IEEE TRANSACTIONS ON COMMUNICATIONS, ACCEPTED FOR PUBLICATION

The Effect of Imperfect Channel Knowledge on a MIMO System with Interference

Namjeong Lee, Student Member, IEEE, Osvaldo Simeone, Member, IEEE, and Joonhyuk Kang, Member, IEEE

Abstract—A common model for transmission over wireless links is that of a multiantenna system affected by an additive interfering signal. In some scenarios of interest, such as when the interferer is located close to the transmitter and performs retransmission, interference may be learned by the transmitter, but remain unknown at the receiver. In this case, it is well known that, if transmitter and receiver have perfect channel state information (CSI), then a technique called Dirty Paper Coding (DPC) is able to fully mitigate the interference.

This paper studies the impact of imperfect CSI on a multipleinput multiple-output (MIMO) system with interference and compares the performance of DPC with that of a scheme where interference is decoded at the receiver, which we refer to as beamforming with joint decoding (BF-JD). Unlike DPC, which models the interference as an "unstructured" random process, BF-JD exploits the fact that the interfering signal is a codeword of the interferer's codebook. It is demonstrated by analysis and numerical results that BF-JD provides advantages over DPC when CSI is imperfect at the transmitter but perfect at the receiver, whereas this is not true for the case of imperfect CSI at both transmitter and receiver.

Index Terms—MIMO system, structured interference, dirty paper coding, beamforming with joint decoding, imperfect CSI.

I. INTRODUCTION

NTERFERENCE is becoming increasingly the main impediment to wireless communication, especially in urban areas of developed countries. To overcome interference, several approaches are currently being implemented, ranging from interference avoidance schemes to interference cancellation (see, e.g., [1]). Since interference is caused by the transmission of other terminals (e.g., mobiles or base stations), it may be learned by some transmitter in the vicinity of the interferer [2], [3]. For instance, assume that the interferer in Fig. 1 uses a retransmission strategy (HARQ), so that a given interfering packet may get retransmitted in successive timeslots. The transmitter, being close to the interferer, can decode the interferer's packet from a given (re)transmission. Once decoding is done, in the next time-slots, the transmitter knows the interferer's codeword and this information can be used for interference management. With respect to the time-slots at

The work of J. Kang was supported by the KCC, Korea, under the R&D program supervised by the KCA (KCA-2012-(09913-05003)). The work of O. Simeone was partially supported by U.S. NSF under grant CCF-0914899.

N. Lee is with the Samsung Advanced Institute of Technology (SAIT), Gyeonggi-do, Korea (e-mail: namjeong.lee@samsung.com).

O. Simeone is with the CWCSPR, New Jersey Institute of Technology, Newark, NJ 07102, USA (e-mail: osvaldo.simeone@njit.edu).

J. Kang is with the Korea Advanced Institute of Science and Technology, Daejeon, Korea (e-mail: jhkang@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TCOMM.2012.061412.110567

hand, the transmitter thus knows the interfering signal "noncausally," i.e., before it is actually transmitted in the current slot. In this scenario, the problem of designing the optimal transmission strategy is about how the transmitter should better use the interference information.

1

If the interference is known non-causally at the transmitter, and if transmitter and receiver have perfect channel state information (CSI), then a technique called Dirty Paper Coding (DPC) is well known to be able to fully mitigate the interference [4], [5]. In other words, even though the receiver does not know the interfering signal, perfect interference cancellation is achieved thanks to the knowledge of the interference at the transmitter only. Moreover, it should be emphasized that the decoder in DPC operates on the interferer's sequence without leveraging the fact that the interferer's signal is a specific codeword of the interferer's codebook. Indeed, the decoder in DPC assumes that the interference is a signal with independent identically distributed (i.i.d.) symbols or arbitrary sequences [6], [7]. What this discussion entails is that leveraging the structure of the interferer's codebook is not necessary with perfect CSI, since optimal (interference-free) performance can be attained by DPC.

In practice, CSI can hardly be considered perfect, especially at the transmitter's side, where it is usually collected via feedback from the receiver¹. This paper studies the impact of imperfect CSI on the performance of a multiantenna system with the aim of assessing whether in this case there are any benefits to be accrued by leveraging the structure of the interference at the decoder. To this end, we compare the performance of DPC with that of a scheme where interference is jointly *decoded* at the receiver along with the desired signal, which we refer to as beamforming with joint decoding (BF-JD). "Beamforming" refers to the fact that the transmitter can allocate some of its power to boost reception of the interferer's signal at the receiver and hence facilitate decoding. Unlike DPC, BF-JD thus exploits the fact that the interfering signal is a codeword of the interferer's codebook. It is demonstrated by analysis and numerical results that BF-JD provides advantages over DPC when CSI is imperfect at the transmitter but perfect at the receiver, whereas this is not true for the case of imperfect CSI at both transmitter and receiver.

The analysis in this paper is inspired by [2], where a general channel operating over a discrete alphabet and with perfect CSI was studied to show the advantages of capitalizing over the structure of the interference signal. In related works, the effect of imperfect CSI on the performance of DPC has

Paper approved by D. J. Love, the Editor for MIMO and Adaptive Techniques of the IEEE Communications Society. Manuscript received August 26, 2011; revised January 13, 2012.

¹In a Time Division Duplex (TDD) system, it can also be collected from the signal received over the reverse link.

2

been studied in [8]–[10] for a multiple-input multiple-output (MIMO) system and in [11] for a single-antenna. References [8]–[10] focus on the analysis of ergodic transmission rates, whereas [11] also studies the outage probability for a non-ergodic setting. An information-theoretic analysis of the non-ergodic setting was first reported in [12]. Overall, this body of work concerning imperfect CSI does not model the interfering signal as structured. The aim of this paper is discussing the advantages of leveraging the interference structure over the standard DPC scheme.

The remainder of the study is organized as follows. A MIMO system affected by structured interference is introduced in Sec. II. The achievable rate of DPC and BF-JD are studied in Sec. III in the case of perfect CSI at the transmitter (CSIT) and at the receiver (CSIR). In Sec. IV-B and IV-C, we provide achievable rate bounds for the case of imperfect CSIT but perfect CSIR, and imperfect CSIT and CSIR, respectively. Sec. V provides some numerical examples to obtain insights into the performance comparison of the considered schemes. Finally, Sec. VI presents the main conclusions of this work.

Notation: A bold face letter denotes a vector or a matrix; $[\cdot]^T$, $[\cdot]^*$, $tr\{\cdot\}$ the transpose, the conjugate transpose, and the trace of a vector or a matrix, respectively; $||\boldsymbol{a}||$ the 2-norm of a vector \boldsymbol{a} ; max $\{x, y\}$ and min $\{x, y\}$ the maximum and minimum element between x and y, respectively; $\mathbb{E}[\cdot]$ an expected value of a vector or a matrix; \boldsymbol{I} the identity matrix; \succeq is used in the matrix positive-semidefinite sense; Mutual information and differential entropy are defined in the standard way following [13]; Distribution of random variable X is denoted as p(x) and similarly for conditional and joint probabilities; For a matrix \boldsymbol{A} , $(\boldsymbol{A})_{ij}$ denotes the (i, j)-th element.

II. SYSTEM MODEL

Consider the $r \times t$ MIMO system affected by an interferer in Fig. 1, where the transmitter has t antennas, the receiver has r antennas, and the interferer has t_S antennas. The received signal at time i with i = 1, ..., n is

$$\boldsymbol{Y}(i) = \boldsymbol{H}\boldsymbol{X}(i) + \boldsymbol{H}_{S}\boldsymbol{S}(i) + \boldsymbol{Z}(i), \qquad (1)$$

where $\boldsymbol{X}(i) \in \mathbb{C}^{t \times 1}$ is the transmitted signal, $\boldsymbol{S}(i) \in \mathbb{C}^{t_S \times 1}$ is the interference, and $\boldsymbol{Z}(i) \sim \mathcal{CN}(0, \boldsymbol{I})$ is the noise at time *i* where $\boldsymbol{Z}(i) \in \mathbb{C}^{r \times 1}$. The power constraint at the transmitter imposes the condition $\mathbb{E}\left[\|\boldsymbol{X}(i)\|^2\right] \leq P_X$ for all *i*. Information about the channel matrices $\boldsymbol{H} \in \mathbb{C}^{r \times t}$ and $\boldsymbol{H}_S \in \mathbb{C}^{r \times t_S}$ may not be fully available at the transmitter and the receiver, as detailed in the next sections.

The interfering sequence S(i), i = 1, ..., n, is assumed to be uniformly drawn from a codebook of 2^{nR_S} codewords, where R_S is the interferer's rate. The codebook is generated randomly such that each letter S(i) of any codeword is i.i.d. with distribution $S(i) \sim C\mathcal{N}(0, \mathbf{K}_S)$ for a given covariance matrix \mathbf{K}_S . The interferer's signal satisfies a perantenna power constraint P_S/t_S so that $(\mathbf{K}_S)_{jj} = P_S/t_S$, $j = 1, ..., t_S$. The interferer's codeword S(i), i = 1, ..., n, is known to the transmitter, so that, when encoding the input sequence $\mathbf{X}(i)$, i = 1, ..., n, the transmitter can use the



Fig. 1. MIMO system with structured interference.

information about the interfering sequence. This can be done in different ways as discussed in the rest of the paper.

A rate R, measured in bit per channel use or equivalently bit per second per Hz (bps/Hz), is said to be achievable if a transmission scheme for X(i), i = 1, ..., n, exists that is able to drive the probability of error to zero on average with respect to all possible choices of the interferer's codebook. This is the same criterion adopted in [3] (see also [2], [14]). It is noted that this condition does not ensure that transmission is reliable for all possible choices of the interfering codebook, but it does guarantee that the probability of observing an interfering codebook for which transmission is not reliable is arbitrarily small. The goal of this paper is to design the transmission scheme for X(i) based on imperfect knowledge of the CSI $\mathcal{H} = [\boldsymbol{H} \boldsymbol{H}_S]$ at either or both transmitter and receiver in order to maximize the achievable rate R. The CSI models will be detailed below. For simplicity, we drop the time index *i* in the following where it does not create confusion.

III. PERFECT CHANNEL STATE INFORMATION

In this section, we recall the DPC and BF-JD schemes and study their performance for the baseline case of perfect CSI at both the transmitter and the receiver. It is well known that, in this case, DPC is able to obtain the same rate as for a channel without interference (i.e., $P_S = 0$) and is thus optimal. Therefore, capitalizing on the structure of the interference through a scheme like BF-JD cannot bring any rate improvement. We will see in the next section that this is not true in the case of imperfect CSI.

A. DPC

DPC is a particular implementation of the so called Gel'fand-Pinsker precoding technique [5]. The idea behind such technique is that of "matching" the transmitted signal to the interfering sequence so as to reduce the uncertainty about the latter at the receiver's side. Using the standard notation in terms of mutual information, the rate achievable with this scheme can be written as [5]

$$R = \max_{p(\boldsymbol{u}|\boldsymbol{s}), \ \boldsymbol{x} = f(\boldsymbol{u}, \boldsymbol{s})} I(\boldsymbol{U}; \boldsymbol{Y}) - I(\boldsymbol{U}; \boldsymbol{S}), \qquad (2)$$

where the optimization is subject to the power constraint $tr\{K_X\} \leq P_X$ with $K_X = \mathbb{E}[XX^*]$. The random variable

U is matrix-valued and accounts for the use of an auxiliary codebook that is used to match the transmitted signal to the interfering sequence. Note that decoding of DPC codeword inherently treats the interference as an unstructured signal. In fact, the decoder essentially recovers a combination of the transmitted signal X and the interference S by decoding the auxiliary codeword U. Its distribution is ruled by the conditional pmf p(u|s), while the transmitted signal X is selected as a function f(u, s) of the auxiliary variable U and the state S. It is shown in [5] that the rate (2) is the capacity of any state-dependent channel where the state sequence S is i.i.d. and thus unstructured.

When applied to the system model (1) with full CSI, it can be shown [15, Ch. 10], following [4], that the following choice of optimization variable

$$\boldsymbol{U} = \boldsymbol{X} + \boldsymbol{W} \boldsymbol{H}_{\boldsymbol{S}} \boldsymbol{S},\tag{3}$$

where \boldsymbol{X} is distributed as $\boldsymbol{X} \sim \mathcal{CN}(0, \boldsymbol{K}_X)$ and independent of \boldsymbol{S} , is optimal if $\boldsymbol{W} \in \mathbb{C}^{t \times r}$ is selected as follows:

$$\boldsymbol{W} = \boldsymbol{K}_{X}\boldsymbol{H}^{*}(\boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + \boldsymbol{I})^{-1}.$$
 (4)

In fact, with these choices, rate (2) reduces to the capacity of a channel with no interference

$$C = \max_{\substack{\boldsymbol{K}_X \succeq 0:\\ tr\{\boldsymbol{K}_X\} = P_X}} \log_2 |\boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_X \boldsymbol{H}^*|, \qquad (5)$$

and cannot thus be improved upon. It is noted that with the choice (4), we have that W(HX + Z) is the minimum meansquared error (MMSE) estimate of X given HX + Z (see [15, Ch. 10]). We also remark that the covariance K_X that maximizes (5) can be easily found by allocating power to the singular values of the channel matrix H via "water-filling" [16].

B. BF-JD

We now consider a scheme, referred to as BF-JD, where the decoder attempts joint decoding of the useful signal and of the interferer. Since the transmitter knows the interfering codeword, it can devote part of its power to boost reception of the interference at the receiver via beamforming. Specifically, from [2], [3], the rate that can be achieved with this scheme is

$$R'_{\text{BF-JD}} = \max_{\substack{p(\boldsymbol{x}|\boldsymbol{s}):\\tr\{\boldsymbol{K}_X\}=P_X}} \min\left\{I\left(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}\right), \ I\left(\boldsymbol{X},\boldsymbol{S};\boldsymbol{Y}\right)-R_S\right\}.$$
(6)

In (6) the distribution $p(\boldsymbol{x}|\boldsymbol{s})$ specifies the correlation between the transmitted signal \boldsymbol{X} and the interfering signal \boldsymbol{S} and thus the amount of resources employed by the transmitter to beamform the interfering signal \boldsymbol{S} to the destination in order to facilitate decoding of the interference at the receiver. The two terms in (6) correspond to the two typical ways in which a decoding error can take place, namely decoding \boldsymbol{X} erroneously when \boldsymbol{S} is decoded correctly and decoding both \boldsymbol{X} and \boldsymbol{S} erroneously (please see [2], [3] for details). While joint decoding is typically advantageous, in some cases it may entail an unnecessary burden on the decoder, and therefore we also allow the decoder to treat interference as noise, which achieves rate

$$R_{\text{BF-JD}}^{\prime\prime} = \max_{\substack{p(\boldsymbol{x}):\\tr\{\boldsymbol{K}_X\}=P_X}} I(\boldsymbol{X};\boldsymbol{Y})$$
(7)

following standard information-theoretic considerations. By selecting the best decoder between the joint decoder and the one that treats interference as noise for the given channel conditions and interference rate R_S , rate

$$R_{\text{BF-JD}} = \max\{R'_{\text{BF-JD}}, R''_{\text{BF-JD}}\}$$
(8)

is achieved.

Evaluating $R_{\text{BF-JD}}^{\prime\prime}$ (7) by using the maximum entropy theorem² [13] leads to

$$R_{\text{BF-JD}}^{\prime\prime} = \max_{\substack{\boldsymbol{K}_{X} \succeq 0:\\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} \log_{2} \left| \boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} \left(\boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} + \boldsymbol{I}\right)^{-1} \right|.$$
(9)

Optimization in (9) can be easily obtained via the water-filling algorithm [16]. Similarly, by the maximum entropy theorem [13], the optimal joint distribution of X and S that maximizes (6) is jointly complex Gaussian with zero mean and covariance matrix

$$\begin{bmatrix} \boldsymbol{K}_X & \boldsymbol{K}_{XS} \\ \boldsymbol{K}_{XS}^* & \boldsymbol{K}_S \end{bmatrix},$$
 (10)

where K_{XS} is the cross-correlation matrix between X and S, which rules the amount of resources used by the transmitter to beamform its signal along the interfering signal. Therefore, the achievable rate (6) becomes, evaluating the mutual information (11). Optimization (11) is non-convex due to the non-concavity of the objective function with respect to K_{XS} . Therefore, its solution requires complex global optimization tools [17]. Due to the complexity of optimization (11), we now obtain an upper bound to the achievable rate (11), which can be evaluated using standard techniques from convex optimization, and is thus easier to calculate. This upper bound, along with lower bounds obtained by evaluating the expression in (11) for given matrices K_X and K_{XS} , allows one to obtain insight into the performance of BF-JD. Our approach follows [18], where a similar method was proposed to obtain an upper bound on the capacity of the MIMO relay channel. The bound in [18], as ours, allows to replace the cross-covariance matrix K_{XS} with a scalar parameter ρ .

Proposition 1: The achievable rate (6) can be upper bounded by

$$R'_{\text{BF-JD}} \leq R^{\text{UB}}_{\text{BF-JD}} = \max_{\substack{\boldsymbol{K}_X \succeq 0, \rho \in [0,1]:\\ tr\{\boldsymbol{K}_X\} = P_X}} \min\left\{R^{\text{UB}}_A, R^{\text{UB}}_B - R_S\right\},$$
(12)

where

$$R_A^{\text{UB}} \triangleq \log_2 \left| \boldsymbol{I} + (1 - \rho^2) \boldsymbol{H} \boldsymbol{K}_X \boldsymbol{H}^* \right|$$
(13)

and

$$R_B^{\text{UB}} \triangleq \inf_{a>0} \log_2 \left| \boldsymbol{I} + \left(1 + \frac{\rho^2}{a} \right) \boldsymbol{H} \boldsymbol{K}_X \boldsymbol{H}^* + (1+a) \boldsymbol{H}_S \boldsymbol{K}_S \boldsymbol{H}_S^* \right.$$
(14)

²The maximum entropy theorem states that for a fixed covariance matrix, say K_Z , the differential entropy h(Z) of a random vector Z, is maximized by a Gaussian distribution. This is enough to conclude that (7) is equal to (9), since we have that $I(X;Y) = h(Y) - \log_2((\pi e)^r | I + H_S K_S H_S^* |)$ from standard calculations and Y has covariance matrix $K_Y = HK_X H^* + H_S K_S H_S^* + I$.

4

IEEE TRANSACTIONS ON COMMUNICATIONS, ACCEPTED FOR PUBLICATION

$$R'_{\text{BF-JD}} = \max_{\substack{\boldsymbol{K}_X \succeq 0, \boldsymbol{K}_{XS}:\\ tr\{\boldsymbol{K}_X\} = P_X}} \min \left\{ \begin{array}{c} \log_2 \left| \boldsymbol{I} + \boldsymbol{H} \left(\boldsymbol{K}_X - \boldsymbol{K}_{XS} \boldsymbol{K}_S^{-1} \boldsymbol{K}_{XS}^* \right) \boldsymbol{H}^* \right|, \\ \log_2 \left| \boldsymbol{I} + \boldsymbol{H} \boldsymbol{K}_X \boldsymbol{H}^* + \boldsymbol{H}_S \boldsymbol{K}_{XS}^* \boldsymbol{H}^* + \boldsymbol{H} \boldsymbol{K}_{XS} \boldsymbol{H}_S^* + \boldsymbol{H}_S \boldsymbol{K}_S \boldsymbol{H}_S^* \right| - R_S \end{array} \right\}.$$
(11)

with an arbitrary positive real number a.

Proof: Please refer to Appendix A.

For given ρ , R_A^{UB} and R_B^{UB} are concave in K_X and thus so is the upper bound (12). Therefore, the upper bound (12) can be obtained by performing a line search methods, such as bisection, over the scalar parameter $\rho \in [0, 1]$, where for each ρ , a convex optimization problem is solved.

IV. IMPERFECT CHANNEL STATE INFORMATION

In this section, the effect of imperfect CSI at the transmitter and the receiver on the transmission techniques discussed above is elaborated on.

A. Channel Estimation Error Model

We consider the following model for imperfect CSI. The receiver estimates the channels obtaining the estimates $\hat{\mathcal{H}}_{R} = [\hat{H}_{R} \ \hat{H}_{S,R}]$ and the transmitter obtains CSI, $\hat{\mathcal{H}}_{T} = [\hat{H}_{T} \ \hat{H}_{S,T}]$, e.g., via feedback from the receiver as in a Frequency Division Duplex (FDD) system or via direct estimation as in a TDD system [19]. The actual channel matrices $\mathcal{H} = [H \ H_{S}]$ are related to the estimates as

$$\mathcal{H} = \hat{\mathcal{H}}_{\mathsf{R}} + \mathcal{E}_{\mathsf{R}} = \hat{\mathcal{H}}_{\mathsf{T}} + \mathcal{E}_{\mathsf{T}},\tag{15}$$

where the CSI errors are $\mathcal{E}_{R} = [E_{R} E_{S,R}]$ and $\mathcal{E}_{T} = [E_{T} E_{S,T}]$. Further statistical characterization of the CSI error matrices will be discussed below for two different models.

B. Imperfect CSIT and Perfect CSIR

In this section, we assume that the receiver has perfect CSI while the transmitter has imperfect CSI. In other words, we have from equation (15) that

$$\mathcal{H} = \hat{\mathcal{H}}_{\mathsf{R}} = \hat{\mathcal{H}}_{\mathsf{T}} + \mathcal{E}_{\mathsf{T}},\tag{16}$$

where the transmitter CSI $\hat{\mathcal{H}}_{\mathsf{T}}$ is fixed (deterministic) and the estimation error matrices \mathcal{E}_{T} at the transmitter, \mathbf{E}_{T} and $\mathbf{E}_{S,\mathsf{T}}$, are independent of each other and have i.i.d. zeromean complex Gaussian entries with variances $\sigma_{E_{\mathsf{T}}}^2$ and $\sigma_{E_{S,\mathsf{T}}}^2$, respectively. It is assumed that the receiver is also informed of the transmitter estimate $\hat{\mathcal{H}}_{\mathsf{T}}$ as it is the case if CSIT is collected via feedback from the receiver. Moreover, the estimation errors vary in an ergodic fashion along the block so as to give operation meaning to the concept of ergodic rates (see, e.g., [15, Ch. 8]). This assumption is implicitly or explicitly adopted in many related works such as [8], [10], [20], [21].

1) DPC: The performance of DPC for the scenario at hand, but with $H = H_S$ in (1) was studied in [8], [9]. Moreover, [22] proposed an implementation of DPC with a general matrix Win (3) based on lattice codes also with $H = H_S$. Here we consider the more general model in (1), where $H \neq H_S$. This was also studied in [10], where an iterative algorithm to find a suboptimal matrix W in (3) was proposed.

To obtain the performance achievable with DPC, we evaluate the general expression (2) using the channel statistics (16) discussed above and recalling the fact that the receiver knows the CSIT error \mathcal{E}_{T} while the transmitter does not. The input variables U and X in (2) are selected according to DPC scheme described in Section III-A and in particular U is selected as in (3) with the estimate $\hat{H}_{S,T}$ instead of the actual matrix H_S . This is because the transmitter is not informed about the channel realization H_S . Overall, we have (17) where we have defined $K_U = K_X + W\hat{H}_{S,T}K_S\hat{H}_{S,T}^*W^*$, $K_{UY} = K_X H^* + W\hat{H}_{S,T}K_SH_S^*$, and $K_Y = HK_X H^* + H_SK_SH_S^* + I$, and we recall that, from (16), we have $H = \hat{H}_T + E_T$ and $H_S = \hat{H}_{S,T} + E_{S,T}$. Note that (17) averages the rate over all CSIT estimation errors \mathcal{E}_T via the operator $\mathbb{E}_{\mathcal{E}_T}[\cdot]$.

Optimization of W in (17) was studied in [8] (see also [9]) for $H = H_S$ and in [10] for the same system model considered in this paper with $H \neq H_S$. Here, instead of the iterative algorithm in [10], we resort to a simple suboptimal choice for W. The choice is done in analogy with the optimal W for the case of perfect CSI. In fact, observe that the received signal can be written as

$$Y = (H_{\mathsf{T}} + E_{\mathsf{T}})X + (H_{S,\mathsf{T}} + E_{S,\mathsf{T}})S + Z$$

= $\hat{H}_{\mathsf{T}}X + \hat{H}_{S,\mathsf{T}}S + \hat{Z},$ (18)

where $\hat{Z} = E_T X + E_{S,T} S + Z$ is the equivalent noise, which is zero-mean with covariance matrix $(1 + \sigma_{E_T}^2 P_X + \sigma_{E_{S,T}}^2 P_S)I$ and not Gaussian distributed. Moreover, the equivalent noise \hat{Z} is dependent but uncorrelated with X and S. Therefore, the suboptimal W is chosen such that $W(\hat{H}_T X + \hat{Z})$ is the linear MMSE estimate of X given $\hat{H}_T X + \hat{Z}$, namely

$$\boldsymbol{W} = \boldsymbol{K}_{X} \hat{\boldsymbol{H}}_{\mathsf{T}}^{*} \left(\hat{\boldsymbol{H}}_{\mathsf{T}} \boldsymbol{K}_{X} \hat{\boldsymbol{H}}_{\mathsf{T}}^{*} + (1 + \sigma_{E_{\mathsf{T}}}^{2} P_{X} + \sigma_{E_{S,\mathsf{T}}}^{2} P_{S}) \boldsymbol{I} \right)^{-1}.$$
(19)

Note that when there is no estimation error, (19) becomes equivalent to (4). In Section V, the performance with choice (19) will be compared with the algorithm in [10], showing that (19) leads to comparable similar performance.

2) BF-JD: The rate achievable by BF-JD is obtained from (6) using the channel statistics mentioned above and with jointly Gaussian X and S as done in Section III-B. Using again the maximum entropy theorem, it can be seen that a zero-mean joint Gaussian distribution for (X, S) with covariance matrix (10) is optimal and thus we obtain

$$R_{\text{BF-JD}} = \max\{R'_{\text{BF-JD}}, R''_{\text{BF-JD}}\},\tag{20}$$

where $R'_{\rm BF-JD}$ and $R''_{\rm BF-JD}$ are given by (21) and (22), respectively.

Note that $\mathcal{H} = \hat{\mathcal{H}}_{T} + \mathcal{E}_{T}$. As discussed in Section III-B, optimization (21) is non-trivial due to the non-concavity of the objective function with respect to K_{XS} while optimization of (22) can be obtained via the water-filling algorithm [16]. Similar to *Proposition 1*, we can obtain an upper bound on (21) as (23). As for *Proposition 1*, such bound has the advantage

LEE et al.: THE EFFECT OF IMPERFECT CHANNEL KNOWLEDGE ON A MIMO SYSTEM WITH INTERFERENCE

$$R_{\text{DPC}} = \max_{\substack{\boldsymbol{K}_{X} \succeq 0, \, \boldsymbol{W}: \\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} I\left(\boldsymbol{U}; \boldsymbol{Y} | \mathcal{E}_{\mathsf{T}}\right) - I\left(\boldsymbol{U}; \boldsymbol{S} | \mathcal{E}_{\mathsf{T}}\right)$$
$$= \max_{\substack{\boldsymbol{K}_{X} \succeq 0, \, \boldsymbol{W}: \\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} |\boldsymbol{K}_{X}| | \boldsymbol{K}_{Y}| - \log_{2} \left| \begin{array}{c} \boldsymbol{K}_{U} & \boldsymbol{K}_{UY} \\ \boldsymbol{K}_{UY}^{*} & \boldsymbol{K}_{Y} \end{array} \right| \right].$$
(17)

$$R'_{\text{BF-JD}} = \max_{\substack{\boldsymbol{K}_{X} \succeq 0, \boldsymbol{K}_{XS}:\\ tr\{\boldsymbol{K}_{X}\}=P_{X}}} \min \left\{ \begin{array}{c} \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} \left| \boldsymbol{I} + \boldsymbol{H}(\boldsymbol{K}_{X} - \boldsymbol{K}_{XS}\boldsymbol{K}_{S}^{-1}\boldsymbol{K}_{XS}^{*}) \boldsymbol{H}^{*} \right| \right], \\ \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} \left| \boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + \boldsymbol{H}\boldsymbol{K}_{XS}\boldsymbol{H}_{S}^{*} + \boldsymbol{H}_{S}\boldsymbol{K}_{XS}^{*}\boldsymbol{H}^{*} + \boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} \right| \right] - R_{S} \end{array} \right\}.$$
(21)

$$R_{\text{BF-JD}}^{\prime\prime} = \max_{\substack{\boldsymbol{K}_{X} \succeq 0:\\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} \left| \boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} \left(\boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} + \boldsymbol{I}\right)^{-1} \right| \right].$$
(22)

$$R_{\text{BF-JD}}^{\prime} \leq R_{\text{BF-JD}}^{\text{UB}} = \max_{\substack{\boldsymbol{K}_{X} \geq 0, \rho \in [0,1]:\\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} \min \left\{ \begin{array}{c} \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} \left| \boldsymbol{I} + (1-\rho^{2}) \boldsymbol{H} \boldsymbol{K}_{X} \boldsymbol{H}^{*} \right| \right], \\ \inf_{a>0} \mathbb{E}_{\mathcal{E}_{\mathsf{T}}} \left[\log_{2} \left| \boldsymbol{I} + \left(1+\frac{\rho^{2}}{a}\right) \boldsymbol{H} \boldsymbol{K}_{X} \boldsymbol{H}^{*} + (1+a) \boldsymbol{H}_{S} \boldsymbol{K}_{S} \boldsymbol{H}_{S}^{*} \right| \right] - R_{S} \end{array} \right\}.$$
(23)

with respect to the exact achievable rate (21) of not requiring optimization over the cross-correlation matrix K_{XS} .

Remark 1: With BF-JD, decoding of S at the receiver allows the decoder to cancel the interference H_SS in (18) since the receiver knows the channel H_S . Instead, DPC can only mitigate the effect of the term $\hat{H}_{S,T}S$ in (18) since the channel matrix H_S is not fully known at the transmitter. Therefore, we expect BF-JD to outperform DPC in the scenario with perfect CSIR but imperfect CSIT. Notice that optimizing W, instead of using the naive choice (19), can generally further improve the performance of DPC [8]–[10]³. However, the general conclusion mentioned above remains valid.

C. Imperfect CSIT and CSIR

In this section, we consider the case where we have imperfect CSI at encoder and decoder. Specifically, in (15) we assume that both transmitter and receiver have the same channel estimates $\hat{\mathcal{H}}_{R} = \hat{\mathcal{H}}_{T}$ which we denote as $\hat{\mathcal{H}}$. Matrices $\hat{\mathcal{H}}$ are fixed (deterministic). As per (15), the channel estimation errors at the transmitter and the receiver are the same and we denote $\mathcal{E}_{R} = \mathcal{E}_{T} = \mathcal{E}$, so that

$$\mathcal{H} = \hat{\mathcal{H}} + \mathcal{E}.$$
 (24)

Similar to the previous section, the CSI error matrices E and E_S are assumed to have i.i.d. zero-mean complex Gaussian entries with variances σ_E^2 and $\sigma_{E_S}^2$, respectively. Moreover, errors are assumed to vary in an ergodic fashion. Note that the received signal can be written as

$$\boldsymbol{Y} = \boldsymbol{\hat{H}}\boldsymbol{X} + \boldsymbol{\hat{H}}_{S}\boldsymbol{S} + \boldsymbol{\hat{Z}}, \qquad (25)$$

where $\hat{Z} = EX + E_S S + Z$ is the equivalent noise, which is zero-mean with covariance matrix $(1 + \sigma_E^2 P_X + \sigma_{E_S}^2 P_S)I$. As it was with (18), \hat{Z} is not Gaussian, and is dependent but uncorrelated with X and S.

 3 DPC with a general W is sometimes referred to as Linear-assignment Gel'fand-Pinsker coding (LA-GPC).

1) DPC: An achievable rate via DPC is given as follows. Proposition 2: The following rate is achievable with DPC

5

$$R_{\text{DPC}} = \max_{\substack{\boldsymbol{K}_{X} \succeq 0:\\ tr\{\boldsymbol{K}_{X}\} = P_{X}}} \log_{2} \left| \boldsymbol{I} + \frac{1}{\left(1 + \sigma_{E}^{2} P_{X} + \sigma_{E_{S}}^{2} P_{S}\right)} \hat{\boldsymbol{H}} \boldsymbol{K}_{X} \hat{\boldsymbol{H}}^{*} \right|$$
(26)

Proof: Since the noise \hat{Z} in (25) is not Gaussian, the proof requires special care. Lemma 1 in [23] shows that, while \hat{Z} is not Gaussian or independent of X and S, the same rate is achievable when \hat{Z} is Gaussian and independent of X and S but for a single-antenna channel. In our work, we extend this result to a MIMO channel. Details of the proof are available in Appendix B.

The covariance K_X that maximizes (26) can be easily found by water-filling technique in [16].

2) *BF-JD*: The following proposition gives an achievable rate via BF-JD.

Proposition 3: The rate achievable by BF-JD is lower bounded by

$$R_{\text{BF-JD}} \ge \max\{R'_{\text{BF-JD}}, R''_{\text{BF-JD}}\},$$
(27)

where

$$R'_{\text{BF-JD}} = \max_{\substack{\boldsymbol{K}_X \succeq 0, \boldsymbol{K}_{XS}:\\ tr\{\boldsymbol{K}_X\} = P_X}} \min\left\{R_A^{\text{LB}}, R_B^{\text{LB}} - R_S\right\}$$
(28)

and $R_{\text{BF-ID}}^{\prime\prime}$ in (29). In (28), we have

$$R_A^{\text{LB}} \triangleq \log_2 \left| \boldsymbol{K}_X - \boldsymbol{K}_{XS} \boldsymbol{K}_S^{-1} \boldsymbol{K}_{XS}^* \right| - \log_2 \left| \boldsymbol{K}_L \right| \quad (30)$$

with

$$\begin{split} \boldsymbol{K}_{L} &= \boldsymbol{K}_{X} - \boldsymbol{K}_{XY} \boldsymbol{A}^{*} - \boldsymbol{K}_{XS} \boldsymbol{B}^{*} - \boldsymbol{A} \boldsymbol{K}_{XY}^{*} - \boldsymbol{B} \boldsymbol{K}_{XS}^{*} \\ &+ \boldsymbol{A} \boldsymbol{K}_{YS} \boldsymbol{B}^{*} + \boldsymbol{B} \boldsymbol{K}_{YS}^{*} \boldsymbol{A}^{*} + \boldsymbol{A} \boldsymbol{K}_{Y} \boldsymbol{A}^{*} + \boldsymbol{B} \boldsymbol{K}_{S} \boldsymbol{B}^{*} \\ \boldsymbol{A} &= (\boldsymbol{K}_{XY} - \boldsymbol{K}_{XS} \boldsymbol{K}_{S}^{-1} \boldsymbol{K}_{YS}^{*}) (\boldsymbol{K}_{Y} - \boldsymbol{K}_{YS} \boldsymbol{K}_{S}^{-1} \boldsymbol{K}_{YS}^{*})^{-1} \\ \boldsymbol{B} &= (\boldsymbol{K}_{XS} - \boldsymbol{K}_{XY} \boldsymbol{K}_{Y}^{-1} \boldsymbol{K}_{YS}) (\boldsymbol{K}_{S} - \boldsymbol{K}_{YS}^{*} \boldsymbol{K}_{Y}^{-1} \boldsymbol{K}_{YS})^{-1} \\ \boldsymbol{K}_{Y} &= \hat{\boldsymbol{H}} \boldsymbol{K}_{X} \hat{\boldsymbol{H}}^{*} + \hat{\boldsymbol{H}} \boldsymbol{K}_{XS} \hat{\boldsymbol{H}}^{*}_{S} + \hat{\boldsymbol{H}}_{S} \boldsymbol{K}_{XS}^{*} \hat{\boldsymbol{H}}^{*} + \hat{\boldsymbol{H}}_{S} \boldsymbol{K}_{S} \hat{\boldsymbol{H}}^{*}_{S} \\ &+ (1 + \sigma_{E}^{2} P_{X} + \sigma_{E_{S}}^{2} P_{S}) \boldsymbol{I} \\ \boldsymbol{K}_{XY} &= \boldsymbol{K}_{X} \hat{\boldsymbol{H}}^{*} + \boldsymbol{K}_{XS} \hat{\boldsymbol{H}}^{*}_{S}, \ \boldsymbol{K}_{YS} = \hat{\boldsymbol{H}} \boldsymbol{K}_{XS} + \hat{\boldsymbol{H}}_{S} \boldsymbol{K}_{S}, \end{split}$$

6

IEEE TRANSACTIONS ON COMMUNICATIONS, ACCEPTED FOR PUBLICATION

$$R_{\text{BF-JD}}^{\prime\prime} = \max_{\substack{\boldsymbol{K}_X \succeq 0:\\ tr\{\boldsymbol{K}_X\} = P_X}} \log_2 \left| \boldsymbol{I} + \hat{\boldsymbol{H}} \boldsymbol{K}_X \hat{\boldsymbol{H}}^* \left((1 + \sigma_E^2 P_X + \sigma_{E_S}^2 P_S) \boldsymbol{I} + \hat{\boldsymbol{H}}_S \boldsymbol{K}_S \hat{\boldsymbol{H}}_S^* \right)^{-1} \right|.$$
(29)

and

$$R_B^{\text{LB}} \triangleq \log_2 |\boldsymbol{K}_S| - \log_2 |\boldsymbol{K}_S - \boldsymbol{K}_{YS}^* \boldsymbol{K}_Y^{-1} \boldsymbol{K}_{YS}| + R_A^{\text{LB}}.$$
(31)

Proof: Since the noise \hat{Z} in (25) is not Gaussian, the proof is non-trivial. Our approach is to extend the technique in [20], where a lower bound of mutual information with channel estimation error is derived for a single-antenna system, to the fading MIMO system with interference (1). Details of the proof are available in Appendix C.

Calculating (28) is non-trivial since the maximization over K_X and K_{XS} is non-convex. Finding upper or lower bounds to deal with this is left as an open problem.

Remark 2: With imperfect CSIT and CSIR, both DPC and BF-JD can only attempt to cancel the interference term $\hat{H}_S S$. In particular, BF-JD is able to do so only when the rate R_S is small enough so that the interference is decodable. Therefore, it is expected that DPC outperforms BF-JD unless the rate R_S is small enough so that BF-JD is also able to effectively mitigate the interference.

V. NUMERICAL RESULTS

In this section, we present simulation results for the achievable rates of DPC and BF-JD. We consider a system with t = r = 2 and we fix $K_S = P_S/t_S I$ for simplicity.

A. Perfect CSI

In Fig. 2, we compare the rate (5) achievable by DPC and the bounds (12) on the performance of BF-JD for the case where CSI is perfect at both the transmitter and the receiver and $R_S = 2$ bps/Hz. We also fix $K_X = P_X/tI$ and $K_{XS} = 0$ in (11) for simplicity. This choice is known to be optimal for high SNR and does not affect significantly the performance comparison among different strategies. As discussed in Section III, it is observed that exploiting interference (BF-JD technique) is generally suboptimal with the perfect CSIT and CSIR.

B. Imperfect CSI

We consider now the case of imperfect CSIT and perfect CSIR. We set the error variances as $\sigma^2 = \sigma_{E_T}^2 = \sigma_{E_{S,T}}^2$ and compare the rate (17) for DPC and (20) for BF-JD. For a reference, we also plot the capacity of a channel with zero interference, which is given by $R_{\text{no-interf}} = \max_{K_X \geq 0: tr\{K_X\}=P_X} \mathbb{E}_{\mathcal{E}_T} [\log_2 |I + HK_X H^*|]$. All rates are av- $K_X \geq 0: tr\{K_X\}=P_X \mathbb{E}_{\mathcal{E}_T} [\log_2 |I + HK_X H^*|]$. All rates are averaged over the distribution of the channel estimates \hat{H}_T and $\hat{H}_{S,T}$ which are assumed to have i.i.d. zero-mean complex Gaussian entries with variance $1 - \sigma^2$. For simplicity, we choose $K_X = P_X/tI$ for all the techniques to be compared. Instead, matrix K_{XS} for BF-JD is obtained via global optimization tools (exhaustive search).

Fig. 3 compares the rates of DPC and BF-JD for varying error variance, while Fig. 4 presents the same comparison



Fig. 2. Achievable rate performance of DPC and BF-JD with perfect CSI $(t = r = 2, K_S = P_S/t_S I, R_S = 2 \text{ bps/Hz}).$



Fig. 3. Achievable rate performance of DPC and BF-JD with imperfect CSIT and perfect CSIR versus the estimation error variance σ^2 (t = r = 2, $K_S = P_S/t_S I$, $K_X = P_X/t I$).

versus the interferer's rate R_S . It is seen that, unlike in the perfect CSI case, exploiting the codebook information via BF-JD provides advantages over a large range of values of estimation errors σ^2 , especially if the secondary rate R_S is small enough, for the reason outlined in *Remark 1*. In particular, it is seen that for small interferer's rates R_S , BF-JD is able to achieve the upper bound of no interference. Instead, for large σ^2 , the rate of BF-JD is outperformed by DPC especially when the interferer's rate R_S is large enough. Moreover, our simulations show that optimizing W following the algorithm in [10] does not change the main conclusions above, while improving somewhat the DPC rate shown here with the MMSE solution (19).

We now consider the case of imperfect CSIR and CSIT



Fig. 4. Achievable rate performance of DPC and BF-JD with imperfect CSIT and perfect CSIR versus the interference rate R_S (t = r = 2, $K_S = P_S/t_S I$, $K_X = P_X/tI$).

in Fig. 5 where the achievable rates are plotted versus the estimation error when $\sigma^2 = \sigma_E^2 = \sigma_{E_S}^2$. As above, rates are averaged over channel estimates \hat{H} and \hat{H}_S which have i.i.d. zero-mean complex Gaussian entries with variance $1 - \sigma^2$. As a reference, we plot the rate achievable in case the interfering term that is known at the transmitter and receiver, namely $\hat{H}_S S$, is absent. This is given by $R_{\text{no-interf}} = \max_{K_X \succeq 0: tr\{K_X\}=P_X} \log_2 |I + \hat{H}K_X\hat{H}^*((1 + \sigma^2 P_X)I)^{-1}|$. Also, $K_X = P_X/tI$ is assumed in all the techniques for simplicity. As we noticed in *Remark 2*, BF-JD has performance similar to DPC as R_S decreases. However, for R_S large enough, BF-JD is outperformed by DPC since the latter can perfectly cancel $\hat{H}_S S$ while this is true for BF-JD only if decoding of the interference is successful.

VI. CONCLUSION

Mitigating interference is becoming an increasingly critical task in wireless systems. Interference mitigation can benefit from side information available at the involved terminals and may take different forms. In this paper, we have studied a MIMO system affected by a multiantenna interferer, where the transmitter is informed about the interferer's codeword. The key design issue is whether the decoder should attempt decoding of the interference or rather the encoder should precode over the interference by using DPC strategies and the decoder treats the interference as unstructured. While with perfect CSI, DPC strategies are known to be optimal, we have shown that with imperfect CSI this is not the case, especially in the common case where the CSI is perfect at the receiver but imperfect at the transmitter.



Fig. 5. Achievable rate performance of DPC and BF-JD with imperfect CSIT and CSIR versus the estimation error variance σ^2 (t = r = 2, $K_S = P_S/t_S I$, $K_X = P_X/tI$).

VII. APPENDIX A: PROOF OF PROPOSITION 1

We upper bound the two terms in (11) as follows. For the first term, we have

$$\log_{2} \left| \boldsymbol{I} + \boldsymbol{H} \left(\boldsymbol{K}_{X} - \boldsymbol{K}_{XS} \boldsymbol{K}_{S}^{-1} \boldsymbol{K}_{XS}^{*} \right) \boldsymbol{H}^{*} \right|$$

$$= \log_{2} \left| \boldsymbol{I} + \boldsymbol{H} \boldsymbol{K}_{X}^{1/2} \left(\boldsymbol{I} - \boldsymbol{A} \boldsymbol{A}^{*} \right) \boldsymbol{K}_{X}^{1/2} \boldsymbol{H}^{*} \right|$$
^(a)
^(a)
^(a)
^(a)
^(b)
^(c)

where $\mathbf{A} = \mathbf{K}_X^{-1/2} \mathbf{K}_{XS} \mathbf{K}_S^{-1/2}$ and $\rho \in [0, 1]$ is a correlation coefficient. Note that (a) is a consequence of the inequality $\mathbf{I} - \mathbf{A}\mathbf{A}^* \leq (1 - \rho^2)\mathbf{I}$ for some $0 \leq \rho \leq 1$ (where \leq represents the inequality with respect to the cone of semi-positive definite matrices), which in turns follows from the fact that $\mathbf{0} \leq \mathbf{I} - \mathbf{A}\mathbf{A}^* \leq \mathbf{I}$ and from the continuity of matrix \mathbf{A} in the vector space (see [18] for details). Moreover, for the second term in (11) we have (33) where (b) is from *Lemma 3.2* in [18].

VIII. APPENDIX B: PROOF OF PROPOSITION 2

Recall from (25), \mathbf{Z} has of zero mean and covariance matrix $(1 + \sigma_E^2 P_X + \sigma_{E_S}^2 P_S)\mathbf{I}$. As explained in Section IV-C, it is not Gaussian and is uncorrelated, but statistically dependent on \mathbf{X} and \mathbf{S} . Let the singular value decomposition (SVD) of matrix $\hat{\mathbf{H}}$ be $\hat{\mathbf{H}} = U\Sigma V^*$, where U is $r \times p$, V is $t \times p$, and Σ is diagonal $p \times p$ where p is the rank of matrix $\hat{\mathbf{H}}$. To convert the MIMO channel $\hat{\mathbf{H}}$ into p-parallel single-antenna channels, the transmitter chooses $\mathbf{X} = V\bar{\mathbf{X}}$ where $\bar{\mathbf{X}}$ is $p \times 1$ vector of transmitted signals and the receiver multiplies U^* to the received signal as

$$\bar{\boldsymbol{Y}} = \boldsymbol{U}^* \left(\boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^* \boldsymbol{V} \bar{\boldsymbol{X}} + \hat{\boldsymbol{H}}_S \boldsymbol{S} + \hat{\boldsymbol{Z}} \right)$$

$$= \boldsymbol{\Sigma} \bar{\boldsymbol{X}} + \boldsymbol{U}^* \hat{\boldsymbol{H}}_S \boldsymbol{S} + \boldsymbol{U}^* \hat{\boldsymbol{Z}}.$$
(34)

Therefore, each element of $\bar{\boldsymbol{Y}} = [\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_p]$ is

$$\bar{Y}_i = \bar{H}_i \bar{X}_i + \bar{H}_{S,i} S_i + \bar{Z}_i, \qquad (35)$$

where $\bar{H}_i = (\Sigma)_{ii}$, $\bar{H}_{S,i} = U_i^* \hat{H}_S$, $\bar{Z}_i = U_i^* \hat{Z}$, and U_i is the *i*-th column of U for $i = 1, \ldots, p$. Since the model

8

IEEE TRANSACTIONS ON COMMUNICATIONS, ACCEPTED FOR PUBLICATION

$$\log_{2} \left| \boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + \boldsymbol{H}_{S}\boldsymbol{K}_{XS}^{*}\boldsymbol{H}^{*} + \boldsymbol{H}\boldsymbol{K}_{XS}\boldsymbol{H}_{S}^{*} + \boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} \right|$$

$$\stackrel{\text{(b)}}{\leq} \inf_{a>0} \log_{2} \left| \boldsymbol{I} + \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + \boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} + \frac{\rho^{2}}{a}\boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + a\boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} \right|$$

$$= \inf_{a>0} \log_{2} \left| \boldsymbol{I} + \left(1 + \frac{\rho^{2}}{a} \right) \boldsymbol{H}\boldsymbol{K}_{X}\boldsymbol{H}^{*} + (1 + a)\boldsymbol{H}_{S}\boldsymbol{K}_{S}\boldsymbol{H}_{S}^{*} \right|.$$
(33)

(35) is a single-antenna channel, from Lemma 1 in [23], the rate $\log_2\left(1 + \frac{\bar{H}_i^2 P_i}{1 + \mathbb{E}[\bar{Z}_i \bar{Z}_i^*]}\right)$ is achievable on each subchannel *i* where P_i is the power allocated to \bar{X}_i , which has to satisfy $\sum_{i=1}^p P_i \leq P_X$. Hence, we have that the rate in (26) is achievable.

IX. APPENDIX C: PROOF OF PROPOSITION 3

We bound the two terms in (6) and system model (25). Starting with the first, we have

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) = h(\boldsymbol{X}|\boldsymbol{S}) - h(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}), \quad (36)$$

where $h(\boldsymbol{X}|\boldsymbol{S}) = \log_2((\pi e)^r |\boldsymbol{K}_{X|S}|)$ with $\boldsymbol{K}_{X|S} = \boldsymbol{K}_X - \boldsymbol{K}_{XS}\boldsymbol{K}_S^{-1}\boldsymbol{K}_{XS}^*$ and, similar to [20], we have

$$h(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}) = h(\boldsymbol{X} - \boldsymbol{A}\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{S}|\boldsymbol{Y},\boldsymbol{S})$$

$$\stackrel{(a)}{\leq} h(\boldsymbol{X} - \boldsymbol{A}\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{S})$$

$$\stackrel{(b)}{\leq} \log_2\left((\pi e)^r |\boldsymbol{K}_L|\right), \quad (37)$$

where (a) is from the fact conditioning reduces entropy and $A, B \in \mathbb{C}^{t \times r}$ are given matrices, and $K_L = K_X - K_{XY}A^* - K_{XS}B^* - AK_{XY}^* - BK_{XS}^* + AK_{YS}B^* + BK_{YS}^*A^* + AK_YA^* + BK_SB^*$. K_{XY}, K_{YS} , and K_Y are as defined after (30); and (b) follows from the maximum entropy theorem. Setting C = [A B] and $\tilde{Y} = [Y S]^T$, (37) can be minimized by choosing C as $C = K_{X\tilde{Y}}K_{\tilde{Y}}^{-1}$ where $K_{X\tilde{Y}}$ and K_Y are the covariance matrices of the corresponding vectors defined above, namely $K_{X\tilde{Y}} = [K_{XY} K_{XS}]$ and

$$\boldsymbol{K}_{\tilde{Y}} = \begin{bmatrix} \boldsymbol{K}_{Y} & \boldsymbol{K}_{YS} \\ \boldsymbol{K}_{YS}^{*} & \boldsymbol{K}_{S} \end{bmatrix}.$$
 (38)

In fact, this way, vector (AY + BS) is the linear MMSE estimate of X in terms of Y and S. From the matrix inversion lemma [24], A and B can be found as

$$A = (K_{XY} - K_{XS}K_S^{-1}K_{YS}^*)(K_Y - K_{YS}K_S^{-1}K_{YS}^*)^{-1}$$

and $B = (K_{XS} - K_{XY}K_Y^{-1}K_{YS})(K_S - K_{YS}^*K_Y^{-1}K_{YS})^{-1}$

Overall, we can bound the first term in (6) as

Ι

$$(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) \ge \log_2 |\boldsymbol{K}_{X|S}| - \log_2 |\boldsymbol{K}_L| = R_A^{LB}.$$
 (39)

For the second term in (6), we proceed as follows,

$$I(\boldsymbol{X}, \boldsymbol{S}; \boldsymbol{Y}) = h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{S})$$

= $h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{S}) + I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{S})$
= $h(\boldsymbol{S}) - h(\boldsymbol{S}|\boldsymbol{Y}) + R_A^{\text{LB}},$ (40)

where $h(\mathbf{S}) = \log_2 ((\pi e)^r |\mathbf{K}_S|), h(\mathbf{S}|\mathbf{Y})$ is upper bounded as $h(\mathbf{S}|\mathbf{Y}) \leq \log_2 ((\pi e)^r |\mathbf{K}_S|_Y|)$ where $\mathbf{K}_{S|Y} = \mathbf{K}_S - \mathbf{K}_S$ $K_{YS}^* K_Y^{-1} K_{YS}$ and I(X; Y|S) satisfies (36). Therefore, we obtain

$$I(\boldsymbol{X}, \boldsymbol{S}; \boldsymbol{Y}) \ge \log_2 |\boldsymbol{K}_S| - \log_2 |\boldsymbol{K}_S - \boldsymbol{K}_{YS}^* \boldsymbol{K}_Y^{-1} \boldsymbol{K}_{YS}| + R_A^{\text{LB}}$$
(41)

which concludes the proof.

REFERENCES

- [1] G. Caire, A. Adhikary, and V. Ntranos, "Cognitive femtocells: breaking the spatial reuse barrier of cellular systems," 2011 ITA Workshop.
- [2] I. Marić, N. Liu, and A. Goldsmith, "Encoding against an interferer's codebook," in *Proc. 2008 Allerton Conf. Commun. Control Comput.*, pp. 523–530.
- [3] O. Simeone, E. Erkip, and S. S. (Shitz), "On exploiting the interference structure for reliable communications," in *Proc. 2010 Conf. Inf. Sci. Syst.*
- [4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [5] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [6] R. Zamir, S. S. (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.
- [7] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.
- [8] C. S. Vaze and M. K. Varanasi, "Dirty paper coding for fading channels with partial transmitter side information," in *Proc. 2008 Asilomar Conf. Signal, Syst. Comput.*, pp. 341–345.
- [9] A. Bennatan and D. Burshtein, "On the fading-paper achievable region of the fading MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 100–115, Jan. 2008.
- [10] C. S. Vaze and M. K. Varanasi, "Dirty paper coding for the MIMO cognitive radio channel with imperfect CSIT," in *Proc. 2009 IEEE Int. Symp. Inf. Theory*, pp. 2532–2536.
- [11] W. Zhang, S. Kotagiri, and J. N. Laneman, "Writing on dirty paper with resizing and its application to quasi-static fading broadcast channels," in *Proc. 2007 IEEE Int. Symp. Inf. Theory*, pp. 381–385.
- [12] P. Mitran, N. Devroye, and V. Tarokh, "On compound channels with side information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1745–1755, Apr. 2006.
- [13] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 572–584, Sep. 1979.
- [14] F. Baccelli, A. E. Gamal, and D. Tse, "Interference networks with pointto-point codes," Feb. 2011. Available: http://arxiv.org/abs/1102.2868
- [15] A. E. Gamal and Y.-H. Kim, "Lecture note on network information theory," June 2010. Available: http://arxiv.org/abs/1001.3404
- [16] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T-BELL Lab. Internal Tech. Memo, June 1995.
- [17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [18] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.
- [19] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [20] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [21] A. D. Dabbagh and D. J. Love, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, Nov. 2008.

LEE et al.: THE EFFECT OF IMPERFECT CHANNEL KNOWLEDGE ON A MIMO SYSTEM WITH INTERFERENCE

- [22] S.-C. Lin, P.-H. Lin, C.-P. Lee, and H.-J. Su, "Filter and nested-lattice code design for fading MIMO channels with side-information," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1489–1494, June 2011.
- [23] S. Yang and J. Belfiore, "The impact of channel estimation error on the DPC region of the two-user Gaussian broadcast channel," in *Proc. 2005 Allerton Conf. Commun. Control Comput.*, pp. 1366–1372.
- [24] D. Bernstein, Matrix Mathematics. Princeton University Press, 2005.



Namjeong Lee (S'09) received her M.S.E. (with honors), and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST) at Daejeon, Korea, in 2007 and 2012, respectively. She is currently working as a research staff member of Samsung Advanced Institute of Technology (SAIT) at Gyeonggi-do, Korea. She was a visiting scholar in the Center for Wireless Communications and Signal Processing Research (CWCSPR) at the New Jersey Institute of Technology (NJIT), Newark, New Jersey, from 2009 to 2010. Her research interest includes

information-theoretic aspects of MIMO channels, cooperative communications, and green communications.



Osvaldo Simeone received the M.Sc. degree (with honors) and the Ph.D. degree in information engineering from Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively. He is currently with the Center for Wireless Communications and Signal Processing Research (CWCSPR), New Jersey Institute of Technology (NJIT), Newark, where he is an Assistant Professor. His current research interests concern the cross-layer analysis and design of wireless networks with emphasis on information-theoretic, signal processing, and queuing aspects.

Specific topics of interest are: cognitive radio, cooperative communications, rate-distortion theory, ad hoc, sensor, mesh and hybrid networks, distributed estimation, and synchronization. Dr. Simeone is a co-recipient of Best Paper Awards of the IEEE SPAWC 2007 and IEEE WRECOM 2007. He currently serves as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS.



Joonhyuk Kang was born in Seoul, KOREA. He received his B.S.E. and M.S.E. degrees from Seoul National University, Seoul, Korea, in 1991 and 1993, respectively. He earned his Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2002. He is currently working as a faculty member of Korea Advanced Institute of Science and Technology (KAIST) at Daejeon, Korea.

From 1993 to 1998, he was a research staff at SAMSUNG Electronics (Suwon, Korea), where he

involved in the development of DSP-based real-time control systems. In 2000, he worked at Cwill Telecommunications (Austin, Texas, U.S.A.), where he participated in the project for multi-carrier CDMA systems with antenna array. He was a visiting scholar in the school of engineering and applied sciences at Harvard University (Cambridge, Massachusetts, U.S.A.) from 2008 to 2009. His research interest includes signal processing for cognitive radio, cooperative communication, physical-layer security, and wireless localization. He is a member of IEEE, KICS (Korea Information and Communications Society), and Tau Beta Pi (The Engineering Honor Society) and a recipient of TxTEC (Texas Telecommunication Consortium) Graduate Fellowship from 2000 to 2002.