Precoding Methods for the MISO Broadcast Channel with Delayed CSIT

Xinping Yi, *Student Member, IEEE*, and David Gesbert, *Fellow, IEEE*

***Abstract*—Recent information theoretic results suggest that precoding on the multi-user downlink MIMO channel with delayed channel state information at the transmitter (CSIT) could lead to data rates much beyond the ones obtained without any CSIT, even in extreme situations when the delayed channel feedback is made totally obsolete by a feedback delay exceeding the channel coherence time. This surprising result is based on the ideas of interference repetition and alignment which allow the receivers to reconstruct information symbols which canceling out the interference completely, making it an optimal scheme in the infinite SNR regime. In this paper, we formulate a similar problem, yet at finite SNR. We propose a first construction for the precoder which matches the previous results at infinite SNR yet reaches a useful trade-off between interference alignment and signal enhancement at finite SNR, allowing for significant performance improvement in practical settings. We present two general precoding methods with arbitrary number of users by means of virtual MMSE and mutual information optimization, achieving good compromise between signal enhancement and interference alignment. Simulation results show substantial improvement due to the compromise between those two aspects.**

***Index Terms*—Multi-user MIMO, Delayed Feedback, Precoding, Interference Alignment**

I. Introduction

Multi-user MIMO systems (or their information-theoretic counterparts "MIMO broadcast channels"), have recently attracted considerable attention from the research community and industry alike. Success is due to their ability to enhance the wireless spectrum efficiency by a factor equal to the number N of antennas installed at the base station, with little restriction imposed on the richness of the multipath channel, the presence or absence of a strong line of sight channel component, and the fact it can easily accommodate single antenna mobile devices. On the downlink of such systems, the ability to beamform (i.e. linearly precode) multiple data streams simultaneously to several users (up to N) comes nevertheless at a price in terms of requiring the base station transmitter to be informed of the channel coefficients of all served users [1]. In frequency division duplex scenarios (the bulk of available wireless standards today), this implies establishing a feedback link from the mobiles to the base station which can carry CSI related information, in quantized format. A common limitation of such an approach, perceived by many to be a key hurdle toward a more widespread use of MU-MIMO methods in real-life networks, lies in the fact that the feedback information typically arrives back to the transmitter with a delay which may cause a severe degradation when comparing the obtained feedback CSIT with the actual current channel state information. Pushed to the extreme, and considering a feedback delay with the same order of magnitude as the coherence period of the channel, the available CSIT feedback becomes completely obsolete (i.e., uncorrelated with the current true channel information) and, seemingly non-exploitable in view of designing the precoding coefficients.

Recently, this commonly accepted viewpoint was challenged by an interesting information-theoretic work which established the usefulness of stale channel state information in designing precoders achieving significantly better rate performance than what is obtained without any CSIT [2]. The premise in [2] is a time-slotted MIMO broadcast channel with a common transmitter serving multiple users and having a delayed version of the correct CSIT, where the delay causes the CSIT to be fully uncorrelated with the current channel vector information. In this situation, it is shown that the transmitter can still exploit the stale channel information: The transmitter tries to reproduce the interference generated to the users in the previous time slots, a strategy we refer to in this paper as interference repetition, while at the same time making sure the forwarded interference occupies a subspace of limited dimension, compatible with its cancelation at the user's side, a method commonly referred to as interference alignment [3, 4]. Building on such ideas, [2] constructs a transmission protocol referred to as the MAT protocol which was shown to achieve the maximum Degrees-of-Freedom (DoF) for the delayed CSIT broadcast MIMO channel. Precoding on delayed CSIT MIMO channels has recently attracted more interesting work, dealing with DoF analysis on extended channels, like the X channel and interference channels [5–7], but also performance analysis including effects of feedback [8] and training [9]. The DoF is a popular information theoretic performance metric indicating the number of interference-free simultaneous data streams which can be communicated over this delayed CSIT channel at infinite SNR, also coinciding with the notion of pre-log factor in the channel capacity expression. In the example of the two-antenna transmitter, two-user channel, the maximum DoF was shown in [2] to be $\frac{4}{3}$, less than the value of 2 which would be obtained with perfect CSIT, but strictly larger than the single DoF obtained in the absence of any CSIT. This means that completely obsolete channel feedback is actually useful.

Manuscript received July 10, 2012; revised November 23, 2012; accepted January 11, 2013. The associate editor coordinating the review of this paper and approving it for publication was Prof. S. Wei.

We acknowledge the partial support of project SHARING, under the CELTIC European funding program. Part of this work has been presented at ICASSP 2012 in Kyoto, Japan.

X. Yi and D. Gesbert are with the Mobile Communications Dept., EURECOM, 06560 Sophia Antipolis, France (email: {xinping.yi, david.gesbert}@eurecom.fr).

Although fascinating from a conceptual point of view, these results are intrinsically focussed on the asymptotic SNR behavior, leaving aside in particular the question of how shall precoding be done practically using stale CSIT at finite SNR. This paper precisely tackles this question. In what follows we obtain the following key results:

- We show finite SNR precoding using delayed CSIT can be achieved using a combination of interference repetition, alignment together with a signal enhancement strategy.
- We propose a precoder construction generalizing the ideas of [2], namely Generalized MAT (GMAT), where a compromise between interference alignment and orthogonality within the desired signal channel matrix is striken, and then generalize it to the scenario with arbitrary number of users.
- The precoder coefficients are interpreted as beamforming vector coefficients in equivalent interference channel scenario, which can be optimized in a number of ways, including using an MMSE metric, and mutual information metric. To the best of our knowledge, the optimization of a finite SNR precoding scheme based on delayed feedback has not yet been addressed.

Numerical evaluation reveals a substantial performance benefit in terms of data rate in the low to moderate SNR region, but coinciding with the performance of [2] when the SNR grows to infinity. Note that a preliminary set of results were reported recently in [10] for the 2-user case, while this paper provides a generalization to the case of arbitrary number of users.

The rest of the paper is organized as follows. In Section II, the channel model of interest is described and the proposed GMAT protocol is detailed first in the 2-user case then is generalized to the K-user case. Section III focuses on the precoder optimization methods based on MMSE and mutual information criteria. Discussion on the multiplexing gain and an interesting interpretation from an equivalent MIMO interference channel is given in Section IV. Numerical examples showing the advantages of the new methods are discussed in Section V. Finally, Section VI concludes the paper.

Notation: Matrices and vectors are represented as uppercase and lowercase letters, and transpose and conjugate transpose of a matrix are denoted as $(\cdot)^T$ and $(\cdot)^H$, respectively. Further, Tr$(\cdot)$, $\|\cdot\|$ and $\|\cdot\|_F$ represent respectively the trace of a matrix, the norm of a vector and a Frobenius norm of a matrix. We reserve $[\mathbf{A}]_{m,n}$ to denote the element at the m-th row and n-th column of matrix $\mathbf{A}$, and $|\mathcal{S}|$ to the cardinality of the set $\mathcal{S}$. Finally, an order-k message denoted by $u_{\mathcal{S}}$ ($|\mathcal{S}| = k$) refers to a linear combination of k distinct symbol vectors intended to k different users in set $\mathcal{S}$.

II. System Model

Consider a K-user MU-MIMO downlink system with a transmitter equipped with K antennas and K single-antenna users. A time slotted transmission protocol in the downlink direction is considered, where the multi-antenna channel vector from the transmitter to i-th user, in the j-th time slot, is denoted by $\mathbf{h}_i^T(j) = [h_{i1}(j) \ \cdots \ h_{iK}(j)]$, with h_{ik} being the channel coefficient from k-th transmit antenna to i-th user. We denote by $\mathbf{x}(j)$ the $K \times 1$ vector of signals sent from the array of K transmit antennas. As in [2], the point made in this paper is that delayed feedback can be of use to the transmitter including the extreme situation where a feedback delay of one unit of time creates a full decorrelation with the current downlink channel. For this reason, we base ourselves on the framework of so-called delayed CSIT [2, 5–9] by which at time j, it is assumed that user-i has perfect knowledge of $\{\mathbf{h}_i(t)\}_{t=1}^{j}$ and of the delayed CSIT of other users $\{\mathbf{h}_k(t)\}_{t=1}^{j-1}, k \neq i$, while the transmitter is informed perfectly $\{\mathbf{h}_i(t)\}_{t=1}^{j-1}, \forall i$. The accessibility of such delayed CSI at other terminals has been justified in previous work such as [8] by the following model. The users feed back their CSI to the transmitter with delays, then the transmitter broadcasts all the CSI to all the users such that all users have access to other users' delayed CSI[1]. Nevertheless, there exists another more efficient scenario for sharing the delayed CSI across users. It is based on the notion of "broadcast uplink feedback", i.e., the terminals broadcast their CSI which is then captured by any overhearing device, which includes both the transmitter and the other terminals. Furthermore, we make no assumption about any correlation between the channel vectors across multiple time slots (could be fully uncorrelated), making it is impossible for the transmitter to use classical MU-MIMO precoding to serve the users, since the transmitter possesses some CSIT possibly independent from the actual channel.

Recently, Maddah-Ali and Tse [2] proposed an algorithm under such a delayed CSIT setting obtaining DoF strictly beyond that obtained without any CSIT, even in extreme situations when the delayed CSIT is made totally obsolete. The key ideas lie in interference repetition and alignment. Doing so, the users are able to reconstruct the signals overheard in previous slots to allow them to cancel out the interference completely. In general, for the K-user case, a K-phase transmission protocol can achieve the maximum DoF $\frac{K}{1+\frac{1}{2}+\cdots+\frac{1}{K}}$. Although such rates are inferior to the ones obtained under the full CSIT setting (cf. K symbols/channel use for K antenna system), they are substantially higher than what was previously reported for the no CSIT case (cf. 1 symbol/channel use regardless of K).

Although optimal in terms of the DoF, at infinite SNR, we point out that the above approach can be substantially improved at finite SNR. The key reason is that, at finite SNR, a good scheme will not attempt to use all DoF to eliminate the interference but will try to strike a compromise between interference canceling and enhancing the detectability of the desired signal in the presence of noise. Taking into account this property of basic receivers leads us to revisit the design of the protocol and in particular the design of the precoding coefficients as functions of the knowledge of past channel vectors under the name of GMAT.

First, we proceed by reviewing the proposed protocol in the 2-user case, highlighting the connections with the original

[1]Clearly, the broadcast phase may introduce some additional delays. The transmitter then exploits the largest delayed version of the CSI, which is common with the one received by the users.

MAT algorithm. We then generalize the protocol to respectively the 3 and K-user cases. In the next section, we then turn to the problem of the optimization of the precoders.

A. GMAT for the 2-user Case

The transmission of GMAT in the first two time slots is identical to the MAT algorithm, with[2]

$$\mathbf{x}(1) = \mathbf{s}_A, \quad \mathbf{x}(2) = \mathbf{s}_B \tag{1}$$

where $\mathbf{x}(t)$ $(t = 1, 2)$ is the 2×1 signal vector sent from the transmitter at time slot t, $\mathbf{s}_A$ and $\mathbf{s}_B$ are 2×1 symbol vectors intended to user A and B, respectively, satisfying $\mathbb{E}\{\mathbf{s}_i\mathbf{s}_i^H\} = \mathbf{I}$. In the third time slot, the transmitter now sends

$$\mathbf{x}(3) = \begin{bmatrix} u_{AB} \\ 0 \end{bmatrix} \tag{2}$$

where u_{AB} corresponds to an order-2 message (i.e., a combination of two individual user messages) in the following form

$$u_{AB} = \mathbf{w}_1^T\mathbf{s}_A + \mathbf{w}_2^T\mathbf{s}_B \tag{3}$$

where $\mathbf{w}_1$ and $\mathbf{w}_2$ are precoding vectors satisfying the power constraint $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq 2$ and can be a function of $\mathbf{h}_i(1)$ and $\mathbf{h}_i(2)$ $(i = A, B)$ according to the delayed CSIT model. Note that this power constraint balances the transmit power used over three time slots. The signal vector received over the three time slots at user A can be rewritten as

$$\bar{\mathbf{y}}_A = \sqrt{\frac{P}{2}}\bar{\mathbf{H}}_{A1}\mathbf{s}_A + \sqrt{\frac{P}{2}}\bar{\mathbf{H}}_{A2}\mathbf{s}_B + \mathbf{n}_A, \tag{4}$$

where $\bar{\mathbf{y}}_A = [y_A(1)\ y_A(2)\ y_A(3)]^T$ is the concatenated received signal vector at user A in overall three time slots, $\mathbf{n}_A = [n_A(1)\ n_A(2)\ n_A(3)]^T$ is the Gaussian noise vector with zero-mean and unit-variance, P is the total transmit power in each time slot, and the effective signal and interference channel matrices are

$$\bar{\mathbf{H}}_{A1} = \begin{bmatrix} \mathbf{h}_A^T(1) \\ \mathbf{0} \\ h_{A1}(3)\mathbf{w}_1^T \end{bmatrix}, \ \bar{\mathbf{H}}_{A2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_A^T(2) \\ h_{A1}(3)\mathbf{w}_2^T \end{bmatrix}, \tag{5}$$

and, by analogy, for user B, we get the interference and signal matrices:

$$\bar{\mathbf{H}}_{B1} = \begin{bmatrix} \mathbf{h}_B^T(1) \\ \mathbf{0} \\ h_{B1}(3)\mathbf{w}_1^T \end{bmatrix}, \ \bar{\mathbf{H}}_{B2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_B^T(2) \\ h_{B1}(3)\mathbf{w}_2^T \end{bmatrix}. \tag{6}$$

1) A Particular Case (MAT Algorithm): We point out that the MAT algorithm [2] can be derived as a particular case of the above method, with $\mathbf{w}_1$ and $\mathbf{w}_2$ specified as

$$\mathbf{w}_1 = \mathbf{h}_B(1), \qquad \mathbf{w}_2 = \mathbf{h}_A(2). \tag{7}$$

The key idea behind the original MAT solution in (7) is that the interference $\mathbf{s}_B$ seen by user A arrives with an effective channel matrix $\bar{\mathbf{H}}_{A2}$ which is of rank one, making it possible for user A to combine the three received signals in order to retrieve $\mathbf{s}_A$ while canceling out $\mathbf{s}_B$ completely. This process is referred to as alignment of interference signal $\mathbf{s}_B$, as it mimics the approach taken in interference channels in e.g., [3]. A similar property is exploited in (6) at user B as well by making $\bar{\mathbf{H}}_{B1}$ be rank 1.

2) Interpretation of GMAT v.s. MAT: A drawback of the original MAT solution in (7) is to optimize the precoders from the point of view of interference alone while the signal matrices $\bar{\mathbf{H}}_{A1}$ and $\bar{\mathbf{H}}_{B2}$ are ignored. Although this approach is optimal from an information theoretic (i.e., DoF) point of view, it is suboptimal at finite SNR.

In contrast, here, the role of introduced beamformer $\mathbf{w}_1$ is to strike a balance between aligning the interference channel of $\mathbf{s}_A$ at user B and enhancing the detectability of $\mathbf{s}_A$ at user A. In algebraic terms, this can be interpreted as having a compromise between obtaining a rank deficient $\bar{\mathbf{H}}_{B1}$ and an orthogonal matrix for $\bar{\mathbf{H}}_{A1}$. When it comes to $\mathbf{w}_2$, the compromise is between obtaining a rank deficient $\bar{\mathbf{H}}_{A2}$ and an orthogonal matrix for $\bar{\mathbf{H}}_{B2}$. How to achieve this trade-off in practice is addressed in Section III.

It is also important to note there might be alternative fashions of constructing finite SNR precoders based on delayed CSIT. For instance, an interesting question is: Can delayed feedback be exploited already in the second time slot with gains on the finite SNR performance? The intuitive answer to this question is yes. However, the use of precoders in the last time slot only generates a strong symmetry and handling of the users, which in turn allows for closed-form and insightful solutions. This symmetric property is also maintained in the MAT algorithm.

B. GMAT for the 3-user Case

Similarly to the MAT algorithm, the proposed GMAT sends 18 symbols in a total of three phases, which include 6, 3, and 2 time slots, respectively, giving an effective rate of $\frac{18}{11}$ symbols/slot. In the first phase, 6 symbol vectors carrying all 18 symbols are sent in 6 consecutive time slots in a way identical to the initial MAT

$$\mathbf{x}(1) = \mathbf{s}_A^1,\ \mathbf{x}(2) = \mathbf{s}_B^1,\ \mathbf{x}(3) = \mathbf{s}_C^1, \tag{8}$$

$$\mathbf{x}(4) = \mathbf{s}_A^2,\ \mathbf{x}(5) = \mathbf{s}_B^2,\ \mathbf{x}(6) = \mathbf{s}_C^2 \tag{9}$$

where $\mathbf{s}_i^1$ and $\mathbf{s}_i^2$ $(i = A, B, C)$ are 3×1 symbol vectors (referred to as the order-1 messages) intended to user-i. As in the 2-user case, we do not introduce channel dependent precoding in the first phase in order to preserve symmetry across the users. Instead, feedback based precoding is introduced in the second phase.

Phase-2 involves 3 time slots, in each of which two order-2 messages (defined as a combination of two order-1 messages) are sent from the first two transmit antennas:

$$\mathbf{x}(7) = \begin{bmatrix} u_{AB}^1 \\ u_{AB}^2 \\ 0 \end{bmatrix}, \ \mathbf{x}(8) = \begin{bmatrix} u_{AC}^1 \\ u_{AC}^2 \\ 0 \end{bmatrix}, \ \mathbf{x}(9) = \begin{bmatrix} u_{BC}^1 \\ u_{BC}^2 \\ 0 \end{bmatrix} \tag{10}$$

where the order-2 messages are constructed by

$$u_{AB}^1 = {\mathbf{w}_{12}^1}^T\mathbf{s}_A^1 + {\mathbf{w}_{21}^1}^T\mathbf{s}_B^1,\ u_{AB}^2 = {\mathbf{w}_{12}^2}^T\mathbf{s}_A^2 + {\mathbf{w}_{21}^2}^T\mathbf{s}_B^2 \tag{11}$$

$$u_{AC}^1 = {\mathbf{w}_{13}^1}^T\mathbf{s}_A^1 + {\mathbf{w}_{31}^1}^T\mathbf{s}_C^1,\ u_{AC}^2 = {\mathbf{w}_{13}^2}^T\mathbf{s}_A^2 + {\mathbf{w}_{31}^2}^T\mathbf{s}_C^2 \tag{12}$$

[2]For the notational simplicity, we use the index exchangeably, where both A and 1 correspond to the first user/component, and so forth.

$$u_{BC}^1 = \mathbf{w}_{23}^{1\,T}\mathbf{s}_B^1 + \mathbf{w}_{32}^{1\,T}\mathbf{s}_C^1,\ u_{BC}^2 = \mathbf{w}_{23}^{2\,T}\mathbf{s}_B^2 + \mathbf{w}_{32}^{2\,T}\mathbf{s}_C^2 \tag{13}$$

where u_{ij}^1 and u_{ij}^2 $(i \neq j)$ are two realizations of the order-2 message dedicated to both user-i and user-j, and $\mathbf{w}_{ji}^1 \in \mathbb{C}^{3\times 1}, \mathbf{w}_{ji}^2 \in \mathbb{C}^{3\times 1}, 1 \le i,j \le 3$ can be arbitrary vector functions of $\mathbf{h}_i(t), i = A, B, C,\ t = 1, \cdots, 6$. The responsibility of Phase-2 is to provide independent equations with regard to $\mathbf{s}_i^1$ (or $\mathbf{s}_i^2$) by utilizing the overheard interferences in the previous phase.

Finally, in the last phase, channel dependent precoding is not introduced as this allows to obtain decoupled optimization problems for each of the $\mathbf{w}_{ji}^l$ as will be made in Section III. In this phase, two order-3 messages are sent at the first transmit antenna within two consecutive time slots, i.e.,

$$\mathbf{x}(10) = \begin{bmatrix} u_{ABC}^1 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{x}(11) = \begin{bmatrix} u_{ABC}^2 \\ 0 \\ 0 \end{bmatrix} \tag{14}$$

where u_{ABC}^l $(l = 1, 2)$ is the order-3 message which is identical to the original MAT algorithm

$$u_{ABC}^l = a_1^l(h_{C1}(7)u_{AB}^1 + h_{C2}(7)u_{AB}^2) + a_2^l(h_{B1}(8)u_{AC}^1 + h_{B2}(8)u_{AC}^2) + a_3^l(h_{A1}(9)u_{BC}^1 + h_{A2}(9)u_{BC}^2)$$

where $\{a_j^l\}$ $(j = 1, 2, 3)$ are chosen in a way similar to the original MAT, i.e., arbitrary yet linearly independent sets of coefficients and known by both transmitter and receivers.

Without loss of generality, we treat user A as the target user, and the compact received signal model in matrix format over the 11 time slots can be given by

$$\bar{\mathbf{y}}_A = \sqrt{\frac{P}{3}}\sum_{l=1}^{2}\bar{\mathbf{H}}_{A1}^l\mathbf{s}_A^l + \sqrt{\frac{P}{3}}\sum_{l=1}^{2}\bar{\mathbf{H}}_{A2}^l\mathbf{s}_B^l + \sqrt{\frac{P}{3}}\sum_{l=1}^{2}\bar{\mathbf{H}}_{A3}^l\mathbf{s}_C^l + \mathbf{n}_A \tag{15}$$

where the equivalent channel matrices can be formulated as

$$\bar{\mathbf{H}}_{A1}^l = \begin{bmatrix} \tilde{\mathbf{H}}_{A1}^l \\ \mathbf{D}_A^l(2)\mathbf{W}_1^l(2) \\ \mathbf{D}_A^l(3)\mathbf{W}_1^l(3) \end{bmatrix}, \bar{\mathbf{H}}_{A2}^l = \begin{bmatrix} \tilde{\mathbf{H}}_{A2}^l \\ \mathbf{D}_A^l(2)\mathbf{W}_2^l(2) \\ \mathbf{D}_A^l(3)\mathbf{W}_2^l(3) \end{bmatrix}, \tag{16}$$

$$\bar{\mathbf{H}}_{A3}^l = \begin{bmatrix} \tilde{\mathbf{H}}_{A3}^l \\ \mathbf{D}_A^l(2)\mathbf{W}_3^l(2) \\ \mathbf{D}_A^l(3)\mathbf{W}_3^l(3) \end{bmatrix} \in \mathbb{C}^{11\times 3} \tag{17}$$

in which

$$\tilde{\mathbf{H}}_{Aj}^l = \begin{bmatrix} \mathbf{0}_{m_1^l\times 3} \\ \mathbf{h}_A(m_1^l + 1) \\ \mathbf{0}_{n_1^l\times 3} \end{bmatrix} \in \mathbb{C}^{6\times 3} \tag{18}$$

with $m_1^l = (3(l-1) + j - 1)$, $n_1^l = 6 - 3(l-1) - j$, $\mathbf{D}_A^l(2) = \mathrm{diag}\{h_{Al}(7), h_{Al}(8), h_{Al}(9)\}$, $\mathbf{D}_A^l(3) = \mathrm{diag}\{h_{A1}(10), h_{A1}(11)\}$, and

$$\mathbf{W}^l(2) = \left[\underbrace{\begin{bmatrix} \mathbf{w}_{12}^{l\,T} \\ \mathbf{w}_{13}^{l\,T} \\ \mathbf{0}_{1\times 3} \end{bmatrix}}_{\mathbf{W}_1^l(2)} \underbrace{\begin{bmatrix} \mathbf{w}_{21}^{l\,T} \\ \mathbf{0}_{1\times 3} \\ \mathbf{w}_{23}^{l\,T} \end{bmatrix}}_{\mathbf{W}_2^l(2)} \underbrace{\begin{bmatrix} \mathbf{0}_{1\times 3} \\ \mathbf{w}_{31}^{l\,T} \\ \mathbf{w}_{32}^{l\,T} \end{bmatrix}}_{\mathbf{W}_3^l(2)}\right] \in \mathbb{C}^{3\times 9} \tag{19}$$

is the global precoding matrix (referred to hereafter as the order-2 message generation matrix) in which $\mathbf{W}_j^l(2)$ is corresponding to user-j.

Given the order-2 message generation matrix $\mathbf{W}_j^l(2) \in \mathbb{C}^{3\times 3}$, the precoding matrix for the third phase (referred to as order-3 message generation matrix) can be recursively obtained by

$$\mathbf{W}_j^l(3) = \mathbf{C}^l(2)\mathbf{\Lambda}^l(2)\mathbf{W}_j^l(2) \in \mathbb{C}^{2\times 3}, \quad j = 1, 2, 3 \tag{20}$$

where $\mathbf{\Lambda}^l(2) = \mathrm{diag}\{h_{Cl}(7), h_{Bl}(8), h_{Al}(9)\}$ is set identically to MAT for simplicity, and

$$\mathbf{C}^l(2) = \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \end{pmatrix} \tag{21}$$

is a constant matrix known by both transmitter and receivers.

1) A Particular Case (MAT Algorithm): The original MAT algorithm can be deduced from the proposed method by selecting

$$\mathbf{W}^1(2) = \begin{bmatrix} \mathbf{h}_B^T(1) & \mathbf{h}_A^T(2) & \mathbf{0}_{1\times 3} \\ \mathbf{h}_C^T(1) & \mathbf{0}_{1\times 3} & \mathbf{h}_A^T(3) \\ \mathbf{0}_{1\times 3} & \mathbf{h}_C^T(2) & \mathbf{h}_B^T(3) \end{bmatrix} \tag{22}$$

and $\mathbf{W}^2(2)$ can be obtained in an analogous way.

Similarly to the 2-user case, interferences carrying unintended symbols $\mathbf{s}_B^l$ and $\mathbf{s}_C^l$ are aligned perfectly at user A, and hence matrices $\bar{\mathbf{H}}_{A2}^l$ and $\bar{\mathbf{H}}_{A3}^l$ are rank deficient with total rank of 5, making the useful symbol $\mathbf{s}_A^l$ retrievable from the left 6-dimensional interference-free subspace. For the proposed GMAT algorithm, we seek to balance signal orthogonality (conditioning of $\bar{\mathbf{H}}_{A1}^l$) and perfect interference alignment by a careful design of $\mathbf{W}^l(2)$.

C. GMAT for the General K-user Case

In K-user case, the maximum achievable DoF is $d = \frac{K}{\sum_{k=1}^{K}\frac{1}{k}}$ [2]. Let $d = \frac{K^2 L}{T}$, where T is an integer representing the overall required time slots and L is the number of repeated transmission to guarantee T to be an integer. Without loss of generality, we assume $L = (K-1)!$. The total T times slots can be divided into K phases. In Phase-1, there consist of LK time slots. As the same way to the MAT algorithm, an order-1 message $\mathbf{x}(t)$ is sent in t-th time slot, i.e.,

$$\mathbf{x}(t) = \mathbf{s}_i^l,\ l = 1, \cdots, L \tag{23}$$

satisfying $t = L(l-1) + i$, where $\mathbf{s}_i^l$ is the $K \times 1$ symbol vector intended to user-i.

From Phase-2 to Phase-K, the transmission of GMAT is similar to MAT algorithm. Phase-k $(2 \le k \le K)$ requires $T_k \triangleq \frac{LK}{k}$ time slots, with each time slot transmitting $K-k+1$ order-k messages from $K-k+1$ transmit antennas, i.e.,

$$\mathbf{x}(t) = \begin{bmatrix} u_{\mathcal{S}_k}^1 & \cdots & u_{\mathcal{S}_k}^{K-k+1} & 0 & \cdots & 0 \end{bmatrix}^T \tag{24}$$

where $u_{\mathcal{S}_k}^j$ $(1 \le j \le K-k+1)$ is the j-th message realization of the order-k message that can be generated by

$$\mathbf{u}_{\mathcal{S}_k}^l = \mathbf{W}^l(k)\mathbf{s}^l \tag{25}$$

where $\mathbf{u}_{\mathcal{S}_k}^l$ is the $Q_k \times 1$ vector $(Q_k \triangleq \binom{K}{k})$ with each element being order-k message that can be interpreted as the combination of any k symbol vectors from $\{\mathbf{s}_i^l\}$ $(1 \le l \le L)$; $\mathcal{S}_k$ is the set of dedicated users and satisfies $|\mathcal{S}_k| = k$; $\mathbf{s}^l = [\mathbf{s}_1^{l\ T} \ \cdots \ \mathbf{s}_K^{l\ T}]^T \in \mathcal{C}^{K^2\times 1}$ is the concatenated symbol vector, and $\mathbf{W}^l(k) \in \mathcal{C}^{Q_k\times K^2}$ is the order-k message generation matrix, whose definition is as follows:

Definition 1 (Order-k Message Generation Matrix). *The order-k message generation matrix $\mathbf{W}^l(k) = \left[\mathbf{W}_1^l(k) \ \cdots \ \mathbf{W}_K^l(k)\right]$ $(2 \le k \le K)$ is a $Q_k \times K^2$ matrix which satisfies:*

1) *it contains k nonzero and $K-k$ zero blocks in each row, where each block is $1\times K$ row vector;*
2) *the positions of nonzero blocks of any two rows are not identical; and*
3) *it contains all possibilities of k nonzero positions out of total K positions in each row.*

We point out that the order-k message is desired by those k users whose symbols are contained, and acts as an interference that will be overheard by other $K-k$ users.

Based on the above definition, the signal model of K-user GMAT protocol can be extended as

$$\bar{\mathbf{y}}_i = \sqrt{\frac{P}{K}}\sum_{l=1}^{L}\bar{\mathbf{H}}_{ii}^l\mathbf{s}_i^l + \sqrt{\frac{P}{K}}\sum_{l=1}^{L}\sum_{j=1,j\neq i}^{K}\bar{\mathbf{H}}_{ij}^l\mathbf{s}_j^l + \mathbf{n}_i \tag{26}$$

where

$$\bar{\mathbf{H}}_{ij}^l = \begin{bmatrix}\tilde{\mathbf{H}}_{ij}^l(1)\\ \vdots \\ \tilde{\mathbf{H}}_{ij}^l(k)\\ \vdots \\ \tilde{\mathbf{H}}_{ij}^l(K)\end{bmatrix} \in \mathbb{C}^{T\times K} \tag{27}$$

with $T = \sum_{i=1}^{K} T_k$, is defined as follows:

- The first submatrix corresponds to the effective channel matrix in Phase-1, which can be given by

$$\tilde{\mathbf{H}}_{ij}^l(1) = \begin{bmatrix}\mathbf{0}_{m_1^l\times K}\\ \mathbf{h}_i(t)\\ \mathbf{0}_{n_1^l\times K}\end{bmatrix} \in \mathbb{C}^{T_1\times K} \tag{28}$$

  where $j = 1,\ldots,K$, $l = 1,\ldots,L$, $m_1^l = (K(l-1) + j - 1)$, $n_1^l = KL - K(l-1) - j$, and $t = m_1^l + 1$;
- The k-th submatrix $(2 \le k \le K-1)$ which corresponds to Phase-k can be formulated as

$$\tilde{\mathbf{H}}_{ij}^l(k) = \begin{bmatrix}\mathbf{0}_{m_k^l\times K}\\ \mathbf{D}_i^l(k)\mathbf{W}_j^l(k)\\ \mathbf{0}_{n_k^l\times K}\end{bmatrix} \in \mathbb{C}^{T_k\times K} \tag{29}$$

  where $m_k^l = \left(\lceil\frac{l\cdot l_k}{L}\rceil - 1\right)Q_k$, $n_k^l = T_k - \lceil\frac{l\cdot l_k}{L}\rceil Q_k$ with $l_k = \frac{T_k}{Q_k}$, and $\mathbf{D}_i^l(k) = \mathrm{diag}\{h_{is}(t)\} \in \mathbb{C}^{Q_k\times Q_k}$ corresponds to the present channel over which the order-k message is sent in Phase-k with $s = ((l\cdot l_k) \mod L) \mod k$ and t being the index of time slot. In general, $\mathbf{W}_j^l(k)$ $(k \ge 2)$ is the order-k message generation matrix specified to user-j, which is recursively defined according to

$$\mathbf{W}_j^l(k+1) = \mathbf{C}^l(k)\mathbf{\Lambda}^l(k)\mathbf{W}_j^l(k) \tag{30}$$

  where $\mathbf{C}^l(k) \in \mathbb{C}^{Q_{k+1}\times Q_k}$ is a constant matrix known by transmitter and all users, satisfying: (1) each row contains $k+1$ nonzero elements, and (2) the positions of nonzero elements of any two rows are different one another; and $\mathbf{\Lambda}^l(k) \in \mathbb{C}^{Q_k\times Q_k}$ is a diagonal matrix whose elements are chosen to be functions of the channel coefficients in Phase-k, so that the interference overheard can be aligned within a limited dimensional subspace. For simplicity, we place emphasis on $\mathbf{W}_j^l(k)$, letting $\mathbf{\Lambda}^l(k)$ be predetermined as the channel coefficients in Phase-k, as did in the original MAT algorithm.
- The last submatrix is corresponding to the last phase, i.e.,

$$\tilde{\mathbf{H}}_{ij}^l(K) = \mathbf{D}_i^l(K)\mathbf{W}_j^l(K) \in \mathbb{C}^{T_K\times K} \tag{31}$$

  where $\mathbf{W}_j^l(K)$ is defined similarly to (30), in which $\mathbf{C}^l(K-1) \in \mathbb{C}^{T_K\times Q_{K-1}}$ is a full rank constant matrix without zero elements, and $\mathbf{D}_i^l(K) = \mathrm{diag}\{h_{i1}(t)\} \in \mathbb{C}^{T_K\times T_K}$ $(t \in [T - T_K + 1, T])$ contains channel coefficients during Phase-K.

For further illustration, we take the 4-user case for example to show its order-2 message generation matrix, i.e.,

$$\mathbf{W}^l(2) = \begin{bmatrix}\mathbf{w}_{12}^{l\ T} & \mathbf{w}_{21}^{l\ T} & \mathbf{0} & \mathbf{0}\\ \mathbf{w}_{13}^{l\ T} & \mathbf{0} & \mathbf{w}_{31}^{l\ T} & \mathbf{0}\\ \mathbf{w}_{14}^{l\ T} & \mathbf{0} & \mathbf{0} & \mathbf{w}_{41}^{l\ T}\\ \mathbf{0} & \mathbf{w}_{23}^{l\ T} & \mathbf{w}_{32}^{l\ T} & \mathbf{0}\\ \mathbf{0} & \mathbf{w}_{24}^{l\ T} & \mathbf{0} & \mathbf{w}_{42}^{l\ T}\\ \mathbf{0} & \mathbf{0} & \mathbf{w}_{34}^{l\ T} & \mathbf{w}_{43}^{l\ T}\end{bmatrix} \tag{32}$$

where $\mathbf{w}_{ji}^l \in \mathbb{C}^{K\times 1}$ is the beamforming vector aiming at the compromise between user-i and user-j. This formulation collapses to (11)-(13) for the 3-user case and to (3) for the 2-user case.

1) A particular Case (MAT Algorithm): Particularly for the 4-user case, the original MAT algorithm is a specialized GMAT algorithm by setting order-2 message generation matrix as

$$\mathbf{W}^1(2) = \begin{bmatrix}\mathbf{h}_B^T(1) & \mathbf{h}_A^T(2) & \mathbf{0} & \mathbf{0}\\ \mathbf{h}_C^T(1) & \mathbf{0} & \mathbf{h}_A^T(3) & \mathbf{0}\\ \mathbf{h}_D^T(1) & \mathbf{0} & \mathbf{0} & \mathbf{h}_A^T(4)\\ \mathbf{0} & \mathbf{h}_C^T(2) & \mathbf{h}_B^T(3) & \mathbf{0}\\ \mathbf{0} & \mathbf{h}_D^T(2) & \mathbf{0} & \mathbf{h}_B^T(4)\\ \mathbf{0} & \mathbf{0} & \mathbf{h}_D^T(3) & \mathbf{h}_C^T(4)\end{bmatrix} \tag{33}$$

for $l = 1$ and similarly for other l. For example, for user A, the interference channels $\bar{\mathbf{H}}_{Aj}^l$ $(j \neq 1)$ are perfectly aligned, leaving $K = 4$ interference-free dimensions for desired signal, and therefore making the intended symbols retrievable at user A. Similarly for other users, all symbols can be recovered. Hence, 96 symbols are delivered within 50 time slots, yielding the sum DoF of $\frac{48}{25}$.

It is worth noting that the higher level messages can be delivered by the combination of lower level messages. For example, from Phase-k to K, the messages delivered to the receivers aim at completely decoding the order-k message. To avoid too many parameters being optimized which requires huge complexity, we will focus merely on the design of the order-2 message generation matrices $\{\mathbf{W}_j^l(2)\}$.

III. GMAT Optimization Design

There exist several distinct avenues for computing the delayed CSIT based precoders (i.e., matrices $\{\mathbf{W}_j^l(2)\}$). Two of them are briefly described in the following subsections. The first is based on the optimization of a virtual MMSE metric, yielding an iterative solution, while the second one considers the maximization of an approximation of the mutual information, yielding suboptimal yet closed-form solutions. Note that none of these approaches have anything in common with finite SNR interference alignment methods with non-delayed CSIT, such as, e.g., [11–13], since the nature of our problem is fully conditioned by the delayed CSIT scenario. In all cases below, the design of the precoders obeys two principles: (i) the precoders are functions of delayed channel feedback, and (ii) the design is based on the exploitation of alignment-orthogonality trade-off that is underpinned by eq-(4), (15), and (26).

A. Virtual MMSE Metric

In the following, we describe an approach based on a virtual MMSE metric (referred to later as "GMAT-MMSE") for the 2-user case, and subsequently generalize it to the K-user case.

1) Special $K = 2$ Case: A difficulty in the precoder design lies in the fact that, since the transmitter does not know $\mathbf{h}_i(3)$ at Slot-3, the optimization of the precoder in (5) and (6) cannot involve such information even though the channel realizations on the third time slot clearly affect the overall rate performance. The question therefore is whether a meaningful criterion can be formulated for the optimization of the precoder that IS NOT a function of the non-delayed CSIT. The answer is positive. In what follows, we first offer an intuitive treatment of this problem. Then, a rigorous mathematical argument is offered for it in the next subsection based on mutual information bounds.

In order to derive an optimization model that does no longer depend on the non-delayed CSIT coefficients $h_{A1}(3)$ and $h_{B1}(3)$, we observe that the key trade-off between alignment of interference and desired signal orthogonality is in fact independent from the realizations of $h_{A1}(3)$ and $h_{B1}(3)$, since such coefficients impact on the amplitudes of the precoders but not on their directions. Hence, it is natural to formulate a virtual signal model that skips dependency on the unknown CSIT yet preserves the above mentioned trade-off:

$$\mathbf{y}_i = \sqrt{\frac{P}{2}}\mathbf{H}_{i1}\mathbf{s}_A + \sqrt{\frac{P}{2}}\mathbf{H}_{i2}\mathbf{s}_B + \mathbf{n}_i, i = A, B \tag{34}$$

where the virtual channel matrices are now modified from (5) and (6) according to:

$$\mathbf{H}_{i1} = \begin{bmatrix} \mathbf{h}_i^T(1) \\ \mathbf{0} \\ \mathbf{w}_1^T \end{bmatrix},\ \mathbf{H}_{i2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_i^T(2) \\ \mathbf{w}_2^T \end{bmatrix},\ i = A, B. \tag{35}$$

Given $\mathbf{w}_1$ and $\mathbf{w}_2$, the optimum RX MMSE filter at user-i over this virtual channel is given by

$$\mathbf{V}_i = \sqrt{\rho}\left(\rho\mathbf{H}_{i1}\mathbf{H}_{i1}^H + \rho\mathbf{H}_{i2}\mathbf{H}_{i2}^H + \mathbf{I}\right)^{-1}\mathbf{H}_{ii} \tag{36}$$

where $\rho = \frac{P}{K}$ (here $K = 2$), and the corresponding minimal MSE is

$$J_i(\mathbf{w}_1, \mathbf{w}_2) = \text{Tr}\left(\mathbf{I} - \rho\mathbf{H}_{ii}^H(\rho\mathbf{H}_{i1}\mathbf{H}_{i1}^H + \rho\mathbf{H}_{i2}\mathbf{H}_{i2}^H + \mathbf{I})^{-1}\mathbf{H}_{ii}\right).$$

Note that we exchangeably use A and 1 to represent the first user, and so forth.

Hence, the optimal $\mathbf{w}_1$, $\mathbf{w}_2$ can be obtained from the following optimization problem, i.e.,

$$\min_{\mathbf{w}_1, \mathbf{w}_2 : \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq 2} J = J_A(\mathbf{w}_1, \mathbf{w}_2) + J_B(\mathbf{w}_1, \mathbf{w}_2).$$

In practice, the gradient based approaches can be used to perform optimization although the convexity of the problem is not guaranteed [14, 15].

2) General K-user Case: In Phase-k, the transmitter does not know $\mathbf{h}_i(t)$ in Slot-t, where $t = \sum_{l=1}^{k-1} T_l + 1, \cdots, \sum_{l=1}^{k} T_l$. Similarly to the 2-user case, the virtual received signal can be generalized as $(i = 1, \cdots, K)$

$$\mathbf{y}_i = \sqrt{\frac{P}{K}}\sum_{l=1}^{L}\mathbf{H}_{ii}^l\mathbf{s}_i^l + \sqrt{\frac{P}{K}}\sum_{l=1}^{L}\sum_{j=1, j\neq i}^{K}\mathbf{H}_{ij}^l\mathbf{s}_j^l + \mathbf{n}_i, \tag{44}$$

where

$$\mathbf{H}_{ij}^l = \left[\tilde{\mathbf{H}}_{ij}^{l\ T}\ \cdots\ \mathbf{0}_{K\times m_k^l}\ \mathbf{W}_j^{l\ T}(k)\ \mathbf{0}_{K\times n_k^l}\ \cdots\ \mathbf{W}_j^{l\ T}(K)\right]^T$$

whose elements are defined in Section II.

Similarly, given $\mathbf{W}_j^l(2)$, the optimum MMSE filter for $\mathbf{s}_i^l$ at user-i becomes

$$\mathbf{V}_i^l = \sqrt{\rho}\left(\rho\sum_{l=1}^{L}\sum_{j=1}^{K}\mathbf{H}_{ij}^l\mathbf{H}_{ij}^{l\ H} + \mathbf{I}\right)^{-1}\mathbf{H}_{ii}^l \tag{45}$$

where $\rho = \frac{P}{K}$ is the normalized transmit power, and the corresponding minimal MSE is

$$\begin{aligned} &J_i^l(\mathbf{W}_j^l(2), j = 1, \cdots, K) \\ &= \text{Tr}\left(\mathbf{I} - \rho\mathbf{H}_{ii}^{l\ H}\left(\rho\sum_{l=1}^{L}\sum_{j=1}^{K}\mathbf{H}_{ij}^l\mathbf{H}_{ij}^{l\ H} + \mathbf{I}\right)^{-1}\mathbf{H}_{ii}^l\right). \end{aligned}$$

The optimal solutions of $\{\mathbf{W}_j^l(2), j = 1, \cdots, K\}$ in the sense of virtual MMSE at receiver side are now given by:

$$\min_{\mathbf{W}_j^l(2), j=1,\cdots,K} J = \sum_{l=1}^{L}\sum_{i=1}^{K} J_i^l(\mathbf{W}_j^l(2)) \tag{46}$$

$$s.t. \quad \sum_{l=1}^{L}\sum_{j=1}^{K}\|\mathbf{W}_j^l(2)\|_F^2 \leq KT_2. \tag{47}$$

As the above optimization does not lend itself easily to a closed-form solution, we propose an iterative procedure, based on the gradient descent of the cost function J, where $\mathbf{W}_j^l(2)$ is iteratively updated according to

$$\hat{\mathbf{W}}_j^l(2)[n+1] = \hat{\mathbf{W}}_j^l(2)[n] - \beta\frac{\partial(J)}{\partial\mathbf{W}_j^l(2)} \tag{48}$$

where n is the iteration index and β is a small step size. The partial derivation is given in the Appendix. Nevertheless, to circumvent non-convexity issue, we explore an alternative optimization method below.

$$I(\mathbf{s}_A;\mathbf{y}_A) = \log\det\left(\mathbf{I} + \left(\mathbf{I} + \rho\bar{\mathbf{H}}_{A2}\bar{\mathbf{H}}_{A2}^H\right)^{-1}\rho\bar{\mathbf{H}}_{A1}\bar{\mathbf{H}}_{A1}^H\right) \tag{37}$$

$$= \log\det\left(\mathbf{I} + \rho\begin{bmatrix}1 & 0\\ 0 & \frac{1+\|\mathbf{h}_A^H(2)\|^2}{\Delta_1(\mathbf{w}_2)}\end{bmatrix}\begin{bmatrix}\|\mathbf{h}_A^H(1)\|^2 & h_{A1}^*(3)\mathbf{w}_1^H\mathbf{h}_A(1)\\ h_{A1}(3)\mathbf{h}_A^H(1)\mathbf{w}_1 & |h_{A1}(3)|^2\|\mathbf{w}_1\|^2\end{bmatrix}\right) \tag{38}$$

$$= \log\left(1 + \rho\|\mathbf{h}_A(1)\|^2 + \frac{\Theta_1(\mathbf{w}_1)}{\Delta_1(\mathbf{w}_2)}\right) \tag{39}$$

B. Mutual Information Metric

Here, we propose an approach based on maximizing an approximation of the mutual information, yielding a convenient closed-form solution for $\{\mathbf{W}_j^l(2)\}$. In the following, we will start with the 2-user case to gain insight, and then generalize it to the K-user case.

1) Special 2-user Case: Recall that

$$\mathbf{y}_A = \sqrt{\rho}\bar{\mathbf{H}}_{A1}\mathbf{s}_A + \sqrt{\rho}\bar{\mathbf{H}}_{A2}\mathbf{s}_B + \mathbf{n}_A \tag{49}$$

where $\rho = \frac{P}{K}$ (here $K = 2$), $\mathbf{w}_1$ and $\mathbf{w}_2$ are functions of $\mathbf{h}_i(j), i = A, B, j = 1, 2$ and satisfy power constraint $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq 2$. Consequently, the exact mutual information of user A can be calculated by the equations on the top of this page. Note that the second line is easily obtained by permuting rows 2 and 3 in $\bar{\mathbf{H}}_{A1}$ and $\bar{\mathbf{H}}_{A2}$, and the third line by the characteristic polynomial equality [16]

$$\det(\mathbf{I} + \rho\mathbf{M}) = 1 + \rho\ \mathrm{Tr}(\mathbf{M}) + \rho^2\det(\mathbf{M}), \tag{50}$$

where $\mathbf{M}$ is a 2×2 Hermitian matrix. By analogy, the mutual information of user B can be given by

$$I(\mathbf{s}_B;\mathbf{y}_B) = \log\left(1 + \rho\|\mathbf{h}_B(2)\|^2 + \frac{\Theta_2(\mathbf{w}_2)}{\Delta_2(\mathbf{w}_1)}\right) \tag{51}$$

where $\Theta_i(\mathbf{w}_j)$ and $\Delta_i(\mathbf{w}_j)$ are defined on the bottom of this page, with

$$C_1 = 1 + \rho\|\mathbf{h}_A(1)\|^2,\quad C_2 = 1 + \rho\|\mathbf{h}_A(2)\|^2 \tag{52}$$

$$C_3 = 1 + \rho\|\mathbf{h}_B(1)\|^2,\quad C_4 = 1 + \rho\|\mathbf{h}_B(2)\|^2. \tag{53}$$

By imposing a symmetric constraint for power allocation between $\mathbf{w}_1$ and $\mathbf{w}_2$, e.g., $\|\mathbf{w}_1\|^2 = \|\mathbf{w}_2\|^2 = 1$ for simplicity, the sum mutual information can be deduced to

$$I(\mathbf{s}_A;\mathbf{y}_A) + I(\mathbf{s}_B;\mathbf{y}_B) = \log\left(1 + \frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\right) + \log\left(1 + \frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right) + \log(C_1C_4) \tag{54}$$

where

$$\mathbf{R}_1 = C_2\left(\mathbf{I} + \rho\mathbf{h}_A^\perp(1)\mathbf{h}_A^{\perp H}(1)\right) \tag{55}$$

$$\mathbf{R}_2 = C_1\left(\gamma_1\mathbf{I} + \rho\mathbf{h}_A^\perp(2)\mathbf{h}_A^{\perp H}(2)\right) \tag{56}$$

$$\mathbf{Q}_1 = C_4\left(\gamma_2\mathbf{I} + \rho\mathbf{h}_B^\perp(1)\mathbf{h}_B^{\perp H}(1)\right) \tag{57}$$

$$\mathbf{Q}_2 = C_3\left(\mathbf{I} + \rho\mathbf{h}_B^\perp(2)\mathbf{h}_B^{\perp H}(2)\right) \tag{58}$$

with

$$\gamma_1 = \frac{1 + \rho\|\mathbf{h}_A(2)\|^2}{\rho|h_{A1}(3)|^2} + 1,\quad \gamma_2 = \frac{1 + \rho\|\mathbf{h}_B(1)\|^2}{\rho|h_{B1}(3)|^2} + 1 \tag{59}$$

and $\mathbf{h}_i^\perp(j) \in \mathbb{C}^{2\times 1}$ is the orthogonal component of $\mathbf{h}_i(j)$ ($i = A, B, j = 1, 2$) satisfying

$$\mathbf{h}_i(j)\mathbf{h}_i^H(j) + \mathbf{h}_i^\perp(j)\mathbf{h}_i^{\perp H}(j) = \|\mathbf{h}_i(j)\|^2\mathbf{I}. \tag{60}$$

The maximization of the mutual information in closed-form is very challenging in the arbitrary SNR regime. In this paper, we investigate the possibility of studying the high-enough SNR regime (i.e., high enough to produce tractable solutions) while maintaining an SNR regime that is finite-enough so as to preserve the key notion of alignment-orthogonality trade-off exposed earlier in Section II. Thus, we approximate the mutual information as

$$I(\mathbf{s}_A;\mathbf{y}_A) + I(\mathbf{s}_B;\mathbf{y}_B) \approx \log\left(\frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right) + \log(C_1C_4) \tag{61}$$

which can be optimized by separately maximizing two Rayleigh Quotients, i.e.,

$$\max_{\|\mathbf{w}_1\|^2=1} \log\left(\frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right) \tag{62}$$

$$= \max_{\|\mathbf{w}_1\|^2=1} \frac{\mathbf{w}_1^H\left(\mathbf{I} + \rho\mathbf{h}_A^\perp(1)\mathbf{h}_A^{\perp H}(1)\right)\mathbf{w}_1}{\mathbf{w}_1^H\left(\gamma_2\mathbf{I} + \rho\mathbf{h}_B^\perp(1)\mathbf{h}_B^{\perp H}(1)\right)\mathbf{w}_1}, \tag{63}$$

$$\max_{\|\mathbf{w}_2\|^2=1} \log\left(\frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\right) \tag{64}$$

$$= \max_{\|\mathbf{w}_2\|^2=1} \frac{\mathbf{w}_2^H\left(\mathbf{I} + \rho\mathbf{h}_B^\perp(2)\mathbf{h}_B^{\perp H}(2)\right)\mathbf{w}_2}{\mathbf{w}_2^H\left(\gamma_1\mathbf{I} + \rho\mathbf{h}_A^\perp(2)\mathbf{h}_A^{\perp H}(2)\right)\mathbf{w}_2}. \tag{65}$$

Hence, we can obtain the optimal solutions $\mathbf{w}_1^{opt}$ and $\mathbf{w}_2^{opt}$, which are given by the dominant generalized eigenvectors of the pairs $(\mathbf{R}_1, \mathbf{Q}_1)$ and $(\mathbf{Q}_2, \mathbf{R}_2)$, respectively.

Nevertheless, these solutions can be found to be dependent on parameters γ_1 and γ_2 which in turn depend on the unknown

$$\Theta_1(\mathbf{w}_1) = \rho C_2|h_{A1}(3)|^2(C_1\|\mathbf{w}_1\|^2 - \rho\mathbf{w}_1^H\mathbf{h}_A(1)\mathbf{h}_A(1)^H\mathbf{w}_1) \tag{40}$$

$$\Delta_1(\mathbf{w}_2) = C_2(1 + \rho|h_{A1}(3)|^2\|\mathbf{w}_2\|^2) - \rho^2|h_{A1}(3)|^2\mathbf{w}_2^H\mathbf{h}_A(2)\mathbf{h}_A(2)^H\mathbf{w}_2 \tag{41}$$

$$\Theta_2(\mathbf{w}_2) = \rho C_3|h_{B1}(3)|^2(C_4\|\mathbf{w}_2\|^2 - \rho\mathbf{w}_2^H\mathbf{h}_B(2)\mathbf{h}_B(2)^H\mathbf{w}_2) \tag{42}$$

$$\Delta_2(\mathbf{w}_1) = C_3(1 + \rho|h_{B1}(3)|^2\|\mathbf{w}_1\|^2) - \rho^2|h_{B1}(3)|^2\mathbf{w}_1^H\mathbf{h}_B(1)\mathbf{h}_B(1)^H\mathbf{w}_1 \tag{43}$$

channel coefficients in Slot-3. Fortunately, it is possible to average their impact and obtain a lower bound on mutual information that no longer depends on such coefficients. Aware of the convexity of mutual information approximation in eq-(61) with regard to $|h_{A1}(3)|^2$ and $|h_{B1}(3)|^2$, we further lower bound it by applying Jensen's inequality, i.e.,

$$\mathbb{E}_{|h_{B1}(3)|^2}\log\left(\frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right) \geq \log\left(\frac{\mathbf{w}_1^H\left(\mathbf{I}+\rho\mathbf{h}_A^{\perp}(1)\mathbf{h}_A^{\perp H}(1)\right)\mathbf{w}_1}{\mathbf{w}_1^H\left(\bar{\gamma}_2\mathbf{I}+\rho\mathbf{h}_B^{\perp}(1)\mathbf{h}_B^{\perp H}(1)\right)\mathbf{w}_1}\right)+\log\left(\frac{C_2}{C_4}\right)$$

$$\mathbb{E}_{|h_{A1}(3)|^2}\log\left(\frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\right) \geq \log\left(\frac{\mathbf{w}_2^H\left(\mathbf{I}+\rho\mathbf{h}_B^{\perp}(2)\mathbf{h}_B^{\perp H}(2)\right)\mathbf{w}_2}{\mathbf{w}_2^H\left(\bar{\gamma}_1\mathbf{I}+\rho\mathbf{h}_A^{\perp}(2)\mathbf{h}_A^{\perp H}(2)\right)\mathbf{w}_2}\right)+\log\left(\frac{C_3}{C_1}\right)$$

with

$$\bar{\gamma}_1 = 1+\|\mathbf{h}_A(2)\|^2+1/\rho, \quad \bar{\gamma}_2 = 1+\|\mathbf{h}_B(1)\|^2+1/\rho$$

being independent of the unknown channel coefficients $h_{A1}(3)$ and $h_{B1}(3)$ where $\mathbb{E}[|h_{A1}(3)|^2] = \mathbb{E}[|h_{B1}(3)|^2] = 1$, such that the original optimization problem can be alternatively done by

$$\max_{\|\mathbf{w}_i\|^2=1} \frac{\mathbf{w}_i^H\left(\mathbf{I}+\rho\mathbf{h}_i^{\perp}(i)\mathbf{h}_i^{\perp H}(i)\right)\mathbf{w}_i}{\mathbf{w}_i^H\left(\bar{\gamma}_{\bar{i}}\mathbf{I}+\rho\mathbf{h}_{\bar{i}}^{\perp}(i)\mathbf{h}_{\bar{i}}^{\perp H}(i)\right)\mathbf{w}_i} \tag{66}$$

where $i, \bar{i} = 1, 2$ and $i \neq \bar{i}$. Note that we exchangeably use A and 1 to represent the first user, and so forth.

Interestingly, the above objective function can be interpreted as dual SINR in a 2-user interference channel. Define

$$\mathrm{DSINR}_i = \frac{\mathbf{w}_i^H\left(\mathbf{I}+\rho\mathbf{h}_i^{\perp}(i)\mathbf{h}_i^{\perp H}(i)\right)\mathbf{w}_i}{\mathbf{w}_i^H\left(\bar{\gamma}_{\bar{i}}\mathbf{I}+\rho\mathbf{h}_{\bar{i}}^{\perp}(i)\mathbf{h}_{\bar{i}}^{\perp H}(i)\right)\mathbf{w}_i} \tag{67}$$

which is referred to as a *regularized* SINR in a dual 2-user interference channel with a desired channel $\mathbf{h}_i^{\perp}$ and interference channel $\mathbf{h}_{\bar{i}}^{\perp}$, where $\mathbf{w}_i$ is interpreted as a receive filter. Thus, the optimization problem in eq-(66) can be equivalently done by maximizing the regularized SINR in the dual MISO interference channels. Note that the regularization lies in not only the interference channels but also the desired channels. This solution is referred to later as "GMAT-DSINR".

2) General K-user Case: Recall that the definition of DSINR in eq-(67) for the 2-user case, where $\mathbf{w}_i$ is determined by the orthogonal channels of itself and also its peers. According to the structure of $\mathbf{W}^l(2)$ for the K-user case, we can follow this approach and design each nonzero submatrices $\mathbf{w}_{ji}^l$ separately. For each $\mathbf{w}_{ji}^l$, the dual interference channel can be constructed by the orthogonal channels. Thus, the regularized dual SINR can be formulated as (e.g., $l = 1$)

$$\mathrm{DSINR}_{ji}^l = \frac{\mathbf{w}_{ji}^{l\ H}\left(\mathbf{I}+\rho\sum_{k\neq i}\mathbf{h}_k^{\perp}(j)\mathbf{h}_k^{\perp H}(j)\right)\mathbf{w}_{ji}^l}{\mathbf{w}_{ji}^{l\ H}\left(\bar{\gamma}_{ji}\mathbf{I}+\rho\mathbf{h}_i^{\perp}(j)\mathbf{h}_i^{\perp H}(j)\right)\mathbf{w}_{ji}^l}, \tag{68}$$

where $j \neq i$, $\mathbf{w}_{ji}^l \in \mathbb{C}^{K\times 1}$ is the i-th (when $i < j$) or $(i-1)$-th (when $i > j$) nonzero block of $\mathbf{W}_j^l(2)$, $\mathbf{h}_i^{\perp}(j) \in \mathbb{C}^{K\times K}$ is one representation of the null space of $\mathbf{h}_i(j)$ with the same norm[3], and

$$\bar{\gamma}_{ji} = \|\mathbf{w}_{ji}^l\|^2 + \|\mathbf{h}_i(j)\|^2 + 1/\rho \tag{69}$$

where $\|\mathbf{w}_{ji}^l\|^2$ can be chosen to satisfy the overall transmit power constraint. Note that the numerator and denominator of DSINR_{ji}^l represent the requirements of signal orthogonality and interference alignment, respectively. While the latter aims at aligning $\mathbf{w}_{ji}^l$ as close as possible to the interference component $\mathbf{h}_i^{\perp}(j)$, the former tries to make $\mathbf{w}_{ji}^l$ as orthogonal as possible to the spanned subspace by all the channel vectors $\mathbf{h}_k^{\perp}(j)$ except $k = i$.

Accordingly, the optimal $\mathbf{w}_{ji}^l$ can be obtained by separately optimizing ($\forall\ i, j,\ i \neq j$)

$$\max_{\mathbf{w}_{ji}^l} \quad \mathrm{DSINR}_{ji}^l, \ j \neq i \tag{70}$$

$$s.t. \quad \sum_{l=1}^{L}\sum_{j=1}^{K}\|\mathbf{W}_j^l(2)\|_F^2 \leq KT_2 \tag{71}$$

where the corresponding solution can be simply obtained by generalized eigenvalue decomposition. By maximizing the dual SINR, $\mathbf{w}_{ji}^l$ is preferred to keep aligned along with $\mathbf{h}_j(j)$ while to be as orthogonal to $\mathbf{h}_k(j)$ ($\forall\ k \neq i$) as possible. Consequently, the optimal solution of $\mathbf{w}_{ji}^l$ balances signal orthogonality with interference alignment between user-j's and other users' dual orthogonal channels at j-th time slot.

IV. Discussion

A. Multiplexing Gain of GMAT

In the following, we show the GMAT algorithm possesses the same multiplexing gain as original MAT. We consider the 2-user case for example. According to equations from (54) to (66), we have

$$\lim_{\rho\to\infty}\frac{\mathbb{E}\log\left(\max_{\|\mathbf{w}_1\|^2=1}\frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right)}{\log\rho} \tag{75}$$

$$= \lim_{\rho\to\infty}\frac{\mathbb{E}\log\left(\frac{\mathbf{w}_1^H\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_1^H\mathbf{Q}_1\mathbf{w}_1}\right)\Big|_{\mathbf{w}_1=\frac{\mathbf{h}_B(1)}{\|\mathbf{h}_B(1)\|}}}{\log\rho} = 1 \tag{76}$$

$$\lim_{\rho\to\infty}\frac{\mathbb{E}\log\left(\max_{\|\mathbf{w}_2\|^2=1}\frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\right)}{\log\rho} \tag{77}$$

$$= \lim_{\rho\to\infty}\frac{\mathbb{E}\log\left(\frac{\mathbf{w}_2^H\mathbf{Q}_2\mathbf{w}_2}{\mathbf{w}_2^H\mathbf{R}_2\mathbf{w}_2}\right)\Big|_{\mathbf{w}_2=\frac{\mathbf{h}_A(2)}{\|\mathbf{h}_A(2)\|}}}{\log\rho} = 1. \tag{78}$$

Thus, together with the fact that $\lim_{\rho\to\infty}\frac{\mathbb{E}\log(C_1C_4)}{\log\rho} = 2$, the multiplexing gain can be achieved with

$$\mathrm{MG} = \lim_{\rho\to\infty}\frac{\mathbb{E}\max_{\|\mathbf{w}_i\|^2=1}(I(\mathbf{s}_A;\mathbf{y}_A)+I(\mathbf{s}_B;\mathbf{y}_B))}{3\log\rho} = \frac{4}{3}$$

which is identical to the original MAT algorithm. Intuitively, at high SNR, the signal orthogonality becomes no relevance, thus our solution naturally seeks perfect interference alignment as in MAT.

[3] We abuse here the vector notation to represent the corresponding orthogonal channel matrix for the sake of consistence.

$$I(\mathbf{s}_A;\mathbf{y}_A)+I(\mathbf{s}_B;\mathbf{y}_B) \tag{72}$$

$$=\log\left(1+\frac{\alpha_1\rho\mathbf{w}_1^H\mathbf{h}_A(1)\mathbf{h}_A^H(1)\mathbf{w}_1+\alpha_2\rho\mathbf{w}_1^H\mathbf{h}_A^{\perp}(1)\mathbf{h}_A^{\perp H}(1)\mathbf{w}_1}{\sigma_1^2+\beta_3\rho\mathbf{w}_2^H\mathbf{h}_A(2)\mathbf{h}_A^H(2)\mathbf{w}_2+\beta_4\rho\mathbf{w}_2^H\mathbf{h}_A^{\perp}(2)\mathbf{h}_A^{\perp H}(2)\mathbf{w}_2}\right) \tag{73}$$

$$+\log\left(1+\frac{\beta_1\rho\mathbf{w}_2^H\mathbf{h}_B(2)\mathbf{h}_B^H(2)\mathbf{w}_2+\beta_2\rho\mathbf{w}_2^H\mathbf{h}_B^{\perp}(2)\mathbf{h}_B^{\perp H}(2)\mathbf{w}_2}{\sigma_2^2+\alpha_3\rho\mathbf{w}_1^H\mathbf{h}_B(1)\mathbf{h}_B^H(1)\mathbf{w}_1+\alpha_4\rho\mathbf{w}_1^H\mathbf{h}_B^{\perp}(1)\mathbf{h}_B^{\perp H}(1)\mathbf{w}_1}\right)+\log(C_1C_4) \tag{74}$$

B. Single-beam MIMO Interference Channel Interpretation

To understand more clearly the roles of desired signal orthogonality and interference alignment, we transform the mutual information equality (54) into another form, and further interpret their relationship from the point of view of a two-user single-beam MIMO interference channel. The strong benefit of this interpretation is that the problem of computing the precoders lends itself to classical precoding techniques in the MIMO interference channel. Based on eq-(54), the sum mutual information equation can be further transformed to the form as shown on the top of this page, where

$$\alpha_1=\frac{\alpha_2}{1+\rho\|\mathbf{h}_A(1)\|^2},\quad \alpha_2=\frac{1+\rho\|\mathbf{h}_A(2)\|^2}{\rho\|\mathbf{h}_A(1)\|^2}, \tag{79}$$

$$\alpha_3=\frac{1}{\rho|h_{B1}(3)|^2\|\mathbf{w}_1\|^2},\quad \alpha_4=\alpha_3+1, \tag{80}$$

$$\beta_1=\frac{\beta_2}{1+\rho\|\mathbf{h}_B(2)\|^2},\quad \beta_2=\frac{1+\rho\|\mathbf{h}_B(1)\|^2}{\rho\|\mathbf{h}_B(2)\|^2}, \tag{81}$$

$$\beta_3=\frac{1}{\rho|h_{A1}(3)|^2\|\mathbf{w}_2\|^2},\quad \beta_4=\beta_3+1, \tag{82}$$

$$\sigma_1^2=\frac{1}{\rho|h_{A1}(3)|^2}+\|\mathbf{w}_2\|^2,\sigma_2^2=\frac{1}{\rho|h_{B1}(3)|^2}+\|\mathbf{w}_1\|^2. \tag{83}$$

According to eq-(73) and eq-(74), the sum mutual information can be treated as that of 2-user MIMO interference channels with 2 antennas at each transmitter and receiver, as shown in Fig. 1. Note that $\mathbf{w}_1$ and $\mathbf{w}_2$ act as the transmit beamformers, where the single beam is transmitted from each transmitter.

Accordingly, the received signals at two receivers can be equivalently expressed as

$$\mathbf{y}_1=\sqrt{\rho}\mathbf{H}_1\mathbf{w}_1s_1+\sqrt{\rho}\mathbf{H}_2\mathbf{w}_2s_2+\mathbf{n}_1 \tag{84}$$

$$\mathbf{y}_2=\sqrt{\rho}\mathbf{G}_2\mathbf{w}_2s_2+\sqrt{\rho}\mathbf{G}_1\mathbf{w}_1s_1+\mathbf{n}_2 \tag{85}$$

where

$$\mathbf{H}_1=\begin{bmatrix}\sqrt{\alpha_1}\mathbf{h}_A^H(1)\\ \sqrt{\alpha_2}\mathbf{h}_A^{\perp H}(1)\end{bmatrix},\quad \mathbf{H}_2=\begin{bmatrix}\sqrt{\beta_3}\mathbf{h}_A^H(2)\\ \sqrt{\beta_4}\mathbf{h}_A^{\perp H}(2)\end{bmatrix}, \tag{86}$$

$$\mathbf{G}_1=\begin{bmatrix}\sqrt{\alpha_3}\mathbf{h}_B^H(1)\\ \sqrt{\alpha_4}\mathbf{h}_B^{\perp H}(1)\end{bmatrix},\quad \mathbf{G}_2=\begin{bmatrix}\sqrt{\beta_1}\mathbf{h}_B^H(2)\\ \sqrt{\beta_2}\mathbf{h}_B^{\perp H}(2)\end{bmatrix} \tag{87}$$

and the noises are distributed with $\mathbf{n}_i\sim\mathcal{CN}(0,\frac{\sigma_i^2}{2}\mathbf{I})$.

Consequently, the received SINR for both users can be written, respectively, as

$$\text{SINR}_1=\frac{\rho\|\mathbf{H}_1\mathbf{w}_1\|^2}{\sigma_1^2+\rho\|\mathbf{H}_2\mathbf{w}_2\|^2}=\frac{\rho\mathbf{w}_1^H\mathbf{H}_1^H\mathbf{H}_1\mathbf{w}_1}{\sigma_1^2+\rho\mathbf{w}_2^H\mathbf{H}_2^H\mathbf{H}_2\mathbf{w}_2} \tag{88}$$

$$\text{SINR}_2=\frac{\rho\|\mathbf{G}_2\mathbf{w}_2\|^2}{\sigma_2^2+\rho\|\mathbf{G}_1\mathbf{w}_1\|^2}=\frac{\rho\mathbf{w}_2^H\mathbf{G}_2^H\mathbf{G}_2\mathbf{w}_2}{\sigma_2^2+\rho\mathbf{w}_1^H\mathbf{G}_1^H\mathbf{G}_1\mathbf{w}_1} \tag{89}$$

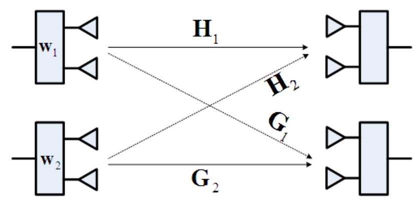

Fig. 1: Interpretation as MIMO Interference Channel.

which are identical to those in eq-(73-74). Note that the approach in the previous section that skips the dependency on the unknown channel coefficients $h_{A1}(3)$, $h_{B1}(3)$ can also be applied. We omit the details here to avoid redundancy. Hence, existing precoder design methods in the two-user single-beam MIMO interference channels with perfect CSIT, e.g., [13, 17–20], can be used here in the context of delayed CSIT precoding. Instead of going into details about those solutions, we take the classic MRT and ZF precoders here for example,

$$\mathbf{w}_1^{MRT}=\mathbf{U}_{\max}(\mathbf{H}_1^H\mathbf{H}_1),\quad \mathbf{w}_2^{MRT}=\mathbf{U}_{\max}(\mathbf{G}_2^H\mathbf{G}_2) \tag{90}$$

$$\mathbf{w}_1^{ZF}=\mathbf{U}_{\min}(\mathbf{G}_1^H\mathbf{G}_1),\quad \mathbf{w}_2^{ZF}=\mathbf{U}_{\min}(\mathbf{H}_2^H\mathbf{H}_2) \tag{91}$$

where $\mathbf{U}_{\max}(\cdot)$ and $\mathbf{U}_{\min}(\cdot)$ are the generalized eigenvectors corresponding to the largest and smallest eigenvalues, respectively. Interestingly, for the first user, it is worth noting that $\alpha_1<\alpha_2$ and therefore $\mathbf{w}_1^{MRT}\rightarrow\mathbf{h}_A^{\perp}(1)$, means perfect orthogonality of desired signal is preferred. On the other hand, $\alpha_3<\alpha_4$, which denotes $\mathbf{w}_1^{ZF}\rightarrow\mathbf{h}_B(1)$, corresponds to the preference of perfect interference alignment. Our proposed GMAT-MMSE and GMAT-DSINR solutions offer a trade-off between them, yielding a better performance at finite SNR.

V. Numerical Results

The effectiveness of the proposed solutions is evaluated in terms of the sum rate per time slot in bps/Hz over a correlated rayleigh fading channel, where the concatenated channel matrix in slot-t can be formulated as

$$\mathbf{H}(t)=\mathbf{R}_r^{1/2}\mathbf{H}_w(t)\mathbf{R}_t^{1/2} \tag{92}$$

where $\mathbf{H}_w(t)$ is normalized i.i.d. rayleigh fading channel matrix, and $\mathbf{R}_t$, $\mathbf{R}_r$ are transmit and receive correlation matrices with (i,j)-th entry being $\tau_t^{|i-j|}$ and $\tau_r^{|i-j|}$ [21, 22], respectively, where τ_t and τ_r are randomly chosen within $[0,1)$. Note that the users' channel vectors are the rows of $\mathbf{H}(t)$.

The parameters in the simulation are set as follows: maximum 500 gradient-descent iterations for the GMAT-MMSE with $\beta=0.01$. The performance is averaged over 1000 channel realizations. Recall that the present channel coefficients

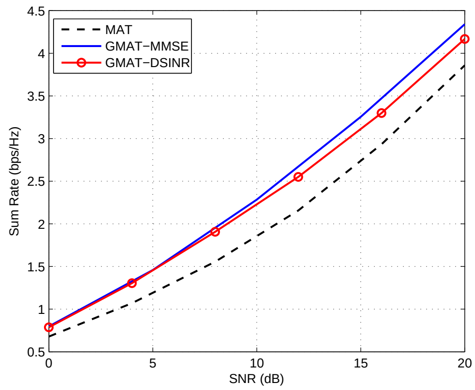

Fig. 2: Sum rate vs. SNR for the 2-user case.

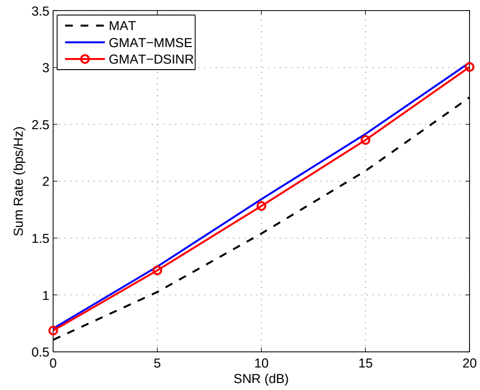

Fig. 3: Sum rate vs. SNR for the 3-user case.

(cf. $\mathbf{D}_i^{(k)}$, e.g., $h_{A1}(3)$ and $h_{B1}(3)$ for the 2-user case) are unknown to the transmitter and therefore are circumvented for transmit precoder design, while they should be taken into account at the receiver. Naturally, such a mismatch would result in performance degradation, but our proposed precoding methods are verified to be always effective thanks to the efficient trade-off between interference alignment and signal enhancement.

We show in Fig. 2 for the 2-user case the sum rate comparison among GMAT-MMSE with the iteratively updated $\mathbf{w}_1$, $\mathbf{w}_2$, GMAT-DSINR with closed-form solutions in eq-(66), and the original MAT algorithm with $\mathbf{w}_1 = \mathbf{h}_B(1)$, $\mathbf{w}_2 = \mathbf{h}_A(2)$, with the same power constraint $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq 2$ for all. Note that the MMSE receiver is used here to set up a reference for comparison together with GMAT-MMSE solution, such that we are able to show how good the closed-form solution can achieve compared to the iterative one at finite SNR. In Fig. 2, the gap of sum rate between GMAT and MAT illustrates improvement of the GMAT-MMSE and GMAT-DSINR algorithms over the initial MAT concept, demonstrating the benefit of the trade-off between interference alignment and desired signal orthogonality enhancement. Compared with the original MAT algorithm, both GMAT approaches have gained great improvement at finite SNR and possessed the same slope, which implies the same multiplexing gain, at high SNR. Interestingly, the closed-form solution performs as well as the iterative one, indicating the effectiveness of the mutual information approximation.

In Fig. 3, we present the similar performance comparison for the 3-user case with *MMSE receiver*. The GMAT-MMSE solution updates order-2 message generation matrix $\{\mathbf{W}_j^l(2)\}$ iteratively, while the original MAT algorithm set it according to eq-(22) and the GMAT-DSINR solution is obtained by optimizing eq-(70) and eq-(71). All these methods hold the same power allocation. With more transmit antennas and users, the same insights regarding the trade-off between signal orthogonality and interference alignment can be always obtained. It is interesting to note that, GMAT-DSINR performs as well as GMAT-MMSE, despite the distributed optimization.

VI. Conclusion

We generalize the concept of precoding over a multi-user MISO channel with delayed CSIT for arbitrary number of users case, by proposing a precoder construction algorithm, which achieves the same DoF at infinite SNR yet reaches a useful trade-off between interference alignment and signal enhancement at finite SNR. Our proposed precoding concept lends itself to a variety of optimization methods, e.g., virtual MMSE and mutual information solutions, achieving good compromise between signal orthogonality and interference alignment. An interesting question is also the diversity gain performance of schemes combining current and delayed CSIT. Clearly our scheme will achieve the same diversity performance as the original MAT since it converges to it in the high SNR regime. The question of whether modified schemes can be devised to address the DoF-diversity trade-off is an interesting open problem.

Appendix

A. Gradient Descent Parameter for GMAT-MMSE

Let $[\mathbf{H}_{ij}^l]_{m,n} = \mathbf{e}_m^H \mathbf{H}_{ij}^l \mathbf{e}_n$ be the m-th row and n-th column element of $\mathbf{H}_{ij}^l$. Particularly,

$$[\mathbf{H}_{ij}^l]_{m,n} = \mathbf{e}_{m'}^H \mathbf{W}_j^l(k)\mathbf{e}_n \tag{96}$$

when $m = \sum_{s=1}^{k-1} T_s + m'$ where $1 \leq m' \leq T_k$ and $1 \leq n \leq K$. Here, $\mathbf{e}_m$ is defined as a binary vector with only one '1' at m-th row. By differentiating over $\mathbf{W}_j^l(2)$, we have the differentiation as shown on the bottom of next page, where

$$\mathbf{Q}^l = \begin{bmatrix} \mathbf{0}_{T_1 \times K} \\ \mathbf{0}_{m_2^l \times K} \\ \mathbf{I} \\ \mathbf{0}_{n_2^l \times K} \\ \vdots \\ \prod_{t=2}^{K-1} \mathbf{C}^l(t)\mathbf{\Lambda}^l(t) \end{bmatrix}. \tag{97}$$

Note that we abuse vector $\mathbf{e}_{m'}$ with various dimensions T_k according to the corresponding matrices $\mathbf{W}_j^l(k)$ for the sake of notational simplicity. Then, it follows that

$$\frac{\partial [\mathbf{H}_{ij}^l]_{m,n}}{\partial [\mathbf{W}_j^{l\ T}(2)]_{p,q}} = \mathbf{e}_m^H \mathbf{Q}^l \mathbf{e}_p \mathbf{e}_q^H \mathbf{e}_n \tag{98}$$

where $1 \le p \le T_k$, $1 \le q \le K$, and we have

$$\frac{\partial \mathbf{H}_{ij}^l}{\partial [\mathbf{W}_j^{l\ T}(2)]_{p,q}} = \mathbf{Q}^l \mathbf{e}_p \mathbf{e}_q^H \quad (99)$$

Finally, according to the chain rule of matrix differentiation [15, 23], we have

$$\frac{\partial \left(J_i^l\right)}{\partial [\mathbf{W}_j^{l\ T}(2)]_{p,q}} = \mathrm{Tr}\left(\left(\frac{\partial J_i^l}{\partial \mathbf{H}_{ij}^l}\right)^T \frac{\partial \mathbf{H}_{ij}^l}{\partial [\mathbf{W}_j^{l\ T}(2)]_{p,q}}\right) \quad (100)$$

$$= \mathrm{Tr}\left(\mathbf{e}_q^H \left(\frac{\partial J_i^l}{\partial \mathbf{H}_{ij}^l}\right)^T \mathbf{Q}^l \mathbf{e}_p\right). \quad (101)$$

So, for the K-user case, the Gaussian descent parameter can be calculated by

$$\frac{\partial (J)}{\partial \mathbf{W}_j^l(2)} = \sum_{i=1}^K \frac{\partial \left(J_i^l\right)}{\partial \mathbf{W}_j^l(2)} = \sum_{i=1}^K \left(\frac{\partial J_i^l}{\partial \mathbf{H}_{ij}^l}\right)^T \mathbf{Q}^l \quad (102)$$

where

$$\left(\frac{\partial J_i^l}{\partial \mathbf{H}_{ii}^l}\right)^T = f\left(\sqrt{\rho}\mathbf{H}_{ii}^l, \rho \sum_{l=1}^L \sum_{j=1, j\neq i}^K \mathbf{H}_{ij}^l \mathbf{H}_{ij}^{l\ H} + \mathbf{I}\right)$$

$$\left(\frac{\partial J_i^l}{\partial \mathbf{H}_{ij}^l}\right)^T = g\left(\sqrt{\rho}\mathbf{H}_{ij}^l, \rho \sum_{l=1}^L \sum_{k=1, k\neq j}^K \mathbf{H}_{ik}^l \mathbf{H}_{ik}^{l\ H} + \mathbf{I}\right)$$

where

$$f(\mathbf{A}, \mathbf{B}) = -\mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H + \mathbf{B}\right)^{-1} \mathbf{B} \left(\mathbf{A}\mathbf{A}^H + \mathbf{B}\right)^{-1}$$
$$g(\mathbf{A}, \mathbf{B}) = \mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H + \mathbf{B}\right)^{-1} (\mathbf{B} - \mathbf{I}) \left(\mathbf{A}\mathbf{A}^H + \mathbf{B}\right)^{-1}.$$

ACKNOWLEDGEMENTS

Fruitful discussions with Sheng Yang, Mari Kobayashi and Paul de Kerret are thankfully acknowledged.

REFERENCES

[1] D. Gesbert, M. Kountouris, R. Heath, C-B. Chae, T. Salzer, "From single user to multiuser communications: shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[2] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *in Proc. of Allerton Conference*, Monticello, IL, Oct. 2010. [Online].Available: http://arxiv.org/abs/1010.1499.

[3] 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.

[4] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[5] H. Maleki, S. A. 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, Jun. 2012.

[6] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "Interference alignment for the MIMO interference channel with delayed local CSIT," Feb. 2011, [Online].Available: http://arxiv.org/abs/1102.5673.

[7] C. S. Vaze, M. K. Varanasi, "The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4396–4417, Jul. 2012.

[8] J. Xu, J. Andrews, and S. Jafar, "MISO broadcast channels with delayed finite-rate feedback: predict or observe?" *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1456–1467, Apr. 2012.

[9] A. Adhikary, H. C. Papadopoulos, S. A. Ramprashad and G. Caire, "Multi-user MIMO with outdated CSI: training, feedback and scheduling," *in Proc. Allerton Conference*, Monticello, IL, Sept. 2011, [Online].Available: http://arxiv.org/abs/1109.6371.

[10] X. Yi, D. Gesbert, "Precoding on the broadcast MIMO channel with delayed CSIT: the finite SNR case," *in Proc. IEEE ICASSP*, Kyoto, Japan, Mar. 2012.

[11] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[12] S. W. Peters and R. W. Heath, Jr., "Interference alignment via alternating minimization," *in Proc. IEEE ICASSP*, Taipei, ROC, Apr. 2009.

[13] F. Negro, S. P. Shenoy, I. Ghauri, and D. Slock, "On the MIMO interference channel," *in Proc. Information Theory and Applications*, San Diego, CA, Feb. 2010.

[14] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[15] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[16] A. C. Aitken, *Determinants and Matrices*, 9th ed. Edinburgh, U.K.: Oliver and Boyd, 1956.

[17] Z. Ho and D. Gesbert, "Balancing egoism and altruism on the interference channel: the MIMO case," *in Proc. IEEE ICC*, Cape Town, South Africa, May 2010.

[18] D. A. Schmidt, W. Utschick, and M. L. Honig, "Beamforming techniques for single-beam MIMO interference networks," *in Proc. Allerton Conference*, Monticello, IL, Sept. 2010.

[19] S. W. Peters and J. R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 206–218, Jan. 2011.

[20] P. Cao, E. A. Jorswieck, and S. Shi, "On the Pareto boundary for the two-user single-beam MIMO interference channel," *submitted to IEEE Trans. Signal Processing*, Dec. 2011, [Online].Available: http://arxiv.org/abs/1202.5474.

[21] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, pp. 369–371, 2001.

[22] C. Xiao, J. Wu, S. Y. Leong, Y. R. Zheng, and K. B. Letaief, "A discrete-time model for triply selective MIMO Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1678–1688, Sept. 2004.

[23] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

$$\frac{\partial [\mathbf{H}_{ij}^l]_{m,n}}{\partial \mathbf{W}_j^{l\ T}(2)} = \left(\frac{\partial [\mathbf{H}_{ij}^l]_{m,n}}{\partial \mathbf{W}_j^l(2)}\right)^T \quad (93)$$

$$= \begin{cases} \mathbf{0} & if\ m \le T_1 \\ \mathbf{e}_n \mathbf{e}_{m'}^H & if\ T_1 + 1 \le m \le T_1 + T_2 \\ \mathbf{e}_n \mathbf{e}_{m'}^H \prod_{t=2}^{k-1} \mathbf{C}^l(t) \mathbf{\Lambda}^l(t) & if\ \sum_{s=1}^{k-1} T_s + 1 \le m \le \sum_{s=1}^{k} T_s\ when\ k \ge 3 \end{cases} \quad (94)$$

$$= \mathbf{e}_n \mathbf{e}_m^H \mathbf{Q}^l \quad (95)$$

Xinping Yi (S'12) received his B.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 2005, and M.E. degree from University of Electronic Science and Technology, Chengdu, China, in 2009, both in Electrical Engineering. Currently, He is pursuing the Ph.D. degree at Mobile Communication Department, EURECOM, Sophia Antipolis, France. From 2009 to 2011, he was with Huawei Technologies, Shenzhen, China, as a research engineer. His current research interests include multiuser information theory, interference alignment and limited feedback.

David Gesbert (F'11) is Professor and Head of the Mobile Communications Department, EURECOM, France. He obtained the Ph.D degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. In 1999, he was a founding engineer of Iospan Wireless Inc, San Jose, Ca.,a startup company pioneering MIMO-OFDM (now Intel). Between 2001 and 2003 he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 200 papers and several patents all in the area of signal processing, communications, and wireless networks.

D. Gesbert was a co-editor of several special issues on wireless networks and communications theory, for JSAC (2003, 2007, 2009), EURASIP Journal on Applied Signal Processing (2004, 2007), Wireless Communications Magazine (2006). He served on the IEEE Signal Processing for Communications Technical Committee, 2003-2008. He authored or co-authored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, 2005 Best Paper (Young Author) Award for Signal Proc. Society journals, and the Best Paper Award for the 2004 ACM MSWiM workshop. He co-authored the book Space time wireless communications: From parameter estimation to MIMO systems, Cambridge Press, 2006.