

On the Superiority of Improper Gaussian Signaling in Wireless Interference MIMO Scenarios

Sandra Lagen, Adrian Agustin, *Member, IEEE*, and Josep Vidal, *Member, IEEE*

Abstract—Recent results have elucidated the benefits of using improper Gaussian signaling (IGS) as compared to conventional proper Gaussian signaling (PGS) in terms of achievable rate for interference-limited conditions. This paper exploits majorization theory tools to formally quantify the gains of IGS along with widely linear transceivers for MIMO systems in interference-limited scenarios. The MIMO point-to-point channel with interference (P2P-I) is analyzed, assuming that received interference can be either proper or improper, and we demonstrate that the use of the optimal IGS when received interference is improper strictly outperforms (in terms of achievable rate and mean square error) the use of the optimal PGS when interference is proper. Then, these results are extended to two practical situations. First, the MIMO Z-interference channel (Z-IC) is investigated, where a trade-off arises: with IGS we could increase the achievable rate of the interfered user while gracefully degrading the rate of the non-interfered user. Second, these concepts are applied to a two-tier heterogeneous cellular network (HCN) where macrocells and smallcells coexist and multiple MIMO Z-IC appear.

Index Terms—Improper Gaussian signaling, widely linear processing, majorization theory, MIMO point-to-point channel with interference, MIMO Z-interference channel, heterogeneous cellular networks.

I. INTRODUCTION

In order to address the massive data demand in wireless networks, today's most advanced cellular networks (as LTE-A [1]) consider the densification of the network with base stations and the spatial re-use of the spectrum [2]. In both cases interference coordination becomes a key enabler. Interference management in wireless communication systems is a challenging task that has been receiving a lot of attention during many years. It is well known that the statistical characteristics of the signals affect the maximum achievable rate of the communication. In this sense, recently, it has been shown that *improper Gaussian signaling* (IGS) is able to improve the achievable rate in interference-limited scenarios [3][4][5].

The key difference between IGS and conventional proper Gaussian signaling (PGS) is the characterization of the second order statistics: while PGS is fully specified by the covariance matrices under the zero-mean assumption, IGS is characterized

not only by the covariance matrices but also by the pseudo-covariance matrices, i.e. the improper Gaussian random vector is correlated with its complex conjugate [6][7].

Conventional PGS has been shown to be optimal in terms of capacity for the Gaussian multiple-input multiple-output (MIMO) point-to-point channel (P2P) [8], and also for the Gaussian MIMO broadcast channel (BC) with dirty paper coding as the capacity achieving strategy [9][10]. For that reason, PGS is usually also assumed in multi-user wireless communications. However, optimality of PGS does not necessarily hold for MIMO BC if transceivers are restricted to be (widely) linear [11]. It is shown there that capacity gains can be obtained by employing IGS because, in the absence of non-linear interference cancellation, the MIMO BC becomes interference-limited [12]. The same applies in the Gaussian MIMO interference channel (IC), where IGS has been shown to: *i*) obtain larger degrees of freedom, i.e. slope of the sum-rate as a function of the signal-to-noise ratio (SNR) at the high SNR regime, [3][5][13] and *ii*) improve the known achievable rates when interference is treated as noise under certain MIMO configurations [4][14][15] and under coexistence of linear and widely linear transceivers in MIMO systems [16][17]. In addition, the positive benefits of IGS have been reported in terms of achievable rates [18] and outage probabilities [19] for underlay cognitive radio systems.

In order to generate and estimate IGS with linear processing, *widely linear precoding* (WLP) [4] and *widely linear estimation* (WLE) [6] need to be adopted at transmitter and receiver sides, respectively. WLE is a generalized concept used in estimation theory whenever improper noise is encountered [20] and in systems that transmit improper signal constellations [21], as it has been already under investigation in 3GPP LTE-A [22]. On the other hand, WLP can be used to generate IGS even when departing from proper signal constellations [11]. Differently from WLE that has been extensively studied in literature, the generation of IGS through WLP to handle interference more effectively is a recent research line.

In this work we focus on gaining more insights on the use of IGS along with widely linear transceivers for MIMO systems in interference-limited conditions. To do so we use the equivalent *composite real representation* [8], whereby real and imaginary parts of the MIMO channel are separated, and exploit majorization theory [23][24] to formally quantify the benefits. We analyze the MIMO P2P channel with interference (MIMO P2P-I), in which a transmitter (TX) equipped with M antennas wish to send information to a receiver (RX) equipped with N antennas that is receiving noise-plus-interference. The optimal signaling (PGS or IGS) and the associated trans-

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. Manuscript received December 2, 2015; revised April 6, 2016, June 7, 2016; accepted June 20, 2016. S. Lagen, A. Agustin and J. Vidal are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain. e-mails: {sandra.lagen, adrian.agustin, josep.vidal}@upc.edu. This work is supported by projects TEC2013-41315-R DISNET ('Ministerio de Economía y Competitividad' and ERDF funds, Spanish Government), 2014SGR-60 (Catalan Government), and FPU12/00828 grant ('Ministerio de Educación, Cultura y Deporte', Spanish Government).

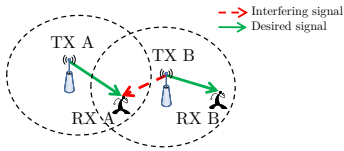


Fig. 1: Example of the MIMO Z-Interference Channel.

mit/receive scheme (linear or widely linear) are well known in the literature and can be obtained from the minimization of any Schur-concave function of the MSE-matrix [25], like maximization of the achievable rate or minimization of the mean square error (MSE) of the transmitted symbols. Their solutions are determined by the statistics of the interference, which can be modeled either as a PGS or IGS depending on the transmission strategy of the interferer: if interference is proper then the optimal signaling is PGS, while if interference is improper then the optimal signaling is IGS. In this sense, we focus on the following question: is it better to use IGS (and improper interference) or to use PGS (and proper interference) in the MIMO P2P-I? By using majorization theory tools [23], we demonstrate that the use of IGS allows obtaining a strict improvement in terms of achievable rate and MSE for $M \geq N$ and full rank channel matrix. The proof sheds light on why the use of IGS is better. Then, we investigate how to exploit these properties in different interference-limited scenarios.

On the one hand, we study the MIMO Z-interference channel (Z-IC) [26], a two-transmitter two-receiver multi-antenna interfering scenario where one of the receivers (RXB) just observes noise while the other (RXA) receives interference. The MIMO Z-IC is present in certain situations in cellular networks, as shown in Fig. 1 where RXA is in the cell-edge. Under the condition of treating interference as noise, the optimal transmission scheme for sum-rate maximization in the Z-IC is only known for the SISO case [27], for which a set of five possible solutions that subsume PGS and IGS depending on the channel conditions is obtained. The MIMO Z-IC is interesting because TXB should apply conventional PGS for optimally transmitting to its intended receiver (RXB) but, according to the properties derived in the sequel for the MIMO P2P-I, it will be beneficial for RXA that TXB employs IGS such that the interference injected onto RXA is improper and TXA can design WLP to improve the performance of RXA (see Fig. 1). Clearly this comes at the cost of reducing the performance of RXB, such that a clear trade-off in the rates of the two receivers arises. In this regard, we analyze the trade-off and propose a simple improper-based scheme for TXB that allows improving the fairness and controlling the sum-rate performance.

On the other hand, we apply the aforementioned properties to heterogeneous cellular networks (HCNs) [2], a promising deployment for future cellular systems. HCNs consist of a multi-tier deployment of macro eNode Bs (MeNBs) and small eNode Bs (SeNBs) that allow boosting the spectral efficiency of the system thanks to the network densification and the spatial re-use of the spectrum among MeNBs and SeNBs. However, HCNs have to deal with a major impairment: the *cross-tier interference*, which can be very detrimental specially in the direction from MeNBs towards the users served by

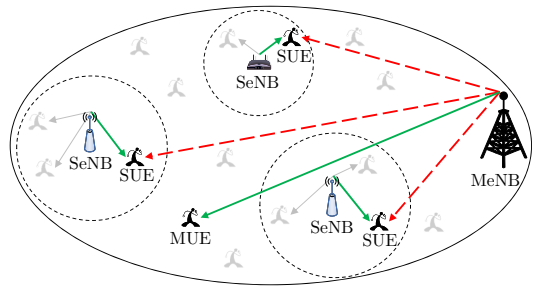


Fig. 2: Predominant cross-tier interference (red dashed lines) in HCNs that lead to multiple MIMO Z-IC. Each SeNB generates a different MIMO Z-IC, which are all coupled by the MeNB.

SeNBs with a lower transmit power [1]¹, as illustrated in Fig. 2. Interestingly, the interfering channel that models the major cross-tier interference impairment in HCNs is the MIMO Z-IC. Furthermore, there will appear as many MIMO Z-IC as the number of SeNBs deployed within the MeNB coverage area, see Fig. 2. Hence, the derived improper-based scheme for the MIMO Z-IC could be applied at the MeNB (corresponding to TXB in Fig. 1) and a performance improvement of all the users served by SeNBs would be guaranteed thanks to the properties derived for the MIMO P2P-I.

To summarize, the main contributions of this work are:

- In the MIMO P2P-I with $M \geq N$, when uniform power allocation (UPA) is used, we demonstrate that the use of the optimal IGS when received interference is improper strictly outperforms in terms of achievable rate and MSE the use of the optimal PGS when interference is proper. If optimal power allocation (OPA) is adopted, then the achievable rate is shown to be equal or larger with IGS, and the strict superiority is shown for a particular case.
- In the MIMO Z-IC, we propose a simple improper-based scheme without any claim about optimality that allows improving the achievable rate of the most impaired user and controlling the rate loss of the non-interfered user. The scheme is characterized by a parameter that provides two extreme solutions: PGS or maximal IGS. If such parameter is fixed (e.g. maximal IGS is used), then its implementation does not require knowledge of the interfering channel either at transmitters or receivers.
- The proposed improper-based scheme is evaluated in an HCN deployment compliant with 3GPP LTE-A specifications [29]. Simulations show significant gains in terms of 5%-tile and mean user throughput as compared to conventional interference coordination techniques.

This paper generalizes our results presented in [30], where superiority (but not strict) of IGS in the MIMO P2P-I was demonstrated in terms of achievable rate for the UPA case. Here, strict superiority is formally proven in terms of achievable rate and MSE for UPA. Also, superiority is shown in terms of achievable rate for OPA. Further, the derived properties are generalized and applied to different interference-limited scenarios, including HCNs.

¹Note that if SeNBs were deployed in concentrated areas then the cross-tier interference would become relevant and additional techniques would be required, see [28].

Organization: The rest of the paper is organized as follows. In Section II the use of IGS for the MIMO P2P-I, as well as the optimal signaling and the associated transmission scheme, are presented. In Section III, the superiority of IGS in the MIMO P2P-I is demonstrated in terms of achievable rate and MSE for UPA and in terms of achievable rate for OPA. In Section IV, a practical improper-based scheme for the MIMO Z-IC is proposed and evaluated. In Section V, the proposed scheme is applied and evaluated in HCN deployments. Finally, concluding remarks are included in Section VI.

Appendix A-A presents preliminaries for improper random vectors. Preliminaries for majorization theory are included in Appendix A-B.

Notation: In this paper, scalars are denoted by italic letters. Boldface lower-case and upper-case letters denote vectors and matrices, respectively. For given scalars a and b , $\min(a, b)$, $(a)^+$, $\log_2(a)$, and $\ln(a)$, denote the minimum between a and b , the maximum between a and 0, the base-2 logarithm, and the natural logarithm, respectively. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ refer to the real and imaginary operators, respectively. For a given vector \mathbf{a} , $\bar{\mathbf{a}}$ denotes a double-sized real-valued vector that stacks its real and imaginary parts as: $\bar{\mathbf{a}} = [\Re\{\mathbf{a}\}^T \Im\{\mathbf{a}\}^T]^T$, and $\|\mathbf{a}\|^2$ refers to the squared 2-norm of vector \mathbf{a} . For given vectors \mathbf{a} and \mathbf{b} , $\mathbf{a} \circ \mathbf{b}$ denotes the Hadamard product (i.e. component-wise product). For a given matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^* , \mathbf{A}^H , and \mathbf{A}^{-1} denote the transpose matrix, the conjugate matrix, the hermitian matrix, and the inverse matrix, respectively. The operators $|\mathbf{A}|$, $\text{Tr}(\mathbf{A})$, $\mathbb{E}\{\mathbf{A}\}$ refer to the determinant, the trace, and the expectation, respectively. $\text{rank}(\mathbf{A})$ denotes the rank of matrix \mathbf{A} . Matrices \mathbf{I}_n and \mathbf{J}_n denote the identity matrix and the anti-identity matrix of size $n \times n$, respectively. $\mathbf{a} = \text{eig}(\mathbf{A})$ refers to a vector that stacks the eigenvalues of matrix \mathbf{A} in decreasing order. \mathbf{a}^{-1} refers to a vector containing the inverse of each element of vector \mathbf{a} , in decreasing order. $\mathbf{A} = \text{diag}(a_1 \dots a_n)$ denotes a $n \times n$ diagonal matrix with values $a_1 \dots a_n$ in its diagonal and 0 in the non-diagonal elements. $\mathbf{a} = \text{diag}(\mathbf{A})$ refers to a vector containing the diagonal values of matrix \mathbf{A} . $\mathbf{a} \succ \mathbf{b}$, $\mathbf{a} \succ_w \mathbf{b}$, and $\mathbf{a} \succ_{w \log} \mathbf{b}$ denote strong majorization, weak majorization, and weak log-majorization, respectively (see definitions in Appendix A-B). $\mathbb{C}^{m \times n}$, $\mathbb{R}^{m \times n}$, and $\mathbb{R}_+^{m \times n}$ denote an m by n dimensional complex space, real space, and real positive space, respectively. The circularly symmetric complex normal distribution is represented by $\mathcal{CN}(\cdot, \cdot)$.

II. SYSTEM MODEL FOR THE MIMO P2P-I

Consider a MIMO P2P-I between one transmitter equipped with M antennas and one receiver with N receive antenna elements, as shown in Fig. 3. The MIMO channel is described by a matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$ containing complex-valued channel gains of the different antenna-pairs. In the following we assume $M \geq N$ and full rank channel matrix \mathbf{H} . Hence, under narrow-band transmissions, the equivalent baseband signal observed at the receiver is expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{s}, \quad \mathbf{s} = \mathbf{i} + \mathbf{n}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the complex-valued transmitted vector and $\mathbf{s} \in \mathbb{C}^{N \times 1}$ denotes the noise-plus-interference vector at the receiver (see Fig. 3), which contains a proper Gaussian noise

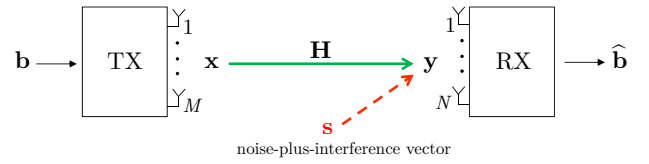


Fig. 3: MIMO Point-to-Point channel with Interference.

$\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ and a Gaussian interference component $\mathbf{i} \in \mathbb{C}^{N \times 1}$ that is caused by an interfering transmitter. Different from the conventional transmission setup where PGS is assumed (i.e. $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_x)$), in this paper the more general IGS is adopted for \mathbf{x} .

Any improper Gaussian signal \mathbf{x} can be generated from a proper Gaussian information-bearing signal $\mathbf{b} \in \mathbb{C}^{N \times 1}$ with unitary power (i.e. $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$) through the use of *widely linear precoding* (WLP) (see Lemma 5 in Appendix A-A), i.e.:

$$\mathbf{x} = \mathbf{W}_1 \mathbf{b} + \mathbf{W}_2 \mathbf{b}^*, \quad (2)$$

where matrices \mathbf{W}_1 and $\mathbf{W}_2 \in \mathbb{C}^{M \times N}$ denote the linear transmit precoders for the information-bearing signal \mathbf{b} and its complex conjugate \mathbf{b}^* . The conventional *linear precoding* (LP) scheme is a special case of WLP in which $\mathbf{W}_2 = \mathbf{0}$ such that \mathbf{x} in (2) is PGS. It is assumed that N parallel data streams are transmitted, although some of the streams can have a rate of zero. Recall that $\mathbb{E}\{\mathbf{b}\mathbf{b}^H\} = \mathbf{I}_N$ and $\mathbb{E}\{\mathbf{b}\mathbf{b}^T\} = \mathbf{0}$. Then, the covariance matrix $\mathbf{C}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ and the pseudo-covariance matrix $\tilde{\mathbf{C}}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^T\}$ of \mathbf{x} in (2) are:

$$\mathbf{C}_x = \mathbf{W}_1 \mathbf{W}_1^H + \mathbf{W}_2 \mathbf{W}_2^H, \quad \tilde{\mathbf{C}}_x = \mathbf{W}_1 \mathbf{W}_2^T + \mathbf{W}_2 \mathbf{W}_1^T. \quad (3)$$

Similarly, as the Gaussian interference component \mathbf{i} in (1) can be either proper or improper depending on the transmission scheme adopted by the interferer, we denote by $\mathbf{C}_s = \mathbb{E}\{\mathbf{s}\mathbf{s}^H\}$ and $\tilde{\mathbf{C}}_s = \mathbb{E}\{\mathbf{s}\mathbf{s}^T\}$ the covariance matrix and the pseudo-covariance matrix of $\mathbf{s} = \mathbf{i} + \mathbf{n}$ in (1), respectively.

A. Composite real representation

The signal model for the MIMO P2P-I can be equivalently expressed using the *composite real representation* [8], in which real and imaginary parts of the signals are separated. Such representation is useful to derive the optimal signaling and the associated transmission scheme in (2). The input-output relation in (1) can be equivalently written as:

$$\bar{\mathbf{y}} = \begin{bmatrix} \Re\{\mathbf{y}\} \\ \Im\{\mathbf{y}\} \end{bmatrix} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{s}}, \quad \bar{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}, \quad (4)$$

where the double-sized real-valued transmitted signal $\bar{\mathbf{x}}$ in (4) (following WLP in (2)) is given by:

$$\bar{\mathbf{x}} = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} = \bar{\mathbf{W}}\bar{\mathbf{b}}, \quad \bar{\mathbf{W}} = \begin{bmatrix} \Re\{\mathbf{W}_1 + \mathbf{W}_2\} & -\Im\{\mathbf{W}_1 - \mathbf{W}_2\} \\ \Im\{\mathbf{W}_1 + \mathbf{W}_2\} & \Re\{\mathbf{W}_1 - \mathbf{W}_2\} \end{bmatrix}. \quad (5)$$

The achievable rate (R) with IGS is obtained as [7]:

$$\begin{aligned} R &= I(\mathbf{x}; \mathbf{y}) = I(\bar{\mathbf{x}}; \bar{\mathbf{y}}) = h(\bar{\mathbf{y}}) - h(\bar{\mathbf{y}}/\bar{\mathbf{x}}) = h(\bar{\mathbf{y}}) - h(\bar{\mathbf{s}}) \\ &= \frac{1}{2} \log_2 |\mathbf{I}_{2N} + \frac{1}{2} \bar{\mathbf{H}} \bar{\mathbf{W}} \bar{\mathbf{W}}^T \bar{\mathbf{H}}^T \mathbf{C}_s^{-1}|, \end{aligned} \quad (6)$$

where $I(\bar{\mathbf{x}}; \bar{\mathbf{y}})$ denotes the mutual information between real-valued random vectors $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, $h(\bar{\mathbf{x}})$ refers to the entropy of $\bar{\mathbf{x}}$ [7], and the $\frac{1}{2}$ factor inside the determinant comes from

$\mathbb{E}\{\bar{\mathbf{b}}\bar{\mathbf{b}}^T\} = \frac{1}{2}\mathbf{I}$. $\mathbf{C}_{\bar{\mathbf{s}}} \in \mathbb{R}^{2N \times 2N}$ accounts for the covariance matrix of $\bar{\mathbf{s}}$ in (4), which can be written in terms of $\mathbf{C}_{\mathbf{s}}$ and $\tilde{\mathbf{C}}_{\mathbf{s}}$ through a transformation matrix $\mathbf{T} \in \mathbb{C}^{2N \times 2N}$ as follows (see Lemma 4 in Appendix A-A):

$$\mathbf{C}_{\bar{\mathbf{s}}} = \mathbb{E}\{\bar{\mathbf{s}}\bar{\mathbf{s}}^T\} = \mathbf{T} \begin{bmatrix} \mathbf{C}_{\mathbf{s}} & \tilde{\mathbf{C}}_{\mathbf{s}} \\ \tilde{\mathbf{C}}_{\mathbf{s}}^* & \mathbf{C}_{\mathbf{s}}^* \end{bmatrix} \mathbf{T}^H, \quad \mathbf{T} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_N & \mathbf{I}_N \\ -j\mathbf{I}_N & j\mathbf{I}_N \end{bmatrix}. \quad (7)$$

The mean square error (MSE) of the transmitted symbols is expressed as: $\epsilon = \text{Tr}(\bar{\mathbf{E}})$, being $\bar{\mathbf{E}} = \mathbb{E}\{(\bar{\mathbf{b}} - \hat{\bar{\mathbf{b}}})(\bar{\mathbf{b}} - \hat{\bar{\mathbf{b}}})^T\}$ the MSE-matrix and $\hat{\bar{\mathbf{b}}} \in \mathbb{R}^{2N \times 1}$ the double-sized real-valued vector of the estimated information-bearing signal at the receiver side. Assuming optimal *widely linear estimation* (see details in [6]), the MSE-matrix results:

$$\bar{\mathbf{E}} = \frac{1}{2} \left(\mathbf{I}_{2N} + \frac{1}{2} \bar{\mathbf{W}}^T \bar{\mathbf{H}}^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \bar{\mathbf{H}} \bar{\mathbf{W}} \right)^{-1}. \quad (8)$$

Therefore, the achievable rate R in (6) is related to the MSE-matrix $\bar{\mathbf{E}}$ in (8) through [30]:

$$R = -\frac{1}{2} \log_2 |2\bar{\mathbf{E}}|. \quad (9)$$

B. MSE-based optimal scheme

Let us present the optimal design for transmit precoder $\bar{\mathbf{W}}$ in (5), a result already known in the literature that is needed to demonstrate the superiority of IGS in next Section III. In the most general MSE-based designs (including minimum MSE and maximum achievable rate designs) [25][31], the optimal transmit precoder $\bar{\mathbf{W}}$ in (5) is obtained from the minimization of a Schur-concave function of the MSE-matrix $\bar{\mathbf{E}}$ in (8) subject to a maximum transmit power constraint:

$$\underset{\bar{\mathbf{W}}}{\text{minimize}} \quad f(\bar{\mathbf{E}}) \quad \text{subject to} \quad \frac{1}{2} \text{Tr}(\bar{\mathbf{W}}\bar{\mathbf{W}}^T) = P^{\max}, \quad (10)$$

where $f(\cdot)$ is any Schur-concave function of $\bar{\mathbf{E}}$ in (8) and P^{\max} is the maximum available power at the transmitter. For instance: $f(\bar{\mathbf{E}}) = \epsilon = \text{Tr}(\bar{\mathbf{E}})$ defines the minimum MSE problem, while $f(\bar{\mathbf{E}}) = -R = \frac{1}{2} \log_2 |2\bar{\mathbf{E}}|$ defines the maximum achievable rate problem.

Consider the following eigenvalue decomposition (EVD):

$$\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \bar{\mathbf{H}} = \bar{\mathbf{Q}} \bar{\mathbf{\Lambda}} \bar{\mathbf{Q}}^T, \quad (11)$$

where $\bar{\mathbf{\Lambda}} = \text{diag}(\bar{\lambda}_1 \dots \bar{\lambda}_{2N})$ is a diagonal matrix containing the positive eigenvalues of $\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \bar{\mathbf{H}}$ and $\bar{\mathbf{Q}} \in \mathbb{R}^{2M \times 2N}$ corresponds to a unitary matrix that contains the associated eigenvectors stacked in columns. Recall that the rank of $\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \bar{\mathbf{H}}$ (i.e. the number of positive eigenvalues) is given by $2N$ and does not depend on the interference statistics, since $\text{rank}(\mathbf{C}_{\bar{\mathbf{s}}}^{-1}) = 2N$, $\bar{\mathbf{H}}$ is a full row rank matrix, and hence $\text{rank}(\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{C}_{\bar{\mathbf{s}}}^{-1} \bar{\mathbf{H}}) = \text{rank}(\mathbf{C}_{\bar{\mathbf{s}}}^{-1}) = 2N$.

Then, the optimal transmit precoder $\bar{\mathbf{W}}^{\text{opt}}$ to problem in (10) presents the following structure [25]:

$$\bar{\mathbf{W}}^{\text{opt}} = \bar{\mathbf{Q}} \bar{\mathbf{P}}^{\frac{1}{2}}, \quad (12)$$

where $\bar{\mathbf{P}} = \text{diag}(\bar{p}_1 \dots \bar{p}_{2N})$ is a diagonal matrix that describes the power allocation per stream, which satisfies $\frac{1}{2} \text{Tr}(\bar{\mathbf{P}}) = P^{\max}$ (see (10)) and depends on the optimization criterion (see Sections III-A and III-B).

The optimal transmit precoder in (12) allows diagonalizing

the MSE-matrix in (8), i.e.:

$$\bar{\mathbf{E}}^{\text{opt}} = \frac{1}{2} (\mathbf{I}_{2N} + \bar{\mathbf{P}} \bar{\mathbf{\Lambda}})^{-1}. \quad (13)$$

So, the optimal achievable rate and the optimal MSE become, respectively:

$$R^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}) = -\frac{1}{2} \log_2 |2\bar{\mathbf{E}}^{\text{opt}}| = \frac{1}{2} \sum_{i=1}^{2N} \log_2 (1 + \bar{p}_i \bar{\lambda}_i), \quad (14)$$

$$\epsilon^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}) = \text{Tr}(\bar{\mathbf{E}}^{\text{opt}}) = \frac{1}{2} \sum_{i=1}^{2N} \frac{1}{1 + \bar{p}_i \bar{\lambda}_i}, \quad (15)$$

where $\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}$ denotes the Hadamard product of vectors $\bar{\mathbf{p}} = \text{diag}(\bar{\mathbf{P}}) = [\bar{p}_1 \dots \bar{p}_{2N}]^T$ and $\bar{\boldsymbol{\lambda}} = \text{diag}(\bar{\mathbf{\Lambda}}) = [\bar{\lambda}_1 \dots \bar{\lambda}_{2N}]^T$. We occasionally use the notations $R^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}})$ and $\epsilon^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}})$ to make their dependencies on the product $\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}$ explicit.

The key aspect of the optimal solution in (12) is the EVD in (11), for which two cases are differentiated: proper interference ($\tilde{\mathbf{C}}_{\mathbf{s}} = \mathbf{0}$) or improper interference ($\tilde{\mathbf{C}}_{\mathbf{s}} \neq \mathbf{0}$). It is well known that if received interference is proper the optimum signaling is PGS (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{0}$), while if received interference is improper the optimum signaling is IGS (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}} \neq \mathbf{0}$). For example, [30] shows how this is concluded from the composite real representation.

III. SUPERIORITY OF IGS FOR THE MIMO P2P-I

In this section we focus on comparing the optimal schemes when interference is either proper or improper so as to determine if IGS is beneficial or not in terms of achievable rate and MSE in the simple MIMO P2P-I. The comparison reduces to relate the eigenvalues (see (14)-(15)) of the following two matrices (which are obtained from (11) by using the structure of $\mathbf{C}_{\bar{\mathbf{s}}}$ in (7)) for a fixed noise-plus-interference covariance matrix $\mathbf{C}_{\bar{\mathbf{s}}}$:

$$\begin{aligned} \text{Proper:} \quad \bar{\boldsymbol{\lambda}}_I &= \text{eig} \left(\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{T}^{-H} \begin{bmatrix} \mathbf{C}_{\mathbf{s}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\mathbf{s}}^* \end{bmatrix}^{-1} \mathbf{T}^{-1} \bar{\mathbf{H}} \right), \\ \text{Improper:} \quad \bar{\boldsymbol{\lambda}}_P &= \text{eig} \left(\frac{1}{2} \bar{\mathbf{H}}^T \mathbf{T}^{-H} \begin{bmatrix} \mathbf{C}_{\mathbf{s}} & \tilde{\mathbf{C}}_{\mathbf{s}} \\ \tilde{\mathbf{C}}_{\mathbf{s}}^* & \mathbf{C}_{\mathbf{s}}^* \end{bmatrix}^{-1} \mathbf{T}^{-1} \bar{\mathbf{H}} \right), \end{aligned} \quad (16)$$

where $\bar{\boldsymbol{\lambda}}_I \in \mathbb{R}_+^{2N \times 1}$ and $\bar{\boldsymbol{\lambda}}_P \in \mathbb{R}_+^{2N \times 1}$ denote the vectors of positive eigenvalues in decreasing order under the reception of improper and proper Gaussian interference, respectively. The comparison in the sequel is based on fixing the same $\mathbf{C}_{\bar{\mathbf{s}}}$, so the level of received interference-plus-noise power is the same but the difference comes from the proper or improper statistics of the interference (i.e. $\tilde{\mathbf{C}}_{\mathbf{s}} = \mathbf{0}$ or $\tilde{\mathbf{C}}_{\mathbf{s}} \neq \mathbf{0}$, see (16)) and the corresponding optimal signaling (i.e. PGS or IGS).

In Lemma 1 we show how the eigenvalues in (16) are related, which will be exploited afterwards.

Lemma 1: For $M \geq N$ and full rank channel matrix \mathbf{H} , the eigenvalues in (16) are related by the following majorization relations:

$$\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1} \quad \text{and} \quad \bar{\boldsymbol{\lambda}}_I \succ_{w \log} \bar{\boldsymbol{\lambda}}_P, \quad (17)$$

where \succ refers to strong majorization, $\succ_{w \log}$ denotes weak log-majorization, $\bar{\boldsymbol{\lambda}}_I^{-1} \in \mathbb{R}_+^{2N \times 1}$ and $\bar{\boldsymbol{\lambda}}_P^{-1} \in \mathbb{R}_+^{2N \times 1}$ refer to vectors whereby each component is obtained from the inverse of the components in $\bar{\boldsymbol{\lambda}}_I$ and $\bar{\boldsymbol{\lambda}}_P$, respectively, in decreasing order.

Proof: See Appendix B-A. ■

The strong majorization result in (17) is equivalent to (see (52) in Definition 5 in Appendix A-B):

$$\begin{aligned} \sum_{i=1}^n \frac{1}{\bar{\lambda}_{I,i}} &\leq \sum_{i=1}^n \frac{1}{\bar{\lambda}_{P,i}} \quad n = 1, \dots, 2N - 1 \\ \text{and} \quad \sum_{i=1}^{2N} \frac{1}{\bar{\lambda}_{I,i}} &= \sum_{i=1}^{2N} \frac{1}{\bar{\lambda}_{P,i}}, \end{aligned} \quad (18)$$

while the weak log-majorization result in (17) is equivalent to (see Definition 7 in Appendix A-B):

$$\begin{aligned} \prod_{i=1}^n \bar{\lambda}_{I,i} &\geq \prod_{i=1}^n \bar{\lambda}_{P,i} \quad n = 1, \dots, 2N \\ \text{and} \quad \sum_{i=1}^n \bar{\lambda}_{I,i} &\geq \sum_{i=1}^n \bar{\lambda}_{P,i} \quad n = 1, \dots, 2N. \end{aligned} \quad (19)$$

Therefore, when using IGS and receiving improper interference in the MIMO P2P-I, the eigenvalues of the equivalent channel are more spread out having an equal or larger sum, an equal or larger product, and an equal sum of the inverses, as compared to using PGS and receiving proper interference. In other words, the arithmetic mean and the geometric mean of the eigenvalues are equal or larger, while the harmonic mean of the eigenvalues is equal.

The proof of the strong majorization result in Lemma 1 (i.e. $\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) is only valid for $M \geq N$ and the channel matrix \mathbf{H} being full rank. Otherwise, the inverse in (65) (see Appendix B-A) does not exist. This is due to the fact that if the desired signal subspace at the receiver has less than N dimensions (which happens either if $M < N$ or if the channel is rank deficient), then the strong majorization result is not valid since the properties of the noise-plus-interference in the noise-plus-interference subspace can be chosen arbitrarily without influencing the desired signal space. In case that either $M < N$ or the channel matrix \mathbf{H} is rank deficient, then the weak-log majorization result in Lemma 1 is satisfied, i.e. $\bar{\lambda}_I \succ_{w \log} \bar{\lambda}_P$ (see [30]), but the strong majorization result in Lemma 1 (which is key to demonstrate the strict superiority of IGS in what follows) is not further valid.

Now we show how Lemma 1 allows determining the superiority of IGS in terms of achievable rate in (14) and MSE in (15). The strong majorization result in (17) allows us to set the strict superiority in terms of achievable rate and MSE when uniform power allocation (UPA) is adopted, while the weak log-majorization result in (17) is used to derive the superiority in terms of achievable rate when optimal power allocation (OPA) is used.

A. Superiority for uniform power allocation

When adopting a UPA strategy, the solution for $\bar{\mathbf{P}} = \text{diag}(\bar{p}_1 \dots \bar{p}_{2N})$ in (12) is given by:

$$\bar{p}_i = \bar{P} = P^{\max}/N, \forall i, \quad (20)$$

such that $\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}} = \bar{P}\bar{\boldsymbol{\lambda}}$. Thus, the optimal achievable rate in (14) and the optimal MSE in (15) become, respectively:

$$R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}) = \frac{1}{2} \sum_{i=1}^{2N} \log_2(1 + \bar{P}\bar{\lambda}_i), \quad (21)$$

$$\epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}) = \frac{1}{2} \sum_{i=1}^{2N} \frac{1}{1 + \bar{P}\bar{\lambda}_i}. \quad (22)$$

Theorem 1: Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix \mathbf{H} , receiving noise-plus-interference with a

given covariance matrix \mathbf{C}_s . When applying the optimal signaling and the associated transmission scheme with UPA, the achievable rate R^{opt} in (14) is strictly increased in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_s \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_s = \mathbf{0}$):

$$R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) + c_R^{\text{upa}} (\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2), \quad (23)$$

with a positive constant c_R^{upa} :

$$c_R^{\text{upa}} = \frac{\bar{P}\lambda_{\min}^3 (1 + 0.5\bar{P}\lambda_{\min})}{2 \ln(2) (1 + \bar{P}\lambda_{\min})^2} > 0, \quad (24)$$

where λ_{\min} denotes the minimum eigenvalue of $\bar{\boldsymbol{\lambda}}_I$ and \bar{P} is defined in (20). The rate gap in (23) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

Proof: See Appendix B-B. The proof departs from the strong majorization result in Lemma 1 and exploits majorization theory on strongly Schur-convex functions (see Definition 9 in Appendix A-B). ■

The rate gap in (23) increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the improperness of the interference increases or, equivalently, as $\tilde{\mathbf{C}}_s$ "increases", see (16)). Further, we can determine the rate gap behavior when varying the desired signal power (i.e. \bar{P}) for a fixed interference-plus-noise power level or, equivalently, when varying the signal-to-interference-plus-noise ratio (SINR):

- At high SINR, c_R^{upa} in (24) scales as $c_R^{\text{upa}} \sim \frac{1}{4 \ln(2)} \lambda_{\min}^2$ such that the rate gap in (23) is constant as the SINR increases. This means that, at the high SINR regime, the rates grow with the same slope with PGS and IGS but there is a constant difference among said achievable rates such that the use of IGS is always beneficial.
- At low SINR, c_R^{upa} in (24) scales as $c_R^{\text{upa}} \sim \frac{1}{2 \ln(2)} \bar{P} \lambda_{\min}^3$ and hence the rate gap in (23) increases with the SINR.

Theorem 2: Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix \mathbf{H} , receiving noise-plus-interference with a given covariance matrix \mathbf{C}_s . When applying the optimal signaling and the associated transmission scheme with UPA, the MSE ϵ^{opt} in (15) is strictly reduced in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_s \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_s = \mathbf{0}$):

$$\epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \leq \epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) - c_e^{\text{upa}} (\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2), \quad (25)$$

with a positive constant c_e^{upa} :

$$c_e^{\text{upa}} = \frac{\bar{P}\lambda_{\min}^3}{2(1 + \bar{P}\lambda_{\min})^3} > 0. \quad (26)$$

The error gap in (25) is strictly positive as long as $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

Proof: See Appendix B-C. The proof departs from the strong majorization result in Lemma 1 and exploits majorization theory on strongly Schur-concave functions (see Definition 9 in Appendix A-B). ■

Similarly as in Theorem 1, the error gap in (25) increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the improperness of the interference increases). The error gap behavior with respect to the SINR is as follows:

- At high SINR, c_e^{upa} in (26) scales as $c_e^{\text{upa}} \sim \frac{1}{2}\bar{P}^{-2}$ such that the error gap in (25) decreases as the SINR increases. This is related to the fact that the rate gap in (23) is constant at high SINR: due to the convex rate-MSE relation in (9), the rate gap being constant implies the error gap being reduced as the SINR increases.
- At low SINR, c_e^{upa} in (26) scales as $c_e^{\text{upa}} \sim \frac{1}{2}\bar{P}\lambda_{\min}^3$ and therefore the error gap in (25) increases with the SINR.

Corollary 1: If UPA is used in a MIMO P2P-I such that $M \geq N$ and \mathbf{H} is full rank, the achievable rate R^{opt} in (14) and the MSE ϵ^{opt} in (15) are strictly outperformed (with a positive gap) when received interference is improper Gaussian distributed (i.e. $\tilde{\mathbf{C}}_s \neq \mathbf{0}$).

Proof: through Theorem 1 and Theorem 2. ■

B. Superiority for optimal power allocation

When adopting an OPA strategy to maximize the achievable rate (i.e. $f(\bar{\mathbf{E}}) = -R = \frac{1}{2}\log_2|2\bar{\mathbf{E}}|$ in (10)), the optimal solution for $\bar{\mathbf{P}} = \text{diag}(\bar{p}_1 \dots \bar{p}_{2N})$ in (12) is given by [32]:

$$\bar{p}_i = (\mu - \bar{\lambda}_i^{-1})^+, \forall i, \quad \mu = \frac{1}{k} (2P^{\text{max}} + \sum_{i=1}^k \bar{\lambda}_i^{-1}), \quad (27)$$

where $k \leq 2N$ is the number of active streams after the water-filling solution in (27). Accordingly, let us denote by $\bar{\mathbf{p}}_P$ and $\bar{\mathbf{p}}_I$ the vectors with power allocation in (27) (ordered in decreasing order) for the proper and improper interference cases, respectively.

Due to the water-filling solution in (27), some streams can have a power equal to 0 and the number of active streams in the proper and improper interference cases might differ. Thus, as the power allocation in (27) depends on the eigenvalues, we cannot exploit majorization theory on strongly Schur-convex functions to guarantee a strict rate improvement with OPA.

Theorem 3: Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix \mathbf{H} , receiving noise-plus-interference with a given covariance matrix \mathbf{C}_s . When applying the optimal signaling and the associated transmission scheme with OPA in (27) for maximum achievable rate, the achievable rate R^{opt} in (14) is equal or larger in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_s \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_s = \mathbf{0}$):

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P). \quad (28)$$

The equality in (28) is satisfied when $\bar{\boldsymbol{\lambda}}_I = \bar{\boldsymbol{\lambda}}_P$, i.e. $\tilde{\mathbf{C}}_s = \mathbf{0}$, such that $\bar{\mathbf{p}}_I = \bar{\mathbf{p}}_P$ (see (27)) and hence $\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I = \bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P$.

Proof: See Appendix B-D. The proof departs from the weak log-majorization result in Lemma 1 and is valid even if the number of streams with OPA in (27) is different in the proper and improper interference cases. ■

There is, however, a particular case in which the strict superiority of IGS over PGS under OPA strategy can be demonstrated through the application of majorization theory tools. In case that all streams are active both with IGS and PGS, i.e. $k_I = k_P = 2N$ in (27), then the water-level μ in (27) is equal with IGS and PGS: $\mu = \mu_I = \mu_P$ because $\sum_{i=1}^{2N} \bar{\lambda}_{I,i}^{-1} = \sum_{i=1}^{2N} \bar{\lambda}_{P,i}^{-1}$ (see (18)). In this case, the optimal

achievable rate in (14) becomes:

$$R^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}) = \frac{1}{2} \sum_{i=1}^{2N} \log_2(\mu \bar{\lambda}_i). \quad (29)$$

Theorem 4: Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix \mathbf{H} , receiving noise-plus-interference with a given covariance matrix \mathbf{C}_s . When applying the optimal signaling and the associated transmission scheme with OPA in (27) for maximum achievable rate and all streams are active for IGS and PGS, the achievable rate R^{opt} in (14) is strictly increased in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_s \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_s = \mathbf{0}$):

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P) + c_R^{\text{opa}} (\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2), \quad (30)$$

with a positive constant c_R^{opa} :

$$c_R^{\text{opa}} = \frac{\lambda_{\min}^2}{4 \ln(2)} > 0. \quad (31)$$

The rate gap in (30) is strictly positive in case $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

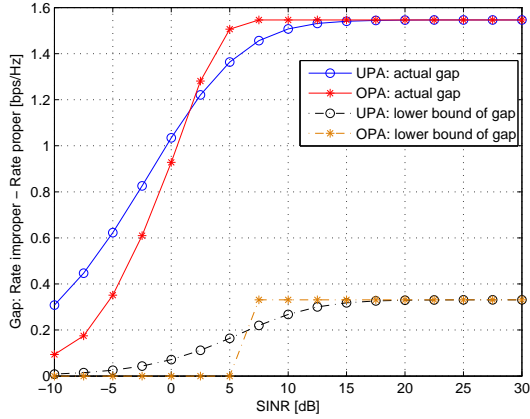
Proof: See Appendix B-E. The proof departs from the strong majorization result in Lemma 1 and exploits majorization theory on strongly Schur-convex functions (see Definition 9 in Appendix A-B). ■

Similarly as in Theorem 1, the rate gap in (30) increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the improperness of the interference increases). However, result in Theorem 4 is only valid when the number of active streams under IGS and PGS is equal to $2N$, which occurs at medium/high SINR regimes. Note also that the derived rate gap under OPA strategy in (31) coincides with the one derived for UPA in (24) at high SINR.

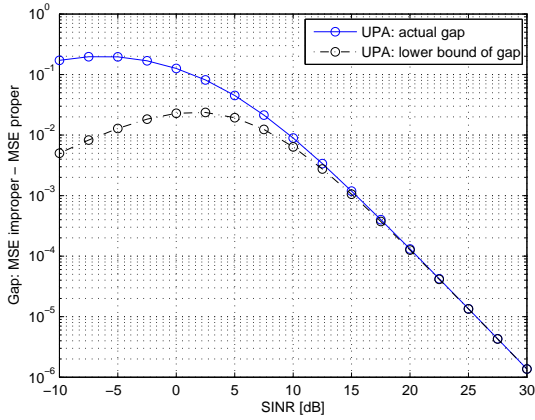
C. Simulation results

First, let us show through simulations the strict rate and MSE improvement of IGS over PGS derived from Theorems 1 and 2 for UPA strategy. We evaluate the gains in the MIMO P2P-I for a single channel realization when interference is either proper or improper Gaussian distributed. Fig. 4.(a) and Fig. 4.(b) depict the actual rate gap and the actual MSE gap versus the SINR, respectively, for an interference-to-noise ratio (INR) of INR=20dB and antenna configuration $M=N=2$ (2×2). The figures also display the lower bound of the gaps presented in (23) for rate and in (25) for MSE. Further, the actual rate gap and the lower bound of gap derived in Theorem 4 under OPA strategy are included in Fig. 4.(a).

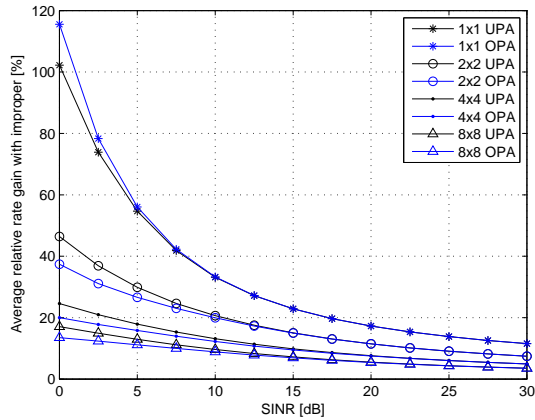
With regard to the rate (see Fig. 4.(a)), it can be observed that the rate gap is constant at high SINR and that it is increasing with the SINR at low SINR. Differently, in terms of the MSE (see Fig. 4.(b)), the error gap is decreasing with the SINR at high SINR and it is increasing with the SINR at low SINR. This corroborates the conclusions drawn from Theorems 1 and 2. The closer the lower bound of the gap is to the actual gap depends on the individual channel realizations, but it is observed to be tighter in the MSE than in the rate. Also, it is observed through simulations that the lower bound of the rate gap derived for UPA strategy is also valid for OPA strategy in all SINR regimes and antenna configurations



(a) gap in the achievable rate



(b) gap in the MSE

 Fig. 4: Gap in the achievable rate and MSE versus SINR of a MIMO P2P-I. 2×2 , INR=20dB. UPA and OPA strategies. (a) rate gap, (b) MSE gap.

 Fig. 5: Average relative rate gain (in %) of IGS over PGS versus SINR for a MIMO P2P-I with different antenna configurations: 1×1 , 2×2 , 4×4 , and 8×8 . INR=20dB. UPA and OPA strategies.

(although this has been mathematically demonstrated only for medium/high SINR regimes, see Theorem 4).

Second, we show how the rate improvement of IGS over PGS scales as the number of transmit/receive antennas increases. 1000 channel realizations of the MIMO P2P-I are used to take statistic results. Fig. 5 displays the average of the relative rate gain (in %) of IGS over PGS versus the SINR

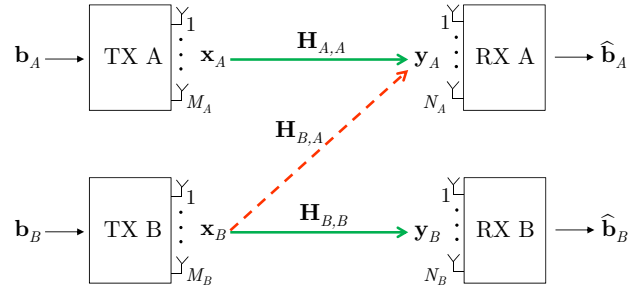


Fig. 6: MIMO Z-Interference Channel.

for INR=20dB and different antenna configurations: 1×1 , 2×2 , 4×4 , and 8×8 . The relative rate gains decrease as the number of transmit/receive antennas increase. This is due to the fact that IGS provides flexibility by splitting one dimension into two halves, which is more useful when the number of transmit/receive antennas is low, otherwise extra dimensions are already added to the system by adding antennas. Even though, it is important to recall that for the 8×8 antenna case, relative rate gains are still obtained for all SINR regimes (see Fig. 5). For SINR=0dB, where the desired signal and the interfering signal have a similar strength because INR=20dB, the relative rate gains are: 102-115% for 1×1 , 46-37% for 2×2 , 24-20% for 4×4 , and 17-13% for 8×8 . In addition, in Fig. 5 we can observe that the relative rate gains are larger for low SINR regimes, in which the MIMO P2P-I is highly limited by interference. However, as the SINR increases, rate gains are still obtained due to the use and reception of IGS.

IV. APPLICATION TO THE MIMO Z-IC

In this section we show how to exploit the benefits of using IGS in the MIMO Z-IC through a simple WLP design. The MIMO Z-IC is a two-transmitter two-receiver multi-antenna interfering scenario, as shown in Fig. 6, where one of the receivers (RXB) just observes noise while the other (RXA) receives interference. The signal model detailed in Section II applies for each of the two MIMO P2P-I links that appear in the MIMO Z-IC. In this regard, we add the subindex $(\cdot)_A$ and $(\cdot)_B$ to refer to the corresponding link, see Fig. 6.

Concerning RXB that is receiving only proper Gaussian noise, the optimal scheme at TXB is a proper-based scheme (i.e. LP) that would be given by $\mathbf{W}_{B,1} = \mathbf{W}_{B,1}^{\text{opt}}$ and $\mathbf{W}_{B,2} = \mathbf{0}$. Such proper-based scheme would generate proper interference onto RXA. However, RXA could benefit in terms of achievable rate from the reception of improper interference from TXB, as is demonstrated in Section III, but the fact that TXB transmits IGS implies a degradation of the rate of RXB due to the sub-optimality of the signaling scheme [8]. Therefore, a clear trade-off arises: with IGS the rate of RXA could be increased but the rate of RXB would be decreased.

Assuming that we can tolerate a certain achievable rate loss at RXB, performance gains at RXA would be guaranteed with IGS provided that TXB uses the same transmit covariance matrix as in the optimum proper scheme for RXB (i.e. $\mathbf{C}_{x_B} = \mathbf{W}_{B,1}^{\text{opt}} \mathbf{W}_{B,1}^{\text{opt}H}$), such that the received noise-plus-interference covariance matrix at RXA ($\mathbf{C}_{s_A} = \mathbf{H}_{B,A} \mathbf{C}_{x_B} \mathbf{H}_{B,A}^H + \sigma^2 \mathbf{I}_{N_A}$) is the same as in the case of

proper interference and, due to the use of WLP with $\tilde{\mathbf{C}}_{\mathbf{x}_B} \neq \mathbf{0}$ (and hence $\tilde{\mathbf{C}}_{\mathbf{s}_A} = \mathbf{H}_{B,A} \tilde{\mathbf{C}}_{\mathbf{x}_B} \mathbf{H}_{B,A}^T \neq \mathbf{0}$), the achievable rate of RXA would be increased by Theorem 1 for UPA and by Theorem 3 for OPA.

A. WLP scheme at TXB

In order to guarantee a rate improvement of RXA we need to construct an improper-based scheme at TXB (characterized by transmit precoders $\mathbf{W}_{B,1}$ and $\mathbf{W}_{B,2}$, see (2)) that maintains the transmit covariance matrix used in the optimal proper-based scheme for RXB (i.e. $\mathbf{C}_{\mathbf{x}_B} = \mathbf{W}_{B,1}^{\text{opt}} \mathbf{W}_{B,1}^{\text{opt}H}$). A practical WLP scheme at TXB that controls the degradation on performance of RXB and generates improper interference onto RXA can be obtained by right-multiplying the transmit precoder $\mathbf{W}_{B,1}^{\text{opt}}$ by two scaling factors $\sqrt{\alpha}$ and $\sqrt{1-\alpha}$ and two unitary matrices $\mathbf{Z}_1 \in \mathbb{C}^{N \times N}$ and $\mathbf{Z}_2 \in \mathbb{C}^{N \times N}$. Thus, the WLP scheme is created as follows:

$$\mathbf{W}_{B,1} = \sqrt{1-\alpha} \mathbf{W}_{B,1}^{\text{opt}} \mathbf{Z}_1, \quad \mathbf{W}_{B,2} = \sqrt{\alpha} \mathbf{W}_{B,1}^{\text{opt}} \mathbf{Z}_2, \quad (32)$$

where $\alpha \in [0, 0.5]$. This improper-based scheme at TXB has the same transmit covariance matrix than the optimal proper-based scheme for RXB provided that \mathbf{Z}_1 and \mathbf{Z}_2 are unitary matrices:

$$\mathbf{C}_{\mathbf{x}_B} = \mathbf{W}_{B,1} \mathbf{W}_{B,1}^H + \mathbf{W}_{B,2} \mathbf{W}_{B,2}^H = \mathbf{W}_{B,1}^{\text{opt}} \mathbf{W}_{B,1}^{\text{opt}H}, \quad (33)$$

but the transmit pseudo-covariance matrix does not vanish to $\mathbf{0}$ and can be tuned as a function of α :

$$\tilde{\mathbf{C}}_{\mathbf{x}_B} = \sqrt{\alpha(1-\alpha)} \mathbf{W}_{B,1}^{\text{opt}} (\mathbf{Z}_1 \mathbf{Z}_2^T + \mathbf{Z}_2 \mathbf{Z}_1^T) \mathbf{W}_{B,1}^{\text{opt}T}. \quad (34)$$

Hence, by performing WLP in (32) at TXB with $\alpha > 0$, the achievable rate of RXB is degraded while the achievable rate of RXA is ensured to be increased as compared to the use of $\mathbf{W}_{B,1} = \mathbf{W}_{B,1}^{\text{opt}}$ and $\mathbf{W}_{B,2} = \mathbf{0}$. This is possible because the WLP scheme in (32) allows maintaining the transmit covariance matrix $\mathbf{C}_{\mathbf{x}_B}$ independently of the values of α , \mathbf{Z}_1 , and \mathbf{Z}_2 , such that the properties in Section III apply.

1) *Parameter α* : The range $\alpha \in [0, 0.5]$ is considered in (32) because $\tilde{\mathbf{C}}_{\mathbf{x}_B}$ has a symmetric shape with respect to α that is centered on $\alpha = 0.5$, i.e. the system performance of the MIMO Z-IC is equivalent when selecting α or $\alpha' = 1 - \alpha$. The suitable selection of parameter α allows controlling the level of *improperness*: if $\alpha = 0$ the generated signal at TXB is proper, while if $\alpha = 0.5$ the generated signal at TXB achieves the maximal level of improperness and improper interference is generated towards RXA.

2) *Spatial unitary matrices \mathbf{Z}_1 and \mathbf{Z}_2* : Matrices \mathbf{Z}_1 and \mathbf{Z}_2 are included in (32) so as to get different MIMO spatial structures of $\mathbf{W}_{B,1}$ and $\mathbf{W}_{B,2}$, although $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{I}_N$ is also an option. It is important to realize from (34) that the performance of the MIMO Z-IC is affected by the term $(\mathbf{Z}_1 \mathbf{Z}_2^T + \mathbf{Z}_2 \mathbf{Z}_1^T)$ rather than by the specific selection of \mathbf{Z}_1 and \mathbf{Z}_2 . So, if we design \mathbf{Z}_1 and \mathbf{Z}_2 such that $\mathbf{Z}_1 \mathbf{Z}_2^T = \mathbf{Z}_2 \mathbf{Z}_1^T$, then a single choice would control the performance of the MIMO Z-IC. For $N = 2$, [30] proposes a design for \mathbf{Z}_1 and \mathbf{Z}_2 that satisfies such condition and depends on a single parameter. For the general MIMO case, \mathbf{Z}_1 and \mathbf{Z}_2 are designed in order

to get two totally opposite behaviors:

$$\begin{aligned} \text{Option 1 (identity):} \quad & \mathbf{Z}_1 \mathbf{Z}_2^T = \mathbf{Z}_2 \mathbf{Z}_1^T = \mathbf{I}_N \\ & \rightarrow \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{I}_N, \end{aligned} \quad (35)$$

$$\begin{aligned} \text{Option 2 (anti-identity):} \quad & \mathbf{Z}_1 \mathbf{Z}_2^T = \mathbf{Z}_2 \mathbf{Z}_1^T = \mathbf{J}_N \\ & \rightarrow \mathbf{Z}_1 = \mathbf{J}_N, \mathbf{Z}_2 = \mathbf{I}_N, \end{aligned} \quad (36)$$

being \mathbf{J}_N the anti-identity matrix² of size N . Both options have a trivial solution for unitary matrices \mathbf{Z}_1 and \mathbf{Z}_2 , as shown in (35)-(36).

B. Trade-off

Let us show that parameter α is the one that allows trading-off on the rates of the MIMO Z-IC system.

1) *Rate improvement at RXA*: The achievable rate improvement of RXA is guaranteed by Theorem 1 for UPA or by Theorem 3 for OPA. The exact rate improvement expression cannot be extracted, however, it is given by how much the eigenvalues in the improper interference case majorize the eigenvalues in the proper interference case, i.e. the difference in the majorization results in (17). Said difference becomes larger as $\tilde{\mathbf{C}}_{\mathbf{s}_A}$ in (16) "increases" because the stronger the off-diagonal blocks in (16) are the more spread out the eigenvalues become (see Lemma 6 in Appendix A-B). Accordingly, the optimum value of α to maximize the achievable rate of RXA (R_A) is $\alpha = 0.5$ (see (34)), as it allows "increasing" $\tilde{\mathbf{C}}_{\mathbf{x}_B}$ and, as consequence, "increasing" the pseudo-covariance matrix of the interference-plus-noise received at RXA: $\tilde{\mathbf{C}}_{\mathbf{s}_A} = \mathbf{H}_{B,A} \tilde{\mathbf{C}}_{\mathbf{x}_B} \mathbf{H}_{B,A}^T$.

2) *Rate degradation at RXB*: Due to the use of WLP at TXB and as RXB only receives proper noise (i.e. $\mathbf{C}_{\mathbf{s}_B} = \sigma^2 \mathbf{I}_N$ and $\tilde{\mathbf{C}}_{\mathbf{s}_B} = \mathbf{0}$), the achievable rate of RXB is degraded. When using the WLP scheme in (32) at TXB, the achievable rate of RXB (R_B) can be decomposed as (derived from [4]):

$$\begin{aligned} R_B = & \underbrace{\log_2 |\mathbf{I}_N + \mathbf{C}_{\mathbf{s}_B}^{-1} \mathbf{H}_{B,B} \mathbf{C}_{\mathbf{x}_B} \mathbf{H}_{B,B}^H|}_{R_B^{\text{opt}}} \\ & + \underbrace{\frac{1}{2} \log_2 |\mathbf{I}_N - \mathbf{C}_{\mathbf{y}_B}^{-1} \tilde{\mathbf{C}}_{\mathbf{y}_B} \mathbf{C}_{\mathbf{y}_B}^{-T} \tilde{\mathbf{C}}_{\mathbf{y}_B}^H|}_{R_B^{\text{penalty}} \leq 0}, \end{aligned} \quad (37)$$

where $\mathbf{C}_{\mathbf{y}_B} = \mathbf{H}_{B,B} \mathbf{C}_{\mathbf{x}_B} \mathbf{H}_{B,B}^H + \sigma^2 \mathbf{I}_N$ and $\tilde{\mathbf{C}}_{\mathbf{y}_B} = \mathbf{H}_{B,B} \tilde{\mathbf{C}}_{\mathbf{x}_B} \mathbf{H}_{B,B}^T$. The first term in (37) corresponds to the achievable rate of RXB if the transmission scheme adopted at TXB was PGS with LP (the optimum one for RXB, with $\tilde{\mathbf{C}}_{\mathbf{x}_B} = \mathbf{0}$). However, due to the use of an improper-based scheme in which $\tilde{\mathbf{C}}_{\mathbf{x}_B} \neq \mathbf{0}$, an achievable rate penalty is obtained at RXB, which is given by the second term in (37) and is strictly negative. The fact that $R_B^{\text{penalty}} \leq 0$ can be shown by using the Fischer's inequality [33, Sect. 8.2] applied over the partitioned matrix $[\mathbf{C}_{\mathbf{y}_B} \quad \tilde{\mathbf{C}}_{\mathbf{y}_B}; \mathbf{C}_{\mathbf{y}_B}^* \quad \tilde{\mathbf{C}}_{\mathbf{y}_B}^*]$ that is positive semidefinite (see Theorem 5 in Appendix A-A), from which it is concluded that $|\mathbf{I}_N - \mathbf{C}_{\mathbf{y}_B}^{-1} \tilde{\mathbf{C}}_{\mathbf{y}_B} \mathbf{C}_{\mathbf{y}_B}^{-T} \tilde{\mathbf{C}}_{\mathbf{y}_B}^H| \leq 1$. Consequently, the closer α is to 0.5 (i.e. $\tilde{\mathbf{C}}_{\mathbf{y}_B}$ in (37) "increases", see (34)) the more the achievable rate of RXB is reduced.

²We refer with anti-identity matrix (also known as reflection matrix) to a square matrix where all the entries are 0 except those on the anti-diagonal that are equal to 1 (going from the lower left corner to the upper right corner).

C. Sum-rate performance

The selection of parameter α has a clear trade-off on the sum-rate performance: R_B in (37) is decreasing while R_A is increasing with $\alpha \in [0, 0.5]$. Let us consider $\tilde{\alpha} = \sqrt{\alpha(1-\alpha)} \in [0, 0.5]$, being $\tilde{\alpha}$ the parameter that impacts on the sum-rate performance of the MIMO Z-IC (see (34)). In this sense, R_B is decreasing and R_A is increasing with respect to $\tilde{\alpha} \in [0, 0.5]$. Let us draw their rate expressions. The rate expression R_B in (37) (using (33) and (34)) as a function of $\tilde{\alpha}$ is:

$$R_B(\tilde{\alpha}) = R_B^{\text{opt}} + \frac{1}{2} \log_2 |\mathbf{I}_N - \tilde{\alpha}^2 \mathbf{X}|, \quad (38)$$

where $\mathbf{X} = \mathbf{C}_{y_B}^{-1} \mathbf{H}_{B,B} \mathbf{W}_{B,1}^{\text{opt}} (\mathbf{Z}_1 \mathbf{Z}_2^T + \mathbf{Z}_2 \mathbf{Z}_1^T) \mathbf{W}_{B,1}^{\text{opt}T} \mathbf{H}_{B,B}^T \mathbf{C}_{y_B}^{-T} \mathbf{H}_{B,B}^* \mathbf{W}_{B,1}^{\text{opt}*} (\mathbf{Z}_2^* \mathbf{Z}_1^H + \mathbf{Z}_1^* \mathbf{Z}_2^H) \mathbf{W}_{B,1}^{\text{opt}H} \mathbf{H}_{B,B}^H$ and R_B^{opt} is shown in (37). On the other hand, the rate expression R_A (derived from (14) and (16)) as a function of $\tilde{\alpha}$ is:

$$R_A(\tilde{\alpha}) = \frac{1}{2} \sum_{i=1}^{2N} \log_2(1 + \bar{p}_i \bar{\lambda}_i), \quad (39)$$

$$\bar{\lambda} = \text{eig} \left(\frac{1}{2} \bar{\mathbf{H}}_{A,A}^T \mathbf{T}^{-H} \begin{bmatrix} \mathbf{C}_{s_A} & \tilde{\alpha} \mathbf{U} \\ \tilde{\alpha} \mathbf{U}^* & \mathbf{C}_{s_A}^* \end{bmatrix}^{-1} \mathbf{T}^{-1} \bar{\mathbf{H}}_{A,A} \right),$$

where $\mathbf{U} = \mathbf{H}_{B,A} \mathbf{W}_{B,1}^{\text{opt}} (\mathbf{Z}_1 \mathbf{Z}_2^T + \mathbf{Z}_2 \mathbf{Z}_1^T) \mathbf{W}_{B,1}^{\text{opt}T} \mathbf{H}_{B,A}^T$.

Then, constrained to the WLP design proposed in Section IV-A, the optimum $\tilde{\alpha} = \sqrt{\alpha(1-\alpha)}$ in order to maximize the sum-rate performance of the MIMO Z-IC is obtained from:

$$\underset{\tilde{\alpha}}{\text{maximize}} R_A(\tilde{\alpha}) + R_B(\tilde{\alpha}) \quad \text{subject to } 0 \leq \tilde{\alpha} \leq 0.5. \quad (40)$$

The optimization problem in (40) can be solved by setting the first order derivative to zero, and selecting the value of $\tilde{\alpha}$ from the candidate points given by the extreme of the domain (0 and 0.5) and the positive real roots of the first order derivative that lie within the domain and have a negative second order derivative (i.e. are maximum points). Through extensive numerical evaluations, we have observed that the optimal value of $\tilde{\alpha}$ always lies in the extremes of the domain ($\tilde{\alpha}^{\text{opt}} = 0$ or $\tilde{\alpha}^{\text{opt}} = 0.5$), which corresponds to $\alpha^{\text{opt}} = 0$ (i.e. PGS) or $\alpha^{\text{opt}} = 0.5$ (i.e. IGS with maximal level of improperness). Intuitively, for low interference regimes the optimal value is $\alpha^{\text{opt}} = 0$ because increasing α leads to an increase of the rate of RXA lower than the reduction of the rate of RXB. On the contrary, for medium/high interference regimes the optimal value is $\alpha^{\text{opt}} = 0.5$ since reducing α leads to a reduction of the rate of RXA larger than the increase of the rate of RXB.

For the MIMO Z-IC, the optimal value of the transmitted power at TXA corresponds always to the maximum one, as R_A is increasing with the available power and no interference is generated onto RXB [27]. Therefore, the operating points of the proposed scheme are obtained by varying the values of α from 0 to 0.5 (at maximum power of TXB) and then varying the transmitted power at TXB from maximum power to 0.

D. Simulation results

The proposed improper-based scheme is evaluated in the MIMO Z-IC scenario displayed in Fig. 6. It is assumed that both transmitters have the same available power P^{max} . Signal-to-noise ratio is defined as $\text{SNR} = P^{\text{max}}/\sigma^2$ and signal-to-interference ratio by $\text{SIR} = 1/\eta$, where factor $\eta \geq 0$ denotes

the average ratio between interfering and direct channel attenuations. Channels are modeled through a Rayleigh distribution, such that $\mathbf{H}_{A,A}, \mathbf{H}_{B,B} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{H}_{B,A} \sim \mathcal{CN}(\mathbf{0}, \eta \mathbf{I})$.

1) *Achievable rate region*: Let us show the achievable rate region of the MIMO Z-IC for a specific channel realization with $\text{SNR}=10\text{dB}$ and $\text{SNR}=20\text{dB}$, $\eta = 1$, $M = N = 2$ (2×2). The MIMO channel matrices are:

$$\mathbf{H}_{A,A} = [1.01e^{-j174.9} \quad 0.74e^{j152.8}; 0.86e^{-j55.5} \quad 0.82e^{j166.7}],$$

$$\mathbf{H}_{B,B} = [0.49e^{j162.3} \quad 1.30e^{-j101.0}; 0.70e^{-j43.8} \quad 0.46e^{j9.8}], \quad (41)$$

$$\mathbf{H}_{B,A} = [1.16e^{-j132.6} \quad 0.90e^{-j88.5}; 0.93e^{-j141.5} \quad 0.70e^{j67.1}].$$

The following transmission schemes are evaluated:

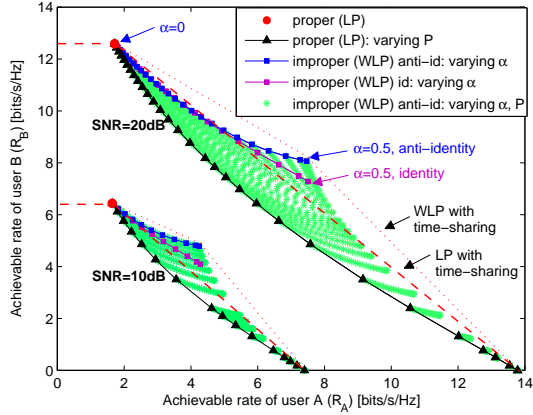
- **proper (LP)**: transmitters use LP,
- **improper (WLP) id: varying α** : transmitters use WLP, whereby TXB uses the improper-based scheme in (32) with different values of α and the identity solution for \mathbf{Z}_1 and \mathbf{Z}_2 in (35),
- **improper (WLP) anti-id: varying α** : transmitters use WLP, and TXB uses the improper-based scheme in (32) with different values of α and the anti-identity solution for \mathbf{Z}_1 and \mathbf{Z}_2 in (36).

Fig. 7 displays the achievable rate region when UPA in (20) is adopted and when OPA in (27) is used. Another way to achieve a similar behavior than the one given by the proposed improper-based scheme (i.e. reduce rate of RXB to increase rate of RXA) is by reducing the power used by TXB ($P \leq P^{\text{max}}$), which can be applied both for LP and WLP cases. Such performance results are depicted in the figures with the label 'varying P '. Note that the optimal solution for TXA corresponds to always using the maximum power P^{max} , as it does not interfere RXB.

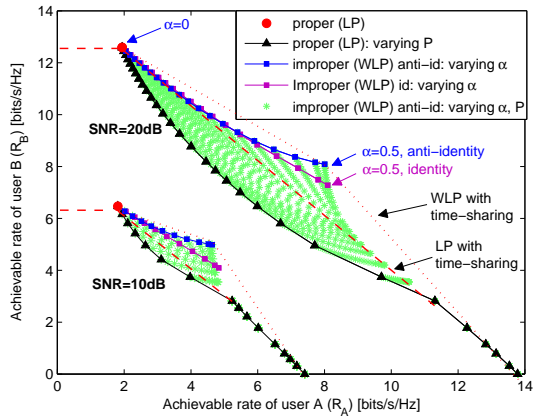
By comparing 'improper (WLP) varying α ' with respect to 'proper (LP)' at $P = P^{\text{max}}$, it is verified that the rate of RXA is always increased when received interference is improper (as is demonstrated by Theorems 1 and 3) at the expenses of a reduced rate for RXB. Accordingly, the proposed WLP scheme is beneficial when interference is non-negligible, such that the gain at RXA is significant. The choice of \mathbf{Z}_1 and \mathbf{Z}_2 (i.e. identity or anti-identity) does not have a significant impact on the rate of RXA, however it affects the rate of RXB, being the anti-identity (Option 2 in (36)) shown to be the best solution. Therefore, in interference-limited scenarios, the proposed improper-based scheme allows trading in transmission fairness mainly through parameter α without adjusting the transmitted power. Even with time-sharing, the improper-based scheme outperforms the proper-based scheme, as is shown with dashed lines in Fig. 7. To conclude, it can be observed that the achievable rate region is enlarged when using IGS through WLP and, additionally, both system sum-rate (i.e. $R_A + R_B$) and system fairness (i.e. $\min(R_A, R_B)$) are improved for this channel realization.

2) *Average sum-rate and min-rate performance*: In this section we show simulation results averaged over 1000 random channel realizations when varying η for $\text{SNR}=10\text{dB}$, OPA, and different antenna configurations. The following schemes are evaluated:

- **proper (LP)**: transmitters use LP (i.e. $\alpha = 0$ at TXB in



(a) UPA



(b) OPA

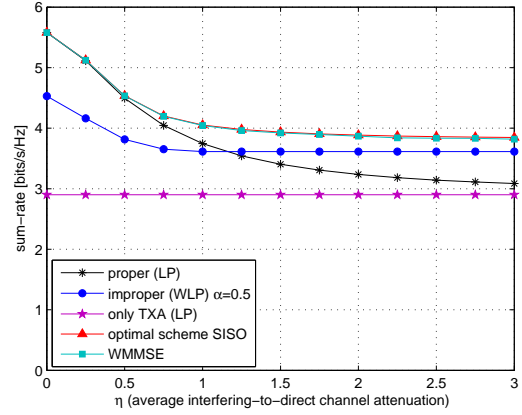
 Fig. 7: Achievable rate region of a MIMO Z-IC with $\mathbf{H}_{A,A}$, $\mathbf{H}_{B,B}$, $\mathbf{H}_{B,A}$ in (41). 2×2 , SNR=10dB and SNR=20dB. (a) UPA, (b) OPA.

(32)) and maximum power.

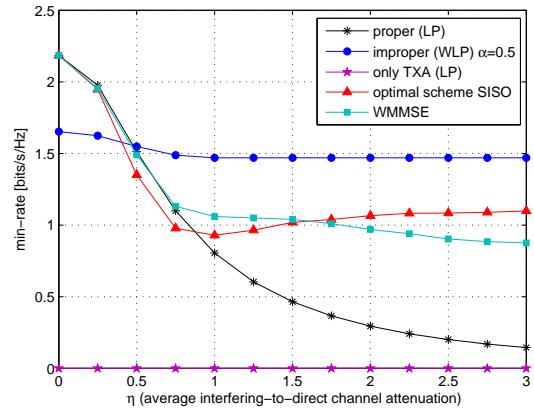
- **improper (WLP) $\alpha=0.5$** : transmitters use WLP and maximum power, whereby for TXB: $\alpha = 0.5$ and the anti-identity solution for \mathbf{Z}_1 and \mathbf{Z}_2 (i.e. Option 2 in (36)) are used in (32).
- **only TXA (LP)**: TXA transmits through optimal LP design, and TXB is turned off.
- **optimal scheme SISO**: optimal sum-rate scheme presented in [27] for the SISO Z-IC. It is valid for $M = N = 1$ and encompasses PGS and IGS solutions.
- **WMMSE**: weighted minimum MSE algorithm in [34] for sum-rate maximization in MIMO IC. To adopt IGS, the complex-valued MIMO Z-IC is transformed into an equivalent double-sized real-valued MIMO Z-IC where the WMMSE algorithm applies. It can be employed for any antenna configuration, encompasses PGS and IGS solutions, and converges to a local optimum.

Fig. 8 displays the sum-rate (i.e. $R_A + R_B$) and the min-rate (i.e. $\min(R_A, R_B)$) versus η for 1×1 . Fig. 9 and Fig. 10 show the same for antenna configurations 2×2 and 4×4 , respectively, so as to see how the system performance scales with the number of transmit/receive antennas.

Note that 'optimal scheme SISO' and 'WMMSE' schemes are used as benchmarks in terms of sum-rate. They involve



(a) sum-rate



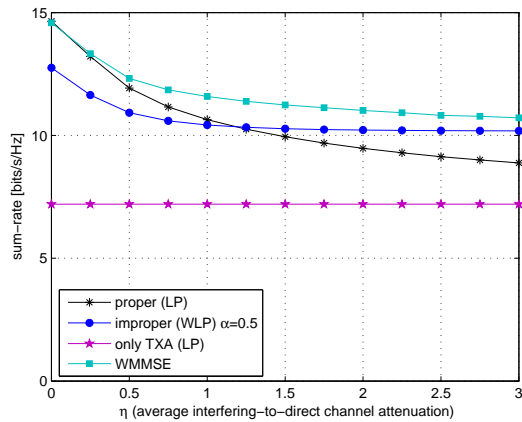
(b) min-rate

 Fig. 8: Average sum-rate and min-rate (in bits/s/Hz) of the MIMO Z-IC in Fig. 6 versus η for 1×1 OPA, SNR=10dB.

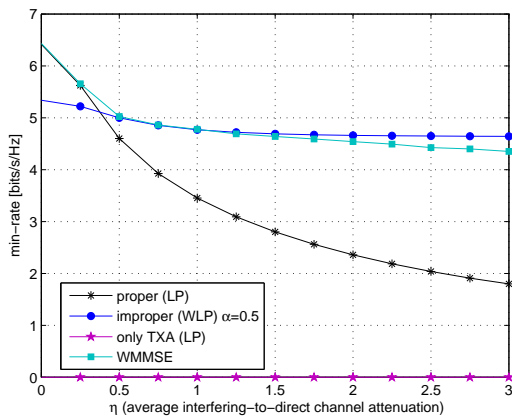
larger complexity, coordination among transmitters and knowledge of all channel matrices. On the other hand, 'proper (LP)', 'improper (WLP) $\alpha=0.5$ ', and 'only TXA' do only require knowledge of the direct channels and no coordination among transmitters is required.

Let us compare the proposed 'improper (WLP) $\alpha=0.5$ ' with 'proper (LP)'. In the 1×1 antenna case (see Fig. 8), it can be observed that the average sum-rate of the system is increased with IGS for values of $\eta \geq 1.25$, but for values of $\eta \geq 0.5$ the use of IGS starts to be important in some channel realizations (as shown by 'optimal scheme SISO'). On the other hand, the average fairness of the system is increased with IGS for values of $\eta \geq 0.5$. In the 2×2 antenna case (see Fig. 9), the average sum-rate is improved with IGS for $\eta \geq 1.5$ and the average fairness for $\eta \geq 0.5$. Finally, in the 4×4 antenna case (see Fig. 10), the average sum-rate is improved for $\eta \geq 2.75$ and the average fairness for $\eta \geq 0.75$.

Therefore, the use of IGS is beneficial when interference exceeds a certain threshold, where the proposed 'improper (WLP) $\alpha=0.5$ ' allows trading in transmission fairness and system sum-rate. Such threshold increases as the number of transmit/receive antennas increase, due to the fact that the larger is the number of transmit/receive antennas the larger are the dimensions of the system and the lower are the gains of the flexibility provided by IGS.



(a) sum-rate



(b) min-rate

 Fig. 9: Average sum-rate and min-rate (in bits/s/Hz) of the MIMO Z-IC in Fig. 6 versus η for 2×2 OPA, SNR=10dB.

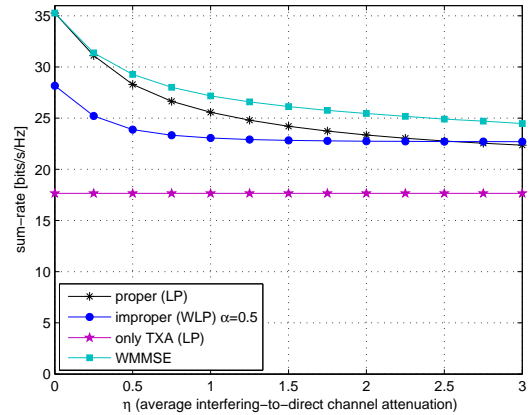
In the SISO case (1×1), the sum-rate performance of the proposed 'improper (WLP) $\alpha = 0.5$ ' is close to the optimal sum-rate performance of 'optimal scheme SISO' (see Fig. 8.(a)), while the average min-rate is improved (see Fig. 8.(b)). Recall that the proposed scheme has much less complexity when α is fixed, as it does not require knowledge of the interfering channel. Further, it can be adopted in any MIMO system such that $M \geq N$.

V. APPLICATION TO HCNs

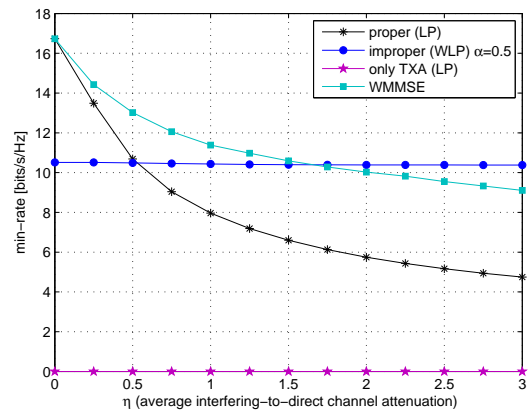
In this section we show how to exploit the benefits of using IGS in HCNs through the simple WLP design presented in Section IV. Let us focus on the deployment shown in Fig. 2 with one MeNB and multiple SeNBs. Assume orthogonal frequency division multiple access (OFDMA), such that on a given time/frequency resource the MeNB serves a single user (denoted by MUE) and each SeNB serves a single user (denoted by SUE). Then, the interference channel towards each SUE can be modeled by the MIMO Z-IC.

More specifically, the HCN in Fig. 2 is related to the MIMO Z-IC in Fig. 6 as follows:

- the MUE (that just observes noise) corresponds to RXB,
- each SUE (that might observe interference due to the active transmission of the MeNB on the same



(a) sum-rate



(b) min-rate

 Fig. 10: Average sum-rate and min-rate (in bits/s/Hz) of the MIMO Z-IC in Fig. 6 versus η for 4×4 OPA, SNR=10dB.

time/frequency resource) corresponds to RXA.

Notice that, due to the deployment of multiple SeNBs within the MeNB coverage area (see Fig. 2), there will appear as many MIMO Z-IC as the number of SUEs within the MeNB coverage area that are being served.

The interesting part is that the proposed improper-based scheme with WLP in subsection IV-A for TXB (i.e. the MeNB) when using maximal IGS (i.e. $\alpha = 0.5$ in (32)) can be applied independently of the interfering channels, as it is based on generating improper interference towards the SUEs (i.e. RXA) but does not depend on the specific SUE that is selected. Accordingly, the proposed improper-based scheme can be easily applied to multi-tier HCN deployments defined in 3GPP LTE-A [35] so as to increase the achievable rate of the SUEs. Furthermore, as multiple MIMO Z-IC may appear, large gains could be obtained with the use of the proposed improper-based scheme because all SUEs would be guaranteed to increase their rate according to Theorem 1 for UPA and Theorem 3 for OPA.

A. Simulation results

The proposed improper-based scheme is evaluated through Monte Carlo simulations in a simulator compliant with 3GPP

LTE-A specifications in [36]. The network consists of a multi-tier deployment where MeNBs and SeNBs use the same carrier frequency of 2GHz with 10MHz bandwidth. The Small Cell Scenario 1 is used, following deployment and simulation parameters specified in [36]. The deployment consists of 1 MeNB and 4 SeNBs that are uniformly distributed within the macrocell area. 60 users (UEs) are deployed per macrocell area, being 2/3 of them placed near the SeNBs, and the remaining UEs are uniformly distributed within the MeNB coverage area. All UEs are placed outdoor. ITU UMa and ITU UMi models with 3D distance are used for path loss and shadowing modeling for MeNB-UE and SeNB-UE links, respectively. The typical urban model is used for frequency-selective fading modeling. Maximum transmit power at MeNB is $P_{\text{MeNB}}^{\text{max}} = 46\text{dBm}$. At SeNBs we use two different maximum transmit power values for simulation: either $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$ or $P_{\text{SeNB}}^{\text{max}} = 15\text{dBm}$. Antenna gains are 17dBi at MeNB, 5dBi at SeNB, and 0dBi at UE. Noise spectral density is -174dBm/Hz . The number of antennas is $M_{\text{MeNB}} = M_{\text{SeNB}} = 2$, and $N_{\text{UE}} = 2$ for all UEs.

Cell selection at each UE is based on best downlink reference signal receive power (RSRP) [36]. A range extension bias (REB) is added at the RSRP received from each SeNB in order to expand its cell-range and offload more UEs to the SeNBs [37]. REB=10dB is used for $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$ and REB=15dB is employed for $P_{\text{SeNB}}^{\text{max}} = 15\text{dBm}$, such that around 73% of the UEs (in mean over different deployments) are offloaded to the SeNBs in both cases. Hence, 73% of the UEs are SUEs and 27% of the UEs are MUEs.

The LTE-A frame composed of 8 downlink sub-frames and 2 uplink sub-frames [35] is assumed, but only downlink is evaluated in the sequel. Full-load traffic model is used, where all UEs in the network have packets to be received. For each frame, the UEs associated to the same MeNB or SeNB are uniformly distributed in frequency domain among the available resource blocks (RBs), such that intra-cell interference is removed and only *inter-cell interference* remains (including cross-tier and co-tier interference). The power available at each MeNB or SeNB is uniformly distributed among the RBs where the UEs have been scheduled.

The following techniques are evaluated on each RB:

- **time-sharing**: time-sharing solution among MeNB and SeNBs. Different muting ratios are used, represented by $X/8$ in the figures, which means that all SeNBs transmit in X sub-frames while MeNB transmits in the $8 - X$ subsequent sub-frames.
- **eICIC ABS**: enhanced inter-cell interference coordination (eICIC) defined in LTE-A with time-division muting based on almost blank sub-frames (ABS) [38]. Different muting ratios are used, represented by $X/8$ in the figures, which means that the MeNB is muted X sub-frames where the SeNBs transmit while in the $8 - X$ subsequent sub-frames all MeNB and SeNBs transmit.
- **FR proper (LP)**: full-reuse (FR) of the frequency band and time slots for MeNB and SeNBs, using proper-based schemes (i.e. LP) at MeNB and SeNBs with OPA in (27).
- **FR improper (WLP)**: FR of the frequency band and time slots for MeNB and SeNBs, using improper-based

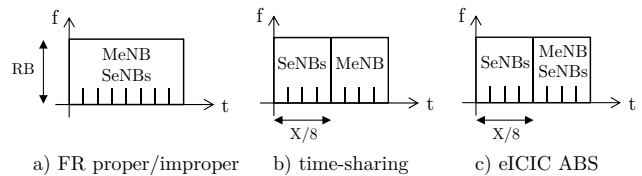


Fig. 11: Transmission schemes considered for HCNs, shown in a frequency/time grid corresponding to 1 RB and 8 downlink sub-frames.

schemes (i.e. WLP) at MeNB and SeNBs with OPA in (27). The proposed scheme in (32) is employed at MeNB with $\alpha = 0.5$ and the anti-identity solution for \mathbf{Z}_1 and \mathbf{Z}_2 in (36).

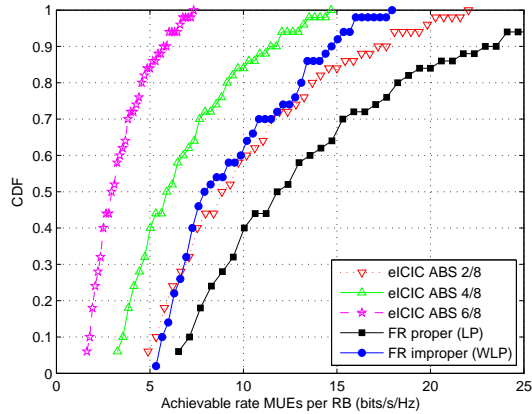
Fig. 11 displays the evaluated transmission schemes in a frequency/time grid corresponding to 1 RB and 8 downlink sub-frames, showing the sub-frames in which MeNB and SeNBs are allowed to transmit.

The performance indicator is the user throughput (UT) measured in Mbits/s. Let us remark that all the evaluated schemes have as main objective dealing with the cross-tier interference, either with time-sharing or with an efficient spatial precoding design. However, the co-tier interference (i.e. interference from SeNB towards SUEs associated to other SeNBs) can degrade the system performance in some cases depending on the deployment and the system parameters. Also, in the case of a full-reuse, the cross-tier interference from SeNB to the MUEs can also be prejudicial in some cases.

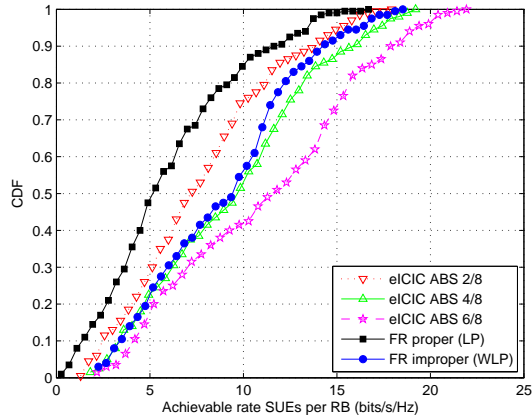
Fig. 12 displays the cumulative distribution function (CDF) of the achievable rates per RB of the MUEs and the SUEs (in bits/s/Hz), separately, for $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$. In the 'eICIC ABS' technique, as the muting ratio ($X/8$) increases, the achievable rate of the MUEs is degraded while the achievable rate of the SUEs is improved because the cross-tier interference is eliminated. As compared to the baseline 'FR proper (LP)', the proposed 'FR improper (WLP)' has a similar behavior than the 'eICIC ABS': it degrades the achievable rate of the MUEs so as to improve the achievable rate of the SUEs (as is also shown in Section IV-D). However, it can be observed that, among all techniques, 'FR improper (WLP)' achieves the fairest performance when considering all MUEs and SUEs.

Fig. 13.(a) displays the 5%-tile UT vs. the mean UT when considering all UEs for the case of $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$. The use of 'time-sharing' or 'eICIC ABS' define an oval area where the system can operate based on the variation of the muting ratio ($X/8$). As the muting ratio increases, the mean UT is increased because the SeNBs have more sub-frames available with less interference (see Fig. 11). On the contrary, there is a limit in terms of fairness (or 5%-tile UT) because when the muting ratio increases the resources devoted to the MUEs decrease and there is a switching point in which the fairness passes from being limited by SUEs to being constrained by the MUEs. However, it is important to realize that, although the 'FR proper (LP)' lies inside the oval areas delimited by the 'time-sharing' or 'eICIC ABS' solutions, the proposed 'FR improper (WLP)' allows to go out of these areas and improve both the system fairness and the system sum-rate performance.

Fig. 13.(b) displays the same as Fig. 13.(a), but for case of $P_{\text{SeNB}}^{\text{max}} = 15\text{dBm}$. Similar conclusions can be extracted.

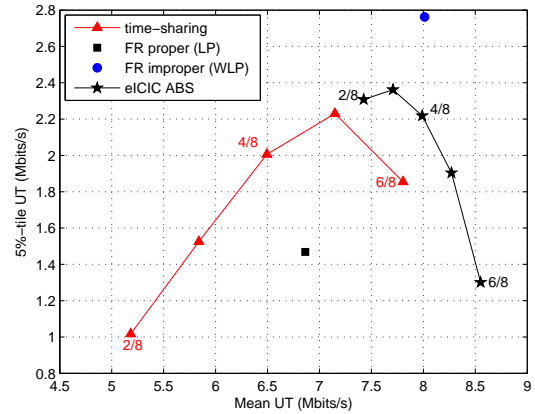


(a) MUEs

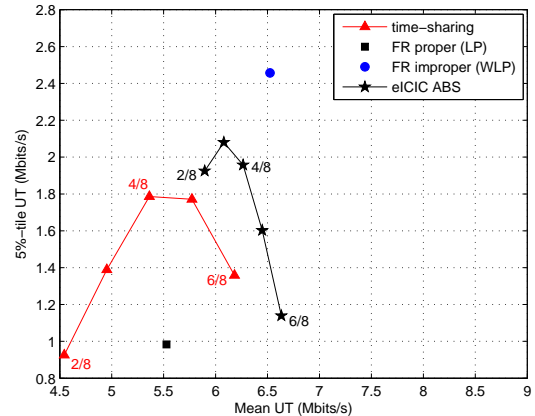


(b) SUEs

Fig. 12: CDF of the achievable rate (in bits/s/Hz) per RB of (a) MUEs and (b) SUEs. $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$.



(a) $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$



(b) $P_{\text{SeNB}}^{\text{max}} = 15\text{dBm}$

Fig. 13: 5%-tile UT vs. mean UT (in Mbits/s). Different muting ratios are displayed for time-sharing and eICIC solutions. (a) $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$, (b) $P_{\text{SeNB}}^{\text{max}} = 15\text{dBm}$.

However, the 5%-tile and the mean UT gains of 'FR improper (WLP)' are larger than in the case of $P_{\text{SeNB}}^{\text{max}} = 24\text{dBm}$, owing to the fact that when reducing the power at SeNBs the interfering-to-direct ratio for the SUEs increases (same interfering power from MeNB, but less direct power from SeNB) and hence the use of IGS provides larger system performance gains.

In addition to the system fairness and performance gains, 'FR improper (WLP)' has the following advantages as compared to 'time-sharing' or 'eICIC ABS' solutions: *i*) there is no need to optimize the muting ratio, and *ii*) the synchronization issues due to the time-sharing and the on/off switching of the MeNB do not need to be tackled. Furthermore, the proposed 'FR improper (WLP)' does not add any complexity in terms of implementability, either in overhead or in additional information required for coordination.

VI. CONCLUSIONS

This paper exploits majorization theory to formally quantify the benefits of IGS in the MIMO P2P-I, and then applies the concepts to the MIMO Z-IC and to HCNs through an efficient design of widely linear transceivers. First, for the MIMO P2P-I, we show that the use of IGS allows: *i*) strictly improving the achievable rate and the MSE and when uniform power

allocation is performed and *ii*) improving the achievable rate when optimal power allocation for maximum rate is adopted. Second, for the MIMO Z-IC, we propose a practical improper-based scheme through a simple WLP design, which allows: *i*) improving the achievable rate of the most impaired user and *ii*) improving the system fairness and controlling the sum-rate performance through a single parameter in interference-limited scenarios. The proposed improper-based scheme is useful provided that the interference levels are high enough, in which case the performance gains are significant. Finally, such benefits are applied to HCNs where multiple MIMO Z-IC appear with high interference levels. In this scenario, 3GPP-compliant simulations show that the 5%-tile and the mean user throughput are improved with the proposed improper-based scheme as compared to conventional time-sharing solutions, proper-based schemes, and the well-known eICIC technique, due to the fact that the proposed scheme can cope with a full reuse of the frequency bands and time slots while providing enough flexibility to combat the predominant cross-tier interference in HCNs.

VII. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and valuable suggestions to

improve the quality of the paper.

APPENDIX A
PRELIMINARIES

A. Preliminaries for improper Gaussian random vectors

In this section we introduce the basic notions of improper random vectors that are used through the paper (see [6] for more details).

Definition 1 ([6]): Given a zero-mean complex random vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$, the matrix $\mathbf{C}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ is the *covariance matrix* of \mathbf{x} and the matrix $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^T\}$ is the *pseudo-covariance matrix* of \mathbf{x} .

By definition, it is easy to check that the covariance matrix $\mathbf{C}_{\mathbf{x}}$ is Hermitian and positive semidefinite while the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}}$ is symmetric.

Definition 2 ([39] - proper): A complex random vector \mathbf{x} is called proper if its pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}}$ vanishes to a zero matrix, otherwise it is called improper.

A more restrictive definition than properness is known as circularly symmetric.

Definition 3 ([6] - circularly symmetric): A complex random vector \mathbf{x} is circularly symmetric if its distribution is rotationally invariant, i.e. if \mathbf{x} and $\hat{\mathbf{x}} = \mathbf{x}e^{j\alpha}$ have the same distribution for any real value α .

For a circularly symmetric random vector \mathbf{x} , we have $\tilde{\mathbf{C}}_{\mathbf{x}} = \tilde{\mathbf{C}}_{\hat{\mathbf{x}}} = \mathbb{E}\{\hat{\mathbf{x}}\hat{\mathbf{x}}^T\} = e^{j2\alpha}\tilde{\mathbf{C}}_{\mathbf{x}}$, which implies $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{0}$. Thus, circularity implies properness, but the converse is not true in general. Nevertheless, if \mathbf{x} is a zero-mean Gaussian random vector, then properness and circularity are equivalent [6], as given by the following lemma.

Lemma 2 ([6]): A complex zero-mean Gaussian random vector \mathbf{x} is circularly symmetric if and only if it is proper.

Hence, the commonly adopted assumption that the noise vector is zero-mean circularly symmetric complex Gaussian (CSCG) is equivalent to say that the noise vector is a proper Gaussian random vector.

For an arbitrary zero-mean complex random vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$, two mathematical tools have been used in the recent literature to represent them: the *composite real representation*, whereby real and imaginary parts of \mathbf{x} are separated: $\bar{\mathbf{x}} = [\Re\{\mathbf{x}\}^T \Im\{\mathbf{x}\}^T]^T \in \mathbb{R}^{2n \times 1}$, and the *augmented complex representation*, which works with the complex vector and its complex conjugate: $\underline{\mathbf{x}} = [\mathbf{x}^T (\mathbf{x}^*)^T]^T \in \mathbb{C}^{2n \times 1}$ [40]. Both representations are mathematically equivalent in the sense that it is equivalent to work with one or the other, since they are related by the following bijective transformation:

$$\bar{\mathbf{x}} = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} = \mathbf{T}\underline{\mathbf{x}}, \quad (42)$$

where

$$\mathbf{T} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \\ -j\mathbf{I}_n & j\mathbf{I}_n \end{bmatrix}. \quad (43)$$

It is very important to recall that although $\bar{\mathbf{x}}$ is a real-valued vector, \mathbf{T} and $\underline{\mathbf{x}}$ are complex-valued. Thus, $\bar{\mathbf{x}}^T = \underline{\mathbf{x}}^H \mathbf{T}^H$.

Lemma 3 ([6]): Any arbitrary zero-mean complex random vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$ is characterized by the covariance matrix of

the augmented vector $\underline{\mathbf{x}}$, which is given by:

$$\mathbf{C}_{\underline{\mathbf{x}}} = \mathbb{E}\{\underline{\mathbf{x}}\underline{\mathbf{x}}^H\} = \begin{bmatrix} \mathbf{C}_{\mathbf{x}} & \tilde{\mathbf{C}}_{\mathbf{x}} \\ \tilde{\mathbf{C}}_{\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}}^* \end{bmatrix}. \quad (44)$$

The augmented covariance matrix $\mathbf{C}_{\underline{\mathbf{x}}}$ in (44) has some built-in redundancy for the second-order characterization of \mathbf{x} . However, it is useful as shown in the following theorem.

Theorem 5 ([6]): $\mathbf{C}_{\mathbf{x}}$ and $\tilde{\mathbf{C}}_{\mathbf{x}}$ are a valid set of covariance and pseudo-covariance matrices, i.e. there exists a random vector \mathbf{x} with covariance and pseudo-covariance matrices given by $\mathbf{C}_{\mathbf{x}}$ and $\tilde{\mathbf{C}}_{\mathbf{x}}$, respectively, if and only if the augmented covariance matrix $\mathbf{C}_{\underline{\mathbf{x}}}$ is positive semidefinite (i.e. $\mathbf{C}_{\underline{\mathbf{x}}} \succeq \mathbf{0}$).

The conditions of the covariance matrix $\mathbf{C}_{\mathbf{x}}$ being Hermitian and positive semidefinite and the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}}$ being symmetric are already implied by the augmented covariance matrix $\mathbf{C}_{\underline{\mathbf{x}}}$ being positive semidefinite [6].

The relation between the covariance matrix of the augmented vector $\underline{\mathbf{x}}$ (i.e. $\mathbf{C}_{\underline{\mathbf{x}}}$ in (44)) and the covariance matrix of the real-valued vector $\bar{\mathbf{x}}$ (i.e. $\mathbf{C}_{\bar{\mathbf{x}}}$) can be derived from (42). It is included in the following lemma.

Lemma 4: The covariance matrix of the real-valued vector $\bar{\mathbf{x}}$ is related to the covariance matrix of the augmented vector $\underline{\mathbf{x}}$ as:

$$\begin{aligned} \mathbf{C}_{\bar{\mathbf{x}}} &= \mathbb{E}\{\bar{\mathbf{x}}\bar{\mathbf{x}}^T\} = \mathbb{E}\{\mathbf{T}\underline{\mathbf{x}}\underline{\mathbf{x}}^H\mathbf{T}^H\} = \mathbf{T}\mathbf{C}_{\underline{\mathbf{x}}}\mathbf{T}^H \\ &= \mathbf{T} \begin{bmatrix} \mathbf{C}_{\mathbf{x}} & \tilde{\mathbf{C}}_{\mathbf{x}} \\ \tilde{\mathbf{C}}_{\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}}^* \end{bmatrix} \mathbf{T}^H, \end{aligned} \quad (45)$$

being matrix \mathbf{T} the one defined in (43). Due to the bijective transformation in (42) and Lemma 3, $\mathbf{C}_{\bar{\mathbf{x}}}$ in (45) characterizes any arbitrary zero-mean complex random vector \mathbf{x} .

In wireless communications, the information-bearing signals are usually selected from a CSCG (i.e. proper Gaussian) codebook. Hence, the problem of how efficiently generate an improper Gaussian signal $\mathbf{x} \in \mathbb{C}^{n \times 1}$ from a proper Gaussian signal $\mathbf{b} \in \mathbb{C}^{m \times 1}$ selected from a CSCG codebook arises. Without loss of generality, assume that $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_m)$, i.e. $\mathbf{C}_{\mathbf{b}} = \mathbf{I}_m$ and $\tilde{\mathbf{C}}_{\mathbf{b}} = \mathbf{0}$. Note that the conventional linear precoding, i.e. $\mathbf{x} = \mathbf{W}\mathbf{b}$ where $\mathbf{W} \in \mathbb{C}^{n \times m}$ is the precoding matrix, is not able to map the proper Gaussian signal \mathbf{b} into an improper Gaussian signal \mathbf{x} , since the pseudo-covariance matrix of \mathbf{x} vanishes: $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{W}\tilde{\mathbf{C}}_{\mathbf{b}}\mathbf{W}^T = \mathbf{0}$. However, widely linear precoding is able to do it [4]. The main result is summarized in the following lemma.

Lemma 5 ([4]): Given a proper Gaussian information-bearing signal $\mathbf{b} \in \mathbb{C}^{m \times 1}$ characterized by $\mathbf{C}_{\mathbf{b}} = \mathbf{I}_m$ and $\tilde{\mathbf{C}}_{\mathbf{b}} = \mathbf{0}$, an improper Gaussian signal $\mathbf{x} \in \mathbb{C}^{n \times 1}$ can be obtained by applying the following *widely linear precoding* (WLP) to the proper Gaussian signal \mathbf{b} :

$$\mathbf{x} = \mathbf{W}_1\mathbf{b} + \mathbf{W}_2\mathbf{b}^*, \quad (46)$$

where $\mathbf{W}_1 \in \mathbb{C}^{n \times m}$ and $\mathbf{W}_2 \in \mathbb{C}^{n \times m}$ are the precoding matrices corresponding to the blocks of $\mathbf{C}_{\underline{\mathbf{x}}}$ in (44):

$$\mathbf{C}_{\underline{\mathbf{x}}}^{\frac{1}{2}} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{W}_2^* & \mathbf{W}_1^* \end{bmatrix}. \quad (47)$$

Therefore, as $\mathbf{C}_{\mathbf{x}}$ characterizes any arbitrary zero-mean complex random vector, any improper Gaussian signal can be generated from a proper Gaussian signal through the use of WLP [4, Sect. II.C]. Further, Lemma 5 states the relation between the covariance and pseudo-covariance matrices of \mathbf{x} ($\mathbf{C}_{\mathbf{x}}$ and $\tilde{\mathbf{C}}_{\mathbf{x}}$) and the precoding matrices (\mathbf{W}_1 and \mathbf{W}_2):

$$\mathbf{C}_{\mathbf{x}} = \mathbf{W}_1 \mathbf{W}_1^H + \mathbf{W}_2 \mathbf{W}_2^H, \quad \tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{W}_1 \mathbf{W}_2^T + \mathbf{W}_2 \mathbf{W}_1^T. \quad (48)$$

Note that if $\mathbf{W}_2 = \mathbf{0}$ (and hence $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{0}$), then (46) reduces to the conventional linear precoding used for PGS.

B. Preliminaries for majorization theory

In this section we introduce the basic notions of majorization theory that are used through the paper. See [23] for a complete reference and [24] for its applications in wireless communications. Majorization theory makes precise the vague notion that the components of a vector \mathbf{y} are "more spread out" or "less nearly equal" than are the components of a vector \mathbf{x} .

Definition 4: For any real-valued vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$, let

$$\mathbf{x}_{[1]} \geq \cdots \geq \mathbf{x}_{[n]} \quad (49)$$

denote the components of vector \mathbf{x} in decreasing order. Similarly, let

$$\mathbf{x}_{(1)} \leq \cdots \leq \mathbf{x}_{(n)} \quad (50)$$

denote the components of vector \mathbf{x} in increasing order.

Under this ordering of vectors, different kinds of majorization relations arise when comparing two vectors: strong majorization, weak majorization, and weak log-majorization, as defined in what follows.

Definition 5 ([23] - strong majorization): For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$, \mathbf{y} majorizes \mathbf{x} (or \mathbf{x} is majorized by \mathbf{y}), written as $\mathbf{y} \succ \mathbf{x}$ (or, equivalently, $\mathbf{x} \prec \mathbf{y}$), if:

$$\begin{aligned} \sum_{i=1}^m \mathbf{y}_{[i]} &\geq \sum_{i=1}^m \mathbf{x}_{[i]} \quad m = 1, \dots, n-1 \\ \text{and} \quad \sum_{i=1}^n \mathbf{y}_{[i]} &= \sum_{i=1}^n \mathbf{x}_{[i]}. \end{aligned} \quad (51)$$

The conditions in (51) are equivalent to:

$$\begin{aligned} \sum_{i=1}^m \mathbf{y}_{(i)} &\leq \sum_{i=1}^m \mathbf{x}_{(i)} \quad m = 1, \dots, n-1 \\ \text{and} \quad \sum_{i=1}^n \mathbf{y}_{(i)} &= \sum_{i=1}^n \mathbf{x}_{(i)}, \end{aligned} \quad (52)$$

Definition 6 ([23] - weak majorization): For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$, \mathbf{y} weakly majorizes \mathbf{x} (or \mathbf{x} is weakly majorized by \mathbf{y}), written as $\mathbf{y} \succ_w \mathbf{x}$ (or, equivalently, $\mathbf{x} \prec_w \mathbf{y}$), if:

$$\sum_{i=1}^m \mathbf{y}_{[i]} \geq \sum_{i=1}^m \mathbf{x}_{[i]} \quad m = 1, \dots, n. \quad (53)$$

Note that $\mathbf{y} \succ \mathbf{x}$ implies $\mathbf{y} \succ_w \mathbf{x}$, so strong majorization is a more restrictive definition than weak majorization.

Definition 7 ([23] - weak log-majorization): For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$, \mathbf{y} weakly log-majorizes \mathbf{x} (or \mathbf{x} is weakly log-majorized by \mathbf{y}), written as $\mathbf{y} \succ_{w \log} \mathbf{x}$ (or, equivalently, $\mathbf{x} \prec_{w \log} \mathbf{y}$), if:

$$\prod_{i=1}^m \mathbf{y}_{[i]} \geq \prod_{i=1}^m \mathbf{x}_{[i]} \quad m = 1, \dots, n. \quad (54)$$

Note that $\mathbf{y} \succ_{w \log} \mathbf{x}$ implies $\mathbf{y} \succ_w \mathbf{x}$, so weak log-majorization is a stronger definition than weak majorization.

Functions that preserve the ordering of majorization are said to be Schur-convex (or Schur-concave if the order is reverted).

Definition 8 ([23] - Schur-convex/concave): A real-valued function ϕ defined on a set $\mathcal{A} \in \mathbb{R}^{n \times 1}$ is said to be Schur-convex on \mathcal{A} if:

$$\mathbf{y} \succ \mathbf{x} \text{ on } \mathcal{A} \Rightarrow \phi(\mathbf{y}) \geq \phi(\mathbf{x}). \quad (55)$$

Similarly, ϕ is said to be Schur-concave on \mathcal{A} if:

$$\mathbf{y} \succ \mathbf{x} \text{ on } \mathcal{A} \Rightarrow \phi(\mathbf{y}) \leq \phi(\mathbf{x}). \quad (56)$$

The Schur-convex or Schur-concave property of a function can be identified as follows.

Proposition 1 ([23]): If $\mathbb{I} \in \mathbb{R}$ is an interval and $g : \mathbb{I} \rightarrow \mathbb{R}$ is a convex function, then:

$$\phi(\mathbf{x}) = \sum_{i=1}^n g(x_{[i]}) \quad (57)$$

is Schur-convex on $\mathbb{I}^{n \times 1}$. Similarly, if $g : \mathbb{I} \rightarrow \mathbb{R}$ is a concave function, then: $\phi(\mathbf{x}) = \sum_{i=1}^n g(x_{[i]})$ is Schur-concave on $\mathbb{I}^{n \times 1}$.

There is a stronger version of Schur-convexity and Schur-concavity for functions that preserve the ordering of majorization with a positive increment, as introduced what follows.

Definition 9 ([41] - Strongly Schur-convex/concave): A real-valued function ϕ defined on a set $\mathcal{A} \in \mathbb{R}^n$ is said to be strongly Schur-convex with modulus $c > 0$ on \mathcal{A} if:

$$\mathbf{y} \succ \mathbf{x} \text{ on } \mathcal{A} \Rightarrow \phi(\mathbf{y}) \geq \phi(\mathbf{x}) + c(\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2). \quad (58)$$

Similarly, ϕ is said to be strongly Schur-concave with modulus $c > 0$ on \mathcal{A} if:

$$\mathbf{y} \succ \mathbf{x} \text{ on } \mathcal{A} \Rightarrow \phi(\mathbf{y}) \leq \phi(\mathbf{x}) - c(\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2). \quad (59)$$

Note that the usual Schur-convexity and Schur-concavity correspond to the case $c = 0$.

The strongly Schur-convex or strongly Schur-concave property of a function can be identified as follows.

Proposition 2 ([41]): If $\mathbb{I} \in \mathbb{R}$ is an interval and $g : \mathbb{I} \rightarrow \mathbb{R}$ is a strongly convex function with modulus $c > 0$, then:

$$\phi(\mathbf{x}) = \sum_{i=1}^n g(x_{[i]}) \quad (60)$$

is strongly Schur-convex with modulus $c > 0$ on $\mathbb{I}^{n \times 1}$. Similarly, if $g : \mathbb{I} \rightarrow \mathbb{R}$ is a strongly concave function with modulus $c > 0$, then: $\phi(\mathbf{x}) = \sum_{i=1}^n g(x_{[i]})$ is strongly Schur-concave with modulus $c > 0$ on $\mathbb{I}^{n \times 1}$.

Now let us present an important majorization result for partitioned Hermitian matrices which states that, for two Hermitian matrices with equal diagonal blocks, the stronger the off-diagonal blocks are the more spread out the eigenvalues become, see [6, Result A3.7]. This result can also be derived from the pinching inequality, see [42].

Lemma 6 ([6] - partitioned Hermitian matrices): For Hermitian matrices \mathbf{A} and \mathbf{B} in the form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^H & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}, \quad (61)$$

the following strong majorization result is fulfilled:

$$\text{eig}(\mathbf{A}) \succ \text{eig}(\mathbf{B}). \quad (62)$$

APPENDIX B PROOFS

A. Proof of Lemma 1

Assume that the channel coefficients of $\mathbf{H} \in \mathbb{C}^{N \times M}$ in (1) follow a Rayleigh distribution (i.e. \mathbf{H} is full rank) and $M \geq N$. Then, let us write the vector of eigenvalues in (16) as follows:

$$\begin{aligned} \bar{\lambda}_P &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_s^* \end{bmatrix}^{-1} \frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H} \right), \\ \bar{\lambda}_I &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s & \tilde{\mathbf{C}}_s \\ \tilde{\mathbf{C}}_s^* & \mathbf{C}_s^* \end{bmatrix}^{-1} \frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H} \right), \end{aligned} \quad (63)$$

in which we have used the fact that $\text{eig}(\mathbf{XY}) = \text{eig}(\mathbf{YX})$ for any matrix \mathbf{X} and \mathbf{Y} . Note that the block-matrices $[\mathbf{C}_s \ \mathbf{0}; \ \mathbf{0} \ \mathbf{C}_s^*]$ and $[\mathbf{C}_s \ \tilde{\mathbf{C}}_s; \ \tilde{\mathbf{C}}_s^* \ \mathbf{C}_s^*]$ in (63) are full rank, so their inverses exist. By using the structure of $\bar{\mathbf{H}}$ in (4), the structure of \mathbf{T} in (7), and some matrix manipulations, it can be shown that $\frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H}$ in (63) is equal to:

$$\frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H} = \begin{bmatrix} \mathbf{H} \mathbf{H}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^* \mathbf{H}^T \end{bmatrix}. \quad (64)$$

This relation is key for the proof. Note that $\frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H}$ in (64) is full rank for $M \geq N$ and $\bar{\mathbf{H}}$ full rank, so its inverse exists and it is equal to $(\frac{1}{2} \mathbf{T}^{-1} \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{T}^{-H})^{-1} = [(\mathbf{H} \mathbf{H}^H)^{-1} \ \mathbf{0}; \ \mathbf{0} \ (\mathbf{H}^* \mathbf{H}^T)^{-1}]$.

The inverses of the positive eigenvalues in (63) (i.e. $\bar{\lambda}_P^{-1}$ and $\bar{\lambda}_I^{-1}$) can be obtained from the positive eigenvalues of the inverse matrix. Then, using the relation in (64), we get:

$$\begin{aligned} \bar{\lambda}_P^{-1} &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_s^* \end{bmatrix} \begin{bmatrix} (\mathbf{H} \mathbf{H}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}^* \mathbf{H}^T)^{-1} \end{bmatrix} \right), \\ \bar{\lambda}_I^{-1} &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s & \tilde{\mathbf{C}}_s \\ \tilde{\mathbf{C}}_s^* & \mathbf{C}_s^* \end{bmatrix} \begin{bmatrix} (\mathbf{H} \mathbf{H}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}^* \mathbf{H}^T)^{-1} \end{bmatrix} \right), \end{aligned} \quad (65)$$

such that, multiplying the block-matrices, we have:

$$\begin{aligned} \bar{\lambda}_P^{-1} &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s (\mathbf{H} \mathbf{H}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_s^* (\mathbf{H}^* \mathbf{H}^T)^{-1} \end{bmatrix} \right), \\ \bar{\lambda}_I^{-1} &= \text{eig} \left(\begin{bmatrix} \mathbf{C}_s (\mathbf{H} \mathbf{H}^H)^{-1} & \tilde{\mathbf{C}}_s (\mathbf{H}^* \mathbf{H}^T)^{-1} \\ \tilde{\mathbf{C}}_s^* (\mathbf{H} \mathbf{H}^H)^{-1} & \mathbf{C}_s^* (\mathbf{H}^* \mathbf{H}^T)^{-1} \end{bmatrix} \right). \end{aligned} \quad (66)$$

Therefore, as the diagonal blocks of the partitioned Hermitian matrices in (66) are equal, we can make use of Lemma 6 in Appendix A-B and hence, from (66), obtain:

$$\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}, \quad (67)$$

which demonstrates the strong majorization result in (17) of Lemma 1.

The weak log-majorization result in (17) of Lemma 1 is demonstrated in continuation by using some useful majorization theory properties that can be derived from (67). The first useful majorization theory property is [23, Sect. 5.A.1.d]: if $\text{eig}(\mathbf{A}) \succ \text{eig}(\mathbf{B})$, then $\prod_{i=k}^K \text{eig}(\mathbf{A})_i \leq \prod_{i=k}^K \text{eig}(\mathbf{B})_i$, $k = 1, \dots, K$, being $\text{eig}(\mathbf{A})_i$ the i -th eigenvalue of \mathbf{A} and K the rank

of \mathbf{A} . Due to the ordering of eigenvalues, the last components of the vectors in (67) correspond to the first components of the eigenvalues in (63), such that as an implication of (67):

$$\prod_{i=1}^n \frac{1}{\bar{\lambda}_{I,i}} \leq \prod_{i=1}^n \frac{1}{\bar{\lambda}_{P,i}} \quad n = 1, \dots, 2N, \quad (68)$$

and hence:

$$\prod_{i=1}^n \bar{\lambda}_{I,i} \geq \prod_{i=1}^n \bar{\lambda}_{P,i} \quad n = 1, \dots, 2N. \quad (69)$$

The second useful majorization theory property is [23, Sect. 5.A.1]: if $\text{eig}(\mathbf{A}) \succ \text{eig}(\mathbf{B})$, then we can apply a convex function $g(\cdot)$ over each component of the vectors and the following is satisfied: $g(\text{eig}(\mathbf{A})) \succ_w g(\text{eig}(\mathbf{B}))$. As $g(x) = \frac{1}{x}$ is a convex function, it follows from (67) that:

$$\sum_{i=1}^n \bar{\lambda}_{I,i} \geq \sum_{i=1}^n \bar{\lambda}_{P,i} \quad n = 1, \dots, 2N. \quad (70)$$

Again, the ordering of eigenvalues is very important here. Finally, from (69) and (70), the weak log-majorization result in (17) is demonstrated (see Definition 7 in Appendix A-B).

B. Proof of Theorem 1

The function $\frac{1}{2} \log_2(1 + \frac{\bar{P}}{x})$ is a convex function on $x \geq 0$ for $\bar{P} > 0$. Thus, as convex functions generate Schur-convex sums (see Proposition 1 in Appendix A-B), the function $\sum_i \frac{1}{2} \log_2(1 + \frac{\bar{P}}{x_i})$ is a Schur-convex function. The achievable rate in (21) can be written in such a form with $x_i = \frac{1}{\lambda_i}$: $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2(1 + \frac{\bar{P}}{1/\lambda_i})$, so it is a Schur-convex function on $\bar{\lambda}^{-1}$ for $\bar{P} > 0$. Therefore, due to Lemma 1 ($\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) and as the achievable rate is a Schur-convex function on $\bar{\lambda}^{-1}$, by majorization theory on Schur-convex functions (see Definition 8 in Appendix A-B) we get:

$$R^{\text{opt}}(\bar{P} \bar{\lambda}_I) \geq R^{\text{opt}}(\bar{P} \bar{\lambda}_P). \quad (71)$$

To further extend this inequality, we make use of the results from strong Schur-convexity (see Definition 9 in Appendix A-B) [41]. Inequality in (23) of Theorem 1 is obtained by showing that the rate expression $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2(1 + \frac{\bar{P}}{x_i})$ with $x_i = \frac{1}{\lambda_i}$ is a strongly Schur-convex function with modulus c_R^{upa} . As strongly convex functions generate strongly Schur-convex sums (see Proposition 2 in Appendix A-B), we need to prove that $\frac{1}{2} \log_2(1 + \frac{\bar{P}}{x})$ is strongly convex³ with modulus c_R^{upa} on interval $x \in [0, x_{\max}]$ or, equivalently, to show that $\frac{1}{2} \log_2(1 + \frac{\bar{P}}{x}) - c_R^{\text{upa}} x^2$ is convex on the interval $x \in [0, x_{\max}]$ [41]. By checking the second order derivative, it is easy to show that $\frac{1}{2} \log_2(1 + \frac{\bar{P}}{x}) - c_R^{\text{upa}} x^2$ is convex for $c_R^{\text{upa}} \leq \frac{\bar{P}(x+0.5\bar{P})}{2 \ln(2) x^2 (x+\bar{P})^2}$. Therefore, as the values of $x_i = \frac{1}{\lambda_i}$ are upper bounded by the minimum positive eigenvalue $x_{\max} = \frac{1}{\min(\lambda_i)}$, there exists an interval $x \in [0, x_{\max}]$ in which $\frac{1}{2} \log_2(1 + \frac{\bar{P}}{x})$ is strongly convex with modulus:

$$c_R^{\text{upa}} = \frac{\bar{P}(x_{\max} + 0.5\bar{P})}{2 \ln(2) x_{\max}^2 (x_{\max} + \bar{P})^2} > 0. \quad (72)$$

Consequently, due to Lemma 1 ($\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) and as the achievable rate is a strongly Schur-convex function on $\bar{\lambda}^{-1}$

³A twice continuously differentiable function $f : (a, b) \rightarrow \mathbb{R}$ is strongly convex with modulus m if and only if $f''(x) \geq m > 0$ for $x \in (a, b)$.

with modulus c_R^{upa} (see Definition 9 in Appendix A-B):

$$R^{\text{opt}}(\bar{P}\bar{\lambda}_I) \geq R^{\text{opt}}(\bar{P}\bar{\lambda}_P) + c_R^{\text{upa}} (\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2). \quad (73)$$

Note that due to the strong majorization result in Lemma 1: $\frac{1}{\min(\bar{\lambda}_P)} \leq \frac{1}{\min(\bar{\lambda}_I)}$, such that the interval $[0, x_{\max}]$ is determined by the IGS case. Therefore, by setting $x_{\max} = \frac{1}{\min(\bar{\lambda}_I)}$ in (72), c_R^{upa} in (24) is derived. Finally, as the squared 2-norm function is a Schur-concave function then $(\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2) \geq 0$ (with equality if and only if $\bar{\lambda}_I = \bar{\lambda}_P$) [41]. So, the rate gap in (23) is strictly positive provided that $\bar{\lambda}_I \neq \bar{\lambda}_P$.

C. Proof of Theorem 2

The function $\frac{1}{2} \frac{x}{(P+x)}$ is a concave function on $x \geq 0$ for $\bar{P} > 0$. Then, as concave functions generate Schur-concave sums (see Proposition 1 in Appendix A-B), the function $\sum_i \frac{1}{2} \frac{x_i}{(1+x_i)}$ is a *Schur-concave* function. The MSE in (22) can be written in such a form with $x_i = \frac{1}{\lambda_i}$: $\epsilon^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \frac{1/\lambda_i}{(P+1/\lambda_i)}$, so it is a Schur-concave function on $\bar{\lambda}^{-1}$ for $\bar{P} > 0$. Therefore, due to Lemma 1 ($\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) and as the MSE is a Schur-concave function on $\bar{\lambda}^{-1}$, by majorization theory on Schur-concave functions (see Definition 8 in Appendix A-B) we have:

$$\epsilon^{\text{opt}}(\bar{P}\bar{\lambda}_I) \leq \epsilon^{\text{opt}}(\bar{P}\bar{\lambda}_P). \quad (74)$$

To further extend this inequality, we make use of the results from strong Schur-concavity (see Definition 9 in Appendix A-B) [41]. Inequality in (25) of Theorem 2 is obtained by showing that the error expression $\epsilon^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \frac{x_i}{(P+x_i)}$ with $x_i = \frac{1}{\lambda_i}$ is a *strongly Schur-concave* function with modulus c_ϵ^{upa} . As strongly concave functions generate strongly Schur-concave sums (see Proposition 2 in Appendix A-B), we need to prove that $\frac{1}{2} \frac{x}{(P+x)}$ is strongly concave⁴ with modulus c_ϵ^{upa} on interval $x \in [0, x_{\max}]$ or, equivalently, to show that $\frac{1}{2} \frac{x}{(P+x)} + c_\epsilon^{\text{upa}} x^2$ is concave on interval $x \in [0, x_{\max}]$ [41]. By checking the second order derivative, one can show that $\frac{1}{2} \frac{x}{(P+x)} + c_\epsilon^{\text{upa}} x^2$ is concave for $c_\epsilon^{\text{upa}} \leq \frac{\bar{P}}{2(x+\bar{P})^3}$. Therefore, as the values of $x_i = \frac{1}{\lambda_i}$ are upper bounded by the minimum positive eigenvalue $x_{\max} = \frac{1}{\min(\lambda_i)}$, there exists an interval $x \in [0, x_{\max}]$ in which the function $\frac{1}{2} \frac{x}{(P+x)}$ is strongly concave with modulus:

$$c_\epsilon^{\text{upa}} = \frac{\bar{P}}{2(x_{\max} + \bar{P})^3} > 0. \quad (75)$$

Consequently, due to Lemma 1 ($\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) and as the MSE is a strongly Schur-concave function on $\bar{\lambda}^{-1}$ with modulus c_ϵ^{upa} (see Definition 9 in Appendix A-B):

$$\epsilon^{\text{opt}}(\bar{P}\bar{\lambda}_I) \leq \epsilon^{\text{opt}}(\bar{P}\bar{\lambda}_P) - c_\epsilon^{\text{upa}} (\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2). \quad (76)$$

By setting $x_{\max} = \frac{1}{\min(\bar{\lambda}_I)}$ in (75), c_ϵ^{upa} in (26) is derived. Also, as $(\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2) \geq 0$, the error gap in (25) is strictly positive provided that $\bar{\lambda}_I \neq \bar{\lambda}_P$.

⁴A twice continuously differentiable function $f : (a, b) \rightarrow \mathbb{R}$ is strongly concave with modulus m if and only if $f''(x) \leq m < 0$ for $x \in (a, b)$.

D. Proof of Theorem 3

In order to prove an achievable rate improvement with the use of IGS when OPA in (27) is adopted, let us focus on demonstrating the following inequalities:

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\lambda}_I) \geq R(\bar{\mathbf{p}}_P \circ \bar{\lambda}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\lambda}_P), \quad (77)$$

where $R(\bar{\mathbf{p}}_P \circ \bar{\lambda}_I)$ refers to the achievable rate in the improper interference case when the power allocation derived from the proper interference case is used (which is not the optimum, but a valid power allocation). Regarding the second inequality in (77), in the following we introduce some interesting properties that would allow us to prove it.

- Property 1: [23, Prop. 3.H.3.b] states that if $\mathbf{a} \succ_w \mathbf{b}$, then $\mathbf{a} \circ \mathbf{u} \succ_w \mathbf{b} \circ \mathbf{u}$ for any $\mathbf{u} \in \mathcal{D}_+$, where \mathcal{D}_+ denotes the set of vectors of length L such that $\{(u_1, \dots, u_L) : u_1 \geq \dots \geq u_L \geq 0\}$. The extension of this proposition to weak log-majorization is straightforward by realizing the properties of the product operation. So we can state that: if $\mathbf{a} \succ_{w \log} \mathbf{b}$, then $\mathbf{a} \circ \mathbf{u} \succ_{w \log} \mathbf{b} \circ \mathbf{u}$ for any $\mathbf{u} \in \mathcal{D}_+$. Accordingly, as $\bar{\mathbf{p}}_P \in \mathcal{D}_+$ and due to Lemma 1 ($\bar{\lambda}_I \succ_{w \log} \bar{\lambda}_P$), we can conclude that when using the power allocation derived from the proper interference case (i.e. $\bar{\mathbf{p}}_P$), the following relation is satisfied:

$$\bar{\mathbf{p}}_P \circ \bar{\lambda}_I \succ_{w \log} \bar{\mathbf{p}}_P \circ \bar{\lambda}_P. \quad (78)$$

- Property 2: From [43, Prop. 1.3] it can be established that if $\mathbf{a} \succ_{w \log} \mathbf{b}$, then $\sum_{i=1}^L \log(1 + a_i) \geq \sum_{i=1}^L \log(1 + b_i)$, being L the vector length, as $\log(1 + x)$ is an increasing function on $x \in [0, \infty)$ and $\log(1 + e^x)$ is a convex function on $x \in [0, \infty)$. Accordingly, as the achievable rate in (14) has such a form, and owing to Property 1 ($\bar{\mathbf{p}}_P \circ \bar{\lambda}_I \succ_{w \log} \bar{\mathbf{p}}_P \circ \bar{\lambda}_P$), we can conclude:

$$R(\bar{\mathbf{p}}_P \circ \bar{\lambda}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\lambda}_P). \quad (79)$$

Hence, the second inequality in (77) is demonstrated. The first inequality in (77) is intrinsic of the water-filling solution for OPA in (27), as for a given $\bar{\lambda}_I$ the optimal power allocation in terms of achievable rate is given by $\bar{\mathbf{p}}_I$, so: $R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\lambda}_I) \geq R(\bar{\mathbf{p}}_P \circ \bar{\lambda}_I)$ and the proof is completed.

E. Proof of Theorem 4

The function $\frac{1}{2} \log_2 \left(\frac{\mu}{x} \right)$ is a convex function on $x \geq 0$ for $\mu > 0$. Thus, as convex functions generate Schur-convex sums (see Proposition 1 in Appendix A-B), the function $\sum_i \frac{1}{2} \log_2 \left(\frac{\mu}{x_i} \right)$ is a *Schur-convex* function. The achievable rate in (29) can be written in such a form with $x_i = \frac{1}{\lambda_i}$: $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2 \left(\frac{\mu}{1/\lambda_i} \right)$, so it is a Schur-convex function on $\bar{\lambda}^{-1}$ for $\mu > 0$. Furthermore, by checking the second order derivative and following similar rationale as in Appendix B-B, it can be observed that the rate expression $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2 \left(\frac{\mu}{x_i} \right)$ is a *strongly Schur-convex* function with modulus c_R^{opa} (i.e. $\frac{1}{2} \log_2 \left(\frac{\mu}{x} \right)$ is strongly convex with modulus c_R^{opa} on interval $x \in [0, x_{\max}]$ or, equivalently, $\frac{1}{2} \log_2 \left(\frac{\mu}{x} \right) - c_R^{\text{opa}} x^2$ is convex on the interval $x \in [0, x_{\max}]$ [41]):

$$c_R^{\text{opa}} = \frac{1}{4 \ln(2) x_{\max}^2} > 0. \quad (80)$$

Consequently, due to Lemma 1 ($\bar{\lambda}_I^{-1} \succ \bar{\lambda}_P^{-1}$) and as the achievable rate is a strongly Schur-convex function on $\bar{\lambda}^{-1}$ with modulus c_R^{opa} (see Definition 9 in Appendix A-B):

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\lambda}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\lambda}_P) + c_R^{\text{opa}} (\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2). \quad (81)$$

By setting $x_{\max} = \frac{1}{\min(\bar{\lambda}_I)}$ in (80), c_R^{opa} in (31) is derived. Finally, as $(\|\bar{\lambda}_I^{-1}\|^2 - \|\bar{\lambda}_P^{-1}\|^2) \geq 0$, the rate gap in (30) is strictly positive provided that $\bar{\lambda}_I \neq \bar{\lambda}_P$.

REFERENCES

- [1] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang, "LTE-Advanced: an operator perspective," *IEEE Commun. Mag.*, vol. 50, pp. 104–114, Feb. 2012.
- [2] N. Bhushan *et al.*, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, pp. 82–89, Feb. 2014.
- [3] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling - Settling the Høst-Madsen-Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4552–4565, Sep. 2010.
- [4] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, pp. 2899–2913, Jun. 2013.
- [5] S. A. Jafar, "Interference alignment: a new look at signal dimensions in a communication network," *Foundations and Trends in Communications and Information Theory*, vol. 7, pp. 1–134, Jun. 2010.
- [6] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of er and noncircular signals*. Cambridge (UK), Cambridge University Press, 2010.
- [7] G. Tauböck, "Complex-valued random vectors and channels: entropy, divergence, and capacity," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2729–2744, May 2012.
- [8] I. Telatar, "Capacity of multi-antenna Gaussian channels," in *AT&T Technical Memorandum*, Jun. 1995.
- [9] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.
- [10] P. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.
- [11] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Lett.*, vol. 20, pp. 1134–1137, Nov. 2013.
- [12] C. Hellings and W. Utschick, "Performance gains due to improper signals in MIMO broadcast channels with widely linear transceivers," *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pp. 4379–4383, May 2013.
- [13] C. Lameiro and I. Santamaria, "Degrees-of-freedom for the 4-user SISO interference channel with improper signaling," *IEEE Int. Conf. Commun.*, pp. 3053–3057, Jun. 2013.
- [14] Z. K. M. Ho and E. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3194–3203, Sep. 2012.
- [15] Y. Zeng, R. Zhang, E. Gunawan, and Y. L. Guan, "Optimized transmission with improper Gaussian signaling in the K-user MISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 6303–6313, Dec. 2013.
- [16] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 64, pp. 652 – 664, Feb. 2016.
- [17] S. Lagen, A. Agustin, and J. Vidal, "Decentralized widely linear precoding design for the MIMO interference channel," *IEEE Global Commun. Conf.*, Dec. 2014.
- [18] C. Lameiro, I. Santamaria, and J. P. Schreier, "Analysis of maximally improper signaling schemes for underlay cognitive radio networks," *IEEE Int. Conf. Commun.*, pp. 1398–1403, Jun. 2015.
- [19] O. Amin, W. Abediseid, and M.-S. Alouini, "Outage performance of cognitive radio systems with Improper Gaussian signaling," *IEEE Int. Symp. Inf. Theory*, pp. 1851–1855, Jun. 2015.
- [20] Y. C. Yoon and H. Leib, "Maximizing SNR in improper complex noise and applications to CDMA," *IEEE Commun. Lett.*, vol. 1, pp. 5–8, Jan. 1997.
- [21] W. Gerstacker, R. Schober, and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, pp. 1512–1523, Sep. 2003.
- [22] 3GPP TR 36.866, *Study on network-assisted interference cancellation and suppression (NAICS) for LTE*, Release 12, v12.0.0, Mar. 2014.
- [23] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*. Springer Series in Statistics, 2009.
- [24] E. Jorswieck and H. Boche, "Majorization and matrix monotone functions in wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 3, pp. 553–701, Jul. 2007.
- [25] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381–2401, Sep. 2003.
- [26] M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, pp. 607–615, Sep. 1985.
- [27] E. Kurniawan and S. Sun, "Improper Gaussian signaling scheme for the Z-interference channel," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 3912 – 3923, Jul. 2015.
- [28] S. Lagen, A. Agustin, and J. Vidal, "Decentralized coordinated precoding for dense TDD small cell networks," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 4546 – 4561, Aug. 2015.
- [29] 3GPP Long Term Evolution (LTE). [Online]. Available, www.3gpp.org/.
- [30] S. Lagen, A. Agustin, and J. Vidal, "Improper Gaussian signaling for the Z-interference channel," *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, pp. 1145–1149, May 2014.
- [31] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4-5, pp. 331–551, 2007.
- [32] S. Barbarossa, *Multiantenna wireless communication systems*. Mobile Communications series, 2005.
- [33] K. M. Abadir and J. R. Magnus, *Matrix algebra*. Cambridge (UK), Cambridge University Press, 2005.
- [34] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, pp. 4331–4340, Sep. 2011.
- [35] 3rd Generation Partnership Project (3GPP). [Online]. Available, www.3gpp.org/.
- [36] 3GPP TR 36.872, *Small cell enhancements for E-UTRA and E-UTRAN - physical layer aspects*, Release 12, v12.1.0, Dec. 2013.
- [37] 3GPP R1-130856, Huawei, HiSilicon, *Evaluation assumptions for small cell enhancements*.
- [38] B. Soret, H. Wang, K. I. Pedersen, and C. Rosa, "Multicell cooperation for LTE-Advanced heterogeneous network scenarios," *IEEE Wireless Commun.*, vol. 20, pp. 27–34, Feb. 2013.
- [39] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1293–1302, Jul. 1993.
- [40] C. Hellings and W. Utschick, "Block-skew-circulant matrices in complex-valued signal processing," *IEEE Trans. Signal Process.*, vol. 63, pp. 2093–2107, Jan. 2015.
- [41] K. Nikodem, T. Rajba, and S. Wasowicz, *Functions generating strongly Schur-convex sums*. Springer International Series of Numerical Mathematics, v. 161, pp. 175–182, 2012.
- [42] R. Bhatia, *Matrix analysis*. Springer, New York, 1997.
- [43] F. Hiai, "Log-majorizations and norm inequalities for exponential operators," *Banach Center Publications*, vol. 38, no. 1, pp. 119–181, 1997.



Sandra Lagen received the M.S. degree in Telecommunication Engineering from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2011. In 2011 she joined the department of Signal Theory and Communications (TSC) at UPC as a research assistant, where she has participated in the EC funded projects FREEDOM and TROPIC. Since 2013 she is working towards her Ph.D. degree within the TSC department with an FPU grant from the Spanish Ministry of Education, Culture and Sport. Her research interests are focused on communication theory and signal processing for interference management in cellular networks.



Adrian Agustin (M'08) received the M.S. degrees in Telecommunication Engineering and Electronic Engineering and the Ph.D. degree from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2000, 2002 and 2008, respectively. From 2000 to 2002, he was with Indra-Espacio, Barcelona, working on research and development of code synchronization techniques for DS-CDMA. In 2002 he joined the Signal Theory and Communications department at UPC as a research assistant. In 2008, he became a research associate. He has participated

in the EC funded projects SATURN, ROMANTIK, FIREWORKS, ROCKET, FREEDOM, TROPIC and TUCAN3G. His research interests include wireless multi-user MIMO, interference management and interference alignment.



Josep Vidal (M'91) received the Telecommunication Engineering and the Ph. D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, where he is Professor at the Signal Theory and Communications department. His research interests are in statistical signal processing, information and communication theory, areas in which he has authored +170 journal and conference papers. Since 2002 has coordinated collaborative EC-funded projects ROMANTIK, FIREWORKS, ROCKET, FREEDOM, TROPIC and TUCAN3G,

belonging to the FP5, FP6 and FP7 programmes, all in different areas of MIMO relay communications, self-organization, cooperative transmission and heterogeneous networks. He has held research appointments with EPF Lausanne, INP Toulouse and University of Hawaii, and has organized several international workshops. From 2011 through 2014 he served as associate editor of IEEE Transactions on Signal Processing. Since 2016 he is member of the IEEE ComSoc Signal Processing for Communications and Electronics Technical Committee.