Distributed CSIT Does Not Reduce the Generalized DoF of the 2-user MISO Broadcast Channel

Antonio Bazco-Nogueras, *Student Member, IEEE,* Paul de Kerret, *Member, IEEE,*
David Gesbert, *Fellow, IEEE,* and Nicolas Gresset, *Senior Member, IEEE*

***Abstract*—In this letter we analyze the high Signal-to-Noise Ratio (SNR) regime performance of the multi-transmitter MISO Broadcast Channel (BC) with so-called distributed Channel State Information at the Transmitters (CSIT), i.e., where CSIT is made transmitter-dependent. Specifically, we show the unexpected result that it is possible to achieve the same Generalized Degrees-of-Freedom (GDoF) independently of whether a channel link estimate is present at one Transmitter (TX), at the other TX, or at both, and regardless of the channel pathloss topology, thanks to a novel transmission scheme. The proposed scheme involves as key ingredient a new precoding scheme coined Sliced Zero-Forcing which efficiently adapts to any distributed CSIT setting.**

***Index Terms*—Generalized Degrees-of-Freedom, Distributed MISO Broadcast Channel, Cooperative Systems, Imperfect CSI.**

I. Introduction

Future wireless networks are expected to be flexible and heterogeneous, with moving or flying base stations, such that the centralization of the signal processing as envisioned in the Cloud Radio Access Network (C-RAN) may not be possible. In such settings, designing an efficient approach to reduce the interference in a decentralized manner becomes necessary.

Several works have focused on such decentralized network deployments, where the TXs are possibly endowed with *different* CSI, from different perspectives. In [1], the erasure interference channel is studied when each TX has local delayed CSIT, showing that *who* has *which* information affects the capacity region. In [2], a hierarchical CSIT setting is assumed in the broadcast channel, where each TX knows the CSIT used by the TXs with worse CSIT than itself, and a robust hierarchical decentralized precoding scheme is presented which obtains significant gains with respect to the previously known schemes. Other works have focused on centralized settings with heterogeneous or alternating CSIT configurations [3].

In this paper, we address the BC with decentralized and non-hierarchical CSIT. This setting models a broad variety of practical deployments where the TXs are not connected with ideal backhaul. To get intuition and analytical results, we consider the high-SNR regime and we consider in particular the Degrees-of-Freedom (DoF) metric, which has already proved instrumental in several key discoveries [4]. Specifically, the DoF is defined as

$$\mathrm{DoF}^{\star} \triangleq \lim_{P\to\infty} \frac{\mathcal{C}(P)}{\log_2(P)}, \tag{1}$$

where $\mathcal{C}(P)$ denotes the sum capacity and P the transmission SNR. However, the DoF analysis exhibits some limitations, especially in scenarios with strong pathloss differences. To circumvent this problem, Etkin *et al.* introduced the GDoF in [5], where the pathloss topology is taken into account by modeling the relative channel strength of each link as a function of P.

Recently, the GDoF of the MISO BC with imperfect, yet centralized[1], CSIT has been derived in [6]. In that work, the estimate of the link between TX k and RX i has an error decreasing in $\bar{P}^{-\alpha_{i,k}}$, where $\alpha_{i,k}$ is the accuracy parameter and $\bar{P} \triangleq \sqrt{P}$. Extending that model to the so-called *Distributed CSIT* setting [7], where the TXs do not access the *same* CSIT estimates, the estimation error at a TX j for the link TX k-RX i is then modeled as $\bar{P}^{-\alpha^{(j)}_{i,k}}$, where the superscript (j) reflects the TX-dependent nature of the CSI quality.

For the 2-user MISO BC, it has been shown in [8] that, for the specific case where $\alpha^{(1)}_{i,k} \geq \alpha^{(2)}_{i,k}$ *for every* i,k, the sum-GDoF solely depends on the CSI quality at the better informed TX. However, in a more realistic scenario, there might not be a notion of "better informed TX" that is valid across all the links.

Our main contribution in this letter is to extend the result of [8] from the "better informed TX" case to *any* CSIT allocation and *any* pathloss topology of the 2-user case. Specifically, we show that it does not matter whether an estimate of a given channel coefficient is known at one TX, at the other TX, or at both. This is surprising as the result in [8] relies on the idea of a (passive) uninformed TX transmitting with fixed coefficients, and an (active) informed TX reducing interference, such that it was believed that this property would not extend to other CSIT configurations. Hence, this work reveals that cooperative settings are much more resilient against CSI mismatches between TXs than commonly thought in the community, what could impact the future design of feedback mechanisms.

Illustrating Example

We introduce in the following a simple example to convey the main intuition of the proposed transmission scheme —called in the following *Sliced Zero-Forcing* (S-ZF) and described in detail in Section IV—, and in order to illustrate

A. Bazco-Nogueras is with the Communication Systems Department, Eurecom, 06410 Biot, France, and also with the Mitsubishi Electric Research and Development Centre Europe, 35708 Rennes, France (bazco@eurecom.fr).

P. de Kerret and D. Gesbert are with the Communication Systems Department, Eurecom, 06410 Biot, France (dekerret@eurecom.fr; gesbert@eurecom.fr).

P. de Kerret and D. Gesbert also acknowledge the support of the ERC 670896.

N. Gresset is with the Mitsubishi Electric Research and Development Centre Europe, 35708 Rennes, France (n.gresset@fr.merce.mee.com).

[1] Centralized refers to a *logically* centralized setting where all the TXs have access to the same, possibly imperfect, CSI.

how the CSIT configuration —i.e., what CSI is known with which quality at which TX— affects the GDoF performance. As aforementioned, each TX has its own estimate of the channel between TX k and RX i with an error scaling as $\bar{P}^{-\alpha^{(j)}_{i,k}}$. In the following example, we consider the conventional DoF, i.e., that the channel pathloss does not scale as the SNR P. Moreover, the CSIT quality allocation is given for $\rho\in[0,1]$ by

$$\text{TX } 1\rightarrow\Big\{\alpha^{(1)}_{1,1}=0.25,\ \alpha^{(1)}_{1,2}=0.25,\ \alpha^{(1)}_{2,1}=0.5,\quad \alpha^{(1)}_{2,2}=0.5\ \Big\},$$
$$\text{TX } 2\rightarrow\Big\{\alpha^{(2)}_{1,1}=\rho,\quad \alpha^{(2)}_{1,2}=\rho,\quad \alpha^{(2)}_{2,1}=1-\rho,\ \alpha^{(2)}_{2,2}=1-\rho\Big\}.$$

Note that as ρ increases, TX 2 becomes better informed about links towards RX 1 and less about links towards RX 2, while TX 1 keeps a fixed estimation quality for each user. In Fig. 1 we show the DoF achieved by the proposed S-ZF scheme as a function of ρ. We compare this DoF with a centralized CSIT setting with CSIT quality $\alpha_{i,k}=\max(\alpha^{(1)}_{i,k},\alpha^{(2)}_{i,k})$, $\forall i,k\in\{1,2\}$, whose DoF is computed in [6], as well as with conventional Zero-Forcing (ZF) and time-sharing.

Surprisingly, it can be seen that the proposed scheme attains the same DoF as the centralized case, whereas conventional ZF, which is optimal in the centralized CSIT scenario, performs poorly when confronted with CSI discrepancies between the TXs. Note that the only case where the conventional ZF scheme is performing as S-ZF is when $\rho = 0.25$, due to the fact that both TXs have the same accuracy for the worse user RX 1.

Notations: The *exponential equality* is denoted as $\doteq$, i.e., $f(P)\doteq P^{\beta}$ is equivalent to $\lim_{P\to\infty}\frac{\log(f(P))}{\log(P)}=\beta$. The *exponential inequalities* $\dot{\leq}$ and $\dot{\geq}$ are defined in the same manner.

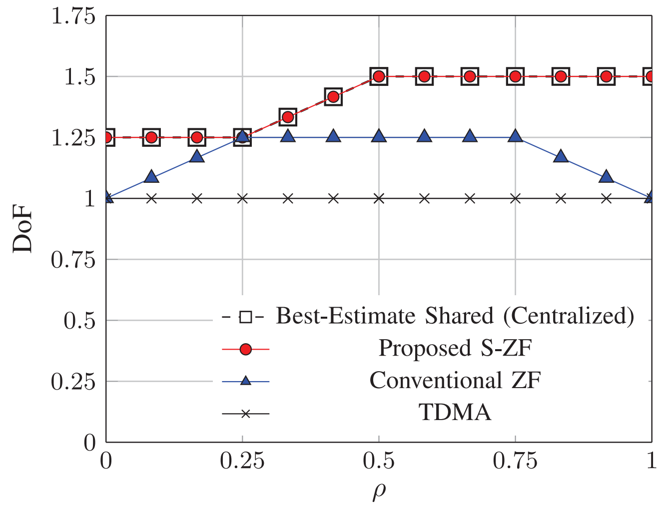

Fig. 1: DoF of the illustrative example setting as function of ρ.

II. System Model

A. Channel Model

We consider single-antenna nodes and we assume that the RXs have perfect CSI. The signal received at RX i is written as

$$y_i\triangleq\boldsymbol{h}^{\mathrm{H}}_i\mathbf{x}+z_i, \tag{2}$$

where $\boldsymbol{h}^{\mathrm{H}}_i\triangleq[h_{i,1},\ h_{i,2}]$ denotes the multi-TX channel to RX i and $h_{i,k}$ denotes the fading channel coefficient from TX k to RX i. The term $z_i\sim\mathcal{N}_{\mathbb{C}}(0,1)$ is the additive Gaussian noise at RX i, and $\mathbf{x}\in\mathbb{C}^{2\times 1}$ is the multi-TX transmitted signal which fulfills the average sum power constraint $\mathbb{E}[\|\mathbf{x}\|^2_2]=P$. $\mathbf{x}$ is generated from the i.i.d. information symbols $s_i\sim\mathcal{N}_{\mathbb{C}}(0,1)$ as

$$\mathbf{x}\triangleq[\boldsymbol{t}_{\mathrm{RX}1}\quad \boldsymbol{t}_{\mathrm{RX}2}]\begin{bmatrix} s_1\\ s_2\end{bmatrix}, \tag{3}$$

where $\boldsymbol{t}_{\mathrm{RX}i}$ denotes the precoder for RX i's symbol s_i. Following the GDoF model [5], the channel coefficient is defined as

$$h_{i,k}\triangleq\bar{P}^{\gamma_{i,k}-1}g_{i,k}, \tag{4}$$

where P is the nominal SNR parameter. The parameter $\gamma_{i,k}\in[0,1]$ is the relative channel strength exponent between TX k and RX i. Intuitively, $\gamma_{i,k}=0$ implies that any received signal lies below the noise floor, i.e., that the channel coefficient is negligible in terms of GDoF, while $\gamma_{i,k}=1$ means that the channel coefficient is not significantly attenuated. Finally, the normalized channel parameters $g_{i,k}$ are mutually independent and drawn from a generic (in the sense that any matrix formed by i.i.d. elements according to this distribution will be full rank) continuous distribution with density and independent of P.

B. Distributed CSIT Model

In the distributed CSIT setting, each TX receives a different estimate of the whole multi-user channel [7]. The channel coefficient estimation at TX j is written as $\hat{h}^{(j)}_{i,k}\triangleq\bar{P}^{\gamma_{i,k}-1}\hat{g}^{(j)}_{i,k}$ where

$$\hat{g}^{(j)}_{i,k}\triangleq g_{i,k}+\bar{P}^{-\alpha^{(j)}_{i,k}}\delta^{(j)}_{i,k}\qquad \forall j\in\{1,2\}, \tag{5}$$

with the estimation noise terms $\delta^{(j)}_{i,k}$ being drawn from a generic continuous distribution with density, independent of P, and independent from TX to TX, and with $\alpha^{(j)}_{i,k}\in[0,\gamma_{i,k}]$ being the *CSIT exponent* at TX j used to parametrize the accuracy of the CSIT. Note that $\alpha^{(j)}_{i,k}\in[0,\gamma_{i,k}]$ because –in terms of GDoF– an estimation with $\alpha^{(j)}_{i,k}=\gamma_{i,k}$ can be intuitively understood as being perfect [6]. Conversely, $\alpha^{(j)}_{i,k}=0$ is intuitively understood as being useless. The set of CSIT exponents of TX j is denoted by $\mathcal{S}^{(j)}_{\alpha}\triangleq\{\alpha^{(j)}_{i,k}|i,k\in\{1,2\}\}$. Importantly, it is assumed that the TXs have the knowledge of all the long-term statistics, i.e., $\gamma_{i,k}$, for all i, k, and both $\mathcal{S}^{(1)}_{\alpha}$, $\mathcal{S}^{(2)}_{\alpha}$.

III. Main Result

We can now state our main result.

Theorem 1. *In the 2-user MISO BC with distributed CSIT exponents $\mathcal{S}^{(1)}_{\alpha}$ and $\mathcal{S}^{(2)}_{\alpha}$, the optimal sum-GDoF denoted by $\mathrm{GDoF}^{\text{D-CSIT}}(\mathcal{S}^{(1)}_{\alpha},\mathcal{S}^{(2)}_{\alpha})$ is given by*

$$\mathrm{GDoF}^{\text{D-CSIT}}(\mathcal{S}^{(1)}_{\alpha},\,\mathcal{S}^{(2)}_{\alpha})\geq\mathrm{GDoF}^{\text{C-CSIT}}(\mathcal{S}^{\max}_{\alpha}), \tag{6}$$

where $\mathrm{GDoF}^{\text{C-CSIT}}(\mathcal{S}^{\max}_{\alpha})$ is the GDoF of the centralized CSIT scenario with a single shared estimate of exponents

$$\mathcal{S}^{\max}_{\alpha}=\Big\{\alpha_{i,k}=\max\Big(\alpha^{(1)}_{i,k},\ \alpha^{(2)}_{i,k}\Big)\Big|i,k\in\{1,2\}\Big\}. \tag{7}$$

The GDoF $\mathrm{GDoF}^{\text{C-CSIT}}(\mathcal{S}^{\max}_{\alpha})$ corresponds to the scenario where both TXs share the best estimate, which is hence a (logically) centralized setting whose GDoF has been derived in [6]. Therefore, the achieved GDoF is only limited by the *most accurate* estimate of each link, no matters which TX has it.

Remark 1. *For completeness, we recall the expression for the Centralized CSIT GDoF derived in [6]:*

$$\mathrm{GDoF}^{\text{C-CSIT}}(\mathcal{S}_{\alpha}^{\max}) \;=\; \min(D_1,D_2), \tag{8}$$

where

$$D_1 \triangleq \max(\gamma_{1,2},\gamma_{1,1}) + \big(\max(\gamma_{2,1}-\gamma_{1,1}+\alpha_1,\ \gamma_{2,2}-\gamma_{1,2}+\alpha_1)\big)^+$$
$$D_2 \triangleq \max(\gamma_{2,2},\gamma_{2,1}) + \big(\max(\gamma_{1,1}-\gamma_{2,1}+\alpha_2,\ \gamma_{1,2}-\gamma_{2,2}+\alpha_2)\big)^+$$

with $\gamma_{i,k}$ being defined in (4) *and α_1 and α_2 given by*

$$\alpha_1 \triangleq \min(\alpha_{1,1},\alpha_{1,2}), \qquad \alpha_2 \triangleq \min(\alpha_{2,1},\alpha_{2,2}). \tag{9}$$

Theorem 1 shows that it is possible to achieve the GDoF of the centralized setting where the best estimate of each link is perfectly shared. It is intuitively expected that the GDoF in this centralized setting is higher than in the decentralized setting considered, i.e., that the inequality in (6) is an equality. Indeed, having one TX with less accurate CSI and having inconsistencies between the TXs is not expected to improve the performance. Yet, rigorously proving the information theoretic upper-bound is tedious and requires additional technical assumptions. For the sake of clarity, and as our main contribution is the achievability, we have left the discussion of the converse for future works. Finally, note that this result can be straightforwardly extended to $M>1$ antennas at each node.

IV. Sliced Zero-Forcing Precoding

We present here the Sliced Zero-Forcing (S-ZF) scheme where, as usual in interference minimizing schemes, the designs of the beamformers towards RX 1 and RX 2 can be *decoupled* [2], [6]. Consequently, we present in the following the beamformer aimed at RX 1 that minimizes the interference to RX 2, whereas the other beamformer follows by a permutation of the indices. Furthermore, we omit hereinafter any index referring to the intended RX (RX 1). Based on the available CSIT, the S-ZF precoder is designed so that the ZF condition is fulfilled, i.e.,

$$\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{t}^{\mathrm{SZF}} = \boldsymbol{h}_2^{\mathrm{H}}\begin{bmatrix}\lambda t_1^{(1)}\\ \lambda t_2^{(2)}\end{bmatrix} = 0, \tag{10}$$

where $t_k^{(j)}$ denotes the precoder coefficient applied at TX k and the super-index (j) means that it is computed locally at TX j on the basis of the local estimate $\hat{\boldsymbol{h}}_2^{(j)}$. Note that one TX does not need to compute/know the coefficient applied at the other TX. λ is a normalization constant chosen to fulfill an average power constraint and is given by

$$\lambda \;\triangleq\; \bar{P}\sqrt{\mathbb{E}\left[\left\|\left[t_1^{(1)}\ t_2^{(2)}\right]^{\mathrm{T}}\right\|_{\mathrm{F}}^2\right]^{-1}} \tag{11}$$

for the constraint $\mathbb{E}[\|\boldsymbol{t}^{\mathrm{SZF}}\|_2^2] \;=\; P$. The normalization constant λ only depends on statistical information and is hence known at both TXs. Given that $\boldsymbol{h}_2^{\mathrm{H}}$ is composed of two coefficients, we can distinguish four different regimes depending on which TX has better knowledge of each link. Those four regimes are shown in Table I above. As illustrated in the table, the four regimes are reduced to three by symmetry. For each of these regimes, we will now describe the S-ZF precoding scheme.

TABLE I: CSIT Configuration Regimes

	$\alpha_{2,2}^{(1)} > \alpha_{2,2}^{(2)}$	$\alpha_{2,2}^{(1)} \le \alpha_{2,2}^{(2)}$
$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$	Most-informed TX (TX 1)	Locally Informed TXs
$\alpha_{2,1}^{(1)} \le \alpha_{2,1}^{(2)}$	Non-locally Informed TXs	Most-informed TX (TX 2)

a) Locally Informed TXs: In this case, each TX has the best estimate of its own channel towards RX 2 (i.e., the interfered user when considering the beamformer aimed at RX 1). The S-ZF precoding coefficient at TX j is then given by[2]

$$t_j^{(j)} = (-1)^j \big(\hat{h}_{2,j}^{(j)}\big)^{-1}, \qquad \forall j \in \{1,2\}. \tag{12}$$

b) Non-locally Informed TXs: In this case, each TX knows more accurately the channel coefficient from the other TX towards RX 2. Upon defining $\ell \triangleq j \pmod 2 + 1$, the precoding coefficient at TX j is

$$t_j^{(j)} = (-1)^j \hat{h}_{2,\ell}^{(j)}, \qquad \forall j \in \{1,2\}. \tag{13}$$

c) Most-informed TX: In this last case, there exists one TX that has the best estimate of both coefficients. The S-ZF precoding then matches the AP-ZF scheme presented in [8]: The TX with less precise CSIT (e.g., TX j) transmits with a constant precoder $t_j^{(j)} \triangleq (-1)^j \lambda$ while the *most-informed TX* precodes with

$$t_\ell^{(\ell)} = (-1)^\ell \big(\hat{h}_{2,\ell}^{(\ell)}\big)^{-1} \hat{h}_{2,j}^{(\ell)}. \tag{14}$$

V. Proof of Theorem 1

Superposition coding schemes have been shown to achieve optimal DoF/GDoF for multiple BC settings with imperfect CSIT [2], [6], [8], [9]. This commonly used transmission structure always fits the expression

$$\mathbf{x} = \boldsymbol{t}_{\mathrm{BC}} s_{\mathrm{BC}} + \boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{ZF}} s_1 + \boldsymbol{t}_{\mathrm{RX2}}^{\mathrm{ZF}} s_2 + \boldsymbol{t}_{\phi} s_{\phi}. \tag{15}$$

Depending on the pathloss topology (i.e., the value of $\gamma_{i,k}$) and the CSIT allocation (i.e., the value of $\alpha_{i,k}^{(j)}$), some of those four symbols may be suppressed. In the general scheme, those symbols form a three-layer structure where each layer has a different power scaling. Specifically:

1) *Low-power layer:* s_ϕ is a non-zero-forced symbol transmitted with power such that it is only received by the intended RX, if the pathloss topology allows for that.
2) *Zero-Forcing layer:* s_i, $i \in \{1,2\}$, is intended to RX i and *canceled* at the other RX using ZF-type precoding. A necessary condition for the optimality of the scheme is that the interference generated by those symbols lies below the noise floor. Therefore, they are transmitted with a power proportional to the accuracy of the CSIT.
3) *Full-power layer:* s_{BC} is a broadcast symbol transmitted with full power, intended to be decoded at both RXs.

In order to perfectly decode every intended symbol, RX i applies successive decoding [8] to first decode s_{BC}, then its intended symbol s_i and finally s_ϕ, if it is intended to RX i.

[2]The precoder can be improved at finite SNR using Regularized ZF. However, regularization is not necessary in terms of GDoF.

Interestingly, the precoders $\boldsymbol{t}_{\mathrm{BC}}$ and $\boldsymbol{t}_{\phi}$ depend only on the long-term statistical information ($\alpha_{i,k}^{(j)}$ and $\gamma_{i,k}$) and are hence not affected by the instantaneous CSI discrepancies between TXs. This implies that, in order to prove Theorem 1, i.e., that it is possible to achieve the same GDoF as in the centralized setting with $\alpha_{i,k}=\max(\alpha_{i,k}^{(1)},\alpha_{i,k}^{(2)})$, we only need to show that S-ZF achieves the same level of interference attenuation as ZF in the centralized setting of reference. This is shown for the interference at RX 2 by means of the following lemma, while the same result holds for RX 1 after permutation of indices.

Lemma 1. *The Sliced ZF achieves the same interference reduction scaling as the conventional ZF precoder computed from the best estimate of each channel coefficient, i.e.,*

$$\mathbb{E}\big[\big|\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}\big|^2\big]\doteq P^{\min(\gamma_{2,1},\gamma_{2,2})-\alpha_2^{opt}}, \tag{16}$$

when $\mathbb{E}\big[\|\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}\|_2^2\big]=P$, *where we have defined the short-hand notation* α_2^{opt} *as*

$$\alpha_2^{opt}\triangleq\min\left(\max_{j\in\{1,2\}}\alpha_{2,1}^{(j)},\max_{j\in\{1,2\}}\alpha_{2,2}^{(j)}\right). \tag{17}$$

Proof. We prove separately each of the regimes of Table I. We need to consider only the first two cases since the result for the *Most-Informed TX* configuration has been already proved in [8].

a) Non-locally Informed TXs: First, it holds from the precoder definition in (13) that $t_1^{(1)}\doteq\bar{P}^{\gamma_{2,2}-1}$ and $t_2^{(2)}\doteq\bar{P}^{\gamma_{2,1}-1}$. Therefore, from the definition of λ we obtain that

$$\lambda\doteq\bar{P}^{2-\max(\gamma_{2,1},\gamma_{2,2})}. \tag{18}$$

Then, the interference term satisfies that

$$\begin{aligned}
&\mathbb{E}\Big[\big|\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}\big|^2\Big]\\
&=\lambda^2\,\mathbb{E}\Big[\big|-h_{2,1}\hat{h}_{2,2}^{(1)}+h_{2,2}\hat{h}_{2,1}^{(2)}\big|^2\Big] &(19)\\
&\overset{(a)}{=}\lambda^2\,\mathbb{E}\Big[\big|-h_{2,1}h_{2,2}-h_{2,1}\bar{P}^{-\alpha_{2,2}^{(1)}+\gamma_{2,2}-1}\delta_{2,2}^{(1)}\\
&\qquad+h_{2,2}h_{2,1}+h_{2,2}\bar{P}^{-\alpha_{2,1}^{(2)}+\gamma_{2,1}-1}\delta_{2,1}^{(2)}\big|^2\Big] &(20)\\
&\overset{(b)}{=}\lambda^2P^{\gamma_{2,1}+\gamma_{2,2}-2}\mathbb{E}\Big[\big|\bar{P}^{-\alpha_{2,1}^{(2)}}g_{2,2}\delta_{2,1}^{(2)}-\bar{P}^{-\alpha_{2,2}^{(1)}}g_{2,1}\delta_{2,2}^{(1)}\big|^2\Big] &(21)
\end{aligned}$$

where (a) holds because, by definition, we can rewrite $\hat{h}_{2,k}^{(j)}$ as

$$\hat{h}_{2,k}^{(j)}\ \triangleq\ h_{2,k}+\bar{P}^{-\alpha_{2,k}^{(j)}+\gamma_{2,k}-1}\delta_{2,k}^{(j)}, \tag{22}$$

and (b) comes from $h_{2,k}\triangleq\bar{P}^{\gamma_{2,k}-1}g_{2,k}$. Focusing on the expectation term in (21), it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big|\bar{P}^{-\alpha_{2,1}^{(2)}}g_{2,2}\delta_{2,1}^{(2)}-\bar{P}^{-\alpha_{2,2}^{(1)}}g_{2,1}\delta_{2,2}^{(1)}\big|^2\Big]\\
&=\bar{P}^{-\alpha_2^{\mathrm{opt}}}\mathbb{E}\Big[\big|\bar{P}^{\alpha_2^{\mathrm{opt}}-\alpha_{2,1}^{(2)}}g_{2,2}\delta_{2,1}^{(2)}-\bar{P}^{\alpha_2^{\mathrm{opt}}-\alpha_{2,2}^{(1)}}g_{2,1}\delta_{2,2}^{(1)}\big|^2\Big] &(23)\\
&\doteq\bar{P}^{-\alpha_2^{\mathrm{opt}}}, &(24)
\end{aligned}$$

where (24) follows from using definition $\alpha_2^{\mathrm{opt}}=\min(\alpha_{2,2}^{(1)},\alpha_{2,1}^{(2)})$ and because $g_{i,k}$ and $\delta_{i,k}^{(j)}$ are independent of P. Including (24) in (21) and substituting λ with (18) yields

$$\begin{aligned}
\mathbb{E}\Big[\big|\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}\big|^2\Big]\ &\doteq\ \lambda^2P^{\gamma_{2,1}+\gamma_{2,2}-2-\alpha_2^{\mathrm{opt}}} &(25)\\
&\doteq\ P^{\min(\gamma_{2,1},\gamma_{2,2})-\alpha_2^{\mathrm{opt}}}. &(26)
\end{aligned}$$

b) Locally Informed TXs: For this scenario, it holds from (12) that $t_1^{(1)}\doteq\bar{P}^{1-\gamma_{2,1}}$ and $t_2^{(2)}\doteq\bar{P}^{1-\gamma_{2,2}}$. Consequently,

$$\lambda\doteq\bar{P}^{\min(\gamma_{2,1},\gamma_{2,2})}. \tag{27}$$

Substituting $\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}$ by its expression in (12) gives

$$\begin{aligned}
&\mathbb{E}\Big[\big|\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{t}_{\mathrm{RX1}}^{\mathrm{SZF}}\big|^2\Big]\\
&=\lambda^2\mathbb{E}\Big[\big|-h_{2,1}(\hat{h}_{2,1}^{(1)})^{-1}+h_{2,2}(\hat{h}_{2,2}^{(2)})^{-1}\big|^2\Big] &(28)\\
&=\lambda^2\mathbb{E}\Big[\big|-\hat{h}_{2,1}^{(1)}\big(\hat{h}_{2,1}^{(1)}\big)^{-1}+\bar{P}^{-\alpha_{2,1}^{(1)}+\gamma_{2,1}-1}\delta_{2,1}^{(1)}\big(\hat{h}_{2,1}^{(1)}\big)^{-1}\\
&\qquad+\hat{h}_{2,2}^{(2)}\big(\hat{h}_{2,2}^{(2)}\big)^{-1}-\bar{P}^{-\alpha_{2,2}^{(2)}+\gamma_{2,2}-1}\delta_{2,2}^{(2)}\big(\hat{h}_{2,2}^{(2)}\big)^{-1}\big|^2\Big] &(29)\\
&=\lambda^2\mathbb{E}\Big[\big|\bar{P}^{-\alpha_{2,1}^{(1)}}\delta_{2,1}^{(1)}\big(\hat{g}_{2,1}^{(1)}\big)^{-1}-\bar{P}^{-\alpha_{2,2}^{(2)}}\delta_{2,2}^{(2)}\big(\hat{g}_{2,2}^{(2)}\big)^{-1}\big|^2\Big] &(30)\\
&\doteq\lambda^2P^{-\alpha_2^{\mathrm{opt}}} &(31)\\
&\doteq P^{\min(\gamma_{2,1},\gamma_{2,2})-\alpha_2^{\mathrm{opt}}}, &(32)
\end{aligned}$$

where (29) follows from $h_{2,k}=\hat{h}_{2,k}^{(j)}-\bar{P}^{-\alpha_{2,k}^{(j)}+\gamma_{2,k}-1}\delta_{2,k}^{(j)}$, (30) comes from substituting $\hat{h}_{2,k}^{(j)}=\bar{P}^{\gamma_{2,k}-1}\hat{g}_{2,k}^{(j)}$, (31) follows from applying the same argument as in (23)-(24), and (32) is obtained after substituting λ with its value in (27). This concludes the proof of Lemma 1, and hence Theorem 1 is proven. □

VI. Conclusion

We have shown that, remarkably, having different CSIT at each TX does not decrease the GDoF of the 2-user MISO Broadcast Channel, *in any channel topology*. Key to this surprisingly good performance is the adaptation of the role of each TX as a function of the multi-TX multi-user CSIT configuration. An interesting future extension is to analyze how this GDoF-based distributed scheme behaves at finite SNR, where different effects that are hidden in the GDoF analysis may appear. Moreover, extending the result to larger values of K is very challenging, and from preliminary results, it is expected that the surprisingly good performance does not extend well to more users.

References

[1] A. Vahid and R. Calderbank, "Two-User Erasure Interference Channels With Local Delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4910–4923, Sept 2016.

[2] P. de Kerret, A. Bazco, and D. Gesbert, "Enforcing Coordination in Network MIMO with Unequal CSIT," in *50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 39–43.

[3] S. Lin and I.-H. Wang, "Gaussian Broadcast Channels With Intermittent Connectivity and Hybrid State Information at the Transmitter," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6362–6383, Sept 2018.

[4] V. R. Cadambe and S. A. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[5] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[6] A. G. Davoodi, B. Yuan, and S. A. Jafar, "GDoF Region of the MISO BC: Bridging the Gap between Finite Precision and Perfect CSIT," *IEEE Trans. Inf. Theory*, pp. 1–1, 2018.

[7] P. de Kerret and D. Gesbert, "Degrees of Freedom of the Network MIMO Channel with Distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[8] A. Bazco, P. de Kerret, D. Gesbert, and N. Gresset, "Generalized Degrees-of-Freedom of the 2-user case MISO Broadcast Channel with Distributed CSIT," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1092–1096.

[9] E. Piovano and B. Clerckx, "Optimal DoF Region of the K-User MISO BC With Partial CSIT," *IEEE Comm. Letters*, vol. 21, no. 11, pp. 2368–2371, Nov 2017.