECOM, Sophia Antipolis, France (email: {chenji, elia}@eurecom.fr)

This paper was presented in part at the 50th Annual Allerton Conference 2012, at ITA-UCSD 2013, and will be presented in part at ISIT 2013.

delayed CSIT may facilitate retrospective compensation for the lack of perfect quality feedback.

Any attempt to capture and meet the tradeoff between performance, and feedback timeliness and quality, must naturally consider the statistics of the channel and of CSIT precision $\{(\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}),(\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$ at any point, about any channel.

D. Notation, conventions and assumptions

We will use the notation

$$\alpha_t^{(1)} \triangleq -\lim_{P\to\infty} \frac{\log \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t}||^2]}{\log P} \tag{4}$$

$$\alpha_t^{(2)} \triangleq -\lim_{P\to\infty} \frac{\log \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t}||^2]}{\log P} \tag{5}$$

to describe the *current quality exponent* for the two users ($\alpha_t^{(1)}$ is for user 1), while we will use

$$\beta_t^{(1)} \triangleq -\lim_{P\to\infty} \frac{\log \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t+\eta}||^2]}{\log P} \tag{6}$$

$$\beta_t^{(2)} \triangleq -\lim_{P\to\infty} \frac{\log \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t+\eta}||^2]}{\log P} \tag{7}$$

- for any sufficiently large but finite integer $\eta > 0$ - to denote the *delayed quality exponents* for each user. The assumption that η is finite, reflects the fact that we only consider delayed CSIT that arrives up to a certain finite time from the moment the channel materializes. In words, $\alpha_t^{(1)}$ measures the quality of the CSIT (about $\boldsymbol{h}_t$) that is available at time t, while $\beta_t^{(1)}$ measures the (best) quality of the CSIT (about $\boldsymbol{h}_t$) which arrives strictly after the channel appears, i.e., strictly after time t (similarly $\alpha_t^{(2)}, \beta_t^{(2)}$ for the channel $\boldsymbol{g}_t$ of the second user).

It is easy to see that without loss of generality, in the DoF setting of interest, we can restrict our attention to the range [1]

$$0 \le \alpha_t^{(i)} \le \beta_t^{(i)} \le 1 \tag{8}$$

where $\beta_t^{(i)} = 1$ corresponds to having (essentially) perfect delayed CSIT for $\boldsymbol{h}_t, \boldsymbol{g}_t$, and where $\alpha_t^{(1)} = \alpha_t^{(2)} = 1$, corresponds to the optimal case of perfect current (full) CSIT.

Furthermore we will use the notation

$$\bar{\alpha}^{(i)} \triangleq \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \alpha_t^{(i)}, \quad \bar{\beta}^{(i)} \triangleq \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \beta_t^{(i)}, \quad i = 1,2 \tag{9}$$

to denote the average of the quality exponents. At this point we note that our results, specifically the achievability part, will hold under the soft assumption that any sufficiently long subsequence $\{\alpha_t^{(1)}\}_{t=\tau}^{\tau+T}$ (resp. $\{\alpha_t^{(2)}\}_{t=\tau}^{\tau+T}, \{\beta_t^{(1)}\}_{t=\tau}^{\tau+T}, \{\beta_t^{(2)}\}_{t=\tau}^{\tau+T}$) has an average that converges to the long term average $\bar{\alpha}^{(1)}$ (resp. $\bar{\alpha}^{(2)}, \bar{\beta}^{(1)}, \bar{\beta}^{(2)}$), for any τ and for some finite T that can be chosen to be sufficiently large to allow for the above convergence.

[1]To see this, we recall from [1], [2] that under a peak-power constraint of P, having CSIT estimation error in the order of P^{-1} causes no DoF reduction as compared to the perfect CSIT case. In our DoF high-SNR setting of interest where $P >> n$, this same observation also holds under an average power constraint of P. The fact that $\alpha_t^{(i)} \le \beta_t^{(i)}$ comes naturally from the fact that one can recall, at a later time, statistically good estimates.

Throughout this paper, $(\bullet)^{\mathsf{T}}$, $(\bullet)^{\mathsf{H}}$ and $||\bullet||_F$ will denote the transpose, conjugate transpose and Frobenius norm of a matrix respectively, while $\mathrm{diag}(\bullet)$ will denote a diagonal matrix, $||\bullet||$ will denote the Euclidean norm, and $|\bullet|$ will denote the magnitude of a scalar. $o(\bullet)$ comes from the standard Landau notation, where $f(x) = o(g(x))$ implies $\lim_{x\to\infty} f(x)/g(x) = 0$. We will also use $\doteq$ to denote *exponential equality*, i.e., we write $f(P) \doteq P^B$ to denote $\lim\limits_{P\to\infty} \dfrac{\log f(P)}{\log P} = B$. Similarly $\dot{\ge}$ and $\dot{\le}$ will denote exponential inequalities. Logarithms are of base 2.

We adhere to the common convention (see [3]–[6]) of assuming perfect and global knowledge of channel state information at the receivers (perfect global CSIR), where the receivers know all channel states and all estimates. We also adhere to the common convention (see [4], [5], [7], [8]) of assuming that the current estimation error is statistically independent of current and past estimates, and consequently the input signal is a function of the message and of the CSIT. This assumption fits well with many channel models spanning from the fast fading channel (i.i.d. in time), to the correlated channel model as this is considered in [8], to the quasi-static block fading model where the CSIT estimates are successively refined while the channel remains static (see [1], see also the discussion in the appendix in Section VIII). Additionally we consider the entries of each estimation error vector $\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}$ (similarly of $\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'}$) to be i.i.d. Gaussian, clarifying though that we are just referring to the M entries in each such specific vector $\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}$, and that we do not suggest that the error entries are i.i.d. in time or across users.

Finally we safely assume that $\mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}||^2] \le \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t''}||^2]$ (similarly $\mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'}||^2] \le \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t''}||^2]$), for any $t' > t''$. This assumption - which simply suggests that one can revert back to past estimates of statistically better quality - is used here for simplicity of notation, and can be removed, after a small change in the definition of the quality exponents, without an effect to the main result.

E. Prior work

The delay-and-quality effects of feedback, naturally fall between the two extreme cases of no CSIT and of full CSIT (immediately available and perfect CSIT), with full CSIT allowing for the optimal 1 DoF per user (cf. [9]), while the absence of any CSIT reduces this to just $1/2$ DoF per user (cf. [10], [11]).

Toward bridging this gap, different works have considered the use of imperfect and delayed feedback. For example, the work by Lapidoth, Shamai and Wigger in [7] considered the case where the amount of feedback is limited to the extend that the channel-estimation error power does not vanish with increasing SNR, in the sense that $\lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t}||^2])/\log P = \lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t}||^2])/\log P = 0$. In this setting - which corresponds to the case here where $\alpha_t^{(1)} = \alpha_t^{(2)} = \beta_t^{(1)} = \beta_t^{(2)} = 0, \ \forall t$ - the work in [7] showed that the symmetric DoF is upper bounded by $2/3$ DoF per user, again under the assumption that the input signaling is independent of the estimation error. It is worth noting that finding the exact DoF in this zero-exponent setting, remains - to the best of our knowledge - an open problem.

At the other extreme, the work by Caire et al. [2] (see also the work of Jindal [1], as well as of Lapidoth and Shamai [12]) showed that having immediately available CSIT estimates with estimation error power that is in the order of P^{-1} - i.e., having $-\lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t}||^2])/\log P = -\lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t}||^2])/\log P = 1$, corresponding here to having $\alpha_t^{(1)} = \alpha_t^{(2)} = 1, \ \forall t$ - causes no DoF reduction as compared to the perfect CSIT case, and can thus achieve the optimal 1 DoF per user.

A valuable tool toward bridging this gap and further understanding the delay-and-quality effects of feedback, came with the work by Maddah-Ali and Tse in [3] which showed that arbitrarily delayed feedback can still allow for performance improvement over the no-CSIT case. In a fast-fading block-fading setting, the work differentiated between current and delayed CSIT - with delayed CSIT defined in [3] as the CSIT which is available after the channel's coherence period - and showed that delayed and completely obsolete CSIT, even without any current CSIT, allows for an improved $2/3$ DoF per user. This setting corresponded to having $\alpha_t^{(1)} = \alpha_t^{(2)} = 0, \ \forall t$, and the optimal $2/3$ DoF was achieved with a scheme that only asked for delayed CSIT for one out of three channels, thus corresponding to

$$\beta_t^{(i)} = \begin{cases} 1 & \text{if } t = i \pmod 3 \\ 0 & \text{otherwise} \end{cases}, \ i = 1, 2 \tag{10}$$

and which allowed for $\bar{\beta}^{(i)} = 1/3$, for user $i = 1, 2$. In the above, where we say $t = i \pmod 3$, we refer to the modulo operation, i.e., we mean that $t = 3k + i$ for some integer k.

Within the same block-fading context of delayed vs. current CSIT, the work by Kobayashi et al., Yang et al., and Gou and Jafar [4], [5], [8], quantified the usefulness of combining delayed and completely obsolete CSIT with immediately available but imperfect CSIT of a certain quality $\alpha = -\lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t}||^2])/\log P = -\lim_{P\to\infty}(\log \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t}||^2])/\log P$ that remained unchanged throughout the communication process. In this setting - which corresponded to having $\alpha_t^{(1)} = \alpha_t^{(2)} = \alpha, \ \forall t$ - the optimal $(2+\alpha)/3$ symmetric DoF was achieved with a scheme that only asked for delayed CSIT for one out of three channels, thus corresponding to

$$\beta_t^{(i)} = \begin{cases} 1 & \text{if } t = i \pmod 3 \\ \alpha & \text{otherwise} \end{cases} \tag{11}$$

and which allowed for $\bar{\beta}^{(i)} = (1 + 2\alpha)/3, \ i = 1, 2$.

An interesting approach was introduced by Tandon et al. in [13] who considered the fast-fading two-user MISO BC setting, where each user's CSIT changes every coherence period by alternating between the three extreme states of perfect current CSIT, perfect delayed CSIT, and no CSIT.

Additionally, Lee and Heath in [14] considered, in the setting of the quasi-static block-fading channel, the possibility that current CSIT may be available only after a certain fraction γ of a finite-duration coherence period T_c.

Other work such as that by Maleki et al. in [6] considered, again in the MISO BC context, an asymmetric setting where both users offered perfect delayed CSIT, but where only one user offered perfect current CSIT while the other user offered no current CSIT. In this setting - which corresponds to having $\alpha_t^{(1)} = \beta_t^{(1)} = 1, \alpha_t^{(2)} = 0, \ \forall t$ - the optimal DoF corner point $(1, 1/2)$ (sum-DoF $d_1 + d_2 = 3/2$) was achieved with a scheme that asked for delayed CSIT for every other channel, specifically corresponding to

$$\beta_t^{(2)} = \begin{cases} 1 & \text{if } t = 0 \pmod 2 \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Other works, in the context of delayed or imperfect CSIT, include [15]–[29].

F. Structure of paper

Section II will give the main result of this work by describing, under the aforementioned common assumptions, the DoF offered by a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^n$ of a certain quality, as this is defined by the statistics of $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t=1,t'=1}^n$. Specifically Theorem 1 provides the optimal DoF for a large range of 'sufficiently good' delayed CSIT, whereas Proposition 1 focuses on the case of low quality delayed CSIT. In the same section, Corollary 1a describes the DoF for the symmetric case where $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)}$ and $\bar{\beta}^{(1)} = \bar{\beta}^{(2)}$, and immediately after that, Corollary 1b explores the benefits of symmetry, by quantifying the extent to which having similar feedback quality for the two users, offers a benefit over the asymmetric case where one user has generally more feedback than the other. Corollary 1c offers insight on the need for delayed CSIT, and shows how, paradoxically, having reduced $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$ allows - to a certain extent - for smaller $\bar{\beta}^{(1)}, \bar{\beta}^{(2)}$. On the other hand, Corollary 1d offers insight on the need for using predicted channel estimates (forecasting channel states in advance), by showing that - in terms of achieving the optimal DoF performance, and in the presence of sufficiently good delayed CSIT - employing predicted CSIT is unnecessary.

Section III highlights the newly considered *periodically evolving feedback* setting over the quasi-static block fading channel, where a gradual accumulation of feedback results in a progressively increasing CSIT quality as time progresses across a finite coherence period. This setting is powerful as it captures the many feedback options that one may have in a block-fading environment where the *statistical* nature of feedback remains largely unchanged across coherence periods. As such, it captures existing settings that have been of particular interest, such as the Maddah-Ali and Tse setting in [3], the Yang et al. and Gou and Jafar setting in [4], [5], the Lee and Heath 'not-so-delayed CSIT' setting in [14], and the asymmetric setting in [6]. In this section we offer examples which - under very clearly specified assumptions - offer insight on how many feedback bits to inject, and when, in order to achieve a certain performance. In the same section, smaller results and examples offer further insight - again in the context of periodically evolving feedback over a quasi-static channel - like for example the result in Corollary 1f which bounds the quality of current and of delayed CSIT needed to achieve a certain target symmetric DoF, and in the process offers insight on when delayed feedback is entirely unnecessary, in the sense that there is no need to wait for feedback that arrives after the end of the coherence period of the channel. Similarly Corollary 1g offers insight on the feedback delays that allow for a given target symmetric DoF in the presence of constraints on current and delayed CSIT qualities. Finally Corollary 1h generalizes the pertinent result in the asymmetric setting in [6].

Section IV corresponds to the achievability part of the proof, and presents the general communication scheme that utilizes the available information of a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$, to achieve the corresponding DoF corner points. This is done - by properly employing different combinations of zero forcing, superposition coding, interference compressing and broadcasting, as well as specifically tailored power and rate allocation - in order to transmit private information, using currently available CSIT estimates to reduce interference, and using delayed CSIT estimates to alleviate the effect of past interference. The scheme has a phase-Markov structure which - in the context of imperfect and delayed CSIT, was first introduced in [30], [31] - and which quantizes the accumulated interference of a certain period of time, broadcasts it in the future, together with common information that will then help resolve the accumulated interference of the past.
After the description of the scheme in its general form, and the explicit description of how the scheme achieves the different DoF corner points, Section IV-D provides example schemes - distilled from the general scheme - for specific settings such as the imperfect-delayed CSIT setting, the (extended) alternating CSIT setting of Tandon et al. [13], as well as discusses schemes with finite and small delay.

Section V provides the details of the outer bound, Section VI offers concluding remarks, the appendices in Section VII and Section IX offer details on the proofs, while the appendix in Section VIII offers a discussion on some of the assumptions employed in this work.

In the end, the above results provide insight on pertinent questions such as:

- What CSIT feedback quality should be provided, and when, in order to achieve a certain target DoF performance? (Theorem 1)
- When is delayed feedback unnecessary? (Corollary 1f)
- Is there any gain in early prediction of future channels? (Corollary 1d)
- What current-CSIT and delayed-CSIT qualities suffice to achieve a certain performance? (Corollary 1f)
- Can imperfect-quality delayed CSIT achieve the same optimality that was previously attributed to sending perfect delayed CSIT? (Corollary 1c)
- How much more valuable are feedback bits that are sent early, than those sent late? (Section III)
- In the quasi-static block-fading case, is it better to send less feedback early, or more feedback later? (Section III)
- What is the effect of having asymmetric feedback links, and when can we have a 'symmetry gain'? (Corollary 1b)

II. DoF REGION OF THE MISO BC

We proceed with the main DoF results, which are proved in Section IV (inner bound) and Section V (outer bound).

We here remind the reader of the sequences $\{\alpha_t^{(1)}\}_{t=1}^n, \{\alpha_t^{(2)}\}_{t=1}^n, \{\beta_t^{(1)}\}_{t=1}^n, \{\beta_t^{(2)}\}_{t=1}^n$ of quality exponents, as these were defined in (4)-(7), as well as of the corresponding averages $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}, \bar{\beta}^{(1)}, \bar{\beta}^{(2)}$ from (9). We also remind the reader that we consider communication over a large time duration n. We henceforth label the users so that $\bar{\alpha}^{(2)} \leq \bar{\alpha}^{(1)}$.

Extending the work in [4] that focused on CSIT with invariant and symmetric quality, we first proceed to construct a new DoF outer bound that supports our setting. The proof can be found in Section V.

Lemma 1: The DoF region of the two-user MISO BC with a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$ of quality $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t=1,t'=1}^{n}$, is upper bounded as

$$d_1 \leq 1, \quad d_2 \leq 1 \tag{13}$$
$$2d_1 + d_2 \leq 2 + \bar{\alpha}^{(1)} \tag{14}$$
$$2d_2 + d_1 \leq 2 + \bar{\alpha}^{(2)}. \tag{15}$$

The following theorem provides the optimal DoF for a large range of 'sufficiently good' delayed CSIT.

Theorem 1: The optimal DoF region of the two-user MISO BC with a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$ of quality $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t=1,t'=1}^{n}$ is given by

$$d_1 \leq 1, \quad d_2 \leq 1 \tag{16}$$
$$2d_1 + d_2 \leq 2 + \bar{\alpha}^{(1)} \tag{17}$$
$$2d_2 + d_1 \leq 2 + \bar{\alpha}^{(2)} \tag{18}$$

for any sufficiently good delayed-CSIT process such that $\min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\} \geq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}, \frac{1+\bar{\alpha}^{(2)}}{2}\}$.

The achievability part of the proof can be found in Section IV which explicitly describes the scheme that achieves the corresponding corner points that match those of the outer bound in Lemma 1.

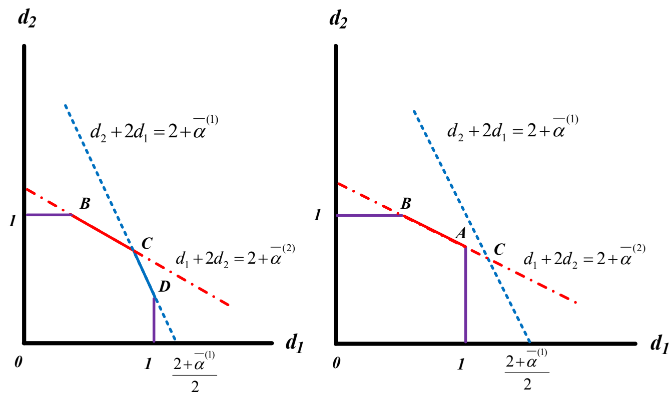
Figure 1 corresponds to the main result in the theorem. The above result is complemented by the following proposition. The proof is in Section IV which describes the scheme that achieves these DoF corner points.

Proposition 1: For a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$ for which $\min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\} < \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}, \frac{1+\bar{\alpha}^{(2)}}{2}\}$, the DoF region is inner bounded by the polygon described by

$$d_1 \leq 1, \quad d_2 \leq 1 \tag{19}$$
$$2d_1 + d_2 \leq 2 + \bar{\alpha}^{(1)} \tag{20}$$
$$2d_2 + d_1 \leq 2 + \bar{\alpha}^{(2)} \tag{21}$$
$$d_1 + d_2 \leq 1 + \min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\}. \tag{22}$$

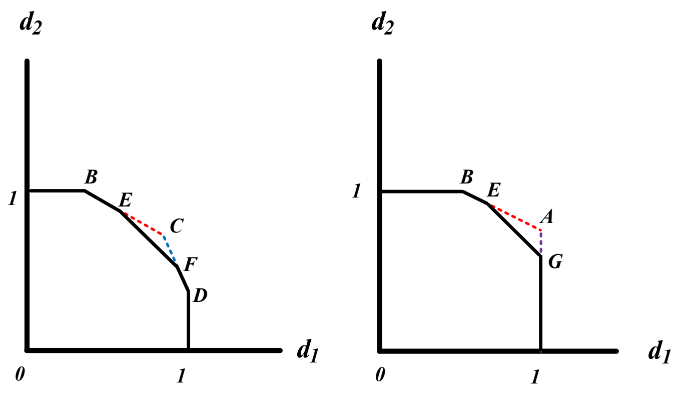
Figure 2 corresponds to the result in Proposition 1.

Before proceeding to specific corollaries that offer further insight, it is worth making a comment on the fact that the entire complexity of the problem is captured in the quality exponents.



(a) Case 1: $2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}<1$ (b) Case 2: $2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}\geq 1$

Fig. 1. Optimal DoF regions for the two-user MISO BC, for the case of $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} \geq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$. The corner points take the following values: $A = (1,\frac{1+\bar{\alpha}^{(2)}}{2})$, $B = (\bar{\alpha}^{(2)},1)$, $C = (\frac{2+2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}}{3},\frac{2+2\bar{\alpha}^{(2)}-\bar{\alpha}^{(1)}}{3})$, and $D = (1,\bar{\alpha}^{(1)})$.



(a) Case 1: $2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}<1$ (b) Case 2: $2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}\geq 1$

Fig. 2. DoF regions for the two-user MISO BC with $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} \leq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$. The corner points take the following values: $E = (2\bar{\delta}-\bar{\alpha}^{(2)},1+\bar{\alpha}^{(2)}-\bar{\delta})$, $F = (1+\bar{\alpha}^{(1)}-\bar{\delta},2\bar{\delta}-\bar{\alpha}^{(1)})$, and $G = (1,\bar{\delta})$, where $\bar{\delta} \triangleq \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)},\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$.

Remark 1: The results suggest that the quality exponents capture - in the setting of interest, and under our assumptions - the effect of the statistics of the CSIT precision $\{(\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}),(\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$. This is indeed the case since the following two hold. Firstly, given the Gaussianity of the estimation errors, the statistics of $\{(\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}),(\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$ are captured by the $2n^2\times 2n^2$ covariance matrix[2] of the $2n^2$-length vector consisting of the elements $\{(\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}),(\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$. The diagonal entries of this covariance matrix are simply $\{\frac{1}{M}\mathbb{E}[||\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}||^2],\frac{1}{M}\mathbb{E}[||\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'}||^2]\}_{t,t'=1}^n$. Secondly, the outer bound has kept open the possibility of any off-diagonal elements (as we will see in (97), (98)), thus the outer bound holds irrespective of the off-diagonal elements of this covariance matrix. Thus, under our assumptions, the essence of the statistics is captured by $\{\mathbb{E}[||\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}||^2],\mathbb{E}[||\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'}||^2]\}_{t,t'=1}^n$, and its effect is captured - in the high-SNR regime - by the quality exponents.

[2]This size of the covariance matrix reflects the fact that the M entries of each $\boldsymbol{h}_t-\hat{\boldsymbol{h}}_{t,t'}$ are i.i.d. (similarly of $\boldsymbol{g}_t-\hat{\boldsymbol{g}}_{t,t'}$). Please note that we refer to independence across the *spatial* dimensions of the channel of one user, and certainly not across time.

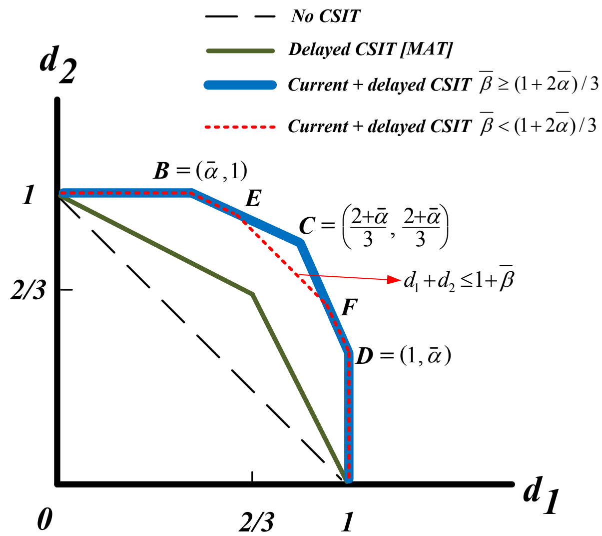


Fig. 3. DoF region of two-user MISO BC with symmetric feedback, $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} = \bar{\alpha}, \bar{\beta}^{(1)} = \bar{\beta}^{(2)} = \bar{\beta}$. The optimal region takes the form of a polygon with corner points $\{(0,0),(0,1),(\bar{\alpha},1),(\frac{2+\bar{\alpha}}{3},\frac{2+\bar{\alpha}}{3}),(1,\bar{\alpha}),(1,0)\}$ for $\bar{\beta} \geq \frac{1+2\bar{\alpha}}{3}$. For $\bar{\beta} < \frac{1+2\bar{\alpha}}{3}$, the derived region takes the form of a polygon with corner points $\{(0,0),(0,1),(\bar{\alpha},1),(2\bar{\beta}-\bar{\alpha},1+\bar{\alpha}-\bar{\beta}),(1+\bar{\alpha}-\bar{\beta},2\bar{\beta}-\bar{\alpha}),(1,\bar{\alpha}),(1,0)\}$.

1) Symmetric vs. asymmetric feedback: We proceed to explore the special case of symmetric feedback where the accumulated feedback quality is similar across users, i.e., where the feedback links of user 1 and user 2 share the same exponent *averages* $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} = \bar{\alpha}$ and $\bar{\beta}^{(1)} = \bar{\beta}^{(2)} = \bar{\beta}$. Most existing works, with an exception in [6], fall under this setting. The following holds directly from Theorem 1 and Proposition 1.

Corollary 1a (DoF with symmetric feedback): The optimal DoF region for the symmetric case takes the form

$$d_1 \leq 1, \quad d_2 \leq 1, \quad 2d_1+d_2 \leq 2+\bar{\alpha}, \quad 2d_2+d_1 \leq 2+\bar{\alpha}$$

when $\bar{\beta} \geq \frac{1+2\bar{\alpha}}{3}$, while when $\bar{\beta} < \frac{1+2\bar{\alpha}}{3}$ this region is inner bounded by the achievable region

$$d_1 \leq 1, \quad d_2 \leq 1 \tag{23}$$
$$2d_1+d_2 \leq 2+\bar{\alpha} \tag{24}$$
$$2d_2+d_1 \leq 2+\bar{\alpha} \tag{25}$$
$$d_2+d_1 \leq 1+\bar{\beta}. \tag{26}$$

Figure 3 depicts the DoF region of the two-user MISO BC with symmetric feedback.

We now quantify the extent to which having symmetric feedback offers a benefit over the asymmetric case where one user has generally more feedback than the other. Different works have identified such instances where having symmetric feedback offers ('symmetry gains') over the asymmetric case (cf. [13], [6]).

The following broad comparison focuses on the case of perfect delayed CSIT, and contrasts the symmetric case $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)}$, to the asymmetric case $\bar{\alpha}^{(1)} \neq \bar{\alpha}^{(2)}$, under an overall feedback constraint $\bar{\alpha}^{(1)} + \bar{\alpha}^{(2)} = \Delta$, for any $\Delta \in (0, 2]$. The comparison is in terms of the optimal sum DoF $d_1 + d_2$, where again we recall that the users are labeled so that $\bar{\alpha}^{(1)} \geq \bar{\alpha}^{(2)}$. The proof is direct from Theorem 1 and Corollary 1a.

Corollary 1b (Symmetric vs. asymmetric feedback): Consider any fixed $\Delta \triangleq \bar{\alpha}^{(1)} + \bar{\alpha}^{(2)}$ in the range $(0, 2]$. If $2\bar{\alpha}^{(1)} - \bar{\alpha}^{(2)} - 1 \leq 0$, having symmetric feedback ($\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)}$) does not offer a sum-DoF gain over the asymmetric feedback case, while if $2\bar{\alpha}^{(1)} - \bar{\alpha}^{(2)} - 1 > 0$, there is a symmetric sum-DoF gain and it takes the form $\frac{2\bar{\alpha}^{(1)} - \bar{\alpha}^{(2)} - 1}{6}$.

Example 1: For example, under the constraint that $\bar{\alpha}^{(1)} + \bar{\alpha}^{(2)} = \Delta = 1$, the asymmetric $\bar{\alpha}^{(1)} = 1, \bar{\alpha}^{(2)} = 0$ gives an optimal sum DoF of $d_1 + d_2 = 3/2$ (Theorem 1 with perfect delayed CSIT), whereas the symmetric $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} = 0.5$ gives $d_1 + d_2 = 5/3$, and a sum-DoF gain of $5/3 - 3/2 = 1/6$.

2) Need for delayed feedback: Imperfect vs. perfect delayed CSIT: We here show that imperfect delayed CSIT can be as useful as perfect delayed CSIT, and provide insight on the overall feedback quality (timely and delayed) that is necessary to achieve a certain DoF performance.

Before proceeding with the result, we recall that the distinction between timely and delayed CSIT, has to do more with feedback timing rather than feedback quality, and that $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$ are more representative of the quality of *timely* feedback, while $\bar{\beta}^{(1)}, \bar{\beta}^{(2)}$ are more representative of the quality of the entirety of feedback (timely plus delayed). In this sense, any attempt to limit the total amount and quality of feedback - that is communicated during a certain communication process - must try to limit $\bar{\beta}^{(1)}, \bar{\beta}^{(2)}$, and not just focus on reducing $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$. For example, having to always send perfect delayed CSIT ($\beta_t^{(1)} = \beta_t^{(2)} = 1, \forall t$) does little to reduce the total amount of feedback, and mainly shifts the time-frame of the problem, again irrespective of possible savings in $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$.

As we will see though, having reduced $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$ translates to needing lesser quality delayed feedback, i.e., having reduced $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}$ paradoxically allows - to a certain extent - for smaller $\bar{\beta}^{(1)}, \bar{\beta}^{(2)}$. This is quantified in the following. The proof is direct, as the following simply restates what is in the Theorem.

Corollary 1c (Imperfect vs. perfect delayed CSIT): A CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$ that offers

$$\min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\} \geq \min\{\frac{1 + \bar{\alpha}^{(1)} + \bar{\alpha}^{(2)}}{3}, \frac{1 + \bar{\alpha}^{(2)}}{2}\} \quad (27)$$

gives the same DoF as a CSIT process that offers perfect delayed CSIT for each channel realization ($\beta_t^{(1)} = \beta_t^{(2)} = 1, \forall t$). For the symmetric case, having

$$\bar{\beta} \geq \frac{1 + 2\bar{\alpha}}{3} \quad (28)$$

guarantees the same.

It is interesting to note that the expressions in the above corollary match the amount of delayed CSIT used by schemes in the past, even though the schemes were not designed with the expressed purpose of reducing the amount of delayed CSIT. For example, the Maddah-Ali and Tse scheme in [3] used delayed CSIT as shown in (10), which happens to match the above expression in (28). This same expression in (28) additionally tells us that any combination of CSIT quality exponents that allows for $\bar{\beta}^{(1)} = \bar{\beta}^{(2)} \geq 1/3$, will allow for the same optimal DoF in [3]. The same observation holds for the schemes in [4], [5] which used delayed CSIT as shown in (11), which again matches (28), which in turn reveals that any combination of CSIT quality exponents that allows for $\bar{\beta}^{(1)} = \bar{\beta}^{(2)} \geq (1 + 2\bar{\alpha})/3$, will achieve the same optimal DoF in [4], [5]. Similarly the asymmetric scheme in [6] used delayed CSIT as shown in (12), which[3] matches (27), which in turn reveals other combinations of CSIT quality exponents that allow for the same optimal DoF.

3) Need for predicted CSIT: We now shift emphasis from delayed CSIT to the other extreme of predicted CSIT. As we recall, we considered a channel process $\{\boldsymbol{h}_t, \boldsymbol{g}_t\}_t$ and a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t,t'}$, consisting of estimates $\hat{\boldsymbol{h}}_{t,t'}$ - available at any time t' - of the channel $\boldsymbol{h}_t$ that materializes at any time t. We also advocated that we can safely assume that $\mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}||^2] \leq \mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t''}||^2]$ (similarly $\mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'}||^2] \leq \mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t''}||^2]$), for any $t' > t''$, simply because one can revert back to past estimates of statistically better quality. This assumption though does not preclude the possible usefulness of early (predicted) estimates, even if such estimates are generally of lesser quality than current estimates (i.e., of lesser quality than estimates that appear during or after the channel materializes). The following gives insight on this.

Corollary 1d (Need for predicted CSIT): For sufficiently good delayed CSIT that guarantees $\min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\} \geq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}, \frac{1+\bar{\alpha}^{(2)}}{2}\}$, there is no DoF gain in using predicted CSIT. Specifically - for sufficiently good delayed CSIT, and in order to achieve the optimal DoF - transmission at a certain time t^*, does not need to consider any estimate $\hat{\boldsymbol{h}}_{t,t'}$ of a future channel $t > t^*$, where this estimate became available - naturally by prediction - at any time $t' \leq t^* < t$.

Proof: The proof is by construction; the designed schemes do not use predicted estimates, while the tight outer bound does not preclude the use of such predicted estimates. ■

III. Periodically evolving CSIT

We here focus on the block fading setting, and consider a gradual accumulation of feedback that results in a progressively increasing CSIT quality as time progresses across the coherence period (T_c channel uses - current CSIT), or at any time after the end of the coherence period (delayed CSIT)[4].

[3]From (12) we can conclude that the scheme in [6] asks for $\bar{\beta}^{(2)} = 1/2, \bar{\beta}^{(1)} = 1$, which matches the expression in (27) since $\min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}\} = \min\{1, 1/2\} = \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}, \frac{1+\min\{\bar{\alpha}^{(1)},\bar{\alpha}^{(2)}\}}{2}\} = \min\{\frac{1+1+0}{3}, \frac{1+0}{2}\} = 1/2$.

[4]This definition of current vs. delayed CSIT, originates from [3], and is the standard definition adopted by most existing works.

Such gradual improvement could be sought in FDD (frequency division duplex) settings with limited-capacity feedback links that can be used more than once during the coherence period to progressively refine CSIT, as well as in TDD (time division duplex) settings that use reciprocity-based estimation that progressively improves over time.

In this setting, the channel remains the same for a finite duration of T_c channel uses, and the time index is arranged so that

$$\boldsymbol{h}_{\ell T_c+1} = \boldsymbol{h}_{\ell T_c+2} = \cdots = \boldsymbol{h}_{(\ell+1)T_c}$$

$$\boldsymbol{g}_{\ell T_c+1} = \boldsymbol{g}_{\ell T_c+2} = \cdots = \boldsymbol{g}_{(\ell+1)T_c}$$

for a non-negative integer ℓ. Furthermore feedback quality is now periodic, as this is reflected in the current-CSIT quality exponents where

$$\alpha_t^{(i)} = \alpha_{\ell T_c+t}^{(i)}, \forall \ell = 0,1,2,\cdots,\ i = 1,2. \tag{29}$$

We focus here - simply for the sake of clarity of exposition - mainly on the symmetric case. In this setting - and after adopting a periodic time index corresponding to having $\ell = 0$ (cf. (29)) - the feedback quality is now represented by the T_c current CSIT quality exponents $\{\alpha_t\}_{t=1}^{T_c}$ and by the delayed CSIT exponent β. Each α_t describes the high SNR quality of the current CSIT estimates at time $t \leq T_c$, whereas β captures the quality of the best CSIT estimates that are received after the channel has elapsed, i.e., after the coherence period of the channel. In this setting we have that

$$0 \leq \alpha_1 \leq \cdots \leq \alpha_{T_c} \leq \beta \leq 1 \tag{30}$$

where any difference between two consecutive exponents is due to feedback that was received during that time slot.

This same setting nicely captures the timing of this periodic feedback. Having for example $\alpha_1 = 1$ simply refers to the case of perfect (full) CSIT, whereas having $\alpha_{T_c} = 0$ simply means that no (or very limited) current feedback is sent during the coherence period of a channel. Similarly having $\alpha_{\gamma T_c} = 0, \gamma \in [0,1]$ simply means that no (or very limited) current feedback is sent during the first γ fraction of the coherence period[5]. For example, having a periodic feedback process that sends refining feedback, let's say, two times per coherence period, at times $t = \gamma_1 T_c + 1, t = \gamma_2 T_c + 1$ and never again about that same channel, will result in having

$$\begin{aligned}\overbrace{0 = \alpha_1 = \cdots = \alpha_{\gamma_1 T_c}}^{\text{Before feedback}} &\leq \overbrace{\alpha_{\gamma_1 T_c+1} = \cdots = \alpha_{\gamma_2 T_c}}^{\text{After first feedback}} \\ &\leq \underbrace{\alpha_{\gamma_2 T_c+1} = \cdots = \alpha_{T_c} = \beta}_{\text{After second feedback}}\end{aligned} \tag{31}$$

whereas if the same feedback system adds some delayed feedback after the channel elapses, then we simply have that $\beta \geq \alpha_{T_c}$.

One can note that reducing α_{T_c} implies a reduced amount of feedback, about a specific channel, that is sent during the coherence period of that same channel, whereas reducing β implies a reduced amount of feedback, during or after the channel's coherence period. Along these lines, reducing $\beta - \alpha_{T_c}$ implies a reduced amount of feedback, about a specific fading coefficient, that is sent after the coherence period of the channel.

[5]Our ignoring here of integer rounding is an abuse of notation that is only done for the sake of clarity, and it carries no real effect on the result.

Our results capture these issues. The results hold directly from the previous results in this work, where we now simply set

$$\bar{\alpha} = \frac{1}{T_c}\sum_{t=1}^{T_c} \alpha_t. \tag{32}$$

The following holds directly from Corollary 1a, for the case of a periodically evolving feedback process over a quasi-static channel.

Corollary 1e (Periodically evolving feedback): For a periodic feedback process with $\{\alpha_t\}_{t=1}^{T_c}$ and perfect delayed CSIT (received at any time after the end of the coherence period), the optimal DoF region over a block-fading channel is the polygon with corner points

$$\{(0,0),(0,1),(\bar{\alpha},1),(\frac{2+\bar{\alpha}}{3},\frac{2+\bar{\alpha}}{3}),(1,\bar{\alpha}),(1,0)\}. \tag{33}$$

This same optimal region can in fact be achieved even with imperfect-quality delayed CSIT, as long as $\beta \geq \frac{1+2\bar{\alpha}}{3}$.

Remark 2 (Feedback quality vs. quantity): While the results here are in terms of feedback *quality* rather than in terms of feedback *quantity*, there are distinct cases where the relationship between the two is well defined. Such is the case when CSIT estimates are derived using basic - and not necessarily optimal - scalar quantization techniques [32]. In such cases, which we mention here simply to offer some insight[6], dedicating $\alpha \log P$ quantization bits to quantize $\boldsymbol{h}$ into an estimate $\hat{\boldsymbol{h}}$, allows for a mean squared error [32]

$$\mathbb{E}\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2 \doteq P^{-\alpha}.$$

Drawing from this, and going back to our previous example, we consider a periodic feedback process that sends refining feedback two times per coherence period, by first sending $\alpha' \log P$ bits of feedback at time $t = \gamma_1 T_c + 1$, then sending extra $\alpha'' \log P$ bits of feedback at time $t = \gamma_2 T_c + 1$, and where it finally sends $\beta - (\alpha' + \alpha'') \log P$ extra bits of refining feedback, at any point after the coherence period of a channel. This would result in having

$$\begin{aligned}&\overbrace{0 = \alpha_1 = \cdots = \alpha_{\gamma_1 T_c}}^{\text{Before feedback}} \leq \overbrace{\alpha' = \alpha_{\gamma_1 T_c+1} = \cdots = \alpha_{\gamma_2 T_c}}^{\text{After first feedback}} \\ &\leq \underbrace{\alpha' + \alpha'' = \alpha_{\gamma_2 T_c+1} = \cdots = \alpha_{T_c}}_{\text{After second feedback, before } T_c} \leq \underbrace{\beta}_{\text{After coherence period}}.\end{aligned} \tag{34}$$

As an example, having periodic feedback that sends $\frac{4}{9}\log P$ bits of feedback at time $t = \frac{1}{3}T_c + 1$, and then sends extra

[6]We clarify that this relationship between CSIT quality and feedback quantity, plays no role in the development of the results, and is simply mentioned in the form of comments that offer intuition. Our focus is on quality exponents, and we make no optimality claim regarding the number of quantization bits.

$\frac{1}{9}\log P$ bits of feedback at time $t = \frac{2}{3}T_c + 1$, will result in

$$\overbrace{0 = \alpha_1 = \cdots = \alpha_{\frac{1}{3}T_c}}^{\text{Before feedback}} \leq \overbrace{\frac{4}{9} = \alpha_{\frac{1}{3}T_c+1} = \cdots = \alpha_{\frac{2}{3}T_c}}^{\text{After first feedback}} \leq \underbrace{\frac{5}{9} = \alpha_{\frac{2}{3}T_c+1} = \cdots = \alpha_{T_c}}_{\text{After second feedback, before } T_c} \quad (35)$$

which gives $\bar{\alpha} = (0 + 4/9 + 5/9)/3 = 1/3$, which in turn gives (Corollary 1e) an optimal DoF region which is defined by the polygon with corner points

$$\{(0,0),(0,1),(1/3,1),(7/9,7/9),(1,1/3),(1,0)\}. \quad (36)$$

One may have noticed that there was no need for extra bits of (delayed) feedback after the end of the coherence period. This is because the existing amount and timing of feedback bits already allowed for

$$\beta = \alpha_{T_c} = 5/9 = \frac{1+2\bar{\alpha}}{3} = \frac{1+2/3}{3}$$

which we have seen in Corollary 1e to be as good as perfect delayed feedback.

Placing our focus back on feedback quality, we proceed with a corollary that offers insight on the question of what CSIT quality and timing suffice to achieve a certain DoF performance. For ease of exposition, we focus on the hardest-to-achieve DoF point $d_1 = d_2 = d$. The proof is again direct.

Corollary 1f (Sufficient feedback for target DoF): Having $\bar{\alpha} \geq 3d - 2$ with $\beta \geq 2d - 1$, or having $\bar{\alpha} \geq 3d - 2$ with $\alpha_{T_c} \geq 2d - 1$ (and no extra delayed feedback), suffices to achieve a symmetric target DoF $d_1 = d_2 = d$.

One can see that having $\bar{\alpha} \geq 3d - 2$ with $\alpha_{T_c} \geq 2d - 1$ simply means that there is not need to send delayed feedback, i.e., there is not need to send feedback after the end of the coherence period.

Another practical aspect that is addressed here, has to do with feedback delays. Such delays might cause performance degradation, which might be mitigated if the feedback, albeit with delays, is of sufficiently good quality. The following corollary provides some insight on these aspects, by describing the feedback delays that allow for a given target symmetric DoF d in the presence of constraints on current and delayed CSIT qualities. We will be specifically interested in the allowable fractional delay of feedback (cf. [14])

$$\gamma \triangleq \arg\max_{\gamma'} \{\alpha_{\gamma' T_c} = 0\} \quad (37)$$

i.e., the fraction $\gamma \leq 1$ for which $\alpha_1 = \cdots = \alpha_{\gamma T_c} = 0, \alpha_{\gamma T_c+1} > 0$. We are also interested to see how this allowable delay reduces in the presence of a constraint $\alpha_t \leq \alpha_{\max} \forall t$ on timely feedback, or a constraint on β which, as we recall, may reflect on a constraint on the total number of (current plus delayed) feedback bits per coherence period. The proof for the following can be found in the appendix of Section IX.

Corollary 1g (Allowable feedback delay): Under a current CSIT quality constraint $\alpha_t \leq \alpha_{\max}\ \forall t$, a symmetric target DoF d can be achieved with any fractional delay

$$\gamma \leq 1 - \frac{3d-2}{\alpha_{\max}}$$

by setting $\alpha_1 = \cdots = \alpha_{\gamma T_c} = 0, \alpha_{\gamma T_c+1} = \cdots = \alpha_{T_c} = \alpha_{\max} = 2d - 1 = \beta$.

Under a delayed CSIT quality constraint $\beta \leq \beta_{\max}$, a target DoF d can be achieved with any

$$\gamma \leq 1 - \frac{3d-2}{\beta_{\max}}$$

by setting $\alpha_1 = \cdots = \alpha_{\gamma T_c} = 0, \alpha_{\gamma T_c+1} = \cdots = \alpha_{T_c} = \beta_{\max} = 2d - 1$. Finally under no specific constraint on CSIT quality, the target DoF d can be achieved with any

$$\gamma \leq 3(1-d)$$

using perfect (but delayed) feedback sent at $t = \gamma T_c + 1$

$$\underbrace{\alpha_1 = \cdots = \alpha_{\gamma T_c} = 0}_{\text{No feedback}}, \underbrace{\alpha_{\gamma T_c+1} = \cdots \alpha_{T_c} = \beta = 1}_{\text{Perfect quality CSIT}}.$$

Example 2: Consider the example where we have a symmetric target DoF $d_1 = d_2 = d = \frac{7}{9}$. This can be achieved with $\gamma = 3(1-d) = 2/3$ if there is no bound on the quality exponents, and with $\gamma = 1 - (3d-2)/\alpha_{\max} = 1/3$ if the feedback link only allows for $\alpha_t \leq \alpha_{\max} = 1/2,\ \forall t$. If on the other hand, feedback timeliness is easily obtained, we can substantially reduce the amount of CSIT and achieve $d = \frac{7}{9}$ with $\alpha_1 = \cdots = \alpha_{T_c} = \bar{\alpha} = 3d - 2 = 1/3$ ($\beta = \frac{1+2\bar{\alpha}}{3} = 2d - 1 = 5/9$).

We now proceed to see how the periodically evolving feedback setting, incorporates different existing settings of interest.

A. The periodically evolving setting as a generalization to existing settings

This periodically-evolving feedback setting is powerful as it captures the many engineering options that one may have in a block-fading setting where the nature of feedback remains largely unchanged across coherence periods. It also captures and generalizes existing settings that have been of particular interest, such as the Maddah-Ali and Tse setting in [3], the Yang et al. and Gou and Jafar setting in [4], [5], the Lee and Heath 'not-so-delayed CSIT' setting in [14], and the asymmetric setting in [6]. We proceed to highlight some of these generalizations for different existing settings of interest.

1) Only delayed CSIT - Maddah-Ali and Tse: The Maddah-Ali and Tse setting in [3] maps to the evolving setting with $\alpha_t = 0, t = 1, 2, \cdots, T_c$ and with perfect delayed CSIT. One direct generalization of [3], that fits into the current evolving setting, is to consider delayed CSIT with reduced quality β. From this generalization, we now know that the same DoF performance of $\{(0,0),(0,1),(\frac{2}{3},\frac{2}{3}),(1,0)\}$, can be achieved even if the delayed CSIT sent, is of imperfect quality, corresponding to any $\beta \geq 1/3$.

2) Delayed CSIT with imperfect current CSIT - Yang et al. and Gou and Jafar: Similarly the Yang et al. and Gou and Jafar setting in [4], [5], maps to the evolving setting with $\alpha_t = \alpha, t = 1, 2, \cdots, T_c$ and with perfect delayed CSIT. As above, one direct generalization is to consider imperfect quality delayed CSIT, and for this we know that the optimal DoF $\{(0,0),(0,1),(\frac{2+\alpha}{3},\frac{2+\alpha}{3}),(1,0)\}$ can be achieved for any combination of CSIT quality exponents that give $\bar{\alpha} = \alpha$, and even with imperfect delayed CSIT quality for any $\beta \geq (1+2\alpha)/3$.

3) 'Not-so-delayed' CSIT - Lee and Heath: The setting in [14] considers perfect delayed feedback and current perfect feedback that comes with a fractional delay $\gamma \in [0, 1]$ of the coherence period, i.e., it considers that perfect feedback always arrives γT_c channel uses after the realization of the channel. This setting - focusing here on the two-user case - then maps to the periodically evolving CSIT setting with perfect delayed CSIT and with

$$\underbrace{\alpha_1 = \cdots = \alpha_{\gamma T_c} = 0}_{\text{No feedback}}, \ \underbrace{\alpha_{\gamma T_c+1} = \cdots = \alpha_{T_c} = 1}_{\text{Perfect CSIT}}. \tag{38}$$

Some practical generalizations were considered in Corollary 1g which describes the maximum possible delay γ needed to achieve a specific DoF performance, under constraints on the CSIT quality.

4) Delayed CSIT with current CSIT for just one user - Maleki, Jafar and Shamai: The evolving setting can be naturally extended to the asymmetric (still periodically-evolving feedback) setting where $\bar{\alpha}^{(1)} \neq \bar{\alpha}^{(2)}$ and where the delayed CSIT exponents $\beta^{(1)}, \beta^{(2)}$ need not be equal. Such asymmetric setting would yield a generalization for the asymmetric setting of Maleki et al. in [6], where both users offered perfect delayed CSIT, and where only the first user offered perfect current CSIT, resulting in an optimal DoF corresponding to DoF corner point $(1, 1/2)$ (sum-DoF $d_1 + d_2 = 3/2$). This setting maps to the periodically evolving CSIT setting with perfect delayed CSIT and with

$$\alpha_t^{(1)} = 1, \alpha_t^{(2)} = 0, \ \forall t.$$

The following corollary offers a broad generalization of the corresponding result in [6]. The proof is direct since the following simply adapts the result in the main theorem, to the periodically evolving setting. This is again for the setting of sufficiently good delayed CSIT for which $\min\{\beta^{(1)}, \beta^{(2)}\} \geq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}, \frac{1+\bar{\alpha}^{(2)}}{2}\}$.

Corollary 1h (Asymmetric and periodic CSIT): The optimal DoF region is defined by corner points $B = (\bar{\alpha}^{(2)}, 1)$, $C = (\frac{2+2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}}{3}, \frac{2+2\bar{\alpha}^{(2)}-\bar{\alpha}^{(1)}}{3})$ and $D = (1, \bar{\alpha}^{(1)})$ whenever $2\bar{\alpha}^{(1)} - \bar{\alpha}^{(2)} < 1$, else by corner points $A = (1, \frac{1+\bar{\alpha}^{(2)}}{2})$ and B.

As an example, we can see that the same DoF corner point $A = (1, 1/2)$ derived in [6], can in fact be achieved with imperfect quality current and delayed CSIT

$$\alpha_t^{(1)} = 1/2, \alpha_t^{(2)} = 0 \ \forall t \ ; \ \beta^{(1)} = \beta^{(2)} = 1/2.$$

Another example could be $\alpha_t^{(1)} = 3/4, \alpha_t^{(2)} = 1/2 \ \forall t, \ \beta^{(1)} = \beta^{(2)} = 3/4$, which corresponds to an optimal DoF corner point $(1, \frac{3}{4})$.

IV. Universal encoding-decoding scheme

We proceed to describe the universal scheme that achieves the aforementioned DoF corner points. The challenge will be to design a scheme of an asymptotically large duration n, that utilizes a CSIT process $\{\hat{\boldsymbol{h}}_{t,t'}, \hat{\boldsymbol{g}}_{t,t'}\}_{t=1,t'=1}^{n}$ of quality defined by the statistics of $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t=1,t'=1}^{n}$. This will be achieved by focusing on the corresponding quality-exponent sequences $\{\alpha_t^{(1)}\}_{t=1}^n, \{\alpha_t^{(2)}\}_{t=1}^n, \{\beta_t^{(1)}\}_{t=1}^n, \{\beta_t^{(2)}\}_{t=1}^n$, as these were defined in (4)-(7). The optimal DoF region in Theorem 1 and the additional corner points in Proposition 1, will be achieved by properly utilizing different combinations of zero forcing, superposition coding, interference compressing and broadcasting, as well as proper power and rate allocation.

As previously suggested, this causal scheme does not require knowledge of future quality exponents, nor does it use predicted CSIT estimates of future channels. The transmitter must know though the long term averages $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}, \bar{\beta}^{(1)}, \bar{\beta}^{(2)}$, which - as is commonly assumed of long term statistics - can be derived.

By 'feeding' this universal scheme with the proper parameters, we can get schemes that are tailored to the different specific settings we have discussed. We will see such examples later in this section.

We remind the reader that the users are labeled so that $\bar{\alpha}^{(2)} \leq \bar{\alpha}^{(1)}$. We also remind the reader of the soft assumption that any sufficiently long subsequence $\{\alpha_t^{(1)}\}_{t=\tau}^{\tau+T}$ (resp. $\{\alpha_t^{(2)}\}_{t=\tau}^{\tau+T}, \{\beta_t^{(1)}\}_{t=\tau}^{\tau+T}, \{\beta_t^{(2)}\}_{t=\tau}^{\tau+T}$) is assumed to have an average that converges to the long term average $\bar{\alpha}^{(1)}$ (resp. $\bar{\alpha}^{(2)}, \bar{\beta}^{(1)}, \bar{\beta}^{(2)}$), for a finite T that can be sufficiently large to allow for this convergence. We briefly note that, as we will see later, in periodic settings such as those described in Section III, T need not be large.

We proceed to describe in Section IV-A the encoding part, and in Section IV-B the decoding part. In Section IV-C we show how the scheme achieves the different DoF corner points of interest. Finally in Section IV-D we provide example instances of our general scheme, for specific cases of particular interest.

For notational convenience, we will use

$$\hat{\boldsymbol{h}}_t \triangleq \hat{\boldsymbol{h}}_{t,t}, \quad \hat{\boldsymbol{g}}_t \triangleq \hat{\boldsymbol{g}}_{t,t}$$

$$\check{\boldsymbol{h}}_t \triangleq \hat{\boldsymbol{h}}_{t,t+\eta}, \quad \check{\boldsymbol{g}}_t \triangleq \hat{\boldsymbol{g}}_{t,t+\eta}$$

to denote the current and delayed estimates of $\boldsymbol{h}_t$ and $\boldsymbol{g}_t$, respectively[7], with corresponding estimation errors being

$$\tilde{\boldsymbol{h}}_t \triangleq \boldsymbol{h}_t - \hat{\boldsymbol{h}}_t, \quad \tilde{\boldsymbol{g}}_t \triangleq \boldsymbol{g}_t - \hat{\boldsymbol{g}}_t \tag{39}$$

$$\ddot{\boldsymbol{h}}_t \triangleq \boldsymbol{h}_t - \check{\boldsymbol{h}}_t, \quad \ddot{\boldsymbol{g}}_t \triangleq \boldsymbol{g}_t - \check{\boldsymbol{g}}_t. \tag{40}$$

[7]Recall that η is a sufficiently large but finite integer, corresponding to the maximum delay allowed for waiting for delayed CSIT.

A. Scheme $\mathcal{X}$: encoding

Scheme $\mathcal{X}$ is designed to have S phases, where each phase has a duration of T channel uses, and where T is finite but - unless stated otherwise - sufficiently large. Specifically each phase s ($s = 1, 2, \cdots, S$) will take place over all time slots t belonging to the set

$$\mathcal{B}_s = \{\mathcal{B}_{s,\ell} \triangleq (s-1)2T + \ell\}_{\ell=1}^{T}, \quad s = 1, \cdots, S. \tag{41}$$

As stated, T is sufficiently large so that

$$\frac{1}{T}\sum_{t\in\mathcal{B}_s} \alpha_t^{(i)} \to \bar{\alpha}^{(i)}, \ \frac{1}{T}\sum_{t\in\mathcal{B}_s} \beta_t^{(i)} \to \bar{\beta}^{(i)}, \ s = 1, \cdots, S \tag{42}$$

$i = 1, 2$. The above allocation in (41) guarantees that there are T channel uses in between any two neighboring phases. Having T being sufficiently large allows for the delayed CSIT corresponding to the channels appearing during phase s, to be available before the beginning of the phase that we label as phase $s + 1$. This implies that $T > \eta$ (cf. (6),(7)), although this assumption can be readily removed[8]. Naturally there is no silent time, and over the remaining channel uses

$$t \in \{(2s-1)T + \ell\}_{\ell=1,s=1}^{\ell=T,s=S}$$

we simply repeat scheme $\mathcal{X}$ with a different message. With n being generally infinite, S is also infinite (except for specific instances, some of which are highlighted in Section IV-D).

Remark 3 (Phase-Markov encoding and decoding): The aforementioned phase-Markov structure of the scheme, asks that the accumulated quantized interference bits of phase s, can be broadcasted to both users inside the common information symbols of the next phase (phase $(s + 1)$), while also a certain amount of common information can be transmitted to both users during phase s, which will then help resolve the accumulated interference of phase $(s - 1)$.

We proceed to give the general description that holds for all phases $s = 1, 2, \cdots, S - 1$, except for the last phase S, which we describe separately afterwards. A brief illustration can be found in Figure 4 and Figure 5.

1) Phase s, for $s = 1, 2, \cdots, S-1$: We proceed to describe the way the scheme, in each phase $s \in [1, S - 1]$, combines zero forcing and superposition coding, power and rate allocation, and interference compressing and broadcasting, in order to transmit private information, using currently available CSIT estimates to reduce interference, and using delayed CSIT estimates to alleviate the effect of past interference.

a) Zero forcing and superposition coding: During phase s, $t \in \mathcal{B}_s$, the transmitter sends

$$\boldsymbol{x}_t = \boldsymbol{w}_t c_t + \hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'} + \hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'} \tag{43}$$

where $a_t, a_t^{'}$ are the symbols meant for user 1, $b_t, b_t^{'}$ for user 2, where c_t is a common symbol, where $\boldsymbol{e}^{\perp}$ denotes a unit-norm vector orthogonal to $\boldsymbol{e}$, and where $\boldsymbol{w}_t$ is a predetermined randomly-generated vector known by all the nodes.

[8] The assumption can be removed because we can, instead of splitting time into two interleaved halves and identifying each half to a message, to instead split time into more parts, each corresponding to a different message. For a sufficiently large number of parts, this would allow for the removal of the assumption that $T \geq \eta$, and the only assumption that would remain would be that T is large enough so that (42) is satisfied. In periodic settings, such T can be small.

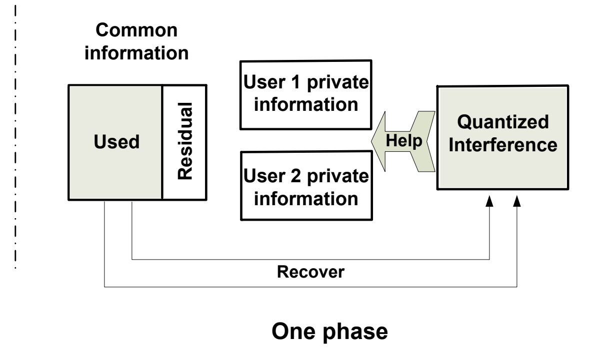


Fig. 5. Illustration of coding over a single phase.

b) Power and rate allocation policy: In describing the power and rates of the symbols in (43), we use the notation

$$P_t^{(x)} \triangleq \mathbb{E}|x_t|^2 \tag{44}$$

to denote the power of x_t corresponding to time-slot t, and we use $r_t^{(x)}$ to denote the prelog factor of the number of bits $r_t^{(x)} \log P - o(\log P)$ carried by symbol x_t at time t.

When in phase s, during time-slot t, the powers and (normalized) rates are set as

$$\begin{array}{ll} P_t^{(c)} \doteq P, & \\ P_t^{(a)} \doteq P^{\delta_t^{(2)}}, & r_t^{(a)} = \delta_t^{(2)} \\ P_t^{(b)} \doteq P^{\delta_t^{(1)}}, & r_t^{(b)} = \delta_t^{(1)} \\ P_t^{(a')} \doteq P^{\delta_t^{(2)}-\alpha_t^{(2)}}, & r_t^{(a')} = (\delta_t^{(2)} - \alpha_t^{(2)})^{+} \\ P_t^{(b')} \doteq P^{\delta_t^{(1)}-\alpha_t^{(1)}}, & r_t^{(b')} = (\delta_t^{(1)} - \alpha_t^{(1)})^{+} \end{array} \tag{45}$$

where $(\bullet)^{+} \triangleq \max\{\bullet, 0\}$.

We design the scheme so that the entirety of common information symbols $\{c_{\mathcal{B}_{s,t}}\}_{t=1}^{T}$, carry

$$T(1 - \bar{\delta}) \log P - o(\log P) \tag{46}$$

bits, and design the power parameters $\{\delta_t^{(1)}, \delta_t^{(2)}\}_{t\in\mathcal{B}_s}$ to satisfy

$$\beta_t^{(i)} \geq \delta_t^{(i)} \quad i = 1, 2, \ t \in \mathcal{B}_s \tag{47}$$

$$\frac{1}{T}\sum_{t\in\mathcal{B}_s} \delta_t^{(1)} = \frac{1}{T}\sum_{t\in\mathcal{B}_s} \delta_t^{(2)} = \bar{\delta} \tag{48}$$

$$\frac{1}{T}\sum_{t\in\mathcal{B}_s} (\delta_t^{(i)} - \alpha_t^{(i)})^{+} = (\bar{\delta} - \bar{\alpha}^{(i)})^{+} \quad i = 1, 2, \tag{49}$$

for some $\bar{\delta}$ that will be bounded by

$$\bar{\delta} \leq \min\{\bar{\beta}^{(1)}, \bar{\beta}^{(2)}, \frac{1 + \bar{\alpha}^{(1)} + \bar{\alpha}^{(2)}}{3}, \frac{1 + \bar{\alpha}^{(2)}}{2}\}. \tag{50}$$

There indeed exist solutions $\{\delta_t^{(1)}, \delta_t^{(2)}\}_{t\in\mathcal{B}_s}$ that satisfy the above, and an explicit solution is shown in Appendix VII-A. Our solution for power and rate allocation allows that, at time t, the transmitter need only acquire knowledge of $\{\alpha_t^{(1)}, \beta_t^{(1)}; \alpha_t^{(2)}, \beta_t^{(2)}\}$, in addition to the derived long-term averages $\bar{\alpha}^{(1)}, \bar{\alpha}^{(2)}, \bar{\beta}^{(1)}, \bar{\beta}^{(2)}$. This nature of the derived solutions is crucial for handling asymmetry ($\alpha_t^{(1)} \neq \alpha_t^{(2)}$, $\beta_t^{(1)} \neq \beta_t^{(2)}$).

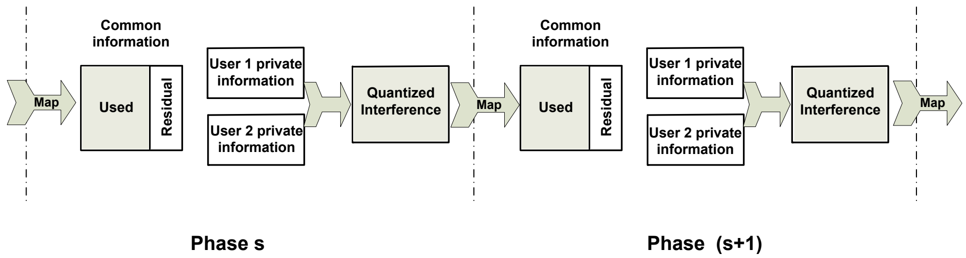


Fig. 4. Illustration of coding across phases.

After transmission, the received signals take the form

$$
\begin{aligned}
y_t^{(1)} = {} & \underbrace{\boldsymbol{h}_t^{\mathsf{T}} \boldsymbol{w}_t c_t}_{P} + \underbrace{\boldsymbol{h}_t^{\mathsf{T}} \hat{\boldsymbol{g}}_t^{\perp} a_t}_{P^{\delta_t^{(2)}}} + \underbrace{\boldsymbol{h}_t^{\mathsf{T}} \hat{\boldsymbol{h}}_t a_t^{'}}_{P^{\delta_t^{(2)}-\alpha_t^{(2)}}} + \underbrace{z_t^{(1)}}_{P^0} \\
& + \underbrace{\overbrace{\check{\boldsymbol{h}}_t^{\mathsf{T}} (\hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'})}^{\check{\iota}_t^{(1)}}}_{P^{\delta_t^{(1)}-\alpha_t^{(1)}}} + \underbrace{\overbrace{\ddot{\boldsymbol{h}}_t^{\mathsf{T}} (\hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'})}^{\iota_t^{(1)}-\check{\iota}_t^{(1)}}}_{P^{\delta_t^{(1)}-\beta_t^{(1)}} \leq P^0}
\end{aligned} \tag{51}
$$

$$
\begin{aligned}
y_t^{(2)} = {} & \underbrace{\boldsymbol{g}_t^{\mathsf{T}} \boldsymbol{w}_t c_t}_{P} + \underbrace{\boldsymbol{g}_t^{\mathsf{T}} \hat{\boldsymbol{h}}_t^{\perp} b_t}_{P^{\delta_t^{(1)}}} + \underbrace{\boldsymbol{g}_t^{\mathsf{T}} \hat{\boldsymbol{g}}_t b_t^{'}}_{P^{\delta_t^{(1)}-\alpha_t^{(1)}}} + \underbrace{z_t^{(2)}}_{P^0} \\
& + \underbrace{\overbrace{\check{\boldsymbol{g}}_t^{\mathsf{T}} (\hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'})}^{\check{\iota}_t^{(2)}}}_{P^{\delta_t^{(2)}-\alpha_t^{(2)}}} + \underbrace{\overbrace{\ddot{\boldsymbol{g}}_t^{\mathsf{T}} (\hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'})}^{\iota_t^{(2)}-\check{\iota}_t^{(2)}}}_{P^{\delta_t^{(2)}-\beta_t^{(2)}} \leq P^0}
\end{aligned} \tag{52}
$$

where

$$
\iota_t^{(1)} \triangleq \boldsymbol{h}_t^{\mathsf{T}} (\hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'}), \;\; \iota_t^{(2)} \triangleq \boldsymbol{g}_t^{\mathsf{T}} (\hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'}) \tag{53}
$$

denote the interference at user 1 and user 2 respectively, and where

$$
\check{\iota}_t^{(1)} \triangleq \check{\boldsymbol{h}}_t^{\mathsf{T}} (\hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'}), \;\; \check{\iota}_t^{(2)} \triangleq \check{\boldsymbol{g}}_t^{\mathsf{T}} (\hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'}) \tag{54}
$$

denote the transmitter's delayed estimates of $\iota_t^{(1)}, \iota_t^{(2)}$. In the above - where under each term we noted the order of the summand's average power - we considered that

$$
\begin{aligned}
\mathbb{E}|\check{\iota}_t^{(1)}|^2 &= \mathbb{E}|\check{\boldsymbol{h}}_t^{\mathsf{T}} \hat{\boldsymbol{h}}_t^{\perp} b_t|^2 + \mathbb{E}|\check{\boldsymbol{h}}_t^{\mathsf{T}} \hat{\boldsymbol{g}}_t b_t^{'}|^2 \\
&= \mathbb{E}|(\hat{\boldsymbol{h}}_t^{\mathsf{T}} + \tilde{\boldsymbol{h}}_t^{\mathsf{T}} - \ddot{\boldsymbol{h}}_t^{\mathsf{T}}) \hat{\boldsymbol{h}}_t^{\perp} b_t|^2 + \mathbb{E}|\check{\boldsymbol{h}}_t^{\mathsf{T}} \hat{\boldsymbol{g}}_t b_t^{'}|^2 \\
&= \mathbb{E}|(\tilde{\boldsymbol{h}}_t^{\mathsf{T}} - \ddot{\boldsymbol{h}}_t^{\mathsf{T}}) \hat{\boldsymbol{h}}_t^{\perp} b_t|^2 + \mathbb{E}|\check{\boldsymbol{h}}_t^{\mathsf{T}} \hat{\boldsymbol{g}}_t b_t^{'}|^2 \\
&\doteq P^{\delta_t^{(1)}-\alpha_t^{(1)}} \\
\mathbb{E}|\check{\iota}_t^{(2)}|^2 &= \mathbb{E}|(\tilde{\boldsymbol{g}}_t^{\mathsf{T}} - \ddot{\boldsymbol{g}}_t^{\mathsf{T}}) \hat{\boldsymbol{g}}_t^{\perp} a_t|^2 + \mathbb{E}|\check{\boldsymbol{g}}_t^{\mathsf{T}} \hat{\boldsymbol{h}}_t a_t^{'}|^2 \\
&\doteq P^{\delta_t^{(2)}-\alpha_t^{(2)}}.
\end{aligned} \tag{55}
$$

c) Quantizing and broadcasting the accumulated interference: After the end of phase s and before the beginning of the next phase - which starts T channel uses after the end of phase s, i.e., after the accumulation of all delayed CSIT - the transmitter reconstructs $\check{\iota}_t^{(1)}, \check{\iota}_t^{(2)}, t \in \mathcal{B}_s$ using its knowledge of delayed CSIT, and quantizes these into

$$
\bar{\iota}_t^{(1)} = \check{\iota}_t^{(1)} - \tilde{\iota}_t^{(1)}, \quad \bar{\iota}_t^{(2)} = \check{\iota}_t^{(2)} - \tilde{\iota}_t^{(2)} \tag{56}
$$

TABLE I

BITS CARRIED BY PRIVATE SYMBOLS, COMMON SYMBOLS, AND BY THE QUANTIZED INTERFERENCE, FOR PHASE s, $s = 1, 2, \cdots, S-1$.

	Total bits ($\times \log P$)
Private symbols for user 1	$T(\bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(2)})^+)$
Private symbols for user 2	$T(\bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(1)})^+)$
Common symbols	$T(1 - \delta)$
Quantized interference	$T((\bar{\delta} - \bar{\alpha}^{(1)})^+ + (\bar{\delta} - \bar{\alpha}^{(2)})^+)$

with $(\delta_t^{(1)} - \alpha_t^{(1)})^+ \log P$ and $(\delta_t^{(2)} - \alpha_t^{(2)})^+ \log P$ quantization bits respectively, allowing for bounded power of quantization noise $\tilde{\iota}_t^{(1)}, \tilde{\iota}_t^{(2)}$, i.e, allowing for

$$
\mathbb{E}|\tilde{\iota}_t^{(2)}|^2 \doteq \mathbb{E}|\tilde{\iota}_t^{(1)}|^2 \doteq 1
$$

since $\mathbb{E}|\check{\iota}_t^{(2)}|^2 \doteq P^{\delta_t^{(2)}-\alpha_t^{(2)}}$, $\mathbb{E}|\check{\iota}_t^{(1)}|^2 \doteq P^{\delta_t^{(1)}-\alpha_t^{(1)}}$ (cf. [32]). Then the transmitter evenly splits the

$$
\sum_{t \in \mathcal{B}_s} \left((\delta_t^{(1)} - \alpha_t^{(1)})^+ + (\delta_t^{(2)} - \alpha_t^{(2)})^+ \right) \log P \tag{57}
$$

quantization bits into the common symbols $\{c_t\}_{t \in \mathcal{B}_{s+1}}$ that will be transmitted during the next phase (phase $s+1$), conveying these quantization bits together with other new information bits for the users.

This transmission of $\{c_t\}_{t \in \mathcal{B}_{s+1}}$ in the next phase, will help each of the users cancel the dominant part of the interference, and it will also serve as an extra observation (which will in turn enable the creation of a corresponding MIMO channel - see (69) later on) that allows for decoding of all private information of that same user. Table I summarizes the number of bits carried by private symbols, common symbols, and by the quantized interference, for phase s, $s = 1, 2, \cdots, S-1$.

We now proceed with the description of encoding over the last phase S.

2) Phase S: The last phase, in addition to communicating new private symbols, conveys the remaining accumulated interference from the previous phase, and does so in a manner that allows for termination at the end of this phase.

During this last phase, the transmitter sends

$$
\boldsymbol{x}_t = \boldsymbol{w}_t c_t + \hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t^{\perp} b_t \tag{58}
$$

$t \in \mathcal{B}_S$, with power and rates set as

$$
\begin{aligned}
P_t^{(c)} \doteq P, \quad P_t^{(a)} &\doteq P^{\alpha_t^{(2)}}, \quad P_t^{(b)} \doteq P^{\delta_t^{(1)}} \\
r_t^{(a)} &= \alpha_t^{(2)}, \quad\;\; r_t^{(b)} = \delta_t^{(1)}.
\end{aligned} \tag{59}
$$

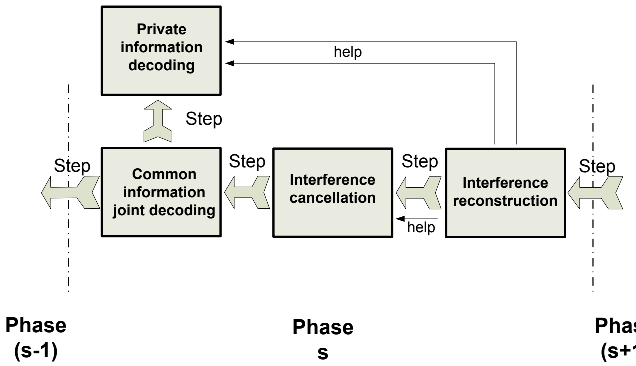


Fig. 6. Illustration of decoding steps.

With the entirety of common information symbols $\{c_{\mathcal{B}_{S,\ell}}\}_{\ell=1}^{T}$ now carrying[9]

$$T(1-\bar{\alpha}^{(2)})\log P - o(\log P) \tag{60}$$

bits, the power parameters $\{\delta_t^{(1)}\}_{t\in\mathcal{B}_S}$ are designed such that

$$\alpha_t^{(1)} \geq \delta_t^{(1)} \quad \forall t \tag{61}$$

$$\frac{1}{T}\sum_{t\in\mathcal{B}_S}\delta_t^{(1)} = \bar{\alpha}^{(2)}. \tag{62}$$

The solution to the above problem is similar to that in (47),(48),(49).

This concludes the part of encoding. After transmission, the received signals are then of the form

$$y_t^{(1)} = \underbrace{\boldsymbol{h}_t^{\intercal}\boldsymbol{w}_t c_t}_{P} + \underbrace{\boldsymbol{h}_t^{\intercal}\hat{\boldsymbol{g}}_t^{\perp} a_t}_{P^{\alpha_t^{(2)}}} + \underbrace{\tilde{\boldsymbol{h}}_t^{\intercal}\hat{\boldsymbol{h}}_t^{\perp} b_t}_{\leq P^0} + \underbrace{z_t^{(1)}}_{P^0} \tag{63}$$

$$y_t^{(2)} = \underbrace{\boldsymbol{g}_t^{\intercal}\boldsymbol{w}_t c_t}_{P} + \underbrace{\tilde{\boldsymbol{g}}_t^{\intercal}\hat{\boldsymbol{g}}_t^{\perp} a_t}_{P^0} + \underbrace{\boldsymbol{g}_t^{\intercal}\hat{\boldsymbol{h}}_t^{\perp} b_t}_{P^{\delta_t^{(1)}}} + \underbrace{z_t^{(2)}}_{P^0}. \tag{64}$$

We now move to describe decoding at both receivers, where this decoding part has a Markov chain structure (see Figure 6), similar to the encoding part.

B. Scheme $\mathcal{X}$: decoding

As it may be apparent (more details will be shown in Section IV-C), the power and rate allocation in (47),(48),(49) guarantees that the quantized interference accumulated during phase s ($s = 1,\cdots,S-1$) has fewer bits than the load of the common symbols transmitted during the next phase (cf. (57)). Consequently decoding of the common symbols during a certain phase, helps recover the interference accumulated during the previous phase. As a result, decoding moves backwards, from the last to the first phase.

[9]We remind the reader of the definition of $\mathcal{B}_{s,\ell}$ (cf. (41)) which denotes the ℓth element of set $\mathcal{B}_s$ consisting of all time indexes of phase s. For example, saying that $t = \mathcal{B}_{1,\ell}$ simply means that $t = \ell$.

1) Phase S: At the end of phase S, we consider *joint* decoding of all common symbols $[c_{\mathcal{B}_{S,1}},\cdots,c_{\mathcal{B}_{S,T}}]^{\intercal}$. Specifically user i, $i = 1,2$, decodes the corresponding common-information vector using its received signal vector $[y_{\mathcal{B}_{S,1}}^{(i)}, y_{\mathcal{B}_{S,2}}^{(i)},\cdots,y_{\mathcal{B}_{S,T}}^{(i)}]^{\intercal}$, and does so by treating the other signals as noise. We now note that the accumulated mutual information satisfies

$$\begin{aligned}
I([c_{\mathcal{B}_{S,1}},\cdots,c_{\mathcal{B}_{S,T}}]^{\intercal};[y_{\mathcal{B}_{S,1}}^{(1)},\cdots,y_{\mathcal{B}_{S,T}}^{(1)}]^{\intercal})& \\
= \log\prod_{t\in\mathcal{B}_S}P^{1-\alpha_t^{(2)}} - o(\log P)& \\
= T(1-\bar{\alpha}^{(2)})\log P - o(\log P)& \\
I([c_{\mathcal{B}_{S,1}},\cdots,c_{\mathcal{B}_{S,T}}]^{\intercal};[y_{\mathcal{B}_{S,1}}^{(2)},\cdots,y_{\mathcal{B}_{S,T}}^{(2)}]^{\intercal})& \\
= \log\prod_{t\in\mathcal{B}_S}P^{1-\delta_t^{(1)}} - o(\log P)& \\
= T(1-\bar{\alpha}^{(2)})\log P - o(\log P)&
\end{aligned} \tag{65}$$

(cf. (61),(62)), to conclude that both users can reliably decode all

$$T(1-\bar{\alpha}^{(2)})\log P - o(\log P) \tag{66}$$

bits in the common information vector $[c_{\mathcal{B}_{S,1}},\cdots,c_{\mathcal{B}_{S,T}}]^{\intercal}$. This is proved in Lemma 2 in the appendix of Section VII-B, which in fact guarantees that both users will be able to decode the amount of feedback bits described in (66), even for finite and small T. This is done to ensure the validity of the schemes also for finite T, and is achieved by employing specific lattice codes that have good properties in the finite-duration high-SNR regime. The details for this step can be found in the aforementioned appendix.

After decoding $[c_{\mathcal{B}_{S,1}},\cdots,c_{\mathcal{B}_{S,T}}]^{\intercal}$, user 1 removes $\boldsymbol{h}_t^{\intercal}\boldsymbol{w}_t c_t$ from the received signal in (63), to decode a_t. Similarly user 2 removes $\boldsymbol{g}_t^{\intercal}\boldsymbol{w}_t c_t$ from its received signal in (64), to decode b_t.

Now we go back one phase and utilize knowledge of $\{c_t\}_{t\in\mathcal{B}_S}$, to decode the corresponding symbols.

2) Phase s, $s = S-1, S-2,\cdots,1$: We here describe, for phase s, the actions of interference reconstruction, interference cancelation, joint decoding of common information symbols, and decoding of private information symbols, in the order they happen.

a) Interference reconstruction: In this phase (phase s), each user employs knowledge of $\{c_t\}_{t\in\mathcal{B}_{s+1}}$ from phase $s+1$, to reconstruct the delayed estimates of all the interference accumulated in phase s, i.e., to reconstruct $\{\bar{\iota}_t^{(2)},\bar{\iota}_t^{(1)}\}_{t\in\mathcal{B}_s}$.

b) Interference cancelation: Now with knowledge of $\{\bar{\iota}_t^{(2)},\bar{\iota}_t^{(1)}\}_{t\in\mathcal{B}_s}$, each user can remove - up to noise level - all the interference $\iota_t^{(i)}$, $t\in\mathcal{B}_s$, by subtracting the delayed interference estimates $\bar{\iota}_t^{(i)}$ from $y_t^{(i)}$.

c) Joint decoding of common information symbols: At this point, user i decodes the common information vector $\boldsymbol{c}_s \triangleq [c_{\mathcal{B}_{s,1}},\cdots,c_{\mathcal{B}_{s,T}}]^{\intercal}$ from its (modified) received signal vector $[y_{\mathcal{B}_{s,1}}^{(i)} - \bar{\iota}_{\mathcal{B}_{s,1}}^{(i)},\cdots,y_{\mathcal{B}_{s,T}}^{(i)} - \bar{\iota}_{\mathcal{B}_{s,T}}^{(i)}]^{\intercal}$ by treating the other signals as noise. The accumulated mutual information then satisfies

$$I(\boldsymbol{c}_s;[y^{(1)}_{\mathcal{B}_{s,1}}-\bar{\iota}^{(1)}_{\mathcal{B}_{s,1}},\cdots,y^{(1)}_{\mathcal{B}_{s,T}}-\bar{\iota}^{(1)}_{\mathcal{B}_{s,T}}]^{\intercal})$$
$$=\log\prod_{t\in\mathcal{B}_s}P^{1-\delta^{(2)}_t}-o(\log P)=T(1-\bar{\delta})\log P-o(\log P)$$
$$I(\boldsymbol{c}_s;[y^{(2)}_{\mathcal{B}_{s,1}}-\bar{\iota}^{(2)}_{\mathcal{B}_{s,1}},\cdots,y^{(2)}_{\mathcal{B}_{s,T}}-\bar{\iota}^{(2)}_{\mathcal{B}_{s,T}}]^{\intercal})$$
$$=\log\prod_{t\in\mathcal{B}_s}P^{1-\delta^{(1)}_t}-o(\log P)=T(1-\bar{\delta})\log P-o(\log P) \quad (67)$$

(cf. (47)-(52)), and we conclude that both users can reliably decode all

$$T(1-\bar{\delta})\log P-o(\log P) \quad (68)$$

bits of the common information vector $\boldsymbol{c}_s$. The details for this step, can again be found in the appendix of Section VII-B.

After decoding $\boldsymbol{c}_s$, user 1 removes $\boldsymbol{h}^{\intercal}_t\boldsymbol{w}_t c_t$ from $y^{(1)}_t-\bar{\iota}^{(1)}_t$, while user 2 removes $\boldsymbol{g}^{\intercal}_t\boldsymbol{w}_t c_t$ from $y^{(2)}_t-\bar{\iota}^{(1)}_t$, $t\in\mathcal{B}_s$.

d) Decoding of private information symbols: After removing the interference, and decoding and subtracting out the common symbols, each user now decodes its private information symbols of phase s. Using knowledge of $\{\bar{\iota}^{(2)}_t,\bar{\iota}^{(1)}_t\}_{t\in\mathcal{B}_s}$, user 1 will use the estimate $\bar{\iota}^{(2)}_t$ (of $\check{\iota}^{(2)}_t$) as an extra observation which, together with the observation $y^{(1)}_t-\boldsymbol{h}^{\intercal}_t\boldsymbol{w}_t c_t-\bar{\iota}^{(1)}_t$, will allow for decoding of both a_t and a'_t, $t\in\mathcal{B}_s$. Specifically user 1, at each instance t, can 'see' a 2×2 MIMO channel of the form

$$\begin{bmatrix} y^{(1)}_t-\boldsymbol{h}^{\intercal}_t\boldsymbol{w}_t c_t-\bar{\iota}^{(1)}_t \\ \bar{\iota}^{(2)}_t\end{bmatrix}=\begin{bmatrix}\boldsymbol{h}^{\intercal}_t\\ \check{\boldsymbol{g}}^{\intercal}_t\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{g}}^{\perp}_t & \hat{\boldsymbol{h}}_t\end{bmatrix}\begin{bmatrix}a_t\\ a'_t\end{bmatrix}+\begin{bmatrix}\tilde{z}^{(1)}_t\\ -\tilde{\iota}^{(2)}_t\end{bmatrix} \quad (69)$$

where

$$\tilde{z}^{(1)}_t=\ddot{\boldsymbol{h}}^{\intercal}_t(\hat{\boldsymbol{h}}^{\perp}_t b_t+\hat{\boldsymbol{g}}_t b'_t)+z^{(1)}_t+\tilde{\iota}^{(1)}_t.$$

The fact that $\mathbb{E}|\tilde{z}^{(1)}_t|^2\doteq 1$, allows for decoding of a_t and a'_t, corresponding to the aforementioned rates $r^{(a)}_t=\delta^{(2)}_t$, $r^{(a')}_t=(\delta^{(2)}_t-\alpha^{(2)}_t)^+$, $t\in\mathcal{B}_s$. Similar actions are taken by user 2, allowing for decoding of b_t and b'_t, again with $r^{(b)}_t=\delta^{(1)}_t$, $r^{(a')}_t=(\delta^{(1)}_t-\alpha^{(1)}_t)^+$, $t\in\mathcal{B}_s$.

At this point, each user has decoded all the information symbols (common and private) corresponding to phase s, goes back one phase (to phase $s-1$) to utilize its knowledge of $\{c_t\}_{t\in\mathcal{B}_s}$, and decodes the common and private symbols of that phase. The whole decoding effort naturally terminates after decoding of the symbols in the first phase.

C. Scheme $\mathcal{X}$: Calculating the achieved DoF

In the following DoF calculation we will consider two separate cases. Case 1 will correspond to

$$2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}<1 \quad (70)$$

which in turn implies that $\bar{\alpha}^{(1)}\le\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}\le\frac{1+\bar{\alpha}^{(2)}}{2}$, while case 2 will correspond to

$$2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}\ge 1 \quad (71)$$

which in turn implies that $\bar{\alpha}^{(1)}\ge\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}\ge\frac{1+\bar{\alpha}^{(2)}}{2}$. We recall that the users are labeled so that $\bar{\alpha}^{(1)}\ge\bar{\alpha}^{(2)}$.

1) Generic DoF point: To calibrate the DoF performance, we first note that for any fixed $\bar{\delta}\le\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)},\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$ (cf. (50)), the rate and power allocation in (47),(48),(49) (as this policy is explicitly described in the appendix of Section VII-A) tells us that, the total amount of information, for user 1, in the *private symbols* of a certain phase $s<S$, is equal to

$$\big(\bar{\delta}+(\bar{\delta}-\bar{\alpha}^{(2)})^+\big)T\log P \quad (72)$$

bits, while for user 2 this is

$$\big(\bar{\delta}+(\bar{\delta}-\bar{\alpha}^{(1)})^+\big)T\log P \quad (73)$$

bits.

The next step is to see how much interference there is to load onto common symbols. Given the power and rate allocation in (47),(48),(49),(50), it is guaranteed that the accumulated quantized interference in a phase $s<S$ (cf. (57)) has $\big((\bar{\delta}-\bar{\alpha}^{(1)})^++(\bar{\delta}-\bar{\alpha}^{(2)})^+\big)T\log P$ bits, which can be carried by the common symbols of the next phase $(s+1)$ since they can carry a total of $\big(1-\bar{\delta}\big)T\log P$ bits (cf. (46)). This leaves an extra space of $\Delta_{\text{com}}T\log P$ bits in the common symbols, where

$$\Delta_{\text{com}}\triangleq 1-\bar{\delta}-(\bar{\delta}-\bar{\alpha}^{(1)})^+-(\bar{\delta}-\bar{\alpha}^{(2)})^+ \quad (74)$$

is guaranteed to be non-negative for any given $\bar{\delta}\le\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)},\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$. This extra space can be split between the two users, by allocating $\omega\Delta_{\text{com}}T\log P$ bits for the message of user 1, and the remaining $(1-\omega)\Delta_{\text{com}}T\log P$ bits for the message of user 2, for some $\omega\in[0,1]$.

Consequently the above, combined with the information stored in private symbols (cf. (72),(73)), allows for

$$d_1=\bar{\delta}+(\bar{\delta}-\bar{\alpha}^{(2)})^++\omega\Delta_{\text{com}} \quad (75)$$
$$d_2=\bar{\delta}+(\bar{\delta}-\bar{\alpha}^{(1)})^++(1-\omega)\Delta_{\text{com}}. \quad (76)$$

The above considers that S is large, and thus removes the effect of having a last phase that carries less new message information. In the following, we will achieve different corner points by accordingly setting the value of $\omega\in[0,1]$ and of $\bar{\delta}\le\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)},\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$.

2) DoF corner points in Theorem 1: To achieved the DoF region in Theorem 1, we will show how to achieve the following DoF corner points (see also Table II)

$$A=\big(1,\frac{1+\bar{\alpha}^{(2)}}{2}\big) \quad (77)$$
$$B=\big(\bar{\alpha}^{(2)},1\big) \quad (78)$$
$$C=\big(\frac{2+2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}}{3},\frac{2+2\bar{\alpha}^{(2)}-\bar{\alpha}^{(1)}}{3}\big) \quad (79)$$
$$D=\big(1,\bar{\alpha}^{(1)}\big). \quad (80)$$

To achieve the DoF region of Theorem 1 we need sufficiently good (but certainly not perfect) delayed CSIT such that

$$\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}\ge\min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\} \quad (81)$$

TABLE II
OPTIMAL CORNER POINTS SUMMARY, FOR SUFFICIENTLY GOOD DELAYED CSIT SUCH THAT $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} \geq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$.

Cases	Corner points	$\bar{\delta}$	ω
Case 1	C	$\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}$	0
	D	$\bar{\alpha}^{(1)}$	1
	B	$\bar{\alpha}^{(2)}$	0
Case 2	B	$\bar{\alpha}^{(2)}$	0
	A	$\frac{1+\bar{\alpha}^{(2)}}{2}$	0

TABLE III
DOF INNER BOUND CORNER POINTS, FOR DELAYED CSIT SUCH THAT $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} < \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$.

Cases	Corner points
Case 1 and case of $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} \geq \bar{\alpha}^{(1)}$	E,F,B,D
Case 1 and case of $\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} < \bar{\alpha}^{(1)}$	B,E,G
Case 2	B,E,G

(cf. Theorem 1), which in turn implies that (cf. (50))

$$\bar{\delta} \leq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}.$$

Under the condition of (81), the DoF corner points are achievable by setting the value of $\omega \in [0,1]$ and of $\bar{\delta} \leq \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\}$ as in Table II.

Specifically when (81) and (70) hold, we achieve DoF point B by setting $\omega = 0, \bar{\delta} = \bar{\alpha}^{(2)}$ which indeed gives (cf. (74),(75),(76))

$$d_1 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(2)})^+ = \bar{\alpha}^{(2)}$$
$$d_2 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(1)})^+ + \Delta_{\text{com}} = \bar{\alpha}^{(2)} + 1 - \bar{\alpha}^{(2)} = 1.$$

To achieve DoF point D we set $\omega = 1$ and $\bar{\delta} = \bar{\alpha}^{(1)}$ and get

$$d_1 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(2)})^+ + \Delta_{\text{com}} = \bar{\alpha}^{(1)} + 1 - \bar{\alpha}^{(1)} = 1$$
$$d_2 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(1)})^+ = \bar{\alpha}^{(1)}$$

while to achieve DoF point C we set $\omega = 0$ and $\bar{\delta} = \frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3}$ and get

$$d_1 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(2)})^+ = \frac{2+2\bar{\alpha}^{(1)}-\bar{\alpha}^{(2)}}{3}$$
$$d_2 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(1)})^+ + \Delta_{\text{com}} = \frac{2+2\bar{\alpha}^{(2)}-\bar{\alpha}^{(1)}}{3}.$$

On the other hand, when (71) (case 2) and (81) hold, to achieve DoF point B we set $\omega = 0$ and $\bar{\delta} = \bar{\alpha}^{(2)}$ as before, while to achieve DoF point A, we set $\omega = 0$ and $\bar{\delta} = \frac{1+\bar{\alpha}^{(2)}}{2}$.

Finally the entire DoF region of Theorem 1 is achieved using time sharing between these corner points.

3) DoF corner points of Proposition 1: Now we focus on the DoF points of Proposition 1 (see Table III). These are the points we label as DoF points B and D, as these were defined in (78) and (80), as well as three new DoF points

$$E = \big(2\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} - \bar{\alpha}^{(2)}, 1 + \bar{\alpha}^{(2)} - \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}\big) \quad (82)$$

$$F = \big(1 + \bar{\alpha}^{(1)} - \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}, 2\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} - \bar{\alpha}^{(1)}\big) \quad (83)$$

$$G = \big(1, \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}\big). \quad (84)$$

As stated in the proposition, we are interested in the regime of reduced-quality delayed CSIT, as this is defined by

$$\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} < \min\{\frac{1+\bar{\alpha}^{(1)}+\bar{\alpha}^{(2)}}{3},\frac{1+\bar{\alpha}^{(2)}}{2}\} \quad (85)$$

and which implies that $\bar{\delta} \leq \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}$ (cf. (50)). In addition to the two cases in (70),(71), we now additionally consider the cases where

$$\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} \geq \bar{\alpha}^{(1)} \quad (86)$$
$$\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} < \bar{\alpha}^{(1)}. \quad (87)$$

When (70),(85) and (86) hold, we set $\omega = 0, \bar{\delta} = \bar{\alpha}^{(2)}$ as before to achieve DoF point B. To achieve point D, we set $\omega = 1$ and $\bar{\delta} = \bar{\alpha}^{(1)}$ as before, whereas to achieve point E, we set $\omega = 0, \bar{\delta} = \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}$ to get (cf. (74), (75), (76))

$$d_1 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(2)})^+ = 2\min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\} - \bar{\alpha}^{(2)}$$
$$d_2 = \bar{\delta} + (\bar{\delta} - \bar{\alpha}^{(1)})^+ + \Delta_{\text{com}} = 1 + \bar{\alpha}^{(2)} - \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}.$$

Finally to achieve DoF point F, we set $\omega = 1$ and $\bar{\delta} = \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}$.

When (70),(85) and (87) hold, we achieve points B and E with the same parameters as before, while to achieve point G, we set $\omega = 1, \bar{\delta} = \min\{\bar{\beta}^{(1)},\bar{\beta}^{(2)}\}$.

Similarly when (71) and (85) hold, we achieve points B,E,G by setting ω and $\bar{\delta}$ as above.

Finally the entire DoF region of Proposition 1 is achieved with time sharing between the corner points.

D. Scheme $\mathcal{X}$: examples

We proceed to provide example instances of our general scheme, for specific cases of particular interest.

1) Fixed and imperfect quality delayed CSIT, no current CSIT: We consider the case of no current CSIT ($\alpha_t^{(i)} = 0, \ \forall t, i$) and of imperfect delayed CSIT of an unchanged quality $\beta \leq 1$. We focus on the case of $\beta = 1/3$. The universal scheme - with these parameters - achieves the optimal DoF by achieving the optimal DoF corner point ($d_1 = \frac{2}{3}, d_2 = \frac{2}{3}$), as in the case of [3] which assumed that the delayed feedback of a channel could be sent with perfect quality.

For this case of $\beta_t^{(i)} = 1/3, \alpha_t^{(i)} = 0$, we have $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} = 0$, $\bar{\beta}^{(1)} = \bar{\beta}^{(2)} = 1/3$. Toward designing the scheme, we set $\bar{\delta} = 1/3$ (cf. (50)). For the case of block fading where we can rewrite the time index to reflect a unit coherence period, delayed CSIT is simply the CSIT that comes during the next coherence period, i.e., during the next time slot. Given the assumption of i.i.d. fading employed in [3], we can set $\eta = 1$ (cf. (6),(7)), which allows for a simpler variant of our scheme where now the phases have duration $T = 1$. In this

TABLE IV
BITS CARRIED BY PRIVATE SYMBOLS, COMMON SYMBOLS, AND BY THE QUANTIZED INTERFERENCE, FOR PHASE $s = 1, 2, \cdots, S-1$.

	Total bits ($\log P$)
Private symbols for user 1	$2/3$
Private symbols for user 2	$2/3$
Common symbols	$2/3$
Quantized interference	$2/3$

TABLE V
BITS CARRIED BY PRIVATE SYMBOLS, COMMON SYMBOLS, AND BY THE QUANTIZED INTERFERENCE, FOR PHASE s, $s = 1, 2, \cdots, S-1$, OF THE ALTERNATING CSIT SCHEME.

	Total bits ($\times \log P$)
Private symbols for user 1	$(7 \times 2)/8$
Private symbols for user 2	$(7 \times 2)/8$
Common symbols	$(1 \times 2)/4$
Quantized interference	$(1 \times 2)/4$

simplified variant, the transmitted signal (cf. (43)) takes the simple form

$$\boldsymbol{x}_t = \boldsymbol{w}_t c_t + \begin{bmatrix} a_t \\ a_t^{'} \end{bmatrix} + \begin{bmatrix} b_t \\ b_t^{'} \end{bmatrix}$$

with the power and rates of the symbols (cf. (45)) set as

$$\begin{gathered} P_t^{(c)} \doteq P, \ r_t^{(c)} = 1 - 1/3 \\ P_t^{(a)} \doteq P_t^{(a')} \doteq P_t^{(b)} \doteq P_t^{(b')} \doteq P^{1/3} \\ r_t^{(a)} = r_t^{(a')} = r_t^{(b)} = r_t^{(b')} = 1/3. \end{gathered} \tag{88}$$

During each phase, the transmitter quantizes - as instructed in (57) - the interference accumulated in that phase, with a quantization rate of $2/3 \log P$, which is mapped into the common symbol c_{t+1} that will be transmitted in the next phase (at time-slot $t+1$). For large enough communication length, simple calculations can show that this can achieve the optimal DoF ($d_1 = \frac{2}{3}, d_2 = \frac{2}{3}$), and can do so with imperfect quality CSIT. Table IV summarizes the rates associated to the symbols in this scheme.

2) Alternating between two current-CSIT states: In the context of the two-user MISO BC with spatially and temporally i.i.d. fading and $M = 2$, the work in [13] considered the *alternating CSIT* setting where CSIT for the two users alternates between perfect current CSIT (labeled here as state P), perfect delayed CSIT (D), or no CSIT (N). In this setting where I_i denoted the CSIT state for the channel of user i at any given time ($I_1, I_2 \in \{P, D, N\}$), the work in [13] considered communication where, for a fraction $\lambda_{I_1 I_2}$ of the time, the CSIT states are equal to I_1, I_2 (state I_1 for the first user, state I_2 for the second user). The same work focused on the symmetric case where $\lambda_{I_1 I_2} = \lambda_{I_2 I_1}$. For $\lambda_P \triangleq \sum_{I_2 \in \{P,D,N\}} \lambda_{P I_2}$ being the fraction of the time where one user has perfect CSIT, and $\lambda_D \triangleq \sum_{I_2 \in \{P,D,N\}} \lambda_{D I_2}$ being the fraction of the time where one user had delayed CSIT, the work in [13] characterized the optimal DoF region to take the form

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

The above setting corresponds to our symmetric setting where $\alpha_t^{(1)}, \beta_t^{(1)}, \alpha_t^{(2)}, \beta_t^{(2)} \in \{0, 1\}$, $\forall t$, and where

$$\lambda_P = \bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} \tag{89}$$

$$\lambda_D = \bar{\beta}^{(1)} - \bar{\alpha}^{(1)} = \bar{\beta}^{(2)} - \bar{\alpha}^{(2)} \tag{90}$$

in which case our DoF inner bound matches the above, and as a result, for any $\bar{\beta} \geq \frac{1+2\bar{\alpha}}{3}$, Theorem 1 generalizes [13] to any set of quality exponents, avoiding the symmetry assumption, as well as easing on the i.i.d. block-fading assumption.

The universal scheme described in this section, can be directly applied to optimally implement more general alternating CSIT settings. We here offer an example where, in the presence of sufficiently good delayed CSIT, the current CSIT of the two users alternates between two quality exponents equal to $\frac{1}{2}$ and $\frac{3}{4}$, i.e.,

$$\begin{array}{lccccc} & t=1 & t=2 & t=3 & t=4 & \cdots \\ \alpha_t^{(1)} = & \frac{1}{2} & \frac{3}{4} & \frac{1}{2} & \frac{3}{4} & \cdots \\ \alpha_t^{(2)} = & \frac{3}{4} & \frac{1}{2} & \frac{3}{4} & \frac{1}{2} & \cdots \end{array}$$

In this case, which corresponds to having $\bar{\alpha}^{(1)} = \bar{\alpha}^{(2)} = 5/8$, we can choose any delayed CSIT process that gives $\bar{\beta}^{(1)} = \bar{\beta}^{(2)} = 3/4$ which suffices (see Corollary 1c) to achieve the optimal DoF region by achieving the optimal DoF point ($d_1 = \frac{7}{8}, d_2 = \frac{7}{8}$).

Toward designing the scheme, we set $\bar{\delta} = 3/4$. For this example, and again considering a block-fading fast-fading setting (unit-length coherence period), the scheme can have phases with duration $T = 2$. The transmitted signal (cf. (43)) now takes the form

$$\boldsymbol{x}_t = \boldsymbol{w}_t c_t + \hat{\boldsymbol{g}}_t^{\perp} a_t + \hat{\boldsymbol{h}}_t a_t^{'} + \hat{\boldsymbol{h}}_t^{\perp} b_t + \hat{\boldsymbol{g}}_t b_t^{'}$$

with power and rates of the symbols being set as instructed in (45). Again as instructed by the general description of the scheme, at the end of phase $s = 1, 2, \cdots, S-1$, the transmitter quantizes the interference accumulated during that phase, and does so using a total of $2(1/8 + 1/8)\log P$ quantization bits (cf. (57)). These bits are then mapped into the common symbols that will be transmitted in the next phase. For a large number of phases, the proposed scheme achieves the optimal DoF point ($d_1 = \frac{7}{8}, d_2 = \frac{7}{8}$). Table V summarizes the rates associated to the symbols in this scheme.

3) Schemes with short duration: We recall that the Maddah-Ali and Tse scheme [3], employs $\beta_1^{(i)} = 1, \beta_2^{(i)} = \beta_3^{(i)} = \alpha_1^{(i)} = \alpha_2^{(i)} = \alpha_3^{(i)} = 0$, $i = 1, 2$, and achieves the optimal DoF with only one phase ($T = 3$ channel uses, i.e., 3 coherence periods). This is done because the information bits of the quantized interference, 'fit' inside the common symbols of a single phase.

There are of course many other cases where this can happen.

One such example would be the case where

	$t=1$	$t=2$	$t=3$	$t=4$	$\cdots$
$\alpha_t^{(1)} =$	0	0	$\frac{1}{4}$	0	$\cdots$
$\beta_t^{(1)} =$	1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\cdots$
$\alpha_t^{(2)} =$	0	$\frac{1}{4}$	0	0	$\cdots$
$\beta_t^{(2)} =$	1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\cdots$

where a single-phase ($T=4$ time-slot) scheme, can achieve the optimal DoF corner point ($d_1 = \frac{11}{16}, d_2 = \frac{11}{16}$), again because the information bits of the quantized interference, can fit in the common symbols of the same phase.

V. Proof of outer bound Lemma

Proof: Let W_1, W_2 respectively denote the messages for the first and second user, and let R_1, R_2 denote the two users' rates. Each user sends their message over n channel uses, where n is large. For ease of exposition we introduce the following notation.

$$\boldsymbol{S}_t \triangleq \begin{bmatrix} \boldsymbol{h}_t^{\mathsf{T}} \\ \boldsymbol{g}_t^{\mathsf{T}} \end{bmatrix}, \quad \check{\boldsymbol{S}}_t \triangleq \begin{bmatrix} \check{\boldsymbol{h}}_t^{\mathsf{T}} \\ \check{\boldsymbol{g}}_t^{\mathsf{T}} \end{bmatrix}, \quad \hat{\boldsymbol{S}}_t \triangleq \begin{bmatrix} \hat{\boldsymbol{h}}_t^{\mathsf{T}} \\ \hat{\boldsymbol{g}}_t^{\mathsf{T}} \end{bmatrix}, \quad \boldsymbol{z}_t \triangleq \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix}$$

$$y_{[n]}^{(i)} \triangleq \{y_t^{(i)}\}_{t=1}^n, \quad i = 1,2$$

$$\Omega_{[n]} \triangleq \{\boldsymbol{S}_t, \check{\boldsymbol{S}}_t, \hat{\boldsymbol{S}}_t\}_{t=1}^n.$$

The first step is to construct a degraded BC by providing the first user with complete and immediately available information on the second user's received signal. In this improved scenario, the following bounds hold.

$$\begin{aligned}
&nR_1 \\
&= H(W_1) \\
&= H(W_1|\Omega_{[n]}) \\
&\le I(W_1; y_{[n]}^{(1)}, y_{[n]}^{(2)}|\Omega_{[n]}) + n\epsilon_n && (91)\\
&\le I(W_1; W_2, y_{[n]}^{(1)}, y_{[n]}^{(2)}|\Omega_{[n]}) + n\epsilon_n \\
&= I(W_1; y_{[n]}^{(1)}, y_{[n]}^{(2)}|W_2, \Omega_{[n]}) + n\epsilon_n \\
&= h(y_{[n]}^{(1)}, y_{[n]}^{(2)}|W_2, \Omega_{[n]}) - \underbrace{h(y_{[n]}^{(1)}, y_{[n]}^{(2)}|W_1, W_2, \Omega_{[n]})}_{no(\log P)} + n\epsilon_n \\
&= \sum_{t=1}^n h(y_t^{(1)}, y_t^{(2)}|y_{[t-1]}^{(1)}, y_{[t-1]}^{(2)}, W_2, \Omega_{[n]}) + no(\log P) + n\epsilon_n && (92)
\end{aligned}$$

where (91) results from Fano's inequality, where $y_0^{(i)}$ was set to zero by convention, and where the last equality follows from the entropy chain rule and the fact that the knowledge of $\{W_1, W_2, \Omega_{[n]})\}$ implies knowledge of $\{y_{[n]}^{(1)}, y_{[n]}^{(2)}\}$ up to noise level.

Similarly

$$\begin{aligned}
&nR_2 \\
&= H(W_2) \\
&\le I(W_2; y_{[n]}^{(2)}|\Omega_{[n]}) + n\epsilon_n && (93)\\
&= \underbrace{h(y_{[n]}^{(2)}|\Omega_{[n]})}_{\le n\log P + no(\log P)} - h(y_{[n]}^{(2)}|W_2, \Omega_{[n]}) + n\epsilon_n && (94)\\
&\le -\sum_{t=1}^n h(y_t^{(2)}|y_{[t-1]}^{(2)}, W_2, \Omega_{[n]}) + n\log P + no(\log P) + n\epsilon_n && (95)\\
&\le -\sum_{t=1}^n h(y_t^{(2)}|y_{[t-1]}^{(1)}, y_{[t-1]}^{(2)}, W_2, \Omega_{[n]}) \\
&\qquad + n\log P + no(\log P) + n\epsilon_n && (96)
\end{aligned}$$

where (95) follows from the entropy chain rule and from the fact that received signals are scalars, while the last step is due to the fact that conditioning reduces entropy.

Now given (92) and (96), we upper bound $R_1 + 2R_2$ as

$$\begin{aligned}
&n(R_1 + 2R_2) \le 2n\log P + no(\log P) + 3n\epsilon_n \\
&+ \sum_{t=1}^n \Big(h(y_t^{(1)}, y_t^{(2)}|U, S_t, \hat{S}_t) - 2h(y_t^{(2)}|U, S_t, \hat{S}_t) \Big) \qquad (97)
\end{aligned}$$

where

$$U \triangleq \{y_{[t-1]}^{(1)}, y_{[t-1]}^{(2)}, W_2, \Omega_{[n]}\} \setminus S_t, \hat{S}_t$$

and where each term $h(y_t^{(1)}, y_t^{(2)}|U, S_t, \hat{S}_t) - 2h(y_t^{(2)}|U, S_t, \hat{S}_t)$ in the summation, can be upper bounded as

$$\begin{aligned}
&h(y_t^{(1)}, y_t^{(2)}|U, S_t, \hat{S}_t) - 2h(y_t^{(2)}|U, S_t, \hat{S}_t) \\
&\le \max_{\substack{P_{X_t} \\ \mathbb{E}[\mathrm{tr}(X_t X_t^{\mathsf{H}})] \le P}} \Big[h(y_t^{(1)}, y_t^{(2)}|U, S_t, \hat{S}_t) - 2h(y_t^{(2)}|U, S_t, \hat{S}_t) \Big] \\
&\le \mathbb{E}_{\hat{S}_t} \max_{\substack{P_{X_t} \\ \mathbb{E}[\mathrm{tr}(X_t X_t^{\mathsf{H}})] \le P}} \mathbb{E}_{S_t|\hat{S}_t} \Big[h(y_t^{(1)}, y_t^{(2)}|U, S_t = \boldsymbol{S}_t, \hat{S}_t = \hat{\boldsymbol{S}}_t) \\
&\qquad - 2h(y_t^{(2)}|U, S_t = \boldsymbol{S}_t, \hat{S}_t = \hat{\boldsymbol{S}}_t) \Big] \\
&= \mathbb{E}_{\hat{S}_t} \max_{\substack{P_{X_t} \\ \mathbb{E}[\mathrm{tr}(X_t X_t^{\mathsf{H}})] \le P}} \mathbb{E}_{\tilde{S}_t} \Big[h(\boldsymbol{S}_t \boldsymbol{x}_t + \boldsymbol{z}_t|U) - 2h(\boldsymbol{g}_t^{\mathsf{T}} \boldsymbol{x}_t + z_t^{(2)}|U) \Big] \\
&= \mathbb{E}_{\hat{S}_t} \max_{\boldsymbol{\Psi} \succeq 0 : \mathrm{tr}(\boldsymbol{\Psi}) \le P} \mathbb{E}_{\tilde{S}_t} \Big[\log\det(I + \boldsymbol{S}_t \boldsymbol{\Psi} \boldsymbol{S}_t^{\mathsf{H}}) - 2\log(1 + \boldsymbol{g}_t^{\mathsf{H}} \boldsymbol{\Psi} \boldsymbol{g}_t) \Big] && (98)\\
&\le \mathbb{E}_{\hat{S}_t} \max_{\boldsymbol{\Psi} \succeq 0 : \mathrm{tr}(\boldsymbol{\Psi}) \le P} \mathbb{E}_{\tilde{S}_t} \Big[\log(1 + \boldsymbol{h}_t^{\mathsf{H}} \boldsymbol{\Psi} \boldsymbol{h}_t) - \log(1 + \boldsymbol{g}_t^{\mathsf{H}} \boldsymbol{\Psi} \boldsymbol{g}_t) \Big]. && (99)
\end{aligned}$$

In the above, (98) uses the results in [33, Corollary 4] that tell us that Gaussian input maximizes the weighted difference of two differential entropies[10], as long as: 1) $y_t^{(2)}$ is a degraded

[10] We note that the results in [33, Corollary 4] are described for the non-fading channel model, however, as argued in the same work in [33, Section V], the results can be readily extended to the fading channel model by linearly transforming the fading channel into an equivalent non-fading channel, with the new channel actually maintaining the same capacity and the same degradedness order.

version of $\{y_t^{(1)}, y_t^{(2)}\}$; 2) U is independent of $z_t^{(1)}, z_t^{(2)}$; 3) the input maximization is done given a fixed fading realization $\hat{\boldsymbol{S}}_t$, and is independent of $\tilde{\boldsymbol{S}}_t$ [11]. Furthermore, in the above, (99) comes from Fischer's inequality which gives that $\det(I + \boldsymbol{S}_t\boldsymbol{\Psi}\boldsymbol{S}_t^{\mathsf{H}}) \leq (1 + \boldsymbol{h}_t^{\mathsf{H}}\boldsymbol{\Psi}\boldsymbol{h}_t)(1 + \boldsymbol{g}_t^{\mathsf{H}}\boldsymbol{\Psi}\boldsymbol{g}_t)$.

At this point we follow the steps involving equation (25) in [4], to upper bound the right hand side of (99) as

$$\mathbb{E}_{\hat{S}_t} \max_{\boldsymbol{\Psi}\succeq 0:\mathrm{tr}(\boldsymbol{\Psi})\leq P} \mathbb{E}_{\tilde{S}_t}\left[\log\left(1 + \boldsymbol{h}_t^{\mathsf{H}}\boldsymbol{\Psi}\boldsymbol{h}_t\right) - \log\left(1 + \boldsymbol{g}_t^{\mathsf{H}}\boldsymbol{\Psi}\boldsymbol{g}_t\right)\right]$$
$$\leq \alpha_t^{(2)} \log P + o(\log P). \tag{100}$$

Combining (97) and (99), gives that

$$n(R_1 + 2R_2) \leq \sum_{t=1}^{n}\left((2 + \alpha_t^{(2)}) \log P + o(\log P) + 3\epsilon_n\right)$$

and consequently that

$$d_1 + 2d_2 \leq 2 + \bar{\alpha}^{(2)}.$$

Similarly, interchanging the roles of the two users, allows for

$$d_2 + 2d_1 \leq 2 + \bar{\alpha}^{(1)}.$$

Finally the fact that each user has a single receive antenna, gives that $d_1 \leq 1, d_2 \leq 1$. ■

VI. Conclusions

The work made progress toward establishing and meeting the limits of using imperfect and delayed feedback. Considering a general CSIT process and a primitive measure of feedback quality, the work provided DoF expressions that are simple and insightful functions of easy to calculate parameters which concisely capture the problem complexity. The derived insight addresses practical questions on topics relating to the usefulness of predicted, current and delayed CSIT, the impact of estimate precision, the effect of feedback delays, and the benefit of having feedback symmetry by employing comparable feedback links across users. Further insight was derived from the introduced *periodically evolving feedback* setting, which captures many of the engineering options relating to feedback, as well as incorporates and generalizes many previously considered settings of interest. For our chosen setting of a small number of users (two in this case), we expect these high-SNR insights to hold for SNR values of operational interest. The nature of the improved bounds and novel constructions, allows for this same insight to hold for a broad family of block fading and non-block fading channel models.

We believe that the adopted approach is fundamental, in the sense that it considers a general fading process, a general CSIT process, and a primitive measure of feedback quality in the form of the precision of estimates at any time about any channel, i.e., in the form of the entire set of estimation errors $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$ at any time about any channel. This set of errors naturally fluctuates depending on the instance of the problem, and as expected, the overall optimal performance is defined by the statistics of this error set. These statistics are mildly constrained to the case of having Gaussian estimation errors which are independent of the prior and current channel estimates. Under these assumptions, the results capture the performance effect of the statistics of feedback. Interestingly this effect - at least for sufficiently good delayed CSIT, and for high SNR - is captured by the averages of the quality exponents. This can be traced back to the assumption that the estimation errors are Gaussian, which means that the statistics of $\{(\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}), (\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'})\}_{t,t'=1}^n$ are captured by a covariance matrix that has diagonal (block) entries of the form $\{\frac{1}{M}\mathbb{E}[||\boldsymbol{h}_t - \hat{\boldsymbol{h}}_{t,t'}||_F^2], \frac{1}{M}\mathbb{E}[||\boldsymbol{g}_t - \hat{\boldsymbol{g}}_{t,t'}||_F^2]\}_{t,t'=1}^n$, and whose off-diagonal entries are not used by the scheme, which though meets an outer bound that has kept open the possibility of any off-diagonal elements. Hence under our assumptions, the essence of the CSIT error statistics is captured by the diagonal block elements (of the aforementioned covariance matrix) whose effects are in turn captured - in the high-SNR regime - by the quality exponents.

This general approach allows for consideration of many facets of the performance-vs-feedback question in the two-user MISO BC setting, accentuating the important facets while revealing the reduced role of other facets. For example, while the approach allows for consideration of predicted CSIT - i.e., of estimates for future channels - the result at the end reveals that such estimates do not provide DoF gains, again under our assumptions. In a similar manner, the result leaves open the possibility of a role in the off-diagonal elements of the aforementioned covariance matrix of estimation errors, but in the end again reveals that these can be neglected without a DoF effect. Similarly, the approach allows for any 'typical' sequence of quality exponents - thus avoiding the need to assume periodic or static feedback processes or a block-fading structure - but despite this generality in the range of the considered exponents, in the end the result reveals that what really matters is the long-term average of each of these sequences of current and delayed CSIT exponents.

Finally we believe the main assumptions here to be mild. Regarding the high SNR assumption, there is substantial evidence that for primitive networks (such as the BC and the IC) with a reasonably small number of users, DoF analysis offers good insight on the performance at moderate SNR. Any possible extensions though to the setting of larger cellular networks, may need to consider saturation effects on the high-SNR spectral efficiency, as these were recently revealed in [28] to hold for settings where communication involves clusters of large size. Furthermore the assumption of having global CSIR, allowed us to focus on the question of feedback to the transmitters, which is a fundamental question on its own. While the overhead of gathering global CSIR must not be neglected, it has been repeatedly shown (cf. [34], [35]) that this overhead is manageable in the presence of a reduced number of users. When considering extensions to other multiuser networks with potentially more users, such analysis may have to be combined with finding ways to disseminate imperfect global CSIR (cf. [27], [34], [35], see also [29], [36]) whose effect increases as the number of users increases. Additionally asking that current estimation errors are independent of current

[11] We recall that $\boldsymbol{x}_t$ is only a function of the messages and of the CSIT (current and delayed) estimates up to time t, and that these CSIT estimates are assumed to be independent of the current estimate errors at time t.

estimates, is a widely accepted assumption. Similarly accepted is the assumption that the estimation error is independent of the past estimates, as this assumption suggests good feedback processes that utilize possible correlations to improve current channel estimates. Finally the requirement that the running average of the quality exponents of a single user, converges to a fixed value after a sufficiently long time, is also believed to be reasonable, as it would hold even if these exponents were themselves treated as random variables from an ergodic process.

VII. Appendix - Further details on the scheme

A. Explicit power allocation solutions under constraints in equations (47),(48),(49)

We remind the reader that, in designing the power allocation policy of the scheme, we must design the power parameters $\{\delta_t^{(1)}, \delta_t^{(2)}\}_{t\in\mathcal{B}_s}$ to satisfy equations (47),(48),(49) which asked that

$$\beta_t^{(i)} \geq \delta_t^{(i)} \quad i=1,2,\ t\in\mathcal{B}_s$$
$$\frac{1}{T}\sum_{t\in\mathcal{B}_s}\delta_t^{(1)} = \frac{1}{T}\sum_{t\in\mathcal{B}_s}\delta_t^{(2)} = \bar{\delta}$$
$$\frac{1}{T}\sum_{t\in\mathcal{B}_s}(\delta_t^{(i)}-\alpha_t^{(i)})^+ = (\bar{\delta}-\bar{\alpha}^{(i)})^+ \quad i=1,2$$

for a given $\bar{\delta}\in[0,1]$. For each phase s, we here explicitly describe such sequence $\{\delta_t^{(1)}, \delta_t^{(2)}\}_{t\in\mathcal{B}_s}$, which is constructed using a waterfilling-like approach.

We first consider the case where $\bar{\delta}\geq\bar{\alpha}^{(i)}$. At any given time $t=\mathcal{B}_{s,1},\mathcal{B}_{s,2},\cdots,\mathcal{B}_{s,T}$, we set

$$\delta_t^{(i)} = \begin{cases} T(\bar{\delta}-\bar{\alpha}^{(i)})-\Delta_{\delta,t}+\alpha_t^{(i)} & \text{if } \beta_t^{(i)}\geq T(\bar{\delta}-\bar{\alpha}^{(i)})-\Delta_{\delta,t}+\alpha_t^{(i)} \\ \beta_t^{(i)} & \text{if } \beta_t^{(i)}< T(\bar{\delta}-\bar{\alpha}^{(i)})-\Delta_{\delta,t}+\alpha_t^{(i)} \end{cases}$$

where $\Delta_{\delta,t}$ is initialized to zero ($\Delta_{\delta,\mathcal{B}_{s,1}}=0$), and is updated each time, so that the calculation of $\delta_{t+1}^{(i)}$, uses

$$\Delta_{\delta,t+1} = \Delta_{\delta,t}+\delta_t^{(i)}-\alpha_t^{(i)}.$$

In the end, the solution takes the form

$$\delta_t^{(i)} = \begin{cases} \beta_t^{(i)}, & t=\mathcal{B}_{s,1},\cdots,\mathcal{B}_{s,\tau'-1} \\ T(\bar{\delta}-\bar{\alpha}^{(i)})+\alpha_t^{(i)}-\sum_{\ell=1}^{\tau'-1}(\beta_{\mathcal{B}_{s,\ell}}^{(i)}-\alpha_{\mathcal{B}_{s,\ell}}^{(i)}), & t=\mathcal{B}_{s,\tau'} \\ \alpha_t^{(i)}, & t=\mathcal{B}_{s,\tau'+1},\cdots,\mathcal{B}_{s,T} \end{cases}$$

where τ' is a function[12] of the quality exponents during phase s. This design of $\{\delta_t^{(i)}\}_{t\in\mathcal{B}_s}$ satisfies (47),(48), as well as (49), since, for the case where $\bar{\delta}-\bar{\alpha}^{(i)}\geq 0$, we deliberate forced $\delta_t^{(i)}-\alpha_t^{(i)}\geq 0,\ t\in\mathcal{B}_s$.

Similarly for $\bar{\delta}\leq\bar{\alpha}^{(i)}$, we set

$$\delta_t^{(i)} = \begin{cases} \alpha_t^{(i)} & \text{if } \alpha_t^{(i)}\leq T\bar{\delta}-\Delta_{\delta,t} \\ T\bar{\delta}-\Delta_{\delta,t} & \text{if } \alpha_t^{(i)}> T\bar{\delta}-\Delta_{\delta,t} \end{cases}$$

where $\Delta_{\delta,t}$ is initialized to zero, and is updated as

$$\Delta_{\delta,t+1} = \Delta_{\delta,t}+\delta_t^{(i)}.$$

[12] Note that there is no need to explicitly describe τ', because the schemes are explicitly described as a function of the above $\delta_t^{(i)}$, which - after calculation - also reveal τ' which - by design - falls within the proper range.

In the end, in this case, the solution takes the form

$$\delta_t^{(i)} = \begin{cases} \alpha_t^{(i)}, & t=\mathcal{B}_{s,1},\cdots,\mathcal{B}_{s,\tau'-1} \\ T\bar{\delta}-\sum_{\ell=1}^{\tau'-1}\alpha_{\mathcal{B}_{s,\ell}}^{(i)}, & t=\mathcal{B}_{s,\tau'} \\ 0, & t=\mathcal{B}_{s,\tau'+1},\cdots,\mathcal{B}_{s,T} \end{cases}$$

where again τ' is a function of the quality exponents during phase s. This satisfies (47),(48), as well as (49) since, for the case where $\bar{\delta}-\bar{\alpha}^{(i)}\leq 0$, we again have $\delta_t^{(i)}-\alpha_t^{(i)}\leq 0,\ t\in\mathcal{B}_s$.

B. Encoding and decoding details for steps in equations (66) , (68)

We here elaborate on how the users will be able to decode the amount of feedback bits described in equations (66) and (68). We first provide the following lemma, which holds for any T.

Lemma 2: Let

$$\bar{y}_t^{(1)} = c_t + P^{\frac{\delta_t^{(2)}}{2}}\bar{z}_t^{(1)}, \tag{101}$$

$$\bar{y}_t^{(2)} = c_t + P^{\frac{\delta_t^{(1)}}{2}}\bar{z}_t^{(2)}, \quad t=1,2,\cdots,T \tag{102}$$

where $\mathbb{E}[|c_t|^2]\leq P$, $Pr(|\bar{z}_t^{(i)}|^2 > P^{\epsilon})\doteq 0$, and $\frac{1}{T}\sum_{t=1}^{T}\delta_t^{(i)}\leq\bar{\delta}^*$ for a given $\bar{\delta}^*\in[0,1]$, $i=1,2$. Also let $r\triangleq 1-\bar{\delta}^*-\epsilon$ for a vanishingly small but positive $\epsilon>0$, and consider communication over T channel uses. Then for any rate up to $R=r\log P-o(\log P)$ (bits/channel use), the probability of error can be made to vanish with asymptotically increasing SNR.

Proof: We will draw each T-length codevector

$$\boldsymbol{c}\triangleq[c_1,\cdots,c_T]^{\intercal}$$

from a lattice code of the form

$$\{\theta\boldsymbol{M}\boldsymbol{q} \mid \boldsymbol{q}\in\aleph\} \tag{103}$$

where $\aleph\subset\mathbb{C}^T$ is the T-dimensional 2^R-QAM constellation, where $\boldsymbol{M}\in\mathbb{C}^{T\times T}$ is a specifically constructed unitary matrix of algebraic conjugates that allows for the *non vanishing product distance* property (to be described later on - see for example [37]), and where

$$\theta = P^{\frac{1-r}{2}} = P^{(\bar{\delta}^*+\epsilon)/2} \tag{104}$$

is designed to guarantee that $\mathbb{E}||\boldsymbol{c}||^2\doteq P$ (to derive this value of θ, just recall the QAM property that $\mathbb{E}||\boldsymbol{q}||^2\doteq 2^R\doteq P^r$). Specifically for any two codevectors $\boldsymbol{c}=[c_1,\cdots,c_T]^{\intercal}, \boldsymbol{c}^{\star}=[c_1^{\star},\cdots,c_T^{\star}]^{\intercal}$, $\boldsymbol{M}$ is designed to guarantee that

$$\prod_{t=1}^{T}|(c_t-c_t^{\star})|^2 \dot{\geq} \theta^{2T}. \tag{105}$$

This can be readily done for all dimensions T by, for example, using the proper roots of unity as entries of a circulant $\boldsymbol{M}$ (cf. [37]), which in turn allows for the above product - before normalization with θ - to take non-zero integer values.

In the post-whitened channel model at user $i=1,2$, we have

$$\begin{aligned}\bar{\bar{\boldsymbol{y}}}^{(1)} &\triangleq \mathrm{diag}(P^{-\delta_1^{(2)}/2},\cdots,P^{-\delta_T^{(2)}/2})\bar{\boldsymbol{y}}^{(1)}\\ &= \mathrm{diag}(P^{-\delta_1^{(2)}/2},\cdots,P^{-\delta_T^{(2)}/2})\boldsymbol{c}+\bar{\boldsymbol{z}}^{(1)}\\ \bar{\bar{\boldsymbol{y}}}^{(2)} &\triangleq \mathrm{diag}(P^{-\delta_1^{(1)}/2},\cdots,P^{-\delta_T^{(1)}/2})\boldsymbol{c}+\bar{\boldsymbol{z}}^{(2)}\end{aligned}$$

where, as we have stated, the noise $\bar{\boldsymbol{z}}^{(i)}$ has finite power in the sense that

$$Pr(||\bar{\boldsymbol{z}}^{(i)}||^2 > P^{\epsilon}) \to 0. \tag{106}$$

At the same time, after whitening at each user, the codeword distance for any two codewords $\boldsymbol{c},\boldsymbol{c}^{\star}$, is lower bounded as

$$\begin{aligned}&||\mathrm{diag}(P^{-\delta_1^{(i)}/2},\cdots,P^{-\delta_T^{(i)}/2})(\boldsymbol{c}-\boldsymbol{c}^{\star})||^2\\ &=\sum_{t=1}^{T}|P^{-\delta_t^{(i)}/2}(c_t-c_t^{\star})|^2\\ &\dot{\geq}\prod_{t=1}^{T}|P^{-\delta_t^{(i)}/2}(c_t-c_t^{\star})|^{2/T} && (107)\\ &=P^{-\frac{1}{T}\sum_{t=1}^{T}\delta_t^{(i)}}\prod_{t=1}^{T}|(c_t-c_t^{\star})|^{2/T}\\ &\dot{\geq}P^{-\frac{1}{T}\sum_{t=1}^{T}\delta_t^{(i)}}\theta^2 && (108)\\ &\geq P^{-\bar{\delta}^*}P^{\bar{\delta}^*+\epsilon}=P^{\epsilon} && (109)\end{aligned}$$

for $i=1,2$, where (107) results from the arithmetic-mean geometric-mean inequality, (108) is due to (105), and where (109) uses the assumption that $\frac{1}{T}\sum_{t=1}^{T}\delta_t^{(i)}\leq\bar{\delta}^*$. Setting ϵ positive but vanishingly small, combined with (106), proves the result. ∎

At this point, we use the lattice code of the above lemma, to design the T-length vector $\boldsymbol{c}$ transmitted during phase s. This encoding guarantees successful decoding of this vector, at both users, at a rate $R=r\log P-o(\log P)$, where $r=1-\bar{\alpha}^{(2)}$ for phase S, else $r=1-\bar{\delta}$ (ϵ is set positive but vanishingly small, recall (66), (68)). We note that for phase S, user $i=1,2$ can linearly transform their signal observations $\{y_t^{(i)}\}_{t\in\mathcal{B}_S}$ (cf. (63),(64)) to take the form in (101),(102), while for phase $s=1,2,\cdots,S-1$, user $i=1,2$ can linearly transform their signal observations $\{y_t^{(i)}-\bar{\iota}_t^{(i)}\}_{t\in\mathcal{B}_s}$ (after removing the interference $\bar{\iota}_t^{(i)}$, cf. (67),(51),(52)), again to take the form in (101) , (102).

Finally we note that the achievable rate is determined by the exponent average $\frac{1}{T}\sum_{t=1}^{T}\delta_t^{(i)}$ and not by the instantaneous exponents $\delta_t^{(i)}$.

VIII. Appendix - Discussion on independence of estimation error and past estimates

The assumption is consistent with a large family of channel models ranging from the fast fading channel (i.i.d in time), to the correlated channel as this was presented in [4][13], and even the quasi-static slow fading model where the CSIT estimates are successively refined over time. Successive CSIT refinement - as this is treated in [1] - considers an incremental amount of quantization bits that progressively improve the CSIT estimates. For example, focusing on the estimates of channel $\boldsymbol{h}_1$, the quality of this estimate would improve in time, with a successive refinement that would entail

$$\begin{aligned}\boldsymbol{h}_1 &= \hat{\boldsymbol{h}}_{1,1}+\tilde{\boldsymbol{h}}_{1,1}\\ &= \underbrace{\hat{\boldsymbol{h}}_{1,1}+\hat{\hat{\boldsymbol{h}}}_{1,1,2}}_{\hat{\boldsymbol{h}}_{1,2}}+\tilde{\boldsymbol{h}}_{1,2}\\ &= \underbrace{\hat{\boldsymbol{h}}_{1,1}+\hat{\hat{\boldsymbol{h}}}_{1,1,2}+\hat{\hat{\boldsymbol{h}}}_{1,2,3}}_{\hat{\boldsymbol{h}}_{1,3}}+\tilde{\boldsymbol{h}}_{1,3}\\ &\vdots\end{aligned}$$

where

$$\tilde{\boldsymbol{h}}_{1,t'}\triangleq\hat{\hat{\boldsymbol{h}}}_{1,t',t''}+\tilde{\boldsymbol{h}}_{1,t''}$$

and where $\hat{\hat{\boldsymbol{h}}}_{1,t',t''}$ denotes the estimate correction that happens between time t' and t''.

Generalizing this to the estimate of any channel $\boldsymbol{h}_t$, and accepting that the estimate correction $\hat{\hat{\boldsymbol{h}}}_{t,t',t''}$ and estimate error $\tilde{\boldsymbol{h}}_{t,t''}$ are statistically independent, allows that the estimation error $\tilde{\boldsymbol{h}}_{t,t''}$ of $\boldsymbol{h}_t$ is independent of the previous and current estimates $\{\hat{\boldsymbol{h}}_{t,\tau}\}_{\tau\leq t''}$, which in turn allows for the aforementioned assumption to hold even for the block fading channel model.

As a side note, even though we consider the quantification of CSIT quality as in (39), we note that our results can be readily extended to the case where we estimate channel directions (phases), in which case we would simply consider $\frac{1}{||\boldsymbol{h}_t||}\boldsymbol{h}_t=\hat{\boldsymbol{h}}_t+\tilde{\boldsymbol{h}}_t,\ \frac{1}{||\boldsymbol{g}_t||}\boldsymbol{g}_t=\hat{\boldsymbol{g}}_t+\tilde{\boldsymbol{g}}_t$ (cf. [19]).

IX. Appendix - Proof of Corollary 1g

In the presence of a constraint on α_{T_c} but not on β, we can raise β such that $\beta\geq\frac{1+2\bar{\alpha}}{3}$, in which case we have that $\bar{\alpha}=3d-2$ (cf. Corollary 1a), and $\frac{1+2\bar{\alpha}}{3}=2d-1$, which allows us to reach

$$\alpha_1=\cdots=\alpha_{\gamma T_c}=0,\alpha_{\gamma T_c+1}=\cdots=\alpha_{T_c}=2d-1=\beta$$

after setting $(1-\gamma)\alpha_{T_c}=\bar{\alpha}=3d-2$.

In the presence of a constraint on β but not on α_{T_c}, when $\beta<\frac{1+2\bar{\alpha}}{3}$, then Corollary 1a gives that $\beta=2d-1$, which means that $\bar{\alpha}\geq\frac{3\beta-1}{2}=3d-2$, which in turn allows us to set $\alpha_{T_c}=\beta=2d-1$ and get

$$\alpha_1=\cdots=\alpha_{\gamma T_c}=0,\alpha_{\gamma T_c+1}=\cdots=\alpha_{T_c}=\beta=2d-1.$$

Finally in the absence of any constraint on α_{T_c} and β, we can set $\alpha_{\gamma T_c+1}=\cdots\alpha_{\gamma T_c}=1=\beta$ for the maximum γ that allows for the desired average to hold.

[13] Note that our assumption is softer than the assumption in [4] where $\{\{\hat{\boldsymbol{h}}_{t,t'},\hat{\boldsymbol{g}}_{t,t'}\}_{t'=1}^{t},\boldsymbol{h}_t,\boldsymbol{g}_t\}_{t=1}^{t^*-1}\leftrightarrow\{\hat{\boldsymbol{h}}_{t^*,t^*},\hat{\boldsymbol{g}}_{t^*,t^*}\}\leftrightarrow\{\boldsymbol{h}_{t^*},\boldsymbol{g}_{t^*}\}$ was assumed to be a Markov chain; an assumption which may not directly hold in block fading settings where for example, having $\boldsymbol{h}_{t^*-1}=\boldsymbol{h}_{t^*}$ (resp. $\boldsymbol{g}_{t^*-1}=\boldsymbol{g}_{t^*}$), breaks the chain $\{\boldsymbol{h}_{t^*-1},\boldsymbol{g}_{t^*-1}\}\leftrightarrow\{\hat{\boldsymbol{h}}_{t^*,t^*},\hat{\boldsymbol{g}}_{t^*,t^*}\}\leftrightarrow\{\boldsymbol{h}_{t^*},\boldsymbol{g}_{t^*}\}$.

REFERENCES

[1] N. Jindal, “MIMO broadcast channels with finite-rate feedback,” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045 – 5060, Nov. 2006.

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845 – 2866, Jun. 2010.

[3] 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.

[4] 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.

[5] 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.

[6] H. Maleki, S. Jafar, and S. Shamai, “Retrospective interference alignment over interference networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 228 – 240, Mar. 2012.

[7] 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.

[8] M. Kobayashi, S. Yang, D. Gesbert, and X. Yi, “On the degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2012.

[9] 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.

[10] 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.

[11] 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.

[12] A. Lapidoth and S. Shamai, “Fading channels: how perfect need “perfect side information” be?” *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118 – 1134, May 2002.

[13] R. Tandon, S. A. Jafar, S. Shamai, and H. V. Poor, “On the synergistic benefits of alternating CSIT for the MISO BC,” Aug. 2012, to appear in *IEEE Trans. Inform. Theory*, available on arXiv:1208.5071.

[14] 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.

[15] 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.

[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. Information Theory (ISIT)*, Jul. 2011.

[17] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, “On the degrees of freedom of three-user MIMO broadcast channel with delayed CSIT,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2011.

[18] A. Ghasemi, A. S. Motahari, and A. K. Khandani, “Interference alignment for the MIMO interference channel with delayed local CSIT,” Feb. 2011, available on arXiv:1102.5673v1.

[19] 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.

[20] 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.

[21] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, “The degrees of freedom region of temporally-correlated MIMO networks with delayed CSIT,” Nov. 2012, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1211.3322.

[22] J. Chen and P. Elia, “MISO broadcast channel with delayed and evolving CSIT,” Nov. 2012, to appear in *ISIT13*, available on arXiv:1211.1622.

[23] J. Chen, S. Yang, and P. Elia, “On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT,” 2013, to appear in *ISIT13*, available on arXiv:1302.0806.

[24] P. de Kerret, X. Yi, and D. Gesbert, “On the degrees of freedom of the K-user time correlated broadcast channel with delayed CSIT,” 2013, to appear in *ISIT13*, available on arXiv:1301.2138.

[25] C. Hao and B. Clerckx, “Imperfect and unmatched CSIT is still useful for the frequency correlated MISO broadcast channel,” Feb. 2013, to appear in *ICC13*, available on arXiv:1302.6521.

[26] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Capacity results for binary fading interference channels with delayed CSIT,” Jan. 2013, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1301.5309.

[27] J. Chen and P. Elia, “MIMO BC with imperfect and delayed channel state information at the transmitter and receivers,” Jun. 2013, to appear in *Proc. IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC13)*.

[28] A. Lozano, R. W. Heath Jr., and J. G. Andrews, “Fundamental limits of cooperation,” 2013, to appear in *IEEE Trans. Inf. Theory*, available on arXiv:1204.0011.

[29] Y. Lejosne, D. Slock, and Y. Yuan-Wu, “NetDoFs of the MISO broadcast channel with delayed CSIT feedback for finite rate of innovation channel models,” 2013, to appear in *ISIT13*.

[30] J. Chen and P. Elia, “Degrees-of-freedom region of the MISO broadcast channel with general mixed-CSIT,” in *Proc. Information Theory and Applications Workshop (ITA)*, Feb. 2013.

[31] ——, “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.

[32] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.

[33] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg, and P. Viswanath, “The capacity region of the degraded multiple-input multiple-output compound broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011 – 5023, Nov. 2009.

[34] A. Adhikary, H. C. Papadopoulos, S. A. Ramprashad, and G. Caire, “Multi-user MIMO with outdated CSI: Training, feedback and scheduling,” in *Proc. Allerton Conf. Communication, Control and Computing*, Sep. 2011.

[35] M. Kobayashi and G. Caire, “On the net DoF comparison between ZF and MAT over time-varying MISO broadcast channels,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2012.

[36] O. El Ayach, A. Lozano, and R. W. Heath Jr., “On the overhead of interference alignment: Training, feedback, and cooperation,” *IEEE Trans. Wireless Commun.*, vol. 58, no. 11, pp. 4192 – 4203, Nov. 2012.

[37] J. Boutros, E. Viterbo, C. Rastello, and J. C. Belfiore, “Good lattice constellations for both Rayleigh fading and Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 501 – 518, Mar. 1996.