Distributed Digital and Hybrid Beamforming Schemes with MMSE-SIC Receivers for the MIMO Interference Channel

Andrea Abrardo, *Senior Member, IEEE*, Gábor Fodor, *Senior Member, IEEE*, Marco Moretti, *Member, IEEE*

***Abstract*—This paper addresses the problem of weighted sum-rate maximization and mean squared error (MSE) minimization for the multiple-input multiple-output *interference* channel. Specifically, we consider a weighted minimum MSE architecture where each receiver employs successive interference cancellation (SIC) to separate the various received data streams and derive a *hybrid beamforming* scheme, where the transmitters operate with a number of radio frequency chains smaller than the number of antennas, particularly suited for millimeter wave channels and 5G applications. To derive our proposed schemes, we first study the relationship between sum-rate maximization and weighted MSE minimization when using SIC receivers assuming fully digital beamforming. Next, we consider the important – and, as it turns out, highly non-trivial – case where the transmitters employ hybrid digital/analog beamforming, developing a distributed joint hybrid precoding and SIC-based combining algorithm. Moreover, for practical implementation, we propose a signaling scheme that utilizes a common broadcast channel and facilitates channel state information acquisition assuming minimal assistance from a central node such as a cellular base station. Numerical results show that both proposed weighted MMSE-SIC schemes exhibits great advantages with respect to their linear counterparts in terms of complexity, feedback information, and performance.**

***Index Terms*—MIMO, weighted MSE minimization, successive interference cancellation, hybrid beamforming.**

I. Introduction

Due to the rapid advancements in device technologies and the growing interest in high bit-rate and highly reliable communication services, wireless communication systems are increasingly supporting advanced multiple antenna techniques, including large scale antenna systems, high accuracy beamforming, spatial multiplexing and interference rejection combining techniques. These schemes are designed to enable high bit-rate services with high spectral and energy efficiency in in-coverage and out-of-coverage situations in various spectrum bands, including below 6 GHz and millimeter wave bands [1]. Among the new high bit-rate services, an important role is played by proximity services based on device-to-device (D2D) communications such as vehicle-to-vehicle communication and intelligent transportation systems applications [2,3]. For such services, multiple-input multiple-output (MIMO) capabilities can be a valuable technology enabler, in particular if the mobile stations (MSs) can leverage on such capabilities with minimal or no infrastructure assistance. Indeed, one of the key requirements on technologies enabling high bit-rate intelligent transportation systems services is that they should be reliable even outside cellular network coverage, and should take advantage of network assistance by the cellular infrastructure whenever it is available [4]–[6]. To this end, there is a need for *distributed* resource allocation schemes that are robust against topological changes and benefit from the use of advanced MIMO precoding and combining schemes at the MSs. In these situations, distributed resource allocation must mainly cope with the problem of interference generated by multiple MIMO transmit-receive pairs communicating in the same or overlapping spectrum.

Recognizing the requirements imposed by high bit-rate and capacity-demanding services, the problem of weighted sum-rate maximization has been extensively researched in the recent literature [7]-[12]. In particular, a key role is played by the works [8], that addresses the problem of maximizing the mutual information in MIMO broadcast channels, and [9], that extends the previous work by maximizing a generic utility function in the broad context of the MIMO interfering broadcast channel. These two papers show that it is possible to obtain a local maximum for the sum-rate by minimizing the mean squared error (MSE) between the transmitted symbols and the received decision variables, when the MSE is multiplied by a properly chosen set of weight coefficients. Despite numerous other works in this area [13]–[16], most of the research has focused on fully digital (FD) precoding and linear receiver architectures. The reason for this choice is intuitive: the authors of [8] have shown that, with some additional processing at the transmitter, it is possible to force the MSE matrix for each user to be diagonal thereby simplifying the receiver architecture.

Nevertheless, there are clear advantages in exploiting the knowledge that the receiver possesses about its intended signal and separating the transmitted streams at the receiver rather than at the transmitter. First, spatially separating the various streams at the transmitter requires the knowledge of the channel covariance matrices, which depend on the cross-channel gains between each transmitter and each receiver [8]. While the acquisition of such information could

Andrea Abrardo (abrardo@dii.unisi.it) is with Dipartimento di Ingegneria dell'Informazione, University of Siena, Italy and with CNIT (Consorzio inter-universitario per le telecomunicazioni). Gábor Fodor (gabor.fodor@ericsson.com and gaborf@kth.se) is with Ericsson Research and Royal Institute of Technology, Stockholm, Sweden. Marco Moretti (marco.moretti@unipi.it) is with Dipartimento di Ingegneria dell'Informazione, University of Pisa, Italy and with CNIT (Consorzio inter-universitario per le telecomunicazioni).

The work of Marco Moretti was supported by the University of Pisa under the PRA 2018-2019 Research Project CONCEPT. G. Fodor would like to acknowledge that part of this work has been performed in the framework of the H2020 project 5GCAR co-funded by the EU.

be considered achievable at the BS, i.e., when considering downlink scenarios as those considered in [8,9], it can be extremely difficult in a distributed MIMO interference channel. In contrast, separating the data streams at the receiver using SIC requires the availability of the channel state information (CSI) only of the direct channel, which can be easily acquired by means of a dedicated control channel between each transmitter-receiver pair. Moreover, separating the various streams at the transmitter requires a FD precoding architecture to guarantee maximum degree of flexibility in the precoding matrices design. However, FD beamforming in large scale antenna systems, as is the case of millimeter wave systems, is expensive and power consuming, since it requires each antenna to be connected with a dedicated radio frequency (RF) chain and a digital-to-analog converter. As a consequence, recent research efforts have been devoted to devising suboptimal but low-complexity hybrid beamforming structures [10,17]–[27].

Hybrid digital-analog large-scale MIMO architectures [19] allow to substantially reduce the number of RF chains and analog-to-digital converters with limited performance penalty in terms of spectral and energy efficiency and are of particular interest in mmWave systems [28]–[31]. For single user systems, [17], [20] and [28] develop transmit precoding and receiver combining that perform close to a fully digital architecture. In particular, the results of [20] and [28] suggest that for single user MIMO systems, an architecture based on reduced number of RF chains can achieve similar sum-rate performance as that of the fully digital architecture. The single cell multiuser MIMO scenario is considered in [10], which develops a hybrid digital and analog beamforming scheme that can approach the performance of the fully digital scheme while substantially reducing the number of RF chains. More recently, a doubly massive multiuser MIMO scheme employing a limited number of RF chains and hybrid digital and analog beamforming architecture is considered in [21]. An important contribution in this context is [32], which designs a SIC-based detection for an analog receiver. This design can be advantageously deployed at base stations employing large antenna arrays operating with analog beamformers. These results indicate an inherent trade-off between architecture costs and complexity in terms of the number of RF chains and the achieved spectral and energy efficiency. However, the above referenced papers do not analyze the multiuser MIMO interference channel where the hybrid precoders and combiners at the multiple transmit-receive pairs must be tuned jointly in order to achieve sum-rate optimality.

In this paper we first develop a *novel* FD beamforming algorithm, indicated as weighted minimum mean squared error (wMMSE)-SIC in the remainder of the paper. The wMMSE-SIC scheme is designed to maximize the sum-rate for the MIMO interference channel by minimizing the weighted MSE when the receivers employ SIC to remove inter-stream interference, and exhibits some nice properties, which make this scheme more attractive than the conventional linear wMMSE algorithm. Next, exploiting the potential of the wMMSE-SIC architecture, we develop a hybrid digital/analog precoding and combining algorithm, which we will denote as wMMSE-SIC-hybrid beamforming (HB).

Regarding the existing literature about hybrid beamforming for the MIMO interference channel, the work in [22] studies a similar problem to the one we address. They study sum-rate maximization in a system where each Tx and Rx node is equipped with multiple antennas and a hybrid MIMO processor, but they only consider linear MMSE receivers. The authors of [23] study the important problem of energy-efficiency maximization for the MIMO interference channel with hybrid beamforming but they also do not include interference cancelation and only perform linear processing at the receiver. A hybrid wMMSE-based precoding/combining design is also presented in [24]. The proposed scheme is shown to have advantages over two-stage hybrid precoding schemes including zero forcing, MMSE and wMMSE precoding/combining designs. However, also in this case, a SIC receiver design and the corresponding iterative precoding/combining algorithms are out of scope. Another set of very recent works is [25]–[27], where the authors develop hybrid schemes for the MIMO interference channel, but do not address the MMSE-SIC receiver structure. Of particular interest is the work in [29], where hybrid beamforming and SIC are combined in a sub-connected architecture, where each RF chain is connected to only a subset of antennas, to achieve the performance of a fully connected architecture, where each RF chain is connected to all antennas via phase shifters. Even if the objective is different from the one we pursue here, this work shows the potential of employing SIC algorithms in the baseband of a hybrid architecture.

Moreover, to practically realize the proposed schemes *in a distributed manner* and to fairly compare their performance, we adapt the well known bi-directional training scheme ([33], [34], [35]) to the proposed MIMO architecture. A thorough comparison of the results for the proposed schemes shows how they depend on the availability of reliable CSI at the receiver.

Note that the proposed architectures and associated signaling schemes are applicable in a variety of typical 5G scenarios, including D2D systems, where the resources allocated to D2D communications are disjunct from those used for traditional cellular communications, and systems that incorporate a mixture of D2D pairs and cellular MSs communicating with their respective serving base station (BS).

The main contributions of the present paper can be summarized as follows:

- We show that the intimate relationship between weighted MSE minimization and sum-rate maximization in MIMO systems can be advantageously extended to the minimum mean squared error (MMSE)-SIC case. Specifically, we prove that if the receiver employs SIC, its MSE matrix is diagonal, and, consequently, the optimal weight matrix in the weighted MSE minimization problem also becomes diagonal.
- We consider two different allocation schemes associated with SIC reception: (i) optimal MMSE design with FD beamforming; (ii) a near-optimal hybrid MMSE precoding scheme. Both schemes can operate in a distributed fashion with minimal assistance by a central entity that can broadcast synchronization signals.
- We propose a practical implementation and signaling

protocol for the proposed allocation schemes.

Both proposed schemes (wMMSE-SIC and wMMSE-SIC-HB) are implemented as iterative algorithms, where each user computes sequentially and autonomously its allocation, while the system converges to a local maximum of the sum-rate.

The remainder of this paper is organized as follows. Section II describes the system model and formulates the problem of sum-rate maximization and weighted MSE minimization. Next, Section IV develops a weighted MSE distributed iterative allocation scheme for fully digital beamforming when employing SIC at the receiver, while Section V considers the hybrid digital/analogue precoding and combining case. Section VI discusses implementation aspects including the implementation differences of the proposed schemes. Section VII presents numerical results and compares the performance of the various schemes. Finally, conclusions are drawn in Section VIII.

II. System Model And Problem Formulation

We consider a wireless network with N_u transmit-receive (tx-rx) pairs with MIMO links, where each transmitter is equipped with L antennas and each receiver is equipped with M antennas, and where the MIMO channels can be assumed constant for the time horizon of radio resource allocation. Let $\mathbf{s}_i = [s_i(1), s_i(2), \ldots, s_i(R)]^T$ be the complex vector collecting the R symbols of user i, with $R \leq \min\{M, L\}$ and $\mathbf{V}_i \in \mathbb{C}^{L\times R}$ be the precoding matrix of the i-th ($1 \leq i \leq N_u$) transmitter. The transmitted vector is

$$\mathbf{x}_i = \mathbf{V}_i \mathbf{s}_i \quad \in \mathbb{C}^{L\times 1}. \tag{1}$$

The information symbols are assumed to be zero-mean independent and identically distributed (i.i.d) random variables, i.e., $\mathbb{E}\left[\mathbf{s}_i\mathbf{s}_i^H\right] = \mathbf{I}_R$ and $\mathbb{E}\left[\mathbf{s}_i\mathbf{s}_j^H\right] = \mathbf{0}_R$ for $i \neq j$. The received signal at the i-th receiver can be expressed as:

$$\mathbf{y}_i = \mathbf{H}_{i,i}\mathbf{x}_i + \sum_{\substack{j=1\\ j\neq i}}^{N_u} \mathbf{H}_{i,j}\mathbf{x}_j + \mathbf{n}_i \quad \in \mathbb{C}^{M\times 1}, \tag{2}$$

where $\mathbf{H}_{i,i} \in \mathbb{C}^{M\times L}$ is the channel matrix between the transmitter and the receiver of the i-th pair, $\mathbf{H}_{i,j} \in \mathbb{C}^{M\times L}$ is the cross channel matrix between the transmitter of the j-th pair and the receiver of the i-th pair and $\mathbf{n}_i \in \mathbb{C}^{M\times 1}$ denotes the additive white Gaussian noise with distribution $\mathcal{CN}\left(0, \sigma_i^2 \mathbf{I}_M\right)$. To recover the transmitted symbol vector $\mathbf{s}_i$, the signals collected by the M antennas of the i-th receiver are combined employing the linear spatial filter $\mathbf{G}_i \in \mathbb{C}^{R\times M}$ to yield the decision variable vector

$$\hat{\mathbf{s}}_i = \mathbf{G}_i^H \mathbf{y}_i. \tag{3}$$

In the interference MIMO channel described above, the achievable rate $R_i = \mathrm{I}(\mathbf{s}_i; \mathbf{y}_i)$ for user i is computed according to equation (4),

$$R_i(\mathcal{V}) = \log\det\left(\mathbf{I}_R + \mathbf{V}_i^H \mathbf{H}_{i,i}^H \bar{\mathbf{J}}_i^{-1} \mathbf{H}_{i,i}\mathbf{V}_i\right), \tag{4}$$

where $\bar{\mathbf{J}}_i = \sum_{j=1, j\neq i}^{N_u} \mathbf{H}_{i,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{i,j}^H + \sigma^2\mathbf{I}_M$ is the covariance matrix of inter-user interference plus noise and $\mathcal{V} = \{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{N_u}\}$ is the set of the precoding matrices of all the N_u users in the system. Our objective is to maximize the *weighted* sum-rate of the system under a power constraint for each user, that is:

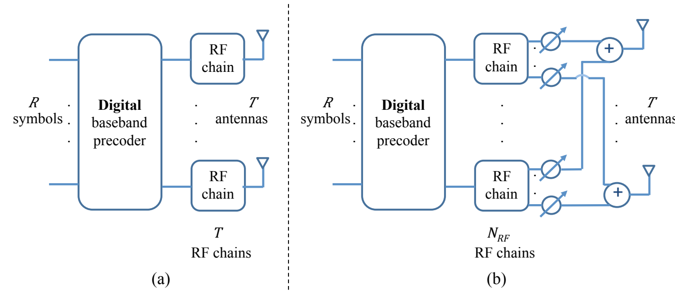

Fig. 1. Fully digital (a) and hybrid digital-analog (b) beamforming architectures. The number of RF chains employed by the hybrid architecture is reduced to $N_{RF} < T$ RF chains.

$$\begin{aligned} &\max_{\mathcal{V}} \sum_{i=1}^{N_u} \alpha_i R_i(\mathcal{V}) \\ &\quad\text{subject to} \\ &\mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^H\right) \leq P_i \quad i = 1, 2, \ldots, N_u, \end{aligned} \tag{5}$$

where the set of weights $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_{N_u}]$ is chosen to enforce a given degree of fairness [36], and a discussion about the strategy with which it is selected is beyond the scope of this work. Note that problem (5) is not convex, and it is therefore difficult to solve.

III. Background

A. Fully digital Weighted MSE Minimization for Linear Receivers

In the seminal paper by Shi *et al.* [9], the authors exploit the close relationship between the signal-to-interference-plus-noise ratio (SINR) and the MSE to find a local solution to the max-rate problem (5). In detail, the paper shows that by adding to the MSE minimization problem a new set of variables that weights the MSE matrix, the spatial transmit filters solution of the minimization problem are also the set of transmit filters that maximize the sum-rate. Accordingly, the set $\mathcal{V}^\star$ of precoding matrices that corresponds to a local optimum point for the sum-rate maximization (5) are obtained by solving the following wMMSE optimization problem:

$$\begin{aligned} &\min_{\mathcal{V},\mathcal{W},\mathcal{G}} \sum_{i=1}^{N_u} \alpha_i \left(\mathrm{tr}\left(\mathbf{W}_i\mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right)\right) - \log\det\left(\mathbf{W}_i\right)\right) \\ &\quad\text{subject to} \\ &\mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^H\right) \leq P_i \quad i = 1, 2, \ldots, N_u, \end{aligned} \tag{6}$$

where $\mathbf{W}_i \succeq 0$ is the matrix of weights for the MSE of user i, while $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{N_u}\}$, and $\mathcal{G} = \{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{N_u}\}$ are the set of all weight matrices and receive filter matrices in the system, respectively. Finally, $\mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right)$ is the $R \times R$ MSE matrix computed for receiver i

as

$$\mathbf{E}_i\left(\mathcal{V},\mathbf{G}_i\right)=\mathbb{E}\left\{\left(\mathbf{s}_i-\hat{\mathbf{s}}_i\right)\left(\mathbf{s}_i-\hat{\mathbf{s}}_i\right)^H\right\}$$
$$=\left(\mathbf{I}_R-\mathbf{G}_i^H\mathbf{H}_{i,i}\mathbf{V}_i\right)\left(\mathbf{I}_R-\mathbf{G}_i^H\mathbf{H}_{i,i}\mathbf{V}_i\right)^H+\mathbf{G}_i^H\bar{\mathbf{J}}_i\mathbf{G}_i. \quad (7)$$

Although problem (6) is still not convex, when two of the three sets of variables are fixed, it is convex in the remaining set of variables. Accordingly, the weighted MSE minimization can be performed by adopting a *block coordinate descent* [37] strategy that iteratively optimizes one of the three sets of variables at a time to find a local optimum. At convergence, the receive filter $\mathbf{G}_i^\star$ is

$$\mathbf{G}_i^\star=\mathbf{J}_i^{-1}\mathbf{H}_{i,i}\mathbf{V}_i^\star, \quad (8)$$

where $\mathbf{J}_i=\sum_{j=1}^{N_u}\mathbf{H}_{i,j}\mathbf{V}_j^\star\mathbf{V}_j^{\star H}\mathbf{H}_{i,j}^H+\sigma^2\mathbf{I}_M=\bar{\mathbf{J}}_i+\mathbf{H}_{i,i}\mathbf{V}_i^\star\mathbf{V}_i^{\star H}\mathbf{H}_{i,i}^H$ is the covariance matrix of the received signal at receiver i. The weight matrix $\mathbf{W}_i^\star$ is

$$\mathbf{W}_i^\star=\mathbf{E}_i\left(\mathcal{V}^\star,\mathbf{G}_i^\star\right)^{-1}, \quad (9)$$

and the precoder filter $\mathbf{V}_i^\star\in\mathcal{V}^\star$ is

$$\mathbf{V}_i^\star=\alpha_i\left(\mathbf{K}_i+\mu_i^\star\mathbf{I}_L\right)^{-1}\mathbf{H}_{i,i}^H\mathbf{G}_i^\star\mathbf{W}_i^\star, \quad (10)$$

where $\mathbf{K}_i=\sum_{j=1}^{N_u}\alpha_j\mathbf{H}_{j,i}^H\mathbf{G}_j^\star\mathbf{W}_j^\star\mathbf{G}_j^{\star H}\mathbf{H}_{j,i}$ and the Lagrange multiplier $\mu_i^\star$ is chosen so that the power constraint in (6) for user i is met.

B. Hybrid Weighted MSE Minimization

In the previous section we have considered a FD system, where there are as many RF chains as antennas. In a *hybrid* beamforming system, such as the one shown in Fig. 1, the number N_{RF} of RF chains is smaller than the number T of active antennas and greater or equal to the number of data streams per user R, i.e., $R\le N_{RF}<T$. In particular, in our case we consider a fully-connected architecture, where each RF chain is connected to all antennas via phase shifters. As for the FD case, our objective is to design a system that maximizes the sum-rate with a fixed power constraint by minimizing the weighted MMSE.

In the following, to avoid confusion we use three different terms for hybrid spatial processing: we use the term *beamformer*, designated by the notation $\mathbf{B}^{(HB)}$, when we consider a generic hybrid spatial filter regardless that it is employed at the transmit or at the receive side, we use the term *precoder*, designated by the notation $\mathbf{V}^{(HB)}$, for the hybrid processing at the transmitter and the term *combiner*, designated by the notation $\mathbf{G}^{(HB)}$, for the hybrid spatial filter at the receiver. The generic hybrid beamformer for the i-th pair $\mathbf{B}_i^{(HB)}$ is a $T\times R$ matrix, and, consistently with the notation previously adopted, it is $T=L$ for the precoder and $T=M$ for the combiner. Following an approach frequently proposed in the literature (see for example [21] and the references therein), the hybrid beamforming matrix can be decomposed into the product of an outer $T\times N_{RF}$ analog matrix $\mathbf{B}_i^{(AB)}=\left[\mathbf{b}_i^{(AB)}(1),\mathbf{b}_i^{(AB)}(2),\ldots,\mathbf{b}_i^{(AB)}(N_{RF})\right]$ and an inner $N_{RF}\times R$ digital matrix $\mathbf{B}_i^{(DB)}=\left[\mathbf{b}_i^{(DB)}(1),\mathbf{b}_i^{(DB)}(2),\ldots,\mathbf{b}_i^{(DB)}(R)\right]$

$$\mathbf{B}_i^{(HB)}=\mathbf{B}_i^{(AB)}\mathbf{B}_i^{(DB)}. \quad (11)$$

The elements of the outer matrix $\mathbf{B}_i^{(AB)}$ must respect a constant-modulus constraint, so that its entries can be modeled as phase shifters that can be implemented directly at RF, i.e.

$$\mathbf{B}_i^{(AB)}\in\mathcal{A}_{T,N_{RF}}, \quad (12)$$

where $\mathcal{A}_{T,N_{RF}}$ is the set of all $T\times N_{RF}$ matrices whose elements have amplitude equal to 1.

In this case, the original wMMSE problem (6) needs to be formulated by taking into account the decomposition $\mathbf{V}_i=\mathbf{V}_i^{(AB)}\mathbf{V}_i^{(DB)}$ and $\mathbf{W}_i=\mathbf{W}_i^{(AB)}\mathbf{W}_i^{(DB)}$ with the two new sets of constraints $\mathbf{V}_i^{(AB)}\in\mathcal{A}_{L,N_{RF}}$ and $\mathbf{W}_i^{(AB)}\in\mathcal{A}_{M,N_{RF}}$.

IV. Fully Digital Weighted MSE Minimization and Weighted Sum-Rate Maximization for Non-linear Receivers

A. Weighted MSE Minimization for Non-Linear Receivers

While the sum-rate maximization problem (5) does not depend on any particular receiver architecture, the authors in [8] have shown that by exploiting the knowledge of some feedback information *at the transmitter*, it is possible to greatly simplify the receiver and to separate the spatial streams arriving at each user by adopting a simple linear strategy. Alternatively, here we aim to derive a wMMSE algorithm based on SIC receivers, in which CSI knowledge available *at the receiver* is exploited to cancel the data of the streams already detected: in this case, the self interference for the rth data stream is limited to the contribution of the data streams with indexes $(r+1)$ to R and $\tilde{\mathbf{y}}_i(k)$, the received signal for data stream k $(k=1,2\ldots,R)$ after cancellation, becomes:

$$\begin{aligned}\tilde{\mathbf{y}}_i(r)&=\mathbf{y}_i-\sum_{l=1}^{r-1}\mathbf{H}_{i,i}\mathbf{v}_i(l)s_i(l)\\&=\mathbf{H}_{i,i}\mathbf{v}_i(r)s_i(r)+\sum_{l=r+1}^{R}\mathbf{H}_{i,i}\mathbf{v}_i(l)s_i(l)\\&\quad+\sum_{\substack{j=1\\j\neq i}}^{N_u}\sum_{l=1}^{R}\mathbf{H}_{i,j}\mathbf{v}_j(l)s_j(l)+\mathbf{n}_i.\end{aligned} \quad (13)$$

Note that the usage of this equation is valid when we consider the mutual information as the parameter of interest, that is when assuming that transmitting at Shannon rate allows to achieve zero error probability by employing very efficient transmission schemes (e.g., turbo or LDPC codes).

In the remainder of the paper, we will denote this approach with wMMSE-SIC, and a tilde will be placed over the letters denoting the variables related to this algorithm. This is to distinguish them from the linear receiver counterpart, indicated as wMMSE. As in the case of (6), the wMMSE-SIC problem, even if not convex, can be solved following an iterative block coordinate descent method.

1) Receive filter design: Having fixed the values of the transmit precoder in $\tilde{\mathcal{V}}$, the receiver filter $\tilde{\mathbf{g}}_i(r)$ that minimizes the MSE, $\mathbb{E}_{\mathbf{s},\mathbf{n}}\left\{|s_i(r) - \tilde{\mathbf{g}}_i^H(r)\tilde{\mathbf{y}}_i(r)|^2\right\}$, is computed locally as

$$\tilde{\mathbf{g}}_i(r) = \tilde{\mathbf{J}}_i(r)^{-1}\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(r), \tag{14}$$

where $\tilde{\mathbf{J}}_i(r) = \mathbf{J}_i - \sum_{l=1}^{r-1}\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(l)\tilde{\mathbf{v}}_i^H(l)\mathbf{H}_{i,i}^H$ is the covariance matrix of the received SIC signal $\tilde{\mathbf{y}}_i(r)$. Hence, the transmitted symbol $s_i(r)$ can be estimated as:

$$\begin{aligned}\hat{s}_i(r) =& \tilde{\mathbf{g}}_i(r)^H\tilde{\mathbf{y}}_i(r) = \tilde{\mathbf{g}}_i(r)^H\mathbf{y}_i - \\ &- \sum_{l=1}^{r-1}\tilde{\mathbf{g}}_i(r)^H\mathbf{H}_{i,i}\mathbf{v}_i(l)s_i(l)\end{aligned} \tag{15}$$

2) Weighting matrices computation: In the block coordinate descent algorithm, after having computed all the spatial filters in $\tilde{\mathcal{G}}$, one evaluates the weighting matrices in $\tilde{\mathcal{W}}$. Employing a non-linear SIC strategy at the receiver i has an important consequence on the structure of the MSE matrix $\mathbf{E}_i\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right)$ and thereby on the optimal weight matrix. Accordingly, we state the following lemma, which turns out to be useful in proving an important theorem.

Lemma 1. *If receiver i employs a SIC strategy, the MSE matrix $\mathbf{E}_i\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right)$ is diagonal.*

Proof: The generic term off the main diagonal of $\mathbf{E}_i\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right)$ is computed as

$$\begin{aligned}&\mathbb{E}\left\{\big(s_i(p) - \hat{s}_i(p)\big)\big(s_i(m) - \hat{s}_i(m)\big)^*\right\} = \\ &= \mathbb{E}\left\{s_i(p)s_i^*(m)\right\} - \mathbb{E}\left\{s_i(p)\hat{s}_i^*(m)\right\} - \mathbb{E}\left\{\hat{s}_i(p)s_i^*(m)\right\} \\ &+ \mathbb{E}\left\{\hat{s}_i(p)\hat{s}_i^*(m)\right\},\end{aligned} \tag{16}$$

with $m \neq p$ and $1 \leq m, p \leq R$. First of all, since the symbols are zero-mean i.i.d. variables, it is

$$\mathbb{E}\left\{s_i(p)s_i^*(m)\right\} = \begin{cases} 1 & \text{if } p = m \\ 0 & \text{otherwise.}\end{cases} \tag{17}$$

Hence, the first term in the right hand side of (16) is null. Moreover, it is not restrictive to assume that $m > p$, and in this case it is

$$\begin{aligned}\mathbb{E}\{s_i(p)\hat{s}_i^*(m)\} &= \mathbb{E}\left\{s_i(p)s_i^*(m)\right\}\tilde{\mathbf{v}}_i^H(m)\mathbf{H}_{i,i}^H\tilde{\mathbf{g}}_i(m) \\ &+ \sum_{l=m+1}^{R}\tilde{\mathbf{v}}_i^H(l)\mathbf{H}_{i,i}^H\tilde{\mathbf{g}}_i(m)\,\mathbb{E}\left\{s_i(p)s_i^*(l)\right\} \\ &+ \sum_{\substack{j=1\\ j\neq i}}^{N_u}\sum_{l=1}^{R}\tilde{\mathbf{v}}_j^H(l)\mathbf{H}_{i,j}^H\tilde{\mathbf{g}}_i(m)\,\mathbb{E}\left\{s_i(p)s_j^*(l)\right\} \\ &+ \mathbb{E}\left\{s_i(p)\mathbf{n}_i^H\right\}\tilde{\mathbf{g}}_i(m) = 0.\end{aligned} \tag{18}$$

For the two remaining terms in (16), it is

$$\mathbb{E}\left\{\hat{s}_i(p)s_i^*(m)\right\} = \tilde{\mathbf{g}}_i^H(p)\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(m), \tag{19}$$

and

$$\mathbb{E}\left\{\hat{s}_i(p)\hat{s}_i^*(m)\right\} = \tilde{\mathbf{g}}_i^H(p)\tilde{\mathbf{J}}_i(m)\tilde{\mathbf{g}}_i(m) = \tilde{\mathbf{g}}_i^H(p)\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(m), \tag{20}$$

where the last equality has been obtained by replacing $\tilde{\mathbf{g}}_i(m)$ with the expression in (14). Combining the results of (19) and (20) yields $\mathbb{E}\left\{\big(s_i(p) - \hat{s}_i(p)\big)\big(s_i(m) - \hat{s}_i(m)\big)^*\right\} = 0.$ ■

The m-th element of the main diagonal of the MSE matrix, which we denote with a slight abuse of notation as $\tilde{e}_i(m)$, is computed as

$$\begin{aligned}\tilde{e}_i(m) &= \mathbb{E}\left\{\big(s_i(m) - \hat{s}_i(m)\big)\big(s_i(m) - \hat{s}_i(m)\big)^*\right\} \\ &= 1 - \tilde{\mathbf{g}}_i^H(m)\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(m).\end{aligned} \tag{21}$$

We are now in the position of proving the following theorem.

Theorem 2. *The optimal weight matrix for problem* (6) *when the receiver implements a SIC strategy is diagonal.*

Proof: Following a block coordinate descent update method, the matrix that minimizes the weighted MSE is

$$\tilde{\mathbf{W}}_i = \mathbf{E}_i^{-1}\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right). \tag{22}$$

In Lemma 1 we have shown that with a SIC receive architecture $\mathbf{E}_i\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right)$ is diagonal. Since the inverse of a diagonal matrix is diagonal, $\tilde{\mathbf{W}}_i$ is diagonal. ■

Note that, just like the MMSE receive filters (14), the matrix $\tilde{\mathbf{W}}_i$ can be computed at receiver i by means of local estimation, as it will be discussed in Section VI.

3) Spatial precoder design: Let us now denote by $\tilde{w}_i(r)$ the r-th element of the main diagonal of $\tilde{\mathbf{W}}_i$. Since the MSE matrix $\mathbf{E}_i\left(\tilde{\mathcal{V}}, \tilde{\mathbf{G}}_i\right)$ depends on the value of the precoding matrices, once the receive filters are fixed, we can rearrange the terms in the objective function in (6) so that the optimization problem can be solved as N_u different independent problems. The precoding problem for user i is formulated as

$$\begin{aligned}\min_{\tilde{\mathbf{V}}_i}\ & \sum_{r=1}^{R}\Big[\alpha_i\tilde{w}_i(r)\left|1 - \tilde{\mathbf{g}}_i(r)^H\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(r)\right|^2 \\ &\qquad + \tilde{\mathbf{v}}_i(r)^H\tilde{\mathbf{K}}_i(r)\tilde{\mathbf{v}}_i(r)\Big] \\ \text{subject to}\ & \\ & \sum_{r=1}^{R}\tilde{\mathbf{v}}_i(r)^H\tilde{\mathbf{v}}_i(r) \leq P_i.\end{aligned} \tag{23}$$

where $\tilde{\mathbf{K}}_i(r) = \mathbf{K}_i - \alpha_i\sum_{l=r+1}^{R}\tilde{w}_i(l)\mathbf{H}_{i,i}^H\tilde{\mathbf{g}}_i(l)\tilde{\mathbf{g}}_i^H(l)\mathbf{H}_{i,i}$. The constrained minimization in (23) can be solved by employing standard convex methods so that the optimal precoding vectors for user i are given by

$$\tilde{\mathbf{v}}_i(r) = \alpha_i\tilde{w}_i(r)\left(\tilde{\tilde{\mathbf{K}}}_i(r) + \mu_i\mathbf{I}_L\right)^{-1}\mathbf{H}_{i,i}^H\tilde{\mathbf{g}}_i(r), \tag{24}$$

and $\tilde{\tilde{\mathbf{K}}}_i(r) = \tilde{\mathbf{K}}_i(r) - \alpha_i\tilde{w}_i(r)\mathbf{H}_{i,i}^H\tilde{\mathbf{g}}_i(r)\tilde{\mathbf{g}}_i^H(r)\mathbf{H}_{i,i}$.The value of the Lagrangian variable μ_i is chosen to satisfy the power budget constraint and can be determined numerically by means of the bisection method [38].

The weighted MSE is reduced at each iteration of the proposed wMMSE-SIC iterative procedure, thus guaranteeing robust convergence to a quasi-optimal solution. The procedure implementing the iterative block descent method is summarized in Algorithm 1, where an apex (q), displaying the specific iteration counter, is added to each variable.

Algorithm 1: Iterative wMMSE-SIC

```
1  Initialize:
2  Generate an initial precoding matrix Ṽ_i^(0), for i = 1, ..., N_u, such
     that the power constraint are met;
3  Set an arbitrarily small value for ε;
4  q ← 1, Δ ← 1;
5  while Δ > ε do
6      for i = 1, ..., N_u do
7          for r = 1, ..., R do
8              Cancel the interference of beams with index
                 1, 2, ..., r − 1;
9              Compute g̃_i^(q)(r) as in (14) employing Ṽ^(q−1) ;
10             Compute ẽ_i^(q)(r) as in (21) employing ṽ_i^(q−1) and g̃_i^(q);
11             Compute w̃_i^(q)(r) = 1/ẽ_i^(q)(r);
12     Having updated the values of G̃ and W̃
13     for i = 1, ..., N_u do
14         for r = 1, ..., R do
15             Compute ṽ_i^(q)(r) (r = 1, ..., R) as in (24);
16     Δ ← max_i ||Ṽ_i^(q) − Ṽ_i^(q−1)||;
17     q ← q + 1
```

B. Weighted Sum-Rate Maximization

We can now state an important property of the proposed iterative wMMSE-SIC algorithm.

Theorem 3. *The wMMSE-SIC algorithm converges to a local optimum of the weighted sum-rate* (5)*.*

Proof: Taking into account the diagonal structure of $\tilde{\mathbf{W}}$ and that $\tilde{w}_i(r) = 1/\tilde{e}_i(r)$ yields

$$\operatorname{tr}\Big(\underbrace{\tilde{\mathbf{W}}_i\tilde{\mathbf{E}}_i\left(\tilde{\mathcal{V}},\tilde{\mathbf{G}}_i\right)}_{\mathbf{I}_R}\Big) - \log\det\left(\tilde{\mathbf{W}}_i\right) = R - \sum_{r=1}^{R}\log\left(\frac{1}{\tilde{e}_i(r)}\right), \tag{25}$$

so that at convergence the wMMSE optimization (6) with respect to the precoder filters is tantamount to

$$\max_{\tilde{\mathcal{V}}} \sum_{i=1}^{N_u}\alpha_i\sum_{r=1}^{R}\log\left(1+\tilde{\gamma}_i(r)\right) \quad \text{subject to} \tag{26}$$
$$\sum_{r=1}^{R}\tilde{\mathbf{v}}_i^H(r)\tilde{\mathbf{v}}_i(r) \le P_i \quad i = 1,2,\ldots,N_u,$$

where the received SINR $\tilde{\gamma}_i(r)$ is

$$\tilde{\gamma}_i(r) = \frac{\left|\tilde{\mathbf{g}}_i^H(r)\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(r)\right|^2}{\mathbf{g}_i^H(r)\tilde{\tilde{\mathbf{J}}}_i(r)\tilde{\mathbf{g}}_i(r)}, \tag{27}$$

where $\tilde{\tilde{\mathbf{J}}}_i(r) = \tilde{\mathbf{J}}_i(r) - \mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(r)\tilde{\mathbf{v}}_i(r)^H\mathbf{H}_{i,i}^H$ and the equivalence $1/\tilde{e}_i(r) = 1+\tilde{\gamma}_i(r)$ is proven in Appendix A.

Hence, we need to prove that a local optimum for (26) is also a local optimum for (5). It is known [39] that MMSE spatial filtering is *information lossless*: the mutual information between the transmitted symbol s and the received vector $\mathbf{y}$ does not change after multiplication with the MMSE spatial filter $\mathbf{g}$, i.e., $\mathrm{I}(s;\mathbf{y}) = \mathrm{I}\left(s;\mathbf{g}^H\mathbf{y}\right)$. Moreover, applying the mutual information chain rule to the i-th term of (5) yields:

$$\begin{aligned}\mathrm{I}(\mathbf{s}_i;\mathbf{y}_i) = {} & \mathrm{I}(s_i(1);\mathbf{y}_i) + \mathrm{I}(s_i(2);\mathbf{y}_i|s_i(1)) + \ldots \\ & + \mathrm{I}(s_i(R);\mathbf{y}_i|s_i(1),s_i(2),...,s_i(R-1)).\end{aligned} \tag{28}$$

Exploiting the conditioning on the symbols $s_i(1), s_i(2), \ldots, s_i(r-1)$, the rth term of the expansion (28) can be rewritten as

$$\begin{aligned}&\mathrm{I}(s_i(r);\mathbf{y}_i|s_i(1),s_i(2),\ldots,s_i(r-1)) = \\ &= \mathrm{I}\left(s_i(r);\tilde{\mathbf{y}}_i(r)\right) = \mathrm{I}\left(s_i(r);\tilde{\mathbf{g}}_i^H(r)\tilde{\mathbf{y}}_i(r)\right) = \log(1+\tilde{\gamma}_i(r)),\end{aligned} \tag{29}$$

where the first equivalence is due to the fact that we have taken advantage of the conditioning on the symbols $s_i(1), s_i(2), \ldots, s_i(r-1)$ to cancel the self interference created by those symbols; the second equivalence exploits the information lossless property of MMSE filtering, and the third equivalence follows from the definition of mutual information. Thus, maximizing the metric $\sum_{r=1}^{R}\log\left(1+\tilde{\gamma}_i(r)\right)$ is equivalent to maximizing $\mathrm{I}(\mathbf{s}_i;\mathbf{y}_i)$, and the equivalence of problems (5) and (26) is proven. ■

A direct consequence of Theorem 3 is the equivalence of the solution of wMMSE-SIC and linear wMMSE, which we highlight in the following corollary.

Corollary 4. *Any $\mathbf{V}_i^\star$ outcome of the wMMSE algorithm, is also a solution of the wMMSE-SIC optimization and vice versa.*

Proof: For the sake of consistency, we just show that $\mathbf{V}_i^\star$ is a local optimum for the wMMSE-SIC algorithm, since proving the vice versa is then straightforward. By *contradiction* we assume that $\mathbf{V}_i^\star$ does not find a local minimum of the wMMSE when the SIC receiver is implemented at the receiver. In this case, initializing Algorithm 1 with $\tilde{\mathbf{V}}_i^{(0)} = \mathbf{V}_i^\star$ would lead to a solution $\tilde{\mathbf{V}}_i^\star$ in the proximity of $\mathbf{V}_i^\star$ which would yield a smaller wMMSE and, consequently, a higher mutual information, which contradicts the hypothesis since $\mathbf{V}_i^\star$ is a local optimum for the mutual information. ■

Thus, while the sets of matrices $\mathcal{W},\mathcal{G}$ and $\tilde{\mathcal{W}},\tilde{\mathcal{G}}$ are different for a given local minimum of the wMMSE, the set of spatial precoders are the same for the wMMSE and wMMSE-SIC algorithms.

A close inspection of the proposed wMMSE-SIC iterative procedure shows that the weighted sum-rate increases at any iteration of the algorithm. Henceforth, the wMMSE-SIC scheme can be seen as a strategic game with simultaneous updates among all tx-rx pairs, where the instructions in lines 9, 11 and 15 of Algorithm 1 correspond to the actions of a generic player i and the set of strategies is given by the precoding vectors $\tilde{\mathbf{V}}_i$. In particular, in a similar way to [40], the proposed approach can be categorized as a *potential game* with better response dynamics, where the potential function is represented by the system weighted sum-rate. Potential games possess important properties that relate the local optima of the potential function to the Nash equilibrium points of the game [41]. Specifically, the set of Nash equilibria corresponds to the set of local optima of the potential function, i.e., the proposed approach is guaranteed to converge to a local optimum of the system sum-rate.

C. Comparison of Linear wMMSE and wMMSE-SIC

One of the main results of wMMSE-SIC is that the MSE matrix is diagonalized with a great simplification of the computational burden. Nevertheless, as pointed out in [8], also in the case of linear wMMSE, for any set of optimal transmit filters it is possible to generate a diagonalizing set which has the same rate-tuple. In fact, replacing in (7) the value of $\mathbf{G}_i$ found in (8), yields

$$\mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right) = \mathbf{I}_R - \mathbf{V}_i^H \mathbf{H}_{i,i}^H \mathbf{J}_i^{-1} \mathbf{H}_{i,i} \mathbf{V}_i, \tag{30}$$

so that, by simple algebraic manipulation, it is possible to show that

$$\mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right)^{-1} = \mathbf{I}_R + \mathbf{V}_i^H \mathbf{H}_{i,i}^H \bar{\mathbf{J}}_i^{-1} \mathbf{H}_{i,i} \mathbf{V}_i, \tag{31}$$

and the rate for user i in (4) can be computed as $R_i\left(\mathcal{V}\right) = \log\det\left(\mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right)^{-1}\right)$. Thus, after having found the optimal set of precoder filters $\mathcal{V}^\star$, one can can compute the eigenvalue decomposition

$$\mathbf{A}_i \mathbf{\Lambda}_i \mathbf{A}_i^H = \mathbf{V}_i^{\star H} \mathbf{H}_{i,i}^H \bar{\mathbf{J}}_i^{-1} \mathbf{H}_{i,i} \mathbf{V}_i^\star \tag{32}$$

and employ the precoder matrix $\mathbf{V}_i^\star \mathbf{A}_i$, which yields the same rate of $\mathbf{V}_i^\star$, to diagonalize the MSE matrix. Even if both techniques are able to diagonalize the MSE matrix, there are important differences between the two schemes due to the fact that the MSE matrix for wMMSE-SIC is inherently diagonal. Furthermore, a few remarks are in order.

- *Computational complexity*. By avoiding the inversion of the full weight matrix $\tilde{\mathbf{W}}_i$, wMMSE-SIC reduces the overall computational burden at each transmitter. Alternatively, if one wants to skip the inversion of $\mathbf{W}_i$ for linear wMMSE, he needs to perform the eigenvalue decomposition in (32), which has an even larger complexity. Moreover, if one does not perform diagonalization at the transmitter, the wMMSE receiver is affected by strong inter-stream interference, which requires very complex receiver design.
- *Amount of feedback information*. As we will see in Section VI-F, the eigenvalue decomposition in (32) requires a larger exchange of feedback information than MMSE-SIC.
- *Flexibility*. Because of the diagonal characteristic of the MSE matrix, the SIC architecture is less affected by the choice of suboptimal precoding and receive matrices such as those employed in the case of hybrid beamforming with a reduced number of RF chains, as it will be shown in the next Section.

V. Hybrid Weighted MSE Minimization for Non-linear Receivers

Designing a HB transceiver architecture to maximize the sum-rate in the interference channel is complicated by the additional decomposition rule (11) and non-convex constraints (12) for the analog beamforming matrices. With these sets of new constraints, the wMMSE problem is

$$\begin{aligned}
&\min_{\mathcal{V},\mathcal{W},\mathcal{G}} \sum_{i=1}^{N_u} \alpha_i \left(\operatorname{tr}\left(\mathbf{W}_i \mathbf{E}_i\left(\mathcal{V}, \mathbf{G}_i\right)\right) - \log\det\left(\mathbf{W}_i\right)\right) \\
&\qquad\qquad \text{subject to} \\
&\operatorname{tr}\left(\mathbf{V}_i \mathbf{V}_i^H\right) \le P_i \quad i = 1, 2, \ldots, N_u, \\
&\mathbf{V}_i = \mathbf{V}_i^{(AB)} \mathbf{V}_i^{(DB)} \quad i = 1, 2, \ldots, N_u, \\
&\mathbf{G}_i = \mathbf{G}_i^{(AB)} \mathbf{G}_i^{(DB)} \quad i = 1, 2, \ldots, N_u, \\
&\mathbf{V}_i^{(AB)} \in \mathcal{A}_{L,N_{RF}} \quad i = 1, 2, \ldots, N_u, \\
&\mathbf{G}_i^{(AB)} \in \mathcal{A}_{M,N_{RF}} \quad i = 1, 2, \ldots, N_u.
\end{aligned} \tag{33}$$

Most of the literature on hybrid beamforming focuses on the single-user case or on multi-user scenarios in the millimeter wave channel with some very specific assumptions so that it is possible to find the HB matrices that provably maximize the sum-rate. Unfortunately, such approaches do not work for the interference channel and here we follow another, heuristic and iterative, line of reasoning.

We solve (33) in threes steps: 1) first we solve one iteration of the FD problem (6) to find the beamforming full digital $T \times R$ matrix $\mathbf{B}_i^{(FD)}$; 2) we design the $T \times N_{RF}$ analog beamforming matrix $\mathbf{B}_i^{(AB)}$ by finding a decomposition that minimizes the mean square error with $\mathbf{B}^{(FD)}$; and 3) we find the $N_{RF} \times R$ digital beamforming matrix $\mathbf{B}_i^{(DB)}$.

A. Analog Beamforming Matrix

The design of $\mathbf{B}_i^{(AB)}$ leverages on the knowledge of the desired fully digital beamforming matrix $\mathbf{B}_i^{(FD)}$: rather than trying to solve directly the nonconvex problem (33), we find the $T \times N_{RF}$ matrix $\mathbf{B}_i^{(AB)}$ and the $N_{RF} \times R$ matrix $\mathbf{P}$ whose product minimizes the mean square error with the FD matrix

$$\mathbf{B}_i^{(AB)} = \arg \min_{\mathbf{B}_i^{(AB)} \in \mathcal{A}_{T,N_{RF}}, \mathbf{P}} \|\mathbf{B}_i^{(FD)} - \mathbf{B}_i^{(AB)} \mathbf{P}\|^2 \tag{34}$$

Problem (34) is still nonconvex but, following the algorithm proposed in [20], a suboptimal solution can be found using a block coordinate descent strategy. The algorithm is iterative and we indicate with $\mathbf{P}_k$ and $\mathbf{Q}_k \in \mathcal{A}_{T,N_{RF}}$ the two matrices at iteration k, respectively. At each iteration we first solve problem (34) with respect to $\mathbf{P}$ having fixed $\mathbf{Q}$ and then with respect to $\mathbf{Q}$ having fixed $\mathbf{P}$. Thus, given $\mathbf{B}_i^{(FD)}$ and $\mathbf{Q}_k$, the decomposition at iteration $k+1$ is found as

$$\begin{aligned}
\mathbf{P}^{(k+1)} &= \left(\mathbf{Q}^{(k)H} \mathbf{Q}^{(k)}\right)^{-1} \mathbf{Q}^{(k)} \mathbf{B}^{(FD)} \\
\mathbf{Q}^{(k+1)} &= \Pi_{\mathcal{A}}\left(\mathbf{B}^{(FD)} \mathbf{P}^{(k+1)H} \left(\mathbf{P}^{(k+1)} \mathbf{P}^{(k+1)H}\right)^{-1}\right)
\end{aligned} \tag{35}$$

where $\Pi_{\mathcal{A}}\left(\mathbf{Q}\right)$ is the (unique) Euclidean projection of $\mathbf{Q}$ on the set $\mathcal{A}$ of matrices with unitary envelope elements, which can easily be obtained by element-wise scaling. At convergence the $\mathbf{Q}$ matrix is the analog beamformer $\mathbf{B}_i^{(AB)}$.

B. Digital Beamforming Matrix

The matrices solution of the decomposition in (34) satisfy the nonconvex constraints of (33) and yield a good approximation of $\mathbf{B}^{(FD)}$ but $\mathbf{P}$ not necessarily is the digital beamforming matrix that minimize the weighted MSE under the original power constraints. Thus, after having found the matrix $\mathbf{B}_i^{(AB)}$, the matrix $\mathbf{B}_i^{(DB)}$ is obtained by minimizing the wMMSE over the equivalent channel $\mathbf{H}_{i,i}\mathbf{B}_i^{(AB)}$.

The digital *precoding* matrix $\tilde{\mathbf{V}}_i^{(DB)}$ is obtained solving

$$\min_{\tilde{\mathbf{V}}_i^{(DB)}} \sum_{r=1}^{R} \Big[\alpha_i \tilde{w}_i(r) \left| 1 - \tilde{\mathbf{g}}_i(r)^H \mathbf{H}_{i,i} \mathbf{B}_i^{(AB)} \tilde{\mathbf{v}}_i^{(DB)}(r) \right|^2 + \tilde{\mathbf{v}}_i^{(DB)}(r)^H (\mathbf{B}_i^{(AB)})^H \tilde{\tilde{\mathbf{K}}}_i(r) \mathbf{B}_i^{(AB)} \tilde{\mathbf{v}}_i^{(DB)}(r) \Big]$$

subject to

$$(\mathbf{B}_i^{(AB)})^H \sum_{r=1}^{R} \tilde{\mathbf{v}}_i^{(DB)}(r)^H \tilde{\mathbf{v}}_i(r) \mathbf{B}_i^{(AB)} \leq P_i. \tag{36}$$

Similarly to (23), the constrained minimization in (36) can be solved by employing standard convex methods to obtain:

$$\tilde{\mathbf{v}}_i^{(DB)}(r) = \alpha_i \tilde{w}_i(r) \left[(\mathbf{B}_i^{(AB)})^H \left(\tilde{\mathbf{K}}_i(r) + \mu_i \mathbf{I}_L \right) \mathbf{B}_i^{(AB)} \right]^{-1} \times (\mathbf{B}_i^{(AB)})^H \mathbf{H}_{i,i}^H \tilde{\mathbf{g}}_i(r), \tag{37}$$

where the value of the Lagrangian variable μ_i allows to satisfy the power budget constraint.

As for the digital *combining* matrix $\tilde{\mathbf{G}}_i^{(DB)}$, elaborating from (14), we get:

$$\tilde{\mathbf{g}}_i^{(DB)}(r) = \left((\mathbf{B}_i^{(AB)})^H \tilde{\mathbf{J}}_i(r) \mathbf{B}_i^{(AB)} \right)^{-1} (\mathbf{B}_i^{(AB)})^H \mathbf{H}_{i,i} \tilde{\mathbf{v}}_i(r). \tag{38}$$

The procedure for finding the hybrid beamforming matrices, denoted by wMMSE-SIC-HB, is described in detail in Algorithm 2 for a single iteration q of Algorithm 1. We have added the apex (q) to keep track of the outer iteration number, elsewhere we have omitted it for notation clarity .

Algorithm 2: Iterative hybrid beamforming of wMMSE-SIC-HB

```
1  Initialize:
2  Compute B_i^{(FD),(q)};
3  Set an arbitrarily small value for ϵ;
4  k ← 0;
5  Δ ← 1;
6  Generate an initial random analog beamforming matrix Q^{(0)};
7  while Δ > ϵ do
8      Compute P^{(k+1)} and Q^{(k+1)} as in (35);
9      Δ ← ||Q^{(k+1)} − Q^{(k)}||^2;
10     k ← k + 1;
11 end;
12 B_i^{(AB),(q)} ← Q^{(k+1)};
13 Compute B_i^{(DB),(q)} (using (37) for precoding and (38) for combining);
14 B_i^{(HB),(q)} ← B_i^{(AB),(q)} B_i^{(DB),(q)};
```

For the procedure described in Algorithm 2 some remarks are in order:

- *Complexity*. Although solving (35) requires the iterative inversion of the $N_{RF} \times N_{RF}$ matrices $\mathbf{Q}^{(k)H}\mathbf{Q}^{(k)}$ and $\mathbf{P}^{(k)}\mathbf{P}^{(k)H}$, the computational burden of these operations is proportional to the number of active RF chains N_{RF}, which is, in general, much smaller than the number of antennas. The same reasoning applies also to (37) and (38), which have a lower dimensionality with respect to their FD counterparts (24) and (14). Thus, it is reasonable to say that each iteration of the HB part of Algorithm 2, i.e. lines 3-14, has a complexity at most comparable to one FD iteration.
- *Convergence*. Theoretically, while problem (34) yields the best hybrid decomposition $\mathbf{B}_i^{(AB)(q)}$ of the FD matrix $\mathbf{B}_i^{(FD)(q)}$ in terms of MSE, it is not necessarily true that the solution of (37) and (38) after plugging in $\mathbf{B}_i^{(AB)(q)}$ will yield a lower wMMSE with respect to the wMMSE computed at iteration $(q-1)$. Such an event, which is extremely rare in our simulations, has the potential to compromise the convergence of the coordinate descent algorithm. In these circumstances, it is easy to show that one can employ as analog beamformer $\mathbf{B}_i^{(AB)(q-1)}$ rather than $\mathbf{B}_i^{(AB)(q)}$ and this chioice will lead in any case to a reduction of the wMMSE so that the convergence property of the algorithm is safeguarded.
- *Comparison of linear wMMSE-HB and wMMSE-SIC-HB*. One of the effects of hybrid beamforming in the wMMSE framework is the loss of performance due to the extra constraints in the design of the analogue beamforming matrix. To elaborate, consider the linear wMMSE scheme with the diagonalization of the MSE matrix at the transmitter. The precoding filter employs the matrix $\mathbf{A}_i$ in (32), computed on the base of the covariance matrix $\bar{\mathbf{J}}_i$, to diagonalize the MSE matrix. As the algorithm proceeds, as soon as the receivers change their filters, $\bar{\mathbf{J}}_i$ changes as well and the MSE matrix is no longer diagonal so that there is a model mismatch in the computation of the weights in (9). In a fully digital system with a large number of antennas, this model mismatch has little effect as the the large number of degrees of freedom available at the receiver side is such to greatly attenuate the residual inter streaming correlation, thus achieving a quasi-diagonal MSE matrix at each iteration. On the contrary, in the presence of hybrid beamforming, the effects of the reduced degrees of freedom is that the MSE matrix looses it diagonal structure, and this affects the final system performance. On the other hand, in the wMMSE-SIC case, diagonalization of the MSE matrix is obtained by construction, both in the fully digital and in the hybrid beamforming schemes. Hence, as it will be investigated in the Results section, wMMSE-SIC-HB is less affected by the reduction of the degrees of freedom of hybrid beamforming and outperforms linear wMMSE-HB.

VI. Distributed Implementation of the Proposed Allocation Schemes

In this section we propose a training scheme designed for the implementation of our wMMSE algorithms, so that each user can compute independently its own precoding and combining filters in a *distributed* manner. We consider a bi-directional training approach where blocks of pilots are alternately transmitted in the forward direction (from transmitters to receivers), and in the reverse direction (from receivers to transmitters). We assume that the system operates in time division duplexing mode so that it can exploit channel reciprocity between the transmitter and the intended receiver. In detail, the channel matrix between the receiver of the j-th pair (acting as a transmitter) and the transmitter of the i-th pair (acting as as a receiver), denoted by $\mathbf{H}'_{i,j}$, is equal to the transpose of the channel matrix between the same nodes when the transmitters and receivers roles are inverted i.e., $\mathbf{H}'_{i,j} = \mathbf{H}^T_{j,i}$. Also, perfect synchronization among different tx-rx pairs is assumed (e.g., with the assistance of the BS), so that all the receivers/transmitters transmit their respective pilot sequences at the same time instances.

Hence, the information needed for updating the precoding vectors and the receive filters can be directly estimated from the received pilot signals without requiring either the estimation of the cross-channel matrixes $\mathbf{H}_{j,i}$, with $j \neq i$, or the exchange of additional information between the different tx-rx pairs. The proposed bi-directional training scheme is inspired by previous works reported in [33], [34] and [35]. However, the presence of multibeam transmissions with SIC at the receivers leads to fundamental differences with respect to the scheme proposed in these previous works.

Specifically, to facilitate the evaluation of the self-interference partial covariance matrices, we assume that the following pieces of control information is exchanged on two separate control channels:

1) a *dedicated control channel* between each tx-rx pair including user-specific reference symbols that allow for the estimation of the direct channel matrices $\mathbf{H}_{i,i}$. In this regard, consecutive training can be used to monotonically improve the channel estimates in one coherence block of the channel, e.g., using iterative techniques. Hence, we can reasonably assume that the direct channel matrices $\mathbf{H}_{i,i}$ can be perfectly estimated, making SIC feasible.
2) a *common control channel* between each node and the BS to exchange control information during the allocation procedure.

The main implementation aspects of the proposed beamforming and power control schemes include the pilot sequence design and the signaling design at both the Rx and Tx sides. Fig. 2 reports the signaling flow diagram of the proposed bi-directional training approach, whose steps are detailed in the following subsections.

A. Pilot Sequence

The pilot sequences are referred to as $\mathbf{\Omega}_i \in \mathbb{C}^{R\times S}$, with elements $\xi_i(r,t)$ representing the t-th symbol of the pilot sequence used by the i-th tx-rx pair for the r-th stream, where S denotes the number of symbols in each pilot sequence. The chosen pilot sequences have unity norm and are orthogonal between beams of the same transmitter. Orthogonality is, however, not fulfilled between different tx-rx pairs since it is not realistic to assume to maintain a symbol level synchronization among nodes. In this case, assuming that each node uses a local scrambling code, the product $\mathbf{\Omega}_i\mathbf{\Omega}_j^H$ can be modelled as a random matrix with elements taken from an i.i.d. zero-mean distribution with variance S, i.e.,

$$\mathbb{E}\left\{\mathbf{\Omega}_i\mathbf{\Omega}_j^H\right\} = \begin{cases} S\mathbf{I}_R & \text{if } i=j \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{39}$$

B. Transmitted Pilot Signals: Receiver Side

To enable the computation of the precoder matrix at the transmitter, every receiver at each iteration transmits a *pilot signal* on the common control channel, which we denote by $\tilde{\mathbf{X}}_i^{rx}$ and is pre-computed as follows:

$$\tilde{\mathbf{X}}_i^{rx} = \left[\mathbf{G}_i\left(\mathbf{W}_i\right)^{1/2}\right]^*\mathbf{\Omega}_i \quad \in \mathbb{C}^{M\times S} \tag{40}$$

where $\mathbf{G}_i \in \mathbb{C}^{M\times R}$ is a matrix with columns $\mathbf{g}_i(r)$ and $\mathbf{W}_i$ represent the MMSE weights, which, for the sake of notations simplicity, encompass also the weights α_i.

Being implemented at the receiver, the computation of the receive MMSE matrix $\mathbf{G}_i$ does not take into account any power constraint and $\mathbf{G}_i$ in certain cases can be extremely large.

To circumvent this problem, it is necessary to scale the signals $\tilde{\mathbf{X}}_i^{rx}$ by a factor that curtails the power required by the exchange of control information.

To this aim, we assume that all receivers evaluate a local scaling factor and they communicate it to the BS. The BS, in turn, takes the minimum of the received terms and informs back all the receivers about the actual scaling factor to be used. Hence, denoting by ρ the *coordinated scaling factor*, we employ the pilot signal $\mathbf{X}_i^{rx} = \rho\tilde{\mathbf{X}}_i^{rx}$, as shown in Figure 2.

C. Estimation of the Required Terms: Transmitter Side

The pilot signal received at the transmitters' side are:

$$\begin{aligned}\mathbf{Y}_i^{tx} = {}& \rho\mathbf{H}^T_{i,i}\left[\mathbf{G}_i\left(\mathbf{W}_i\right)^{1/2}\right]^*\mathbf{\Omega}_i + \\ &+ \rho\sum_{\substack{j=1\\ j\neq i}}^{N_u}\mathbf{H}^T_{j,i}\left[\mathbf{G}_j\left(\mathbf{W}_j\right)^{1/2}\right]^*\mathbf{\Omega}_j + \mathbf{N}_i \quad \in \mathbb{C}^{L\times S}.\end{aligned} \tag{41}$$

Assuming perfect direct channel estimation and that the receivers communicate the MMSE filters and the MMSE weights to the intended transmitters on the dedicated control channel, the transmitters can cancel the signal received from the intended receivers obtaining:

$$\begin{aligned}\hat{\mathbf{Y}}_i^{tx} &= \mathbf{Y}_i^{tx} - \rho\mathbf{H}^T_{i,i}\left[\mathbf{G}_i\left(\mathbf{W}_i\right)^{1/2}\right]^*\mathbf{\Omega}_i = \\ &= \rho\sum_{\substack{j=1\\ j\neq i}}^{N_u}\mathbf{H}^T_{j,i}\left[\mathbf{G}_j\left(\mathbf{W}_j\right)^{1/2}\right]^*\mathbf{\Omega}_j + \mathbf{N}_i.\end{aligned} \tag{42}$$

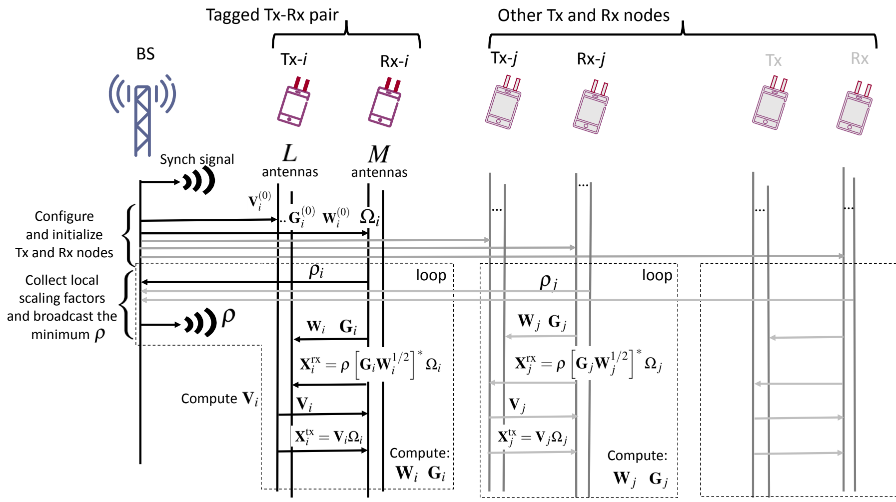

Fig. 2. Network assisted alternating bidirectional signaling to support the proposed distributed iterative schemes. The network (cellular base station) provides synchronization signals and configures the mobile stations. Mobile stations then utilize channel reciprocity and dedicated control channels to collect the terms that are necessary to configure the precoder and receiver matrices.

Employing $\hat{\mathbf{Y}}_i^{tx}$ the transmitters are able compute an estimate of the transmitter covariance matrix as

$$\widehat{\widehat{\mathbf{K}}}_i = \frac{1}{\rho^2}\left(\frac{1}{S}\left[\hat{\mathbf{Y}}_i^{tx}\left(\hat{\mathbf{Y}}_i^{tx,q}\right)^H\right]^* - \sigma_i^2\mathbf{I}_L\right). \tag{43}$$

Since the expectation of $\widehat{\widehat{\mathbf{K}}}_i$ yields

$$\mathbb{E}_{\mathbf{\Omega},\mathbf{n}}\left[\widehat{\widehat{\mathbf{K}}}_i\right] = \sum_{\substack{j=1\\ j\neq i}}^{N_u} \mathbf{H}_{j,i}^H\mathbf{G}_j\mathbf{W}_j\mathbf{G}_j^H\mathbf{H}_{j,i} = \bar{\mathbf{K}}_i. \tag{44}$$

$\widehat{\widehat{\mathbf{K}}}_i$ is an unbiased estimator of the covariance matrix $\bar{\mathbf{K}}_i$ needed at the transmitters side to compute the precoder filters. Exploiting the direct channel an estimate of the full covariance matrix can be computed as

$$\widehat{\mathbf{K}}_i = \widehat{\widehat{\mathbf{K}}}_i + \mathbf{H}_{i,i}^H\mathbf{G}_i\mathbf{W}_i\mathbf{G}_i^H\mathbf{H}_{i,i}. \tag{45}$$

As for the wMMSE-SIC and the hybrid schemes, it is worth noting that estimates of the covariance matrixes $\tilde{\mathbf{K}}_i(r)$ and $\tilde{\tilde{\mathbf{K}}}_i(r)$ ($r = 1, 2, \ldots, R$ and $i = 1, 2, \ldots, N_u$) can be easily computed by exploiting $\widehat{\mathbf{K}}_i$ and the knowledge of the direct channel gains.

D. Transmitted Pilot Signals: Transmitters Side

The pilot signals transmitted from the transmitters to the receivers at each iteration, denoted by $\tilde{\mathbf{X}}_i^{tx}$, are computed as follows:

$$\mathbf{X}_i^{tx} = \mathbf{V}_i\mathbf{\Omega}_i. \tag{46}$$

E. Estimation of the Required Terms: Receivers Side

The pilot signal received at the receivers' side are:

$$\mathbf{Y}_i^{rx} = \mathbf{H}_{i,i}\mathbf{V}_i\mathbf{\Omega}_i + \sum_{\substack{j=1\\ j\neq i}}^{N_u}\mathbf{H}_{i,j}\mathbf{V}_j\mathbf{\Omega}_j + \mathbf{N}_i \quad \in \mathbb{C}^{M\times S}. \tag{47}$$

Assuming perfect direct channel estimation, and assuming that the transmitters communicate to the intended receivers the precoding vectors $\mathbf{v}_i(k)$, the receivers can cancel the signal received from the intended transmitters thus obtaining:

$$\hat{\mathbf{Y}}_i^{rx} = \mathbf{Y}_i^{rx} - \mathbf{H}_{i,i}\mathbf{V}_i\mathbf{\Omega}_i = \sum_{\substack{j=1\\ j\neq i}}^{N_u}\mathbf{H}_{i,j}\mathbf{V}_j\mathbf{\Omega}_j + \mathbf{N}_i. \tag{48}$$

The receivers are able to estimate the covariance matrix as

$$\widehat{\widehat{\mathbf{J}}}_i = \frac{1}{S}\left[\hat{\mathbf{Y}}_i^{rx}\left(\hat{\mathbf{Y}}_i^{rx}\right)^H\right]. \tag{49}$$

Replacing (48) in (49) and computing the expectation shows that $\widehat{\widehat{\mathbf{J}}}_i$ is an unbiased estimator of $\bar{\mathbf{J}}_i$

$$\mathbb{E}_{\mathbf{\Omega},\mathbf{n}}\left[\widehat{\widehat{\mathbf{J}}}_i\right] = \sum_{\substack{j=1\\ j\neq i}}^{N_u}\mathbf{H}_{i,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{i,j}^H + \sigma_i^2\mathbf{I}_M = \bar{\mathbf{J}}_i. \tag{50}$$

Also at the receiver it is possible to exploit the knowledge of the direct channel for computing an unbiased estimate of $\mathbf{J}_i$

$$\widehat{\mathbf{J}}_i = \widehat{\widehat{\mathbf{J}}}_i + \mathbf{H}_{i,i}\mathbf{V}_i\mathbf{V}_i^H\mathbf{H}_{i,i}^H. \tag{51}$$

In the presence of SIC, i.e.,in the wMMSE-SIC and wMMSE-SIC-HB cases, $\widehat{\widehat{\mathbf{J}}}_i$ and the knowledge of the direct channel can be used to compute the various covariance matrices required to compute the receive filters.

As a final step, the receivers must estimate the MMSE weights $\mathbf{W}_i$. In the wMMSE case, the receivers first estimate the MMSE matrix $\widehat{\mathbf{E}}_i = \mathbf{I} - \mathbf{V}_i^H\mathbf{H}_{i,i}^H\widehat{\mathbf{J}}_i^{-1}\mathbf{H}_{i,i}\mathbf{V}_i$ to obtain $\widehat{\mathbf{W}}_i = \left(\widehat{\mathbf{E}}_i\right)^{-1}$. In the presence of SIC, the receivers estimate the SINRs $\gamma_i(r)$ given in (27), and then exploit the equivalence $1/\tilde{e}_i(r) = 1 + \tilde{\gamma}_i(r)$ proven in Appendix A to get the weights $\tilde{w}_i(l) = 1/\tilde{e}_i(r)$.

F. Comparison of Linear wMMSE and wMMSE-SIC in terms of feedback information

All through this section we have seen that the amount of feedback required by linear wMMSE and the proposed

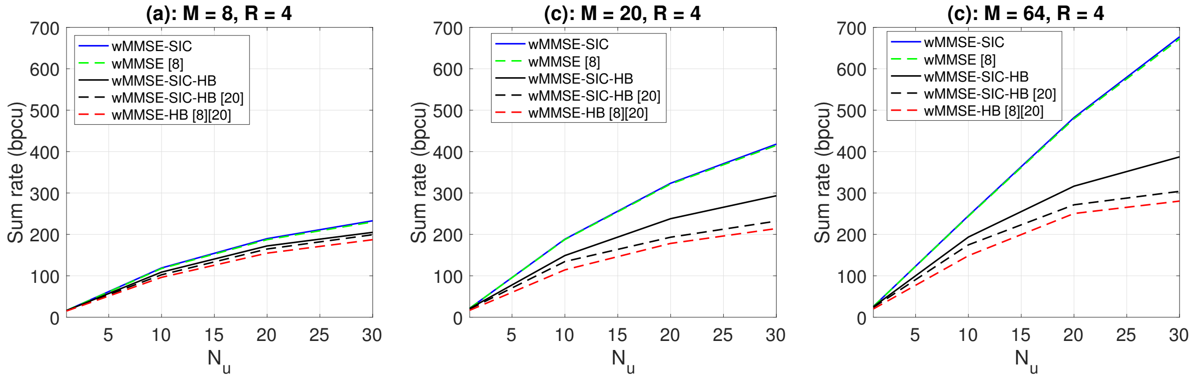

Fig. 3. Sum-rate versus number of users, i.e., tx-rx pairs, with ideal estimation for $SNR = 12$ dB, $R = 4$, $M = 8$ (a), $M = 20$ (b) and $M = 64$ (c).

wMMSE-SIC schemes is the same. Nevertheless, this is valid only for the standard linear wMMSE algorithm and does not apply when one wants to diagonalize the MMSE matrix for the linear wMMSE scheme. In fact, computing the eigenvalue decomposition in (32) requires at the transmitter the explicit knowledge of the precoding vectors of all users in the system. This type of information is not possessed by the transmitters and is not exchanged on any of the control channels and, accordingly, would require its own signaling, thereby noticeably increasing the amount of exchanged information.

VII. Numerical Results

We consider a cellular system with a circular coverage area of radius 500 m serving $N_u = 1, 10,$ or 20 tx-rx pairs deployed randomly in the coverage area. The algorithms' performance are evaluated via Monte Carlo simulations, where at each simulation instance the distance between the tx and rx nodes of each pair is uniformly distributed in the interval $[D_{min}, D_{max}]$, with $D_{min} = 50$ m and $D_{max} = 200$ m.

The channel attenuation is due to path loss – proportional to the distance between the transmitters and receivers –, shadowing and fading. The path loss exponent is $\alpha = 4$, while shadowing is assumed log-normally distributed with standard deviation $\sigma = 8$ dB. The transmitters and receivers are assumed to have the same number of antennas, i.e., $L = M$, denoted by M in the following, and a fixed number of $N_{RF} = 4$ RF chains. The number of streams per use is $R = N_{RF}$. We use an uncorrelated fading channel model with channel coefficients generated from the complex Gaussian distribution $\mathcal{CN}(0, 1)$. The variance σ_i^2 of the additive zero-mean Gaussian noise, which includes the interference from the cellular network, and the maximum power budget $P_{i,}$ are assumed to be the same for all tx-rx pairs.

The performance measure of interest is the achieved system sum-rate representing the sum of the bits per channel use (bpcu) by each transmitter. Each point in the curves is obtained by averaging over 1000 Monte Carlo drops, where at each drop a new instance of nodes' positions, as well as large and small scale fading are generated.The signal-to-noise ratio (SNR) of a specific simulation is obtained by properly setting the maximum power budgets and the noise variance.

In all figures, we plot the performance of the FD wMMSE-SIC (blue solid lines) and wMMSE (green dashed lines) schemes together with the hybrid beamforming schemes wMMSE-SIC-HB (black solid lines) and wMMSE-HB schemes (red dashed lines). The wMMSE algorithm is implemented employing the linear diagonalizing procedure described in [8]. The wMMSE-HB scheme is implemented considering the classical block coordinate descent approach proposed in [20]. In this case, as discussed in Section V, convergence of the wMMSE algorithm cannot be ensured and hence we consider as final result the best total sum-rate obtained after 100 iterations. Finally, to show the benefits of the HB algorithm we propose, we also show the performance of the wMMSE-SIC-HB scheme with hybrid beamforming implemented following [20] (black dashed line).

Fig. 3 plots the sum-rate as a function of the number of tx-rx pairs N_u for a fixed $SNR = 12$ dB in the case of perfect estimation of all parameters, i.e., assuming $S = \infty$. The results have been obtained for three different number of antennas: $M = 8$ (a), $M = 20$ (b), and $M = 64$ (c). As remarked in Sect. V, in the FD scenario the proposed wMMSE-SIC yields similar performance to the wMMSE scheme, while in the HB scenario the proposed wMMSE-SIC-HB clearly outperforms the schemes based on the HB approach proposed in [20]. As expected, the hybrid beamforming schemes show inferior sum-rate performance at all SNR values with respect to the unconstrained fully digital ones. The amount of performance loss mainly depends on the number of nodes, i.e., the amount of interference, and, to a lesser extent, on the number of antennas and on the SNR. The loss of beamforming accuracy in the HB case makes more difficult to suppress other users' interference, which leads to degraded performance. However, such a performance degradation is compensated by the possibility of increasing the number M of antennas at a very low cost, maintaining the same number N_{RF} of digital chains. Notice that the wMMSE-SIC-HB scheme with $M = 64$ performs as the FD wMMSE scheme with $M = 20$.

Figs. 4-5 report the sum-rate achieved as a function of the SNR for $S = \infty$, $M = 8$ (a), $M = 20$ (b), $M = 64$ (c) number of antennas, and for $N_u = 10$ and $N_u = 20$ tx-rx pairs, respectively. Once again, linear wMMSE and wMMSE-SIC have the same performance and the sum-rate of both schemes

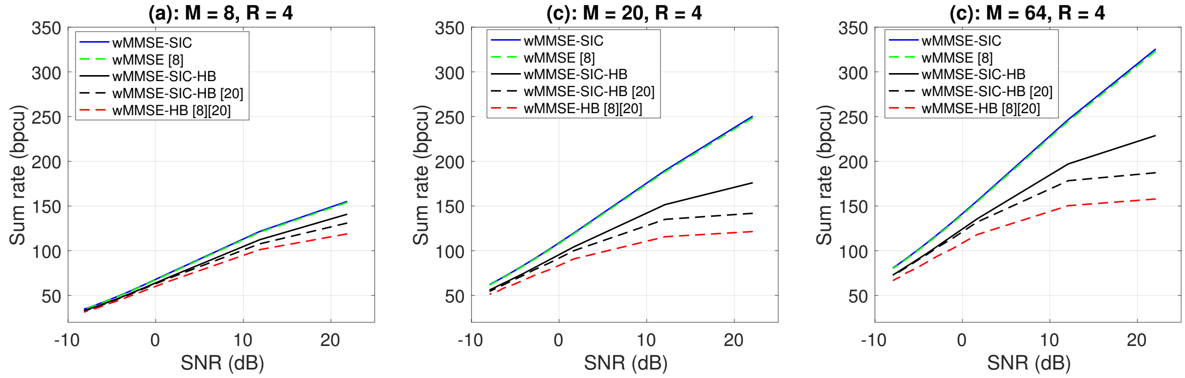

Fig. 4. Sum-rate versus SNR in the case of ideal estimation for $N_u = 10$, $R = 4$, $M = 8$ (a), $M = 20$ (b) and $M = 64$ (c).

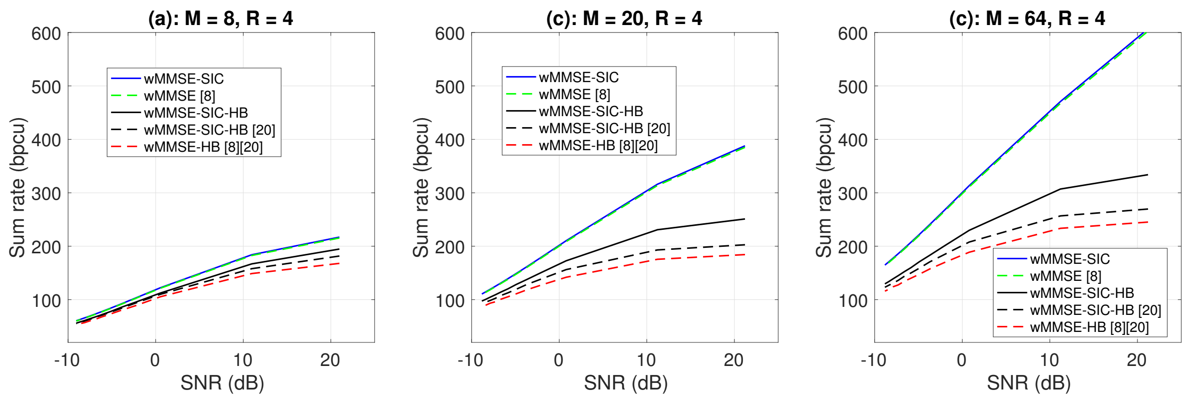

Fig. 5. Sum-rate versus SNR in the case of ideal estimation for $N_u = 20$, $R = 4$, $M = 8$ (a), $M = 20$ (b) and $M = 64$ (c).

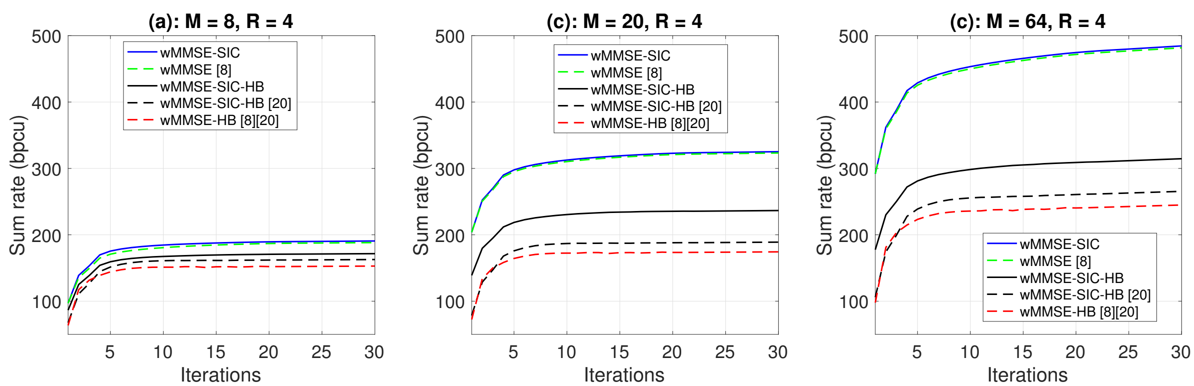

Fig. 6. Convergence behavior of wMMSE-SIC, wMMSE, wMMSE-SIC-HB and wMMSE-HB schemes for $SNR = 12$ dB, $N_u = 20$, $R = 4$, $M = 8$ (a), $M = 20$ (b) and $M = 64$ (c).

increases in the same manner as we increase the number of antennas. On the other hand, the proposed wMMSE-SIC-HB approach outperforms wMMSE-HB. These results are in line with the remarks of Sect.V, and show that wMMSE-SIC tends to have better performance than WMMSE when the degrees of freedom are limited. Thus, SIC is more beneficial in the hybrid scheme than in the FD scheme. It is worth noting that also for this set of results hybrid beamforming schemes are able to compensate the hardware limitations on the maximum number of RF chains as the number of available antennas increases.

Fig. 6 reports the sum-rate versus the number of iterations for all the proposed iterative schemes adopting the parameter setting as in Fig. 3, in the case of $N_u = 20$. The sum-rates are computed by averaging over 1000 simulation runs. Notice that the number of required iterations is low for all schemes and, in particular, at most 20 iterations are required. Although not explicitly shown in these figures, a similar behavior is observed with different SNR values and different N_u as well.

Fig. 7 plots the performance of the considered allocation schemes as a function of the length S of the training sequence for a fixed number of tx-rx pairs, namely $N_u = 20$ and for

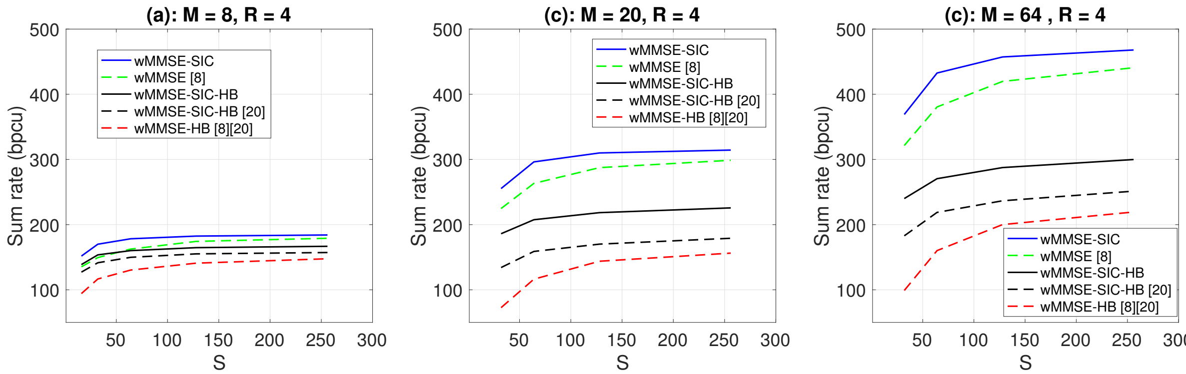

Fig. 7. Sum-rate versus versus the pilot sequences' length obtained for 1000 independent channel realizations and for $M = 8$ (a), $M = 20$ (b), $M = 64$ (c), and $SNR = 12$ dB.

$SNR = 12$ dB. The reported sum-rates are collected after 20 iterations, i.e., in this case the number of iterations is fixed without requiring convergence of the iterative procedures. In the presence of channel variability, the training phase must be limited to the minimum necessary, and 20 iterations have been found to be sufficient to approach convergence in all cases. Notice that for small S (short training sequence), all schemes show poor performance because of two reasons. First, short pilot sequences imply low pilot energy leading to poor SNR for channel estimation. Second, too short pilot sequences lead to losing orthogonality between tx-rx pairs, which is especially harmful in the presence of a large number of tx-rx pairs in a limited geographical area. However, the SIC strategy shows higher robustness towards such impairments with respect to wMMSE, thus confirming the validity of the proposed approach. As an example, with $S = 64$, the wMMSE-SIC schemes approach the case of perfect parameters' estimation, as can be seen by comparing the results in Fig. 7 with those plotted in Fig. 3. On the contrary, the linear wMMSE schemes require a much higher value of S.

VIII. Concluding Remarks

We presented three novel distributed allocation schemes for the MIMO interference channel based on MMSE beamforming. To deal with inter-stream interference, all three transmit schemes operate jointly with SIC reception, and result in joint iterative transmit-receive schemes. In particular, we first proposed an allocation scheme that aims at optimally setting the transmitter precoding vectors in the case of fully digital beamforming. Then, we proposed two hybrid schemes designed to deal with a limited number of RF chains compared to the number of antennas. Then, for practical implementation, we proposed a signaling scheme that takes advantage of the presence of a common broadcast channel and facilitates CSI acquisition in a typical cellular scenario with minimal infrastructure assistance. Numerical results demonstrate the effectiveness of the proposed schemes both in terms of high system spectral efficiency achievement and practical viability. Moreover, although the two hybrid schemes incur higher inter-stream interference than the fully digital scheme, they require a low number of iterations and are robust against CSI errors and could be of particular interest in millimeter wave applications.

Appendix A: Equivalence $1/\tilde{e}_i(k) = 1 + \gamma_i(k)$

Let us focus on user i and define, for the sake of notation clarity, $\mathbf{r}_i(m) = \mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(m)$ and $\mathbf{r}_j(m) = \mathbf{H}_{i,j}\tilde{\mathbf{v}}_j(m)$. We can now rewrite (14) as

$$\tilde{\mathbf{g}}_i(k) = \left(\mathbf{M}_i(k) + \mathbf{r}_i(k)\mathbf{r}_i^H(k)\right)^{-1}\mathbf{r}_i(k). \tag{A.1}$$

where it is

$$\mathbf{M}_i(k) = \left(\sum_{l=k+1}^{L}\mathbf{r}_i(l)\mathbf{r}_i^H(l) + \sum_{j=1,j\neq i}^{N_u}\sum_{l=1}^{L}\mathbf{r}_j(l)\mathbf{r}_j^H(l) + \sigma^2\mathbf{I}_L\right). \tag{A.2}$$

By applying the Woodbury matrix identity to (A.1), one obtains

$$\begin{aligned}\tilde{\mathbf{g}}_i(k) &= \mathbf{M}_i^{-1}(k)\mathbf{r}_i(k) - \mathbf{M}_i^{-1}(k)\mathbf{r}_i(k)\cdot \\ &\cdot\left(1 + \mathbf{r}_i^H(k)\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k)\right)^{-1}\mathbf{r}_i^H(k)\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k) = \\ &= \left(1 - \frac{\beta(k)}{1+\beta(k)}\right)\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k) \\ &= \frac{1}{1+\beta(k)}\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k).\end{aligned} \tag{A.3}$$

where $\beta(k) = \mathbf{r}_i^H(k)\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k)$. Thus, employing (A.3), the useful part of the filtered signal $\tilde{\mathbf{g}}_i^H(m)\mathbf{H}_{i,i}\tilde{\mathbf{v}}_i(m)$ can be rewritten as

$$\tilde{\mathbf{g}}_i^H(k)\mathbf{r}_i(k) = \frac{1}{1+\beta(k)}\mathbf{r}_i^H(k)\mathbf{M}_i^{-1}(k)\mathbf{r}_i(k) = \frac{\beta(k)}{1+\beta(k)}. \tag{A.4}$$

Since the received SINR $\tilde{\gamma}_i(k)$ is

$$\tilde{\gamma}_i(k) = \frac{|\tilde{\mathbf{g}}_i^H(k)\mathbf{r}_i(k)|^2}{\tilde{\mathbf{g}}_i^H(k)\mathbf{M}_i(k)\tilde{\mathbf{g}}_i(k)}, \tag{A.5}$$

we can replace $\tilde{\mathbf{g}}_i^H(k)\mathbf{r}_i(k)$ in the numerator with the expression found in (A.4) and, taking into account that $\mathbf{g}_i^H(k)\mathbf{M}_i(k)\mathbf{g}_i(k) = \frac{\beta(k)}{(1+\beta(k))^2}$, one obtains

$$\tilde{\gamma}_i(k) = \frac{\beta^2(k)}{(1+\beta(k))^2}\frac{(1+\beta(k))^2}{\beta(k)} = \beta(k). \tag{A.6}$$

Since from (21) it is

$$\tilde{e}_i(m) = 1 - \tilde{\mathbf{g}}_i^H(m)\mathbf{r}_i(m) = 1 - \frac{\beta(m)}{1+\beta(m)} = \frac{1}{1+\beta(m)}, \tag{A.7}$$

one immediately finds that

$$\frac{1}{\tilde{e}_i(m)} = 1 + \beta(m) = 1 + \gamma_i(m). \tag{A.8}$$

References

[1] K. Venugopal, M. C. Valenti, and R. W. Heath, "Device-to-device millimeter wave communications: Interference, coverag, rate, and finite topologies," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, Sep. 2016.

[2] C. Campolo, A. Molinaro, G. Araniti, and A. O. Berthet, "Better platooning control toward autonomous driving – an LTE device-to-device communications strategy that meets ultralow latency requirements," *IEEE Vehic. Techn. Mag.*, Mar. 2017.

[3] 3GPP, "Study on enhancement of 3GPP support for 5G V2X services (Release 15)," www.3gpp.org, Technical Report TR 22.886 V15.0.0, 2016.

[4] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, and Q. L. ad N. Brahmi, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, Dec. 2014.

[5] H. Seo, K.-D. Lee, S. Yasukawa, Y. Peng, and P. Sartori, "LTE evolution for vehicle-to-everything services," *IEEE Commun. Mag.*, vol. 54, no. 6, Jun. 2016.

[6] S. Chen, J. Hu, Y. Shi, Y. Peng, J. Fang, R. Zaho, and L. Zhao, "Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G," *IEEE Commun. Standards Mag.*, Jun. 2017.

[7] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process*, vol. 51, no. 11, pp. 2839–2848, 2003.

[8] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[9] 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 Transactions on Signal Processing*, vol. 59, no. 9, Sep. 2011.

[10] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, April 2016.

[11] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Transaction on Signal Processing*, vol. 62, no. 3, pp. 641–656, February 2014.

[12] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, and O.-S. Shin, "Sum rate maximization based on sub-array antenna selection in a full-duplex system," in *IEEE GLOBECOM*, 2017.

[13] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 206–218, Jan 2011.

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

[15] J. Shin and J. Moon, "Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2776–2783, October 2012.

[16] X. Jing, L. Mo, H. Liu, and C. Zhang, "Linear space-time interference alignment for K-user MIMO interference channels," *IEEE Access*, vol. 6, pp. 3085–3095, 2018.

[17] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[18] S. Han, C.-L. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[19] R. Mendez-Rial, C. Rusu, N. Gonzales-Prelic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, no. 1, pp. 247–267, March 2016.

[20] H. Gauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave mimo systems," *IEEE Journal on Selectect Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, April 2016.

[21] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmwave systems," *IEEE Trans. on Green Communications and Networking*, vol. 2, no. 2, pp. 385–396, June 2018.

[22] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2714–2720, June 2015.

[23] C. Ma, J. Shi, N. Huang, and M. Chen, "Energy-efficient hybrid precoding for millimeter wave systems in MIMO interference channels," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, May 2016.

[24] D. H. N. Nguyen, L. B. Le, T. L. Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19 167–19 181, October 2017.

[25] C. K. Thomas and D. Slock, "Mixed time scale weighted sum rate maximization for hybrid beamforming in multi-cell MU-MIMO systems," in *IEEE Globecom Workshops*, Singapore, December 2017.

[26] ——, "Hybrid beamforming design in multi-cell MU-MIMO systems with per-RF or per-antenna power constraints," in *88th IEEE Vehicular Technology Conference VTC-2018 Fall*, Chicago, IL, USA, August 2018.

[27] ——, "Deterministic annealing for hybrid beamforming design in multi-cell MU-MIMO systems," in *19th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018.

[28] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, April 2016.

[29] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[30] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[31] A. Li and C. Masouros, "Hybrid analog-digital millimeter-wave MU-MIMO transmission with virtual path selection," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 438–441, February 2017.

[32] J. Choi, "Analog beamforming for low-complexity multiuser detection in mm-wave systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6747–6752, August 2016.

[33] C. Shi, R. A. Berry, and M. L. Honig, "Bi-directional training for adaptive beamforming and power control in interference networks," *IEEE Transaction on Signal Processing*, vol. 62, no. 3, pp. 607–618, February 2014.

[34] J. Kaleva, A. Tolli, M. Juntti, R. Berry, and M. Honig, "Decentralized joint precoding for WSRMax with pilot aided beamformer estimation," in *IEEE GLOBECOM*, 2017.

[35] R. Brandt and M. Bengtsson, "Distrubuted CSI acquisition and coordinated precoding for TDD multicell MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2890–2906, May 2016.

[36] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, *Weighted Sum-Rate Maximization in Wireless Networks: A Review*, N. Publishing, Ed. Foundations and Trends in Networking, 2011, vol. 6, no. 1-2, dOI: 10.1561/1300000036.

[37] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

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

[39] G. D. Forney Jr, "Shannon meets Wiener ii: On MMSE estimation in successive decoding schemes," *arXiv preprint cs/0409011*, 2004.

[40] A. Abrardo and M. Moretti, "Distributed power allocation for D2D communications underlaying/overlaying OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1466–1479, March 2017.

[41] D. Monderer and L. S. Shapley, *Potential Games*, I. Academic Press, Ed. *Games and Economic Behavior*, 1996, vol. 14, no. 1.