

A rank-constrained coordinate ascent approach to hybrid precoding for the downlink of wideband massive MIMO systems

José P. González-Coma, Óscar Fresnedo, *Member, IEEE*, Luis Castedo, *Senior Member, IEEE*

Abstract—An innovative approach to hybrid analog-digital precoding for the downlink of wideband massive MIMO systems is developed. The proposed solution, termed Rank-Constrained Coordinate Ascent (RCCA), starts seeking the full-digital precoder that maximizes the achievable sum-rate over all the frequency subcarriers while constraining the rank of the overall transmit covariance matrix. The frequency-flat constraint on the analog part of the hybrid precoder and the non-convex nature of the rank constraint are circumvented by transforming the original problem into a more suitable one, where a convenient structure for the transmit covariance matrix is imposed. Such structure makes the resulting full-digital precoder particularly adequate for its posterior analog-digital factorization. An additional problem formulation to determine an appropriate power allocation policy according to the rank constraint is also provided. The numerical results show that the proposed method outperforms baseline solutions even for practical scenarios with high spatial diversity.

Index Terms—massive MIMO, hybrid precoding, wideband, rank constraints.

I. INTRODUCTION

MASSIVE MIMO is a fundamental technology to reach the foreseen challenging requirements of data throughput, spectral efficiency and user accommodation of the future wireless communication systems [1, 2]. In this context, a large range of communication scenarios can benefit from using large arrays of antennas, including mobile networks [3, 4], vehicular-to-infrastructure (V2I) communications [5, 6], control of unnamed aerial vehicles (UAVs) [7, 8], or machine-to-machine (M2M) communications [9]. However, unlike conventional MIMO, having one dedicated radio frequency (RF) chain per antenna element is unaffordable in massive MIMO because of its extremely high cost and power consumption [3]. Consequently, massive MIMO transceivers usually have a hybrid digital-analog structure which combine an analog RF precoding network, with a limited number of RF chains, and a digital baseband processing unit [10–12]. In this work, we consider hybrid precoding for the downlink of a wideband massive MIMO system. The design of high performance hybrid precoders is significantly more challenging in wideband situations because analog RF precoders are frequency-flat, i.e., their frequency response is constant over all the available bandwidth [13].

In the literature, wideband hybrid precoding has been addressed for both single-user [14–17] and multiuser [13, 18–21] massive MIMO systems. For instance, the authors in [13] exploit the similarity between the channel responses at different subcarriers for millimeter-wave (mmWave) systems with a large number of antennas at both ends. Thus, the performance of the solutions in [13] strictly relies on the assumption of high similarity among the structure of channel matrices at different subcarriers. In [18], hybrid wideband precoders are designed using an alternating optimization algorithm based on the equivalence between sum-rate optimization and weighted sum mean squared error (MSE) minimization. This solution, however, is limited to single-stream transmissions and single-antenna users. The authors in [19] extend the use of linear successive allocation (LISA) to the design of wideband hybrid precoders by exploiting the common subspace structure for the subcarriers of each user. A new stream is allocated to a given user at each iteration of the procedure, which causes less interference to the remaining users. However, this approach degrades severely for channels with high frequency selectivity, as it quickly consumes all the available spatial degrees of freedom. In [20], the authors propose an alternating algorithm to minimize the Frobenius norm between an unconstrained full-digital solution and the wideband hybrid precoders, but this approach only applies under the assumptions of single-stream transmissions and equal power allocation on each subcarrier. The work in [21] proposes a design of the hybrid precoders for wideband mmWave massive MIMO systems based on maximizing the mutual information with a block diagonalization procedure. In this case, the frequency-dependent baseband precoders must be able to completely cancel the multiuser interference, which is a difficult task in general because of the low dimensions of these precoding matrices for a reasonable number of RF chains. This effect also presents in other methods pursuing the same of strategy of interference removal in the baseband stage, e.g. [13, 22, 23]. Two-timescale hybrid precoding approaches are considered in [24–26], where the frequency-flat analog precoding network is designed according to the channel statistics, whereas the digital baseband matrices are updated according to the time-varying Channel State Information (CSI) for each channel realization. The performance of these strategies for wideband scenarios relies on the assumption that the channel statistics are similar over the whole bandwidth, and hence the resulting RF precoder is adequate for all the subcarriers. This assumption, however, is unrealistic in large bandwidth scenarios.

On the other hand, most of the previous work assume line of sight (LOS) channel models with a limited number of reflection paths [13, 19, 27], which makes the inter-user

Copyright (c) 2015 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.

J.P. González-Coma is with Defense University Center at the Spanish Naval Academy, Plaza de España s/n, 36920, Marín, Spain.

O. Fresnedo and L. Castedo are with Dept. Computer Engineering & CITIC Research Center, University of A Coruña, Campus de Elviña s/n, 15071, A Coruña.

interference easier to handle. These assumptions, however, are not practical in mmWave indoor and sub-6 GHz scenarios, where the channel matrices are spatially rich. This effect hinders the similarity of the channel spatial features over different subcarriers, and it is even more pronounced when considering large bandwidths [28]. In these scenarios, the design of hybrid precoders with a frequency-flat analog component is a challenging problem.

In this paper, we address the general case of hybrid precoding for the downlink of a wideband massive MIMO system. In particular, we develop an innovative strategy based on limiting the rank of the full-digital precoders while exploiting the spatial features of the wideband channel matrices. This decision is motivated by the lack of accuracy of the methods that approximate the digital solutions for multiuser wideband scenarios, which suffer from a drastic dimensional reduction that severely limits the achievable performance. Accordingly, we design digital precoders with a rank restriction which effectively reduces the number of allocated data streams. The aim is to obtain digital precoders more suitable for the subsequent hybrid factorization. In this sense, since the subspace spanned by the rank-constrained digital design has a limited dimension, the corresponding hybrid design will be capable of handling wideband multiuser channels more effectively. Few existing works have considered the incorporation of rank constraints to address the limited number of available RF chains [29, 30]. Indeed, the authors in [29] showed a relationship between the rank of the full-digital precoders and the accuracy of its corresponding hybrid factorization. Building on this idea, if the rank of a full-digital solution is adjusted according to the number of available RF chains, its corresponding analog and digital counterparts would achieve similar performance results.

However, the incorporation of rank constraints into the problem formulation for wideband scenarios significantly complicates the design of hybrid precoders due to the flat-frequency nature of the analog RF part. To circumvent this issue, we propose an original approach termed as Rank-Constrained Coordinate Ascent (RCCA), where all the user transmit covariance matrices are alternatively updated while constraining the rank of the overall wideband transmit covariance matrix. This is accomplished thanks to the use of a common frequency-flat structure for each user transmit covariance matrix, combined with a power allocation that is jointly designed for all the users and subcarriers. While the common transmit covariance matrix structure accounts for the inter-user interference at all the subcarriers, the power allocation is intended to select the most promising data streams of each user.

The main contributions of this work are summarized as follows:

- An RCCA algorithm for the design of hybrid precoders for the downlink of wideband massive MIMO systems is developed. Unlike existing approaches, this algorithm does not employ a codebook or minimizes the distance to optimized full-digital solutions according to some metric. Instead, it determines full-digital solutions with a particular structure, which is suitable for their posterior analog-digital factorization. The proposed design also aims at exploiting the potential similarity among the user

subcarriers. However, unlike other previous approaches [13, 21, 27], the proposed solution addresses a general channel model, since it does not make explicit assumptions on the frequency response of the subcarriers (e.g., mmWave scenarios can assume that the channel response at different subcarriers is rather similar).

- The initial rank-constrained optimization problem is transformed into a sequence of steps which provide a useful insight about the desired structure of the user transmit covariance matrices. Also, the transmit power distribution among users and subcarriers is specifically designed to incorporate the rank constraint, and to consider the inter-user interference across the entire channel bandwidth.
- Under the assumption of a rank constraint design of the hybrid precoders, we prove that the allocation of exactly the same number of streams as that of available RF chains leads to the optimal system performance in terms of achievable sum-rate.

II. SYSTEM MODEL

Let us consider the downlink of a wideband massive MIMO system where a base station (BS) with M antennas communicates to U users with R antennas each. The number of RF chains at the BS is limited to $N_{\text{RF}} \ll M$. As in [10–12], we also assume that $R \ll M$. Orthogonal Frequency Division Multiplexing (OFDM) modulation with K subcarriers is utilized for transmission over the wideband channels.

Let $s_u[k]$ be the data symbols transmitted to user u over subcarrier k , with $u = 1, \dots, U$ and $k = 1, \dots, K$. We assume that the BS allocates $N_{s,u}$ data streams to user u . Hence, $s_u[k] \in \mathbb{C}^{N_{s,u} \times 1}$ and the total number of streams transmitted by the BS is $N_s = \sum_{u=1}^U N_{s,u}$. We also assume that $\mathbb{E}[s_u[k]s_j^H[k]] = \mathbf{0}$ for $j \neq u$, and $\mathbb{E}[s_u[k]s_u^H[l]] = \mathbf{0}$ for $k \neq l$. The user data is linearly precoded using a limited number of RF chains at the BS, following a hybrid digital-analog architecture, with $\mathbf{P}_{\text{BB}}^u[k] \in \mathbb{C}^{N_{\text{RF}} \times N_{s,u}}$ being the digital baseband precoder for user u at the k -th subcarrier. Hence, the discrete-time digitally precoded OFDM symbol vector $s_u^n \in \mathbb{C}^{N_{\text{RF}}}$ reads as

$$s_u^n = \sum_{k=1}^K \mathbf{P}_{\text{BB}}^u[k] s_u[k] e^{j \frac{2\pi(k-1)n}{K}}, \quad n = 0, 1, \dots, K-1. \quad (1)$$

Note the number of OFDM symbols in the time domain is also K . We consider that data symbols s_u^n in (1) modulate with the square-root raised cosine pulse-shaping filter $p_{\text{src}}(t)$ such that $p_{\text{src}}(t) * p_{\text{src}}(-t) = p_{\text{rc}}(t)$ is the raised cosine pulse signal which fulfills the Nyquist zero inter-symbol interference (ISI) criterion. Next, the resulting continuous-time vector signal inputs the analog precoding stage modeled by the matrix $\mathbf{P}_{\text{RF}} \in \mathcal{P}^{M \times N_{\text{RF}}}$. This matrix is frequency-flat and common to all users and \mathcal{P} represents the set of feasible RF precoder matrices with unit-modulus entries. Correspondingly, the time-domain signal vector for user u is given by

$$s_u(t) = \sum_{n=0}^{N_c-1} \sum_{k=1}^K \mathbf{P}_{\text{RF}} \mathbf{P}_{\text{BB}}^u[k] s_u[k] p_{\text{src}}(t - nT_s) e^{j \frac{2\pi(k-1)n}{K}}, \quad (2)$$

where T_s is the sampling period. The linear combination of the signals $s_u(t)$ for the different users is transmitted through the wireless channel, for which we consider the following general geometrical model for the channel responses [17, 31, 32]

$$\mathbf{H}_u(t)[k] = \gamma \sum_{p=1}^{N_p} \alpha_{u,p} \delta(t - \tau_{u,p}) \mathbf{a}_{R,u}(\theta_{u,p})[k] \mathbf{a}_T^H(\phi_{u,p})[k], \quad (3)$$

with $\gamma = \sqrt{MR/N_p}$ being the power normalization factor, N_p the number of channel paths, $\alpha_{u,p}$ the complex gain, $\tau_{u,p}$ the delay, and $\mathbf{a}_{R,u}(\theta_{u,p})[k] \in \mathbb{C}^{R \times 1}$ and $\mathbf{a}_T(\phi_{u,p})[k] \in \mathbb{C}^{M \times 1}$ the frequency dependent array response vectors of the user and the BS at the corresponding angular directions $\theta_{u,p}$ and $\phi_{u,p}$ for subcarrier k , respectively. Therefore, the discrete-time equivalent channel response at the d -delay tap, which includes the effects of the transmit and receive pulse-shaping filters, can be written as

$$\mathbf{H}_u^d[k] = \gamma \sum_{p=1}^{N_p} \alpha_{u,p} p_{rc}(dT_s - \tau_{u,p}) \mathbf{a}_{R,u}(\theta_{u,p})[k] \mathbf{a}_T^H(\phi_{u,p})[k], \quad (4)$$

where $d \in \{1, \dots, D\}$, being D the maximum number of delay taps. From the former equation, the spatial relationship among the different subcarriers is apparent, specially for those with similar indices k . Moreover, the usage of uniform linear arrays (ULAs) at both ends is considered here, although the proposed scheme is applicable to other configurations. Therefore, the array response vectors are given by [19, 23]

$$\mathbf{a}_T(\phi)[k] = \frac{1}{\sqrt{M}} \left[1, e^{j \frac{2\pi c}{f[k]} d \sin \phi}, \dots, e^{j \frac{2\pi c}{f[k]} d(M-1) \sin \phi} \right]^T, \quad (5)$$

where $f[k]$ is the frequency at the k -th subcarrier, c is the speed of light, and d is the inter-antenna spacing. The carrier frequency for the k -th sub-band is given by $f[k] = f_c + \xi[k]$, where f_c is the central frequency and $\xi[\ell]$ represents an offset with respect to f_c . If we assume a signal bandwidth of B , the frequency offset for the k -th subcarrier reads as $\xi[k] = (k-1 - \frac{K-1}{2}) \frac{B}{K}$. Note that the expression for $\mathbf{a}_R(\theta)[k]$ is similar to that of $\mathbf{a}_T(\phi)[k]$, and thus omitted for brevity. When the signal bandwidth satisfies $B \ll f_c$, $f[k] \approx f_c$ and the response vectors are not affected by the frequency offsets. On the contrary, when B is comparable to f_c the beam-squint effect arises in (5).

In order to suppress ISI and inter-carrier interference (ICI), a cyclic prefix large enough is added to the OFDM symbol in (1). Moreover, the basis functions employed to build the time-domain signal in (2) are also chosen to constitute an orthogonal set on the subcarrier index k . Combining these two facts, the discrete-time equivalent of the signal received by user u over subcarrier k , $\mathbf{x}_u[k] \in \mathbb{C}^{R \times 1}$, can be expressed as [33]

$$\begin{aligned} \mathbf{x}_u[k] &= \sum_{d=0}^{D-1} \mathbf{H}_u^d[k] \sum_{i=1}^U \mathbf{P}_{RF} \mathbf{P}_{BB}^i[k] s_i[k] e^{-j \frac{2\pi(k-1)d}{K}} + \mathbf{n}_u \\ &= \mathbf{H}_u[k] \sum_{i=1}^U \mathbf{P}_{RF} \mathbf{P}_{BB}^i[k] s_i[k] + \mathbf{n}_u, \end{aligned} \quad (6)$$

where $\mathbf{H}_u[k] \in \mathbb{C}^{R \times M}$ is the u -th user channel frequency response matrix at subcarrier k , $\mathbf{H}_u[k] = \sum_{d=0}^{D-1} \mathbf{H}_u^d[k] e^{-j \frac{2\pi(k-1)d}{K}}$,

and $\mathbf{n}_u \sim \mathcal{N}_C(\mathbf{0}, \sigma_n^2 \mathbf{I}_R)$ is the receiving additive white Gaussian noise (AWGN). The available transmit power at the BS is P_{tx} , and hence the system Signal-to-Noise Ratio (SNR) is defined as $\text{SNR} = P_{tx}/\sigma_n^2$. Without loss of generality, we assume $\sigma_n^2 = 1$ for simplicity.

III. INITIAL PROBLEM FORMULATION

Considering the communication model described in the previous section, we are interested in exploring the design of hybrid precoders which maximize the theoretically achievable sum-rate. This metric is extensively used in the literature to measure the overall system performance [12, 13, 20, 22, 34]. Assuming Gaussian signaling and superposition coding with successive interference cancellation, the achievable rate for user u over subcarrier k is given by

$$R_u[k] = \log_2 \det \left(\mathbf{I}_R + \mathbf{X}_u^{-1}[k] \mathbf{H}_u[k] \mathbf{P}_{\text{H}}^u[k] \mathbf{P}_{\text{H}}^{u,H}[k] \mathbf{H}_u^H[k] \right),$$

with $\mathbf{X}_u[k] = \sum_{i>u} \mathbf{H}_u[k] \mathbf{P}_{\text{H}}^i[k] \mathbf{P}_{\text{H}}^{i,H}[k] \mathbf{H}_u^H[k] + \mathbf{I}_R$ containing the interference plus noise terms for a given user decoding order, and with $\mathbf{P}_{\text{H}}^i[k] = \mathbf{P}_{\text{RF}} \mathbf{P}_{\text{BB}}^i[k]$ the hybrid precoding matrix for the i -th user at subcarrier k . Note that the expression above represents the individual rates which could be obtained theoretically with a configuration based on successive interference cancellation, and therefore they would actually lead to an upper bound for the achievable sum-rate in practical systems. This assumption also matches the common strategy in which the precoder is first designed assuming an ideal receiver, and then the corresponding receiver is calculated for this precoder, thus enabling decoupled precoding and decoding problems [11, 13, 23, 35–37].

As already mentioned, we aim at determining the precoder matrices \mathbf{P}_{RF} and $\mathbf{P}_{\text{BB}}^u[k]$, $u = 1, \dots, U$, $k = 1, \dots, K$, which maximize the achievable sum-rate given by

$$R = \frac{1}{K} \sum_{k=1}^K \sum_{u=1}^U R_u[k],$$

for a fixed transmit power P_{tx} . Hence, the corresponding constrained optimization problem can be formulated as

$$\max_{\mathbf{P}_{\text{RF}}, \mathbf{P}_{\text{BB}}^u[k]} R \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{u=1}^U \|\mathbf{P}_{\text{RF}} \mathbf{P}_{\text{BB}}^u[k]\|_F^2 \leq P_{tx}, \quad \mathbf{P}_{\text{RF}} \in \mathcal{P}. \quad (7)$$

This difficult optimization problem can be addressed by first determining the full-digital precoders which maximize the achievable sum-rate R , and next factorizing them into their RF and baseband counterparts. Thus, by defining the overall hybrid transmit covariance matrix as

$$\begin{aligned} \mathbf{Q}_{\text{H}} &= \mathbf{P}_{\text{RF}} \left[\mathbf{P}_{\text{BB}}^1[1] \mathbf{P}_{\text{BB}}^{1,H}[1], \dots, \mathbf{P}_{\text{BB}}^1[K] \mathbf{P}_{\text{BB}}^{1,H}[K], \right. \\ &\quad \left. \dots, \mathbf{P}_{\text{BB}}^U[K] \mathbf{P}_{\text{BB}}^{U,H}[K] \right] \left(\mathbf{I}_{UK} \otimes \mathbf{P}_{\text{RF}}^H \right), \end{aligned} \quad (8)$$

we can observe that a side effect of using hybrid precoding is the imposition of the restriction $\text{rank}(\mathbf{Q}_{\text{H}}) \leq N_{\text{RF}}$. Building on this idea and contrary to approaches like, e.g. [20, 37], we impose a rank constraint on the design of the full-digital

precoders with the aim of obtaining a more accurate factorization in a later step. For that, we first express the achievable sum-rate in the downlink as follows

$$R^D = \frac{1}{K} \sum_{u=1}^U \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_R + \mathbf{Y}_u^{-1}[k] \mathbf{H}_u[k] \mathbf{Q}_u[k] \mathbf{H}_u^H[k] \right),$$

where the matrix $\mathbf{Q}_u[k]$ is introduced to represent the full-digital transmit covariance matrix for the k -th subcarrier of user u , and $\mathbf{Y}_u[k] = \sum_{i>u} \mathbf{H}_i[k] \mathbf{Q}_i[k] \mathbf{H}_i^H[k] + \mathbf{I}_R$. Note that the super-index D refers to the downlink. Next, we introduce the following rank-constrained formulation

$$\max_{\{\mathbf{Q}_u[k]\}_{u=1, k=1}^{U, K}} R^D \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{Q}_u[k]) \leq P_{\text{tx}},$$

$$\text{rank}(\mathbf{Q}) \leq N_{\text{RF}}, \quad (9)$$

where $\mathbf{Q} = [\mathbf{Q}_1[1], \dots, \mathbf{Q}_1[K], \dots, \mathbf{Q}_U[K]] \in \mathbb{C}^{M \times MUK}$ is the overall transmit covariance matrix. In particular, the rank constraint in (9) forces the precoder design to satisfy $\text{rank}(\mathbf{Q}) \leq N_{\text{RF}}$; that is, the columns of \mathbf{Q} will span a subspace of dimension smaller than or equal to N_{RF} on the vector space \mathbb{C}^M . Observe that the maximum subspace dimension achievable by the span of \mathbf{Q}_H in (8) also satisfies this restriction. Therefore, our strategy limits the performance of the full-digital solution to ensure that it can be accurately approximated by means of a factorization procedure. Since we are imposing a rank constraint but not restricting the entries of \mathbf{Q} in any way, the problem formulation in (9) can be interpreted as a relaxed version of the original one in (7), and hence the resulting sum-rate will actually be an upper bound to that obtained in (7). Note also that, contrary to other approaches like e.g. [29, 30], we do not focus on selecting the best subset of users since all of them will be served if $K \leq N_{\text{RF}}$ holds.

IV. REFORMULATION OF THE OPTIMIZATION PROBLEM

Solving the optimization problem (9) is challenging due to the coupling between the covariance matrices for different users and subcarriers and the presence of the rank restriction. These constraints are non-convex, and hence it is difficult to deal with them. For this reason, it is common to transform rank-constrained problems into other ones that are easier to handle by means of relaxed formulations or convex surrogates [38, 39]. In this case, we start leveraging the duality between the achievable sum-rate for the downlink and a virtual uplink [40]. Assuming successive decoding, the achievable sum-rate for the dual virtual uplink is given by

$$R^U = \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_M + \sum_{u=1}^U \mathbf{H}_u^H[k] \mathbf{S}_u[k] \mathbf{H}_u[k] \right), \quad (10)$$

where $\mathbf{H}_u^H[k]$ and $\mathbf{S}_u[k]$ are the uplink channel response and the uplink transmit covariance matrices for user u over the k -th subcarrier, respectively. Observe that no particular user decoding order is assumed in (10). Therefore, different orderings differ on the individual achievable rates while leading to the same achievable sum-rate. Moreover, the result in [40] guarantees that the uplink sum-rate R^U is achieved in the

downlink by selecting the transmit covariance matrices as $\mathbf{Q}_u[k] = \Delta_u[k] \mathbf{S}_u[k] \Delta_u^H[k]$, where

$$\Delta_u[k] = \mathbf{B}_u^{-1/2}[k] \mathbf{F}_u[k] \mathbf{G}_u^H[k] \mathbf{A}_u^{1/2}[k] \quad (11)$$

are the uplink-downlink conversion matrices. These matrices are determined via the auxiliary matrices

$$\mathbf{A}_u[k] = \mathbf{I}_R + \sum_{i=1}^{u-1} \mathbf{H}_i[k] \mathbf{Q}_i[k] \mathbf{H}_i^H[k]$$

$$\mathbf{B}_u[k] = \mathbf{I}_M + \sum_{i=u+1}^U \mathbf{H}_i^H[k] \mathbf{S}_i[k] \mathbf{H}_i[k] \quad (12)$$

where $\mathbf{F}_u[k] \in \mathbb{C}^{M \times R}$ and $\mathbf{G}_u[k] \in \mathbb{C}^{R \times R}$ result from the singular value decomposition (SVD) $\mathbf{F}_u[k] \mathbf{\Xi}_u[k] \mathbf{G}_u^H[k] = \mathbf{B}_u^{-1/2}[k] \mathbf{H}_u^H[k] \mathbf{A}_u^{-1/2}[k]$. Notice that $\mathbf{A}_u[k]$ and $\mathbf{B}_u[k]$ are obtained in sequential order, i.e., for $u = 1, 2, \dots, U$, and independently for each subcarrier.

By means of this duality, the optimization problem in (9) is reformulated as

$$\max_{\{\mathbf{S}_u[k] \geq 0\}_{u=1, k=1}^{U, K}} R^U \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{S}_u[k]) \leq P_{\text{tx}},$$

$$\text{rank}(\mathbf{W}) \leq N_{\text{RF}}, \quad (13)$$

with $\mathbf{W} = [\Delta_1[1] \mathbf{S}_1[1] \Delta_1^H[1], \dots, \Delta_U[K] \mathbf{S}_U[K] \Delta_U^H[K]]$. Recall that the result in [40] guarantees that the achievable sum-rate R^U obtained with a set of uplink transmit covariance matrices $\mathbf{S}_u[k] \forall u, k$ is also achievable in the downlink using $\mathbf{Q}_u[k] = \Delta_u[k] \mathbf{S}_u[k] \Delta_u^H[k] \forall u, k$, i.e., $R^U = R^D$ under the same total power constraint $\sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{S}_u[k]) = \sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{Q}_u[k])$. Moreover, (13) includes the uplink-downlink conversion matrices, $\Delta_u[k] \forall k, u$, in the restriction $\text{rank}(\mathbf{W}) \leq N_{\text{RF}}$ to ensure that the downlink rank restriction holds. Therefore, $\text{rank}(\mathbf{W}) \leq N_{\text{RF}}$ in (13) is equivalent to $\text{rank}(\mathbf{Q}) \leq N_{\text{RF}}$ in (9). By virtue of the total power and the rank conditions, optimal solutions of (13) also lead to optimal solutions for the problem in (9). Accordingly, we first solve (13) in the dual virtual uplink and next obtain the corresponding downlink precoders with the help of the uplink-downlink conversion matrices. Towards this aim, we develop two intermediate steps to circumvent the rank restriction in (13). In the first step, a common spatial structure for the covariance matrices is proposed for all the frequencies of each user u and, based on that decision, the second step determines the power allocation for each subcarrier k .

A. Frequency-flat transmission covariance matrix structure

To address the first step, we must consider the involved structure of \mathbf{W} in (13). A design of $\mathbf{S}_u[k]$ independent for each subcarrier would definitely lead to large values for the rank of \mathbf{W} . Moreover, in a more practical sense, a design of $\mathbf{S}_u[k]$ exploiting all the available degrees of freedom is hardly achievable with a frequency-flat analog precoder design. Therefore, we will start by providing some insight regarding the structure of a frequency-flat transmit covariance matrix for a particular user.

For that, let us first rewrite (10) as $R^U = \sum_{u=1}^U R_u^U$. Moreover, we assume that user u is decoded first, without loss of optimality, to obtain

$$R_u^U = \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_M + \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u[k] \mathbf{H}_{\bar{u}}[k] \right), \quad (14)$$

where $\mathbf{H}_{\bar{u}}[k] = \mathbf{H}_u[k] (\sum_{i \neq u} \mathbf{H}_i^H[k] \mathbf{S}_i[k] \mathbf{H}_i[k] + \mathbf{I}_M)^{-\frac{1}{2}}$ represents the u -th user effective channel for subcarrier k . With the purpose of studying the desirable properties of a frequency-flat covariance matrix, we substitute $\mathbf{S}_u[k]$ by its frequency-flat version \mathbf{S}_u in (14), leading to

$$\bar{R}_u^U = \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_M + \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u \mathbf{H}_{\bar{u}}[k] \right). \quad (15)$$

Fixing the transmit covariance matrices of the remaining users, i.e. $\mathbf{S}_i[k] \forall k$ with $i \neq u$, the matrix \mathbf{S}_u which maximizes (15) can be determined independently, thus ensuring an increase on the achievable sum-rate given by $\bar{R}^U = \sum_{i \neq u} R_i^U + \bar{R}_u^U$.

Under the previous considerations, we can approach the problem in (13) piecemeal by stating the next optimization problem for the update of the u -th covariance matrix,

$$\max_{\mathbf{S}_u \geq \mathbf{0}} \bar{R}_u^U \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{i \neq u} \text{tr}(\mathbf{S}_i[k]) + \text{tr}(\mathbf{S}_u) \leq P_{\text{tx}}, \quad (16)$$

where the rank constraint is for the time being disregarded. The formulation in (16) leads to the following Lagrangian function

$$L(\mathbf{S}_u, \lambda_u, \mathbf{\Lambda}_u) = \bar{R}_u^U + \text{tr}(\mathbf{\Lambda}_u \mathbf{S}_u) - \lambda_u \left(\sum_{k=1}^K \sum_{i \neq u} \text{tr}(\mathbf{S}_i[k]) + \text{tr}(\mathbf{S}_u) - P_{\text{tx}} \right), \quad (17)$$

with $\lambda_u \geq 0$ and $\mathbf{\Lambda}_u \geq \mathbf{0}$ an auxiliary matrix to guarantee the constraints on \mathbf{S}_u . In this scenario, we can readily obtain the Karush-Kuhn-Tucker (KKT) conditions as

$$\frac{\partial L(\mathbf{S}_u, \lambda_u, \mathbf{\Lambda}_u)}{\partial \mathbf{S}_u} = \mathbf{\Lambda}_u - \lambda_u \mathbf{I}_R + \sum_{k=1}^K \left(\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u \right)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k], \quad (18)$$

with $\mathbf{S}_u \mathbf{\Lambda}_u = \mathbf{\Lambda}_u \mathbf{S}_u = \mathbf{0}$, and

$$\lambda_u \left(\sum_{k=1}^K \sum_{i \neq u} \text{tr}(\mathbf{S}_i[k]) + \text{tr}(\mathbf{S}_u) - P_{\text{tx}} \right) = 0.$$

Now, multiplying the expression in (18) times \mathbf{S}_u we get

$$\left(\sum_{k=1}^K \left(\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u \right)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k] - \lambda_u \mathbf{I}_R \right) \mathbf{S}_u = \mathbf{0}. \quad (19)$$

Proposition 1. *The KKT condition in (19) holds for $K = 1$ and \mathbf{S}_u fulfilling $(\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k] > \mathbf{0}$.*

Proof. See Appendix A. \square

It is apparent that finding a frequency-flat matrix \mathbf{S}_u fulfilling Prop. 1 simultaneously for all the subcarriers is difficult when $K > 1$. Notwithstanding,

if these conditions approximately hold for each individual subcarrier, i.e., we select a matrix \mathbf{S}_u which approximates $(\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k]$ to a positive-definite matrix for $k = 1, \dots, K$, the sum $\sum_{k=1}^K (\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k]$ will also approach a positive-definite matrix. For that reason, we propose to select a covariance matrix \mathbf{S}_u satisfying

$$\sum_{k=1}^K \left(\mathbf{I}_R + \mathbf{H}_{\bar{u}}^H[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{S}_u \right)^{-1} \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k] \approx \mathbf{Z}_u \mathbf{D}_u \mathbf{Z}_u^H, \quad (20)$$

where \mathbf{D}_u is a diagonal matrix and \mathbf{Z}_u is a unitary matrix. In this case, the KKT condition in (19) will approximately be met.

Applying this procedure iteratively to update the covariance matrix for each of the users will increase the performance metric, as we discuss in Section V. Moreover since $\mathbf{S}_u, \forall u$ are frequency-flat matrices, if these transmit covariance matrices are also feasible solutions of (13), that they actually provide a lower bound for the achievable performance in (13).

B. Frequency-selective power allocation

As explained in Prop. 1, it is advisable to rewrite the u -th user frequency-flat covariance matrix as $\mathbf{S}_u = \mathbf{U}_u \mathbf{\Psi}_u \mathbf{U}_u^H$, with \mathbf{U}_u unitary and $\mathbf{\Psi}_u$ diagonal, to deal with the requirement of a common analog precoder for all the subcarriers. In addition, utilizing a common basis \mathbf{U}_u for the K subcarriers facilitates achieving a low rank condition for \mathbf{W} in (13). However, taking into account the flexibility provided by the frequency-selective baseband precoders, it is reasonable to independently adjust the diagonal matrix $\mathbf{\Psi}_u$ for each subcarrier. That is, we propose to design the frequency-selective covariance matrices $\mathbf{S}_u[k]$ as

$$\mathbf{S}_u[k] = \mathbf{U}_u \mathbf{P}_u[k] \mathbf{U}_u^H, \quad (21)$$

with $\mathbf{P}_u[k] \geq \mathbf{0}$ a diagonal matrix with the power allocation for user u and subcarrier k . The motivation for the design strategy given by (21) is threefold:

- 1) The similarity among the row subspaces for channel matrices corresponding to neighboring subcarriers is leveraged. Accordingly, \mathbf{U}_u is a basis used to span a vector space and $\mathbf{P}_u[k]$ selects some of the columns to define an appropriate subspace for each subcarrier k . Thus, if channel subspace similarity is strong, (20) approximately holds individually for most of the subcarriers.
- 2) The conversion matrices in (11) depend on the structure of the channel matrices (cf. (12)). As a consequence, the similarity among the channel matrix subspaces, together with the use of the common basis \mathbf{U}_u , enables us to obtain a low rank \mathbf{W} in (13) even when K is large, i.e.,

$$\text{rank}(\mathbf{W}) \approx \text{rank}([\mathbf{\Delta}_1[K_c] \mathbf{U}_1, \dots, \mathbf{\Delta}_U[K_c] \mathbf{U}_U]), \quad (22)$$

where the approximation comes from considering $\mathbf{\Delta}_u[k] \approx \mathbf{\Delta}_u[K_c], \forall k, u$, where K_c is the central subcarrier. This subcarrier is the one presenting the minimum distance, in terms of frequency offset, with respect to the most distant subcarriers. Therefore, we can expect

that the spatial features of $\Delta_u[k]$ with $k \neq K_c$ are similar to those of $\Delta_u[K_c]$.

- 3) Recall from (7) that the transmit power in a hybrid architecture is distributed via the baseband precoders $\mathbf{P}_{\text{BB}}^u[k]$. Hence, we can use the frequency-selective matrices $\mathbf{P}_u[k]$ in (21) to adjust the rank of the covariance matrices $\mathbf{S}_u[k]$ by selecting the more promising per-user streams in a per-subcarrier basis.

According to the spatial characteristics described in Sec. IV-A, we will use the covariance matrices in (21) to address the optimization problem in (13) by considering the relaxed restriction in (22). Thus, we will address the problem by assuming that the rank constraint is imposed by means of the right term in (22). In this context, it is still necessary to determine the power allocation matrices $\mathbf{P}_u[k]$. It is worth remarking that the power allocation policy must be designed to effectively ensure that

$$\text{rank}([\Delta_1[K_c]\mathbf{S}_1[1]\Delta_1^H[K_c], \dots, \Delta_U[K_c]\mathbf{S}_U[K]\Delta_U^H[K_c]) \leq N_{\text{RF}},$$

while searching for maximizing the sum-rate in (13). In the following, we propose a strategy for obtaining $\mathbf{P}_u[k]$ when \mathbf{U}_u is already given for $u = 1, \dots, U$.

Let us start introducing the spatially truncated channels $\mathbf{H}_{\bar{u}}(\beta)[k] = (\mathbf{I}_R - \sum_{n=1}^N \mathbf{u}_{u,i_n} \mathbf{u}_{u,i_n}^H) \mathbf{H}_{\bar{u}}[k]$, where $\mathbf{u}_{u,c}$ denotes the c -th column of \mathbf{U}_u . Recall that $\mathbf{H}_{\bar{u}}[k]$ represents the u -th user effective channel for subcarrier k . To define the column index set $\mathcal{I} = \{i_1, \dots, i_N\}$, consider the entries of the diagonal matrix

$$\boldsymbol{\Sigma}_u[k] = \text{diag}(\mathbf{U}_u^H \mathbf{H}_{\bar{u}}[k] \mathbf{H}_{\bar{u}}^H[k] \mathbf{U}_u) \quad (23)$$

as the potential equivalent channel gains for user u at subcarrier k . Hence, we build \mathcal{I} by establishing priorities for the columns of the basis \mathbf{U}_u as

$$\begin{cases} i \in \mathcal{I}, & \text{if } \prod_{k=1}^K [\boldsymbol{\Sigma}_u[k]]_{i,i} \geq \beta, \\ i \notin \mathcal{I}, & \text{otherwise.} \end{cases} \quad (24)$$

That is, the column i of the matrix \mathbf{U}_u is selected when the product of the potential gains across all the subcarriers is larger than some threshold β . Otherwise, the stream associated to such a column is discarded for the whole frequency band. Accordingly, the value of β is closely related to the rank expression in (22). Building on this idea, we propose to find the power allocation matrices $\mathbf{P}_u[k] \forall u, k$ for given $\mathbf{U}_u \forall u$ by solving the optimization problem

$$\begin{aligned} & \max_{\{\mathbf{P}_u[k]\}_{u,k=1}^{U,K}} \sum_{k=1}^K \sum_{u=1}^U \log_2 \det \left(\mathbf{I}_M + (\mathbf{V}_u(\beta)[k])^H \mathbf{P}_u[k] \mathbf{V}_u(\beta)[k] \right) \\ & \text{s.t. } \mathbf{P}_u[k] \in \mathcal{D}^+, \forall u, k \text{ and } \sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{P}_u[k]) \leq P_{\text{tx}}, \quad (25) \end{aligned}$$

where $\mathbf{V}_u(\beta)[k] = \mathbf{U}_u^H \mathbf{H}_{\bar{u}}(\beta)[k]$ and \mathcal{D}^+ is the set of diagonal matrices with non-negative entries. Note that the problem in (25) uses the spatial truncated channels $\mathbf{H}_{\bar{u}}(\beta)[k]$ to address the rank constraint, which leads to a significant reduction in the number of equivalent channel gains considered for the

power allocation procedure. Accordingly, a value of β high enough entails power allocation matrices $\mathbf{P}_u[k]$ such that the relaxed rank constraint in (22) is satisfied. However, since the optimal value of β is unknown beforehand, its determination will be included in our algorithmic solution. Notice also that the covariance matrices obtained from (25) are subject to more restrictive constraints than those in (13), due to the imposed frequency-flat basis \mathbf{U}_u .

In summary, to tackle the optimization problem in (13), we impose the particular structure for the user covariance matrices $\mathbf{S}_u[k]$ in (21), according to the KKT conditions for the frequency-flat basis \mathbf{U}_u . Next, we propose the optimization problem (25) to determine the frequency-selective $\mathbf{P}_u[k]$ power allocation matrices. In the ensuing section, we describe the proposed algorithm to implement this procedure.

V. PROPOSED ALGORITHM

In this section we develop an algorithmic approach to solve the problem in (13). At each iteration ℓ , the structure of the covariance matrices $\mathbf{S}_u^{(\ell)}[k]$ is updated considering the matrices for the remaining users fixed, according to the strategy explained in Sec. IV-A. To that end, the frequency-flat basis $\mathbf{U}_u^{(\ell)}$ is determined as the left singular vectors (LSV) of the matrix $[\mathbf{H}_{\bar{u}}^{(\ell)}[1], \mathbf{H}_{\bar{u}}^{(\ell)}[2], \dots, \mathbf{H}_{\bar{u}}^{(\ell)}[K]]$, with the aim of capturing the common structure for the effective channels corresponding to user u across the whole bandwidth. Note that this approach aims at satisfying the condition in (20). Indeed, when the row subspaces for the equivalent channels are similar, the condition approximately holds. Moreover, assuming similar row subspaces for neighboring subcarriers is reasonable in the proposed scenario [27, 32, 41]. In the following, we will omit the super-index (ℓ) for notation simplicity.

Next, in order to determine the power distribution among the users and subcarriers, i.e., $\mathbf{P}_u[k]$ in (21) for $u = 1, \dots, U$ and $k = 1, \dots, K$, we compute the equivalent channel gains $\boldsymbol{\Sigma}_u[k]$ in (23). Note also that we propose to determine the power allocation according to (25). Therefore, it is necessary to perform a search on β and solve (25) to obtain a power allocation that maximizes the objective function and fulfills the rank constraint on \mathbf{W} . To avoid this time-consuming search, we propose an alternative strategy. For that, it is important to remark that the objective function of (25) is upper bounded as follows

$$\begin{aligned} & \sum_{u=1}^U \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_M + (\mathbf{V}_u(\beta)[k])^H \mathbf{P}_u[k] \mathbf{V}_u(\beta)[k] \right) \\ & \leq \sum_{u=1}^U \log_2 \prod_{k=1}^K \prod_{i=1}^R \left(1 + [\mathbf{P}_u[k]]_{i,i} [\boldsymbol{\Sigma}_u[k]]_{i,i} \right), \quad (26) \end{aligned}$$

where $[\mathbf{P}_u[k]]_{i,i}$ represents the i -th entry of the power allocation matrix $\mathbf{P}_u[k]$. Hence, we can use an iterative procedure where the largest equivalent gain, considering all the subcarriers jointly, is selected at each iteration. That is, $(j, i) = \max_{u,r} [\boldsymbol{\Upsilon}_u]_{r,r}$, with $\boldsymbol{\Upsilon}_u = \prod_{k=1}^K \boldsymbol{\Sigma}_u[k]$, $u = 1, \dots, U$. In this way, the tuples of indices (j, i) corresponding to the largest wideband channel gains are sequentially added to the set \mathcal{S} , such that the index j refers to the selected user and

Algorithm 1 Rank-Constrained Coordinate Ascent (RCCA)

```

1: Initialize:  $\ell \leftarrow 0$ ,  $\{\mathbf{S}_u^{(0)}[k]\}_{u=1,k=1}^{U,K}$ ,  $K_c \leftarrow \lfloor K/2 \rfloor$ ,
2: repeat
3:   for all  $u \in \{1, \dots, U\}$  do
4:      $\mathbf{H}_u^{(\ell)}[k] \leftarrow$  Update with  $\mathbf{H}_u^{(\ell)}[k]$  from (15),  $\forall k$ 
5:      $\Delta_u^{(\ell)}[k] \leftarrow$  Compute using (11),  $\forall k$ 
6:      $\mathbf{U}_u^{(\ell)} \leftarrow$  LSV of  $[\mathbf{H}_u^{(\ell)}[1], \mathbf{H}_u^{(\ell)}[2], \dots, \mathbf{H}_u^{(\ell)}[K]]$ 
7:      $\Sigma_u^{(\ell)}[k] \leftarrow \text{diag}(\mathbf{U}_u^{(\ell)\text{H}} \mathbf{H}_u^{(\ell)}[k] \mathbf{H}_u^{(\ell)}[k] \mathbf{U}_u^{(\ell)})$ ,  $\forall k$ 
8:      $\Upsilon_u^{(\ell)} \leftarrow \prod_{k=1}^K \Sigma_u^{(\ell)}[k]$ 
9:   end for
10:   $r^{(\ell)} \leftarrow 0$ ,  $\mathbf{T}^{(\ell)} \leftarrow []$ ,  $\mathcal{S}^{(\ell)} \leftarrow \emptyset$ 
11:  while  $r^{(\ell)} < N_{\text{RF}}$  do
12:     $(j, i) \leftarrow \text{argmax}_{j,i} [\Upsilon_j^{(\ell)}]_{i,i}$ 
13:     $\mathbf{T}^{(\ell)} \leftarrow [\mathbf{T}^{(\ell)} \ \Delta_j^{(\ell)}[K_c][\mathbf{U}_j^{(\ell)}]_{:,i}]$ 
14:     $r^{(\ell)} \leftarrow \text{rank}(\mathbf{T}^{(\ell)})$ 
15:     $\mathcal{S}^{(\ell)} \leftarrow \mathcal{S}^{(\ell)} \cup \{(j, i)\}$ 
16:     $[\Upsilon_j^{(\ell)}]_{i,i} \leftarrow 0$ 
17:  end while
18:   $\mathbf{p}^{(\ell)} \leftarrow$  waterf. over  $[\Sigma_j^{(\ell)}[k]]_{i,i}$ ,  $\forall k$  with  $(j, i) \in \mathcal{S}^{(\ell)}$ 
19:   $\{\mathbf{S}_u^{(\ell+1)}[k]\}_{u=1}^U \leftarrow$  Compute using (21) with  $\mathbf{U}_u^{(\ell)}$ ,  $\mathbf{p}^{(\ell)}$ 
20:   $\ell \leftarrow \ell + 1$ 
21: until Convergence criterion is met
22:  $\{\Theta_u[k]\}_{u=1,k=1}^{U,K} \leftarrow$  Prec. matrices for  $\{\mathbf{S}_u^{(\ell)}[k]\}_{u=1,k=1}^{U,K}$ 
23:  $\mathbf{P}_{\text{RF}}$ ,  $\{\mathbf{P}_{\text{BB}}^u[k]\}_{u=1,k=1}^{U,K} \leftarrow$  [42, Alg. 1] with  $\{\Theta_u[k]\}_{u=1,k=1}^{U,K}$ 

```

the index i to the corresponding stream. Therefore, the i -th column of \mathbf{U}_j is considered as an adequate candidate to allocate a certain amount of power in the subsequent design of $\mathbf{P}_u[k]$. Furthermore, in order to account for the actual rank restriction in the downlink design, the rank constraint is evaluated over the columns of $\Delta_j[K_c][\mathbf{U}_j]$, that is, over the dual downlink covariance matrix given by the right term in (22). As mentioned, we will only consider the conversion matrix for the central subcarrier $\Delta_j[K_c]$.

The procedure stops when N_{RF} linearly independent vectors are selected, i.e., when $|\mathcal{S}| \geq N_{\text{RF}}$, and hence

$$\text{rank}([\Delta_{j_1}[K_c][\mathbf{U}_{j_1}]_{:,i_1}, \dots, \Delta_{j_{|\mathcal{S}|}}[K_c][\mathbf{U}_{j_{|\mathcal{S}|}}]_{:,i_{|\mathcal{S}|}}]) = N_{\text{RF}},$$

with the set $\mathcal{S} = \{(j_1, i_1), \dots, (j_{|\mathcal{S}|}, i_{|\mathcal{S}|})\}$. This assignment strategy allows us to obtain solutions for a relaxed version of the original formulation in (13) where the rank constraint is the right term in (22). With the proposed scheme, we seek to minimize the loss of the posterior factorization procedure to obtain the hybrid precoder components.

Once the set \mathcal{S} is determined, we calculate the power allocation vector $\mathbf{p} \geq \mathbf{0}$ by using waterfilling over the equivalent channel gains $[\Sigma_j[k]]_{i,i}$ in (23), just for the candidate tuples $(j, i) \in \mathcal{S}$. Using \mathbf{p} , we compute the frequency-selective power allocation matrices $\mathbf{P}_u[k] \forall u, k$, allowing us a flexible power distribution among the users and subcarriers. Indeed, the matrices $\mathbf{P}_u[k]$ would switch off some of the users and/or subcarriers if their equivalent channel gains are poor.

An important remark is that, when the KKT condition in (19) is satisfied, the upper bound of (26) is tight. In that case, the power allocation procedure is hence equivalent to performing the search on β and solve (25) for that value.

The proposed method, which we will term RCCA, is a variation of a block coordinate ascent algorithm where the structure of the covariance matrices are independently determined by considering the remaining ones as fixed, and computing the power allocation jointly for all of them. This coupling in the power restriction, together with the absence of convex and closed properties for the feasible set imposed by the rank restriction, makes impossible to ensure monotonically increasing iterations [43]. Nevertheless, it is possible to employ numerical techniques to ensure the convergence to a local optimum, like those employed in [44]. Still, numerical results show good convergence behavior for this insightful algorithm.

The proposed algorithm implicitly selects the number of data streams transmitted to each user at each subcarrier according to the power allocation procedure, considering the largest equivalent channel gains and satisfying the rank constraint with equality, according to the ensuing proposition

Proposition 2. *The optimum overall transmit covariance matrix \mathbf{Q} which maximizes the achievable sum-rate in (6) satisfies $\text{rank}(\mathbf{Q}) = N_{\text{RF}}$.*

Proof. See Appendix B. \square

To determine the overall computational complexity of Alg. 1, we analyze the main steps of the procedure. The computational cost of obtaining the uplink-downlink conversion matrices in (11) (steps 4 and 5), is bounded by $\mathcal{O}(M^2R)$ due to the computation of the matrix products, while the SVD in step 6 and the product in step 7 present complexity orders of $\mathcal{O}(KMR^2)$ and $\mathcal{O}(MR^2)$, respectively. Next, finding the number of linearly independent columns in $\mathbf{T}^{(\ell)}$ (step 14) is computationally cheap with order $\mathcal{O}(MN_{\text{RF}})$, whereas determining the uplink covariance matrices in step 19 presents a cost of $\mathcal{O}(R^3)$. Finally, obtaining the downlink precoding matrices of step 22 requires again computing matrix products, with a cost of about $\mathcal{O}(M^2R)$, and $\mathcal{O}(MN_{\text{RF}}^2)$ operations are necessary for the factorization procedure of step 23. If we consider only the dominant computational components, we obtain an overall computational complexity of $\mathcal{O}(\max\{KMR^2, M^2R\})$.

VI. SIMULATION RESULTS

In this section, the results of computer experiments carried out to illustrate the performance of the proposed RCCA algorithm are presented. We also make comparisons with four benchmark approaches, namely: 1) hybrid LISA [19], 2) the alternating minimization (AM)-based algorithm proposed in [20], 3) the hybrid precoding design of [13], and 4) a wideband extension of the iterative trace-based algorithm (ITA) proposed in [30] for narrowband scenarios. The first one was chosen because it provides superior performance to traditional approaches for wideband hybrid precoding. The second and third solutions represent low-complexity alternatives for wideband hybrid precoding. Finally, the latter is a rank-constrained approach that iteratively solves a series of convex problems (cf. [30]). In its wideband extension, ITA selects the best N_{RF} -dimensional subspace from the common space spanned by all the subcarriers, although the working principles significantly differ from the RCCA algorithm. In addition, we assume

TABLE I
 SIMULATION PARAMETER SETTINGS.

Parameter	Value
Number of users	$U = 6$
Central subcarrier frequency	$f_c = 5$ GHz
Bandwidth	$B = 200$ MHz
Number of BS antennas	$M = 32$
Distance between antennas	$d = \frac{\lambda}{2}$ m
Number of antennas at the users	$R = 2$
Delay taps	$D = 8$
Number of subcarriers	$K = 64$
Number of RF chains	$N_{\text{RF}} = 4$

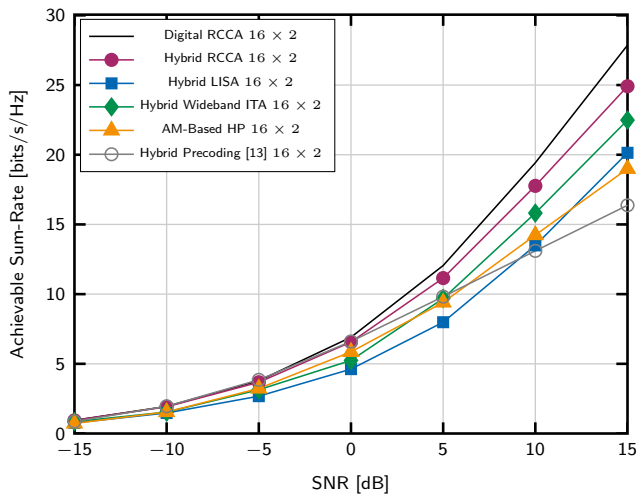


Fig. 1. Achievable sum-rates of the different algorithms for a downlink scenario with the following configuration: $M = 16$, $R = 2$, $U = 6$, $N_{\text{RF}} = 4$, $N_p = 16$ and $K = 16$.

successive interference cancellation at reception for all the approaches.

Unless explicitly stated, the main simulation parameters are those in Table I. The results were averaged over 200 channel realizations generated according to the geometrical channel model described in (4). For all the experiments, we initialize the covariance matrices $\{\mathbf{S}_u^{(0)}[k]\}_{u=1, k=1}^{U, K}$ in Alg. 1 to the scaled identity matrix, such that the power constraint holds with equality.

Figures 1 and 2 show the system performance of the different approaches for two downlink setups. In particular, the setup in Fig. 1 uses $M = 16$, $N_{\text{RF}} = 4$, and $K = 16$ to enable comparisons with the high-complexity ITA solution. We have included the achievable sum-rates obtained with both the rank-constrained digital precoders provided by the RCCA algorithm (Digital RCCA) and after their corresponding factorization to obtain the hybrid version (Hybrid RCCA). We consider a relatively large number of channel paths, $N_p = \{16, 32\}$, as it is more realistic in a practical scenarios, and makes the precoder design more difficult because of the larger differences between the channel responses corresponding to

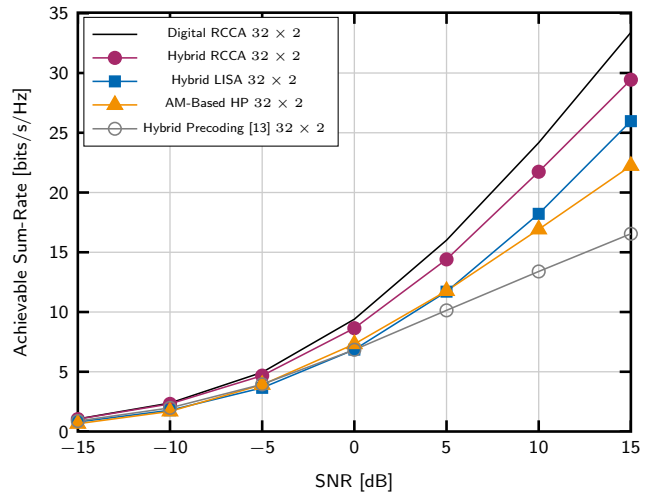


Fig. 2. Achievable sum-rates of the different algorithms for a downlink scenario with the following configuration: $M = 32$, $R = 2$, $U = 6$, $N_{\text{RF}} = 4$, $N_p = 32$ and $K = 64$.

distant subcarriers. Moreover, the multiuser interference also increases with this parameter. As observed, the proposed RCCA algorithm outperforms the other approaches, especially for medium and high SNR regimes. The superior performance of RCCA with respect to Hybrid LISA increases for the largest SNR values in both setups. This is because RCCA considers the interference from all the other users at each algorithm step, whereas LISA only selects one candidate user at each step and removes the interference caused by that user to the previously selected ones. LISA is suitable for low scattering scenarios where user subcarriers experience similar channel responses but quickly consumes the available spatial degrees of freedom. Hybrid wideband ITA achieves an intermediate performance at the expense of higher complexity. On the other hand, the AM-based and the hybrid precoding procedure of [13] are competitive at the low SNR regime but they clearly degrade as the SNR increases. For the AM-based alternative, this is due to the use of the maximum ratio transmission (MRT) strategy that ignores interference, and the hindrance of imposing a per-subcarrier power constraint. In the case of the method of [13], the authors again rely on an analog precoding design that neglects interference, thus limiting its usability in mid and high SNR scenarios. As observed, the performance degradation of this method for high SNR values is particularly remarkable as the number of transmit antennas and subcarriers grows.

Comparing Digital RCCA and Hybrid RCCA, we observe that the performance loss caused by the factorization is relatively small. This confirms that the assumption in (22) holds in practice, i.e., the use of a common structure to limit the rank of the solutions leads to covariance matrices with approximately the same rank in the downlink. Larger losses for higher SNR values are related to the limitations of the factorization algorithms to keep the structure of the digital counterpart [45].

A second computer experiment was carried out to evaluate the impact on performance of the number of channel paths N_p . Figure 3 shows the achievable sum-rates for RCCA, Hybrid

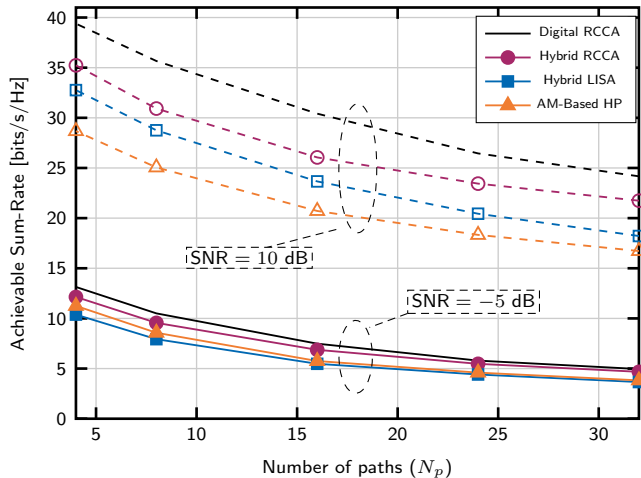


Fig. 3. Achievable sum-rates vs. number of reflection paths for the different approaches and two SNR values: -5 dB and 10 dB. The configuration setup is: $M = 32$, $R = 2$, $U = 6$, $N_{RF} = 4$ and $K = 64$.

LISA and AM-based for the setup in Table I and different number of paths $N_p \in \{4, 8, 16, 24, 32\}$. In addition, two different SNR values are considered, namely -5 dB and 10 dB. As observed, RCCA provides the highest achievable sum-rate even for those situations more favorable to Hybrid LISA, i.e., those with smaller number of channel paths N_p . In addition, the proposed algorithm efficiently exploits the similarity between the subcarriers even for large values of N_p . Note, however, that the performance of all approaches decreases as N_p increases due to the different spatial features among channels for distant subcarriers, and the stronger multiuser interference.

In the next experiment, we are interested in evaluating the behavior of the proposed algorithm depending on the number of RF chains available at the BS. Figure 4 shows the achievable sum-rates of the different approaches considered for the hybrid design of the downlink precoders for a challenging scenario with $U = 12$ users and $N_p = 32$ paths. We have decided to increase the number of users with respect to Table I to better illustrate the dependency on the number of RF chains. Like in the previous experiment, we focus on two representative SNR values: -5 dB and 10 dB. As observed, the obtained results show the superior performance of the proposed design strategy RCCA regardless of the number of RF chains. The advantages of using the RCCA-based approach are remarkable for the scenario with SNR = 10 dB, i.e., high SNR regime, and a large number of RF chains. In this case, RCCA provides a gain, in terms of achievable sum-rate, of about 8 bits/s/Hz with respect to Hybrid LISA. Again, this positive behavior of the RCCA algorithm is related to a more exhaustive management of the user interference, which is more apparent at the high SNR regime. On the other hand, AM-based strategy provides an acceptable performance in the scenario with the lowest SNR value but its performance significantly degrades as the SNR increases for any number of RF chains.

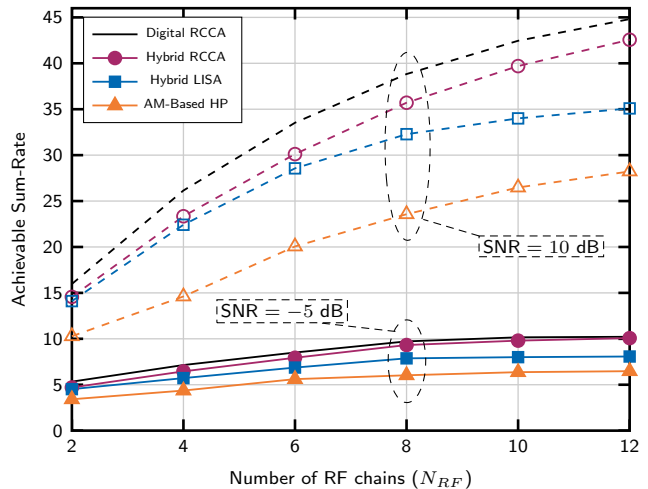


Fig. 4. Achievable sum-rates vs. number of RF chains at the BS for the different approaches and two SNR values: -5 dB and 10 dB. The configuration setup is: $M = 32$, $R = 2$, $U = 12$, $K = 64$ and $N_p = 32$.

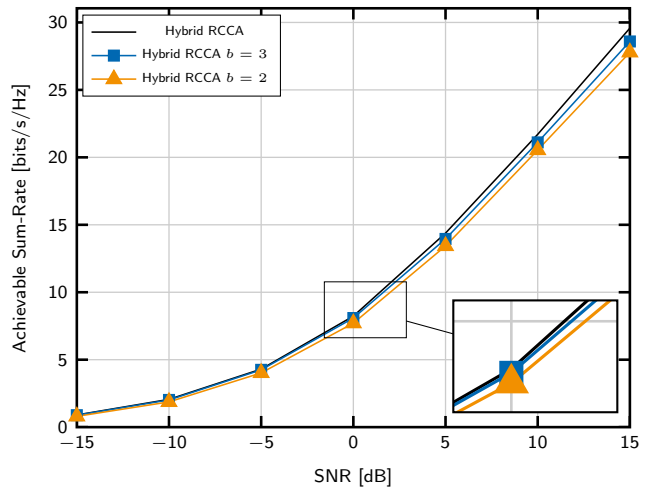


Fig. 5. Achievable sum-rates of RCCA for a downlink scenario with the following configuration: $M = 32$, $R = 2$, $U = 6$, $N_{RF} = 4$, $N_p = 32$ and $K = 64$ and different number of resolution bits $b = \infty, 3, 2$.

A. Performance under practical impairments

In this section, we evaluate the performance impact caused by two effects that are present in practical setups. More specifically, we first consider an analog hybrid precoding with quantized phases and, second, the more general scenario where the array response vectors are affected by beam squint.

In Fig. 5 we evaluate the impact of reducing the resolution of the phase shifters employed to build the analog network of the hybrid architecture. Under this assumption, the entries of the analog precoding matrix \mathbf{P}_{RF} can only take values from a quantized phases set. In particular, assuming b quantization bits, the available phases are 2^b . According to the obtained results, we can conclude that the proposed method is robust against phase quantization, as the performance loss is moderate even for $b = 2$ quantization bits (about 1.5 bits/s/Hz).

In the next experiment, we employ the configuration setup of Table I adapting the parameters to a mmWave scenario

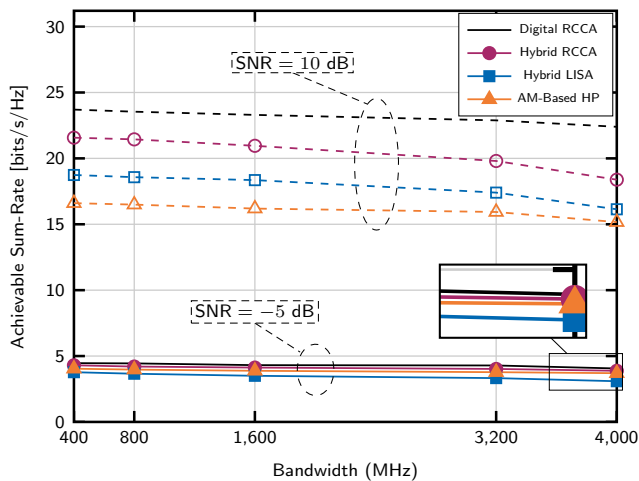


Fig. 6. Achievable sum-rates vs. bandwidth for the different approaches and two SNR values: -5 dB and 10 dB. The configuration setup is: $M = 32$, $R = 2$, $U = 12$, $N_{\text{RF}} = 4$, $K = 64$ and $N_p = 32$.

where the array response vectors of (5) become frequency dependent and the beam squint effect is more remarkable [28]. In particular, we consider a central frequency $f_c = 28$ GHz and different signal bandwidths $B = \{400, 800, 1600, 3200, 4000\}$ MHz. From the results observed in Fig. 6, we can conclude that the influence of beam squint is significant for a SNR value of 10 dB and moderate when the SNR is -5 dB. This is motivated because the spatial similarity of the channel subspaces for different frequencies is compromised by beam squint. Accordingly, the performance losses due to hybrid factorization increase with the bandwidth B , as the hybrid approach is less effective when handling multiuser interference.

B. Complexity and convergence analysis

In this subsection, we first compare the computational complexity of the proposed RCCA algorithm with respect to the competitors. Table II shows the complexity order of these five wideband hybrid precoding approaches: *I*) The complexity of RCCA is mainly determined by the step 6 in Algorithm 1, where the calculation of the SVD of the matrix which results from stacking the equivalent channels corresponding to all the subcarriers for the u -th user is required. The complexity of this step is in the order of $O(KMR^2)$ for the considered setups. *II*) The most computational expensive operation in Hybrid LISA is the calculation of the precoding candidates to select the best user at each algorithm step [19]. The complexity order of this operation is $O(\min\{KM^2, K^2M\})$ since all the frequency equivalent channels for each user are considered. *III*) The complexity order of the AM-based method was determined in [20], and it comes from the computation of the matrix product of the digital precoders for all the subcarriers and users times the analog precoding matrix. As observed, the complexity order of the AM-based method is actually lower than that of the other approaches since it is linear on the number of transmit antennas and subcarriers. *IV*) The hybrid precoding design of [13] also presents a computationally efficient solution, as the

TABLE II
COMPUTATIONAL COMPLEXITY OF WIDEBAND HYBRID PRECODERS.

Algorithm	Complexity order
RCCA	$O(\max\{KMR^2, M^2R\})$
Hybrid LISA	$O(\min\{KM^2, K^2M\})$
AM-Based HP	$O(KMUN_{\text{RF}})$
Wideband ITA	$O(K^3M^6)$
Hybrid Precoding in [13]	$O(M^2)$

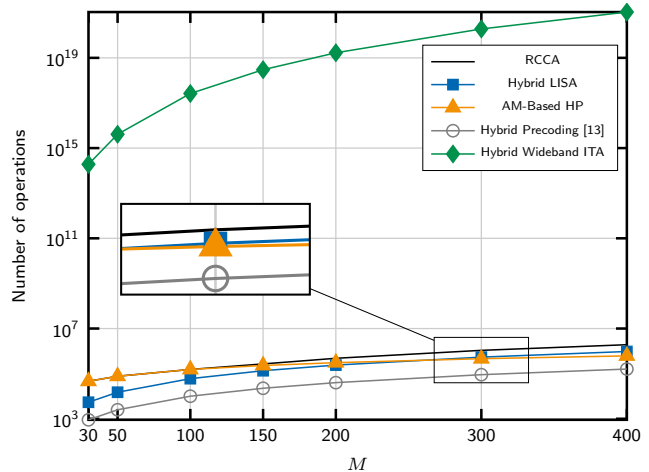


Fig. 7. Computational complexity vs. number of antennas at the BS M for the different approaches. The configuration setup is: $R = 2$, $U = 6$, $N_{\text{RF}} = 4$ and $K = 64$.

complexity order $O(M^2)$ comes from the computation of the average covariance matrices for the wideband channels. These products can be obtained in parallel for all the subcarriers. *V*) The wideband extension of ITA has a complexity order significantly higher than the other approaches. This is because the general-purpose solvers for semidefinite programming (SDP) problems are based on interior point methods whose complexity is in the order of $O(mn^2 + n^3)$ for a SDP problem with m linear matrix inequalities and n variables [46].

In order to illustrate the influence of the number of antennas M on the overall computational costs of the the hybrid precoder designs considered in this section, we provide in Fig. 7 an approximation to the number of operations required for each method in a logarithmic scale. The result shows that Hybrid LISA and the posed RCCA method increase their computational costs faster than other computationally cheap solutions. In addition, Hybrid Wideband ITA becomes hardly applicable even for moderate numbers of antennas.

Now, we empirically analyze the convergence of the RCCA algorithm. As commented, the particular structure of the user covariance matrices and the non-convex nature of the problem imposed by the rank restrictions make it impossible to theoretically guarantee the convergence of the RCCA algorithm. However, the obtained simulation results and the empirical convergence analysis carried out during the computer experiments suggests that the RCCA algorithm presents good convergence properties. This is in part motivated by the use of numerical techniques which ensure the convergence of the

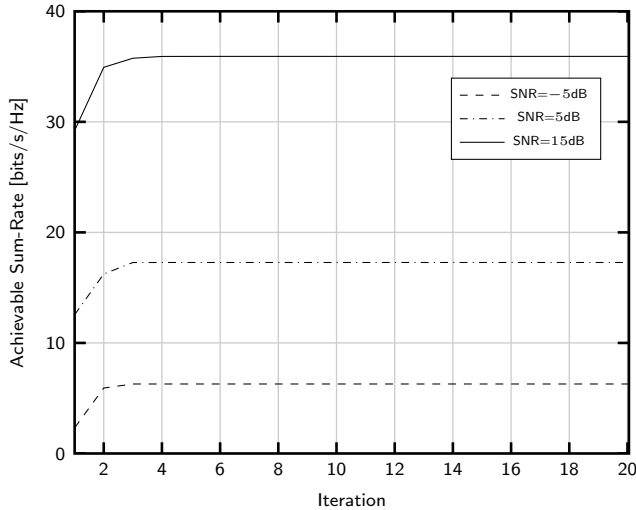


Fig. 8. Achievable sum-rates vs. number of iterations for the RCCA algorithm and SNR values: -5 dB, 10 dB, and 15 dB. The configuration setup is: $M = 32$, $R = 2$, $U = 6$, $N_{RF} = 6$ and $K = 64$.

algorithm to a local optimum, following a similar approach to [44]. An example of this is observed in Figure 8 which shows the achievable sum-rates obtained with the RCCA algorithm at the end of each iteration for three different SNR values, namely -5 dB, 5 dB and 15 dB. As observed, the algorithm rapidly converges after a few iterations in the three SNR cases. Usually, the number of iterations required to achieve convergence increases with the number of channel paths N_p , the number of users U , and the SNR; that is, for scenarios where interference management is difficult. As already mentioned, due to the complicated structure of the search space of the original problem formulation in (4), the optimum achievable sum-rate remains unknown. Thus, it is impossible to check if convergence to the global optimum has been reached.

VII. CONCLUSIONS

An innovative approach to hybrid precoding at the BS of a downlink massive MIMO wideband system, termed RCCA, has been developed. The approach aims at exploiting the spatial similarity among the different OFDM subcarriers by forcing digital solutions with a low rank. In addition, the inter-user interference is considered to approximate equivalent channel gains which in turn are used to update the transmit covariance matrices. The numerical experiments show that the proposed RCCA approach outperforms both direct extensions of previous rank-constrained designs and other benchmark solutions such as hybrid LISA or AM-based algorithms. In this sense, RCCA has been shown to be a robust solution for a wide range of practical SNR values and scenarios with different number of reflection channel paths.

ACKNOWLEDGMENTS

This work has been supported in part by grants ED431C 2020/15 and ED431G 2019/01 (to support the Centro de Investigación de Galicia “CITIC”) funded by Xunta de Galicia and ERDF Galicia 2014-2020, and by grants

PID2019-104958RB-C42 (ADELE) and PID2020-118139RB-I00 funded by MCIN/AEI/10.13039/501100011033. The authors thank the Defense University Center at the Spanish Naval Academy (CUD-ENM) for all the support provided for this research.

APPENDIX A

FREQUENCY-FLAT COVARIANCE MATRIX STRUCTURE

We focus on the case where the covariance matrix has to be designed for a single subcarrier, $K = 1$. Then, the KKT condition in (19) is equivalent to

$$\left(\mathbf{I} + \mathbf{H}_{\bar{u}} \mathbf{H}_{\bar{u}}^H \mathbf{S}_u\right)^{-1} \mathbf{H}_{\bar{u}} \mathbf{H}_{\bar{u}}^H - \lambda_u \mathbf{I} = \mathbf{0}. \quad (27)$$

As observed, if the structure of the covariance matrix \mathbf{S}_u forces the first term (27) to be positive definite, then the subspace spanned by the columns of \mathbf{S}_u achieves the condition in (27) with equality for certain spectrum of \mathbf{S}_u . To show this statement we introduce the decomposition $\mathbf{H}_{\bar{u}} \mathbf{H}_{\bar{u}}^H = \mathbf{U} \mathbf{\Gamma} \mathbf{U}^H$, with unitary matrix \mathbf{U} and $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_R)$ with $\gamma_i \geq 0$, $\forall i \in \{1, \dots, R\}$. We consider that the elements γ_1 to γ_j are strictly larger than 0 without loss of generality. Now, we rewrite the left-hand side of (27) as

$$\begin{aligned} & \left(\mathbf{I} + \mathbf{U} \mathbf{\Gamma} \mathbf{U}^H \mathbf{S}_u\right)^{-1} \mathbf{U} \mathbf{\Gamma} \mathbf{U}^H + \mathbf{U} \mathbf{L} \mathbf{U}^H - \mathbf{U} \mathbf{L} \mathbf{U}^H - \lambda_u \mathbf{I} \\ & = \left(\mathbf{U} (\mathbf{\Gamma} + \mathbf{L})^{-1} \mathbf{U}^H + \mathbf{U} (\mathbf{\Gamma} + \mathbf{L})^{-1} \mathbf{\Gamma} \mathbf{U}^H \mathbf{S}_u\right)^{-1} \\ & \quad - \mathbf{U} \mathbf{L} \mathbf{U}^H - \lambda_u \mathbf{I}, \end{aligned} \quad (28)$$

where we have also introduced the auxiliary diagonal matrix $\mathbf{L} = \text{diag}(0, \dots, 0, l_{j+1}, \dots, l_R)$. Next, we select the following structure for the covariance matrix $\mathbf{S}_u = \mathbf{U} \mathbf{\Psi} \mathbf{U}^H$ to get

$$\begin{aligned} & \left(\mathbf{U} \left[(\mathbf{\Gamma} + \mathbf{L})^{-1} + \tilde{\mathbf{I}} \mathbf{\Psi}\right] \mathbf{U}^H\right)^{-1} - \mathbf{U} (\mathbf{L} + \lambda_u \mathbf{I}) \mathbf{U}^H \\ & = \mathbf{U} \left(\left[(\mathbf{\Gamma} + \mathbf{L})^{-1} + \tilde{\mathbf{I}} \mathbf{\Psi}\right]^{-1} - (\mathbf{L} + \lambda_u \mathbf{I}) \right) \mathbf{U}^H, \end{aligned} \quad (29)$$

with $\tilde{\mathbf{I}} = \text{diag}(1, \dots, 1, 0, \dots, 0)$ having j ones in the diagonal, followed by $R - j$ zeros. In addition, note that $\mathbf{\Psi} = \text{diag}(\psi_1, \dots, \psi_R)$. The expression in (29) will be equal to the zero matrix if the two terms inside the brackets cancel each other out, that is, if

$$(\mathbf{\Gamma} + \mathbf{L})^{-1} + \tilde{\mathbf{I}} \mathbf{\Psi} = (\mathbf{L} + \lambda_u \mathbf{I})^{-1}.$$

We now distinguish between the first j elements and the remaining ones, leading to the following conditions

$$\begin{cases} \frac{1}{\gamma_i} + \psi_i = \frac{1}{\lambda_u} & \text{if } i \leq j \\ \frac{1}{\gamma_i} = \frac{1}{\lambda_u + l_i} & \text{otherwise.} \end{cases} \quad (30)$$

While the first condition can be achieved by setting $\psi_i = \frac{1}{\lambda_u} - \frac{1}{\gamma_i}$, the second condition approximately holds when we choose $\tilde{\gamma}_i \gg \lambda_u$, $\forall i$. We hence conclude that, selecting the structure of the user covariance matrix as $\mathbf{S}_u = \mathbf{U} \mathbf{\Psi} \mathbf{U}^H$, the first term in (27) results in a positive definite matrix, which is meaningful in a per-carrier basis.

APPENDIX B

 ANALYSIS OF THE OPTIMALITY OF $N_s = N_{\text{RF}}$

In this appendix, we evaluate the effect of including an additional stream when $N_s = N_{\text{RF}}$ in (9) holds to show that, even under the assumption of perfect decoding, this will lead to a reduction in the achievable sum-rate.

We consider that the matrices $\mathbf{S}_u[k]$ allocate streams to a certain set of users \mathcal{U} such that the rank of the overall transmit covariance is N_{RF} , while providing the largest equivalent channel gains, according to Sec. IV-B. The achievable sum-rate obtained under the power constraint $\sum_{k=1}^K \sum_{u=1}^U \text{tr}(\mathbf{S}_u[k]) \leq P_{\text{tx}}$ is denoted by R , and for simplicity we assume $\sigma_n^2 = 1$. In this context, if a new stream is allocated to user $j \notin \mathcal{U}$, the achievable sum-rate including the new stream reads as

$$R' = \sum_{u \in \mathcal{U}} R'_u + R'_j, \quad (31)$$

where R'_u is the achievable rate for user u when considering the appropriate set of transmit covariance matrices $\mathbf{S}'_u[k]$, and R'_j is the achievable rate for the j -th user which is given by

$$R'_j = \sum_{k=1}^K \log_2 \det \left(\mathbf{I} + \mathbf{H}_{\bar{j}}^H[k] \mathbf{S}'_j[k] \mathbf{H}_{\bar{j}}[k] \right), \quad (32)$$

with $\mathbf{H}_{\bar{j}}[k]$ being the effective channels. As we keep the rank of the overall transmit covariance matrix constant in the uplink, the j -th user covariance matrix is given by

$$\mathbf{S}'_j[k] = \sum_{u \in \mathcal{U}} \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H,$$

where $\mathbf{S}'_u[k] = \mathbf{V}_u \mathbf{P}_u[k] \mathbf{V}_u^H$ has the structure in (21) for all users and subcarriers, with semi-unitary $\mathbf{V}_u \in \mathbb{C}^{M \times N_{s,u}}$, and $\mathbf{P}_{j,u}[k] \geq \mathbf{0}$ are the diagonal matrices that determine the power allocation for the j -th user.

By employing the effective channel $\mathbf{H}_{\bar{j}}[k]$, the KKT condition derived in App. A can be extended to consider multiple carriers as [cf. (27)]

$$\left(\mathbf{I} + \mathbf{H}_{\bar{j}}[k] \mathbf{H}_{\bar{j}}^H[k] \mathbf{S}'_j[k] \right)^{-1} \mathbf{H}_{\bar{j}}[k] \mathbf{H}_{\bar{j}}^H[k] - \lambda_j[k] \mathbf{I} = \mathbf{0}.$$

Note that the equation above cannot be fulfilled in general due to the lack of flexibility in the design of $\mathbf{S}'_j[k]$. The reason behind this claim is that the subspaces spanned by the rows and columns of $\mathbf{S}'_j[k]$ are rotations of those for \mathbf{V}_u , and they are not specifically designed for user j . Accordingly, the KKT conditions would not be reachable simultaneously for all the users and, therefore, better power distributions would be possible. This situation remains the same as long as user j is active, i.e., the performance metric can be improved by assigning all the available power to the users in the set \mathcal{U} . Therefore, we conclude that $R' \leq R$.

Moreover, for the sake of completeness, we investigate the impact of the achievable sum-rate for a more practical approach, i.e., when successive cancellation is not considered. In this case, Zero-Forcing (ZF) precoding achieves channel capacity in the high SNR regime [47, 48]. Therefore, we consider a set of downlink transmit covariance matrices $\mathbf{Q}_u[k] = \mathbf{V}_u \mathbf{P}_u[k] \mathbf{V}_u^H$, with semi-unitary $\mathbf{V}_u \in \mathbb{C}^{M \times N_{s,u}}$

and $\mathbf{P}_u[k] \in \mathbb{C}^{N_{s,u} \times N_{s,u}}$, such that the equivalent channel gains for users $u \in \mathcal{U}$ are larger than the gains for the remaining ones $j \notin \mathcal{U}$ for given covariance matrices $\mathbf{Q}_u[k]$, i.e., $\mathbf{H}_u[k] \mathbf{V}_u \mathbf{V}_u^H \mathbf{H}_u^H[k] \geq \mathbf{H}_{\bar{j}}[k] \mathbf{V}_u \mathbf{V}_u^H \mathbf{H}_{\bar{j}}^H[k]$, and $\mathbf{H}_i[k] \mathbf{Q}_u[k] \mathbf{H}_i^H[k] = \mathbf{0}, \forall k$, with $u, i \in \mathcal{U}$ and $j \notin \mathcal{U}$. In other words, the set of covariance matrices $\mathbf{Q}_u[k]$, for $u \in \mathcal{U}$, are optimal under the assumption $N_s = N_{\text{RF}}$. The achievable sum-rate is then a simplified version of that in (9), i.e.,

$$R = \sum_{k=1}^K \sum_{u \in \mathcal{U}} \log_2 \det \left(\mathbf{I} + \mathbf{H}_u[k] \mathbf{Q}_u[k] \mathbf{H}_u^H[k] \right).$$

When a new stream is allocated to user $j \notin \mathcal{U}$, and with the aim of satisfying the rank constraint, the covariance matrix for user j reads as

$$\mathbf{Q}_j[k] = \sum_{u \in \mathcal{U}} \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H. \quad (33)$$

In this scenario, the zero-forcing condition for user j does not hold, leading to the following achievable rate for the user u and the subcarrier k

$$R'_u[k] = \log_2 \frac{\det \left(\mathbf{I} + \mathbf{H}_u[k] (\mathbf{Q}_u[k] + \mathbf{Q}_j[k]) \mathbf{H}_u^H[k] \right)}{\det \left(\mathbf{I} + \mathbf{H}_u[k] \mathbf{Q}_j[k] \mathbf{H}_u^H[k] \right)}.$$

Using the definition in (33) and the zero interference assumption, the former expression is rewritten as

$$R'_u[k] = \log_2 \frac{\det \left(\mathbf{I} + \mathbf{H}_u[k] \mathbf{V}_u (\mathbf{P}_u[k] + \mathbf{P}_{j,u}[k]) \mathbf{V}_u^H \mathbf{H}_u^H[k] \right)}{\det \left(\mathbf{I} + \mathbf{H}_u[k] \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H \mathbf{H}_u^H[k] \right)}.$$

Therefore, since the power available for the users $u \in \mathcal{U}$ is smaller due to the incorporation of the new stream, the achievable sum-rate for $u \in \mathcal{U}$ can be upper bounded as follows

$$\sum_{k=1}^K \sum_{u \in \mathcal{U}} R'_u[k] \leq R - \sum_{k=1}^K \sum_{u \in \mathcal{U}} \log_2 \det \left(\mathbf{I} + \mathbf{H}_u[k] \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H \mathbf{H}_u^H[k] \right). \quad (34)$$

For the new user j , it is also possible to obtain upper bounds for $R'_j[k]$ as follows

$$\begin{aligned} R'_j[k] &= \log_2 \det \left(\mathbf{I} + \mathbf{H}_{\bar{j}}[k] \sum_{u \in \mathcal{U}} \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H \mathbf{H}_{\bar{j}}^H[k] \right) \\ &\leq \log_2 \det \left(\mathbf{I} + \sum_{u \in \mathcal{U}} \mathbf{H}_u[k] \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H \mathbf{H}_u^H[k] \right), \end{aligned} \quad (35)$$

where we have used that the covariance matrices are designed for users $u \in \mathcal{U}$. Thus the equivalent channel gains are larger than for users not included in the set \mathcal{U} . Combining the bounds in (34) and (35), and introducing the auxiliary positive semi-definite matrices $\mathbf{A}_u[k] = \mathbf{H}_u[k] \mathbf{V}_u \mathbf{P}_{j,u}[k] \mathbf{V}_u^H \mathbf{H}_u^H[k]$, we obtain

$$\sum_{k=1}^K \sum_{u \in \mathcal{U}} R'_u + R'_j \leq R + \sum_{k=1}^K \log_2 \frac{\det \left(\mathbf{I} + \sum_{u \in \mathcal{U}} \mathbf{A}_u[k] \right)}{\prod_{u \in \mathcal{U}} \det \left(\mathbf{I} + \mathbf{A}_u[k] \right)}.$$

Now, the achievable sum-rate not increasing after allocating a new stream is equivalent to proving that

$$\det \left(\mathbf{I} + \sum_{u=1}^{|\mathcal{U}|} \mathbf{A}_u[k] \right) \leq \prod_{u=1}^{|\mathcal{U}|} \det \left(\mathbf{I} + \mathbf{A}_u[k] \right). \quad (36)$$

Rewriting the left-hand side of this inequality, we get

$$\prod_{u=1}^{|\mathcal{U}|} \det \left(\mathbf{I} + \left(\mathbf{I} + \sum_{i=1}^{u-1} \mathbf{A}_i[k] \right)^{-1} \mathbf{A}_u[k] \right) \leq \prod_{u=1}^{|\mathcal{U}|} \det \left(\mathbf{I} + \mathbf{A}_u[k] \right).$$

This last inequality is proven by showing that, for positive semidefinite matrices \mathbf{A} and \mathbf{B} , the following inequality holds

$$\det \left(\mathbf{I} + \left(\mathbf{I} + \mathbf{B} \right)^{-1} \mathbf{A} \right) \leq \det \left(\mathbf{I} + \mathbf{A} \right). \quad (37)$$

Denoting as $\lambda_n(\mathbf{X})$ the n -th largest eigenvalue of \mathbf{X} , we have that [49, p. 423]

$$\lambda_n \left(\left(\mathbf{I} + \mathbf{B} \right)^{-1} \mathbf{A} \right) \leq \lambda_1 \left(\left(\mathbf{I} + \mathbf{B} \right)^{-1} \right) \lambda_n(\mathbf{A}). \quad (38)$$

We therefore conclude that the eigenvalues of $\left(\mathbf{I} + \mathbf{B} \right)^{-1} \mathbf{A}$ are smaller than or equal to those of \mathbf{A} . Correspondingly, the inequalities in (37) and (36) hold, thus completing the proof.

REFERENCES

- [1] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [2] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.
- [3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [4] S. A. Busari, S. Mumtaz, S. Al-Rubaye, and J. Rodriguez, "5G millimeter-wave mobile broadband: Performance and challenges," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 137–143, 2018.
- [5] A. Pfadler, C. Ballesteros, J. Romeu, and L. Jofre, "Hybrid massive MIMO for urban V2I: Sub-6 ghz vs mmWave performance assessment," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4652–4662, 2020.
- [6] R. Zhang, Z. Zhong, J. Zhao, B. Li, and K. Wang, "Channel measurement and packet-level modeling for V2I spatial multiplexing uplinks using massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7831–7843, 2016.
- [7] A. Garcia-Rodriguez, G. Geraci, D. Lopez-Perez, L. G. Giordano, M. Ding, and E. Björnson, "The essential guide to realizing 5G-connected UAVs with massive MIMO," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 84–90, 2019.
- [8] P. Chandhar, D. Danev, and E. G. Larsson, "Massive MIMO for communications with drone swarms," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1604–1629, 2018.
- [9] B. M. Lee, "Energy-efficient operation of massive MIMO in industrial internet-of-things networks," *IEEE Internet of Things Journal*, vol. 8, no. 9, pp. 7252–7269, 2021.
- [10] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, 2014.
- [11] F. Sofrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.
- [12] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.
- [13] F. Sofrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1432–1443, 2017.
- [14] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801–1818, 2016.
- [15] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, 2016.
- [16] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2907–2920, 2017.
- [17] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1847–1860, 2021.
- [18] L. Kong, S. Han, and C. Yang, "Wideband hybrid precoder for massive MIMO systems," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 305–309.
- [19] J. P. González-Coma, W. Utschick, and L. Castedo, "Hybrid LISA for wideband multiuser millimeter-wave communication systems under beam squint," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1277–1288, 2019.
- [20] H. Yuan, J. An, N. Yang, K. Yang, and T. Q. Duong, "Low complexity hybrid precoding for multiuser millimeter wave systems over frequency selective channels," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 983–987, 2019.
- [21] S. Gao, X. Cheng, and L. Yang, "Mutual information maximizing wideband multi-user (wMu) mmWave massive MIMO," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3067–3078, 2021.
- [22] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, 2015.
- [23] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave mimo systems," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8315–8328, 2019.
- [24] A. Liu, V. K. N. Lau, and M.-J. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive mimo," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 432–444, 2018.
- [25] A. Liu, V. Lau, and Y. Teng, "Power minimization for massive mimo systems with two-timescale hybrid precoding," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.
- [26] R. Mai, T. Le-Ngoc, and D. H. N. Nguyen, "Two-timescale hybrid rf-baseband precoding with mmse-vp for multi-user massive mimo broadcast channels," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4462–4476, 2018.
- [27] K. Venugopal, N. González-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 330–333, 2017.
- [28] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5893–5908, 2019.
- [29] T. E. Bogale, L. B. Le, A. Haghghat, and L. Vandendorpe, "On the Number of RF Chains and Phase Shifters, and Scheduling Design With Hybrid Analog Digital Beamforming," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3311–3326, May 2016.
- [30] J. P. González-Coma, O. Fresnedo, and L. Castedo, "Rank Constrained Precoding for the Downlink of mmWave Massive MIMO Hybrid Systems," *IEEE Access*, vol. 9, pp. 28 459–28 470, 2021.
- [31] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [32] K. Venugopal, N. González-Prelcic, and R. W. Heath, "Optimal frequency-flat precoding for frequency-selective millimeter wave channels," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5098–5112, 2019.
- [33] A. Rodríguez, F. González, J. Sueiro, R. Valcarce, C. Nartallo, and F. Cruz, *COMUNICACIONES DIGITALES*. Pearson Educación/Prentice Hall, 2007. [Online]. Available: <https://books.google.es/books?id=0Rn3PQAACAAJ>
- [34] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1,

- pp. 552–563, 2020.
- [35] Y.-Y. Lee, C.-H. Wang, and Y.-H. Huang, “A hybrid rf/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave mimo systems,” *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 305–317, 2015.
- [36] 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.
- [37] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, “Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.
- [38] H. Yu and V. K. N. Lau, “Rank-Constrained Schur-Convex Optimization With Multiple Trace/Log-Det Constraints,” *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 304–314, January 2011.
- [39] S. Ma, D. Goldfarb, and L. Chen, “Fixed point and Bregman iterative methods for matrix rank minimization,” *Mathematical Programming*, vol. 128, no. 1, pp. 321–353, June 2011. [Online]. Available: <https://doi.org/10.1007/s10107-009-0306-5>
- [40] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.
- [41] Z. Gao, L. Dai, Z. Wang, and S. Chen, “Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO,” *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, December 2015.
- [42] J. P. González-Coma, J. Rodríguez-Fernandez, N. González-Prelcic, L. Castedo, and R. W. Heath, “Channel estimation and hybrid precoding for frequency selective multiuser mmwave mimo systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 353–367, 2018.
- [43] D. Bertsekas, *Nonlinear programming*. Belmont, Massachusetts: Athena Scientific, 1999.
- [44] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, “Sum power iterative water-filling for multi-antenna Gaussian broadcast channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.
- [45] C. Rusu, R. Mendez-Rial, N. González-Prelcic, and R. W. Heath, “Low complexity hybrid precoding strategies for millimeter wave communication systems,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8380–8393, 2016.
- [46] K.-C. Toh, R. Tütüncü, and M. Todd, “On the implementation and usage of SDPT3 – a Matlab software package for semidefinite-quadratic-linear programming, version 4.0,” *Handbook on Semidefinite, Conic and Polynomial Optimization*, vol. 166, 2012.
- [47] Taesang Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.
- [48] C. Guthy, W. Utschick, and G. Dietl, “Low-Complexity Linear Zero-Forcing for the MIMO Broadcast Channel,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 6, pp. 1106–1117, December 2009.
- [49] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.