Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems

José P. González-Coma*, Javier Rodríguez-Fernández†, Nuria González-Prelcic†,
Luis Castedo* and Robert W. Heath Jr.‡
*Universidade da Coruña, CITIC, Spain †Universidade de Vigo, Spain
‡The University of Texas at Austin, USA
Email:{jose.gcoma,luis}@udc.es,
{jrodriguez,nuria}@gts.uvigo.es, rheath@utexas.edu

***Abstract*—Configuring the hybrid precoders and combiners in a millimeter wave (mmWave) multiuser (MU) multiple-input multiple-output (MIMO) system is challenging in frequency selective channels. In this paper, we develop a system that uses compressive estimation on the uplink to configure precoders and combiners for the downlink (DL). In the first step, the base station (BS) simultaneously estimates the channels from all the mobile stations (MSs) on each subcarrier. To reduce the number of measurements required, compressed sensing techniques are developed that exploit common support on the different subcarriers. In the second step, exploiting reciprocity and the channel estimates, the base station designs hybrid precoders and combiners. Two algorithms are developed for this purpose, with different performance and complexity tradeoffs: 1) a factorization of the purely digital solution, and 2) an iterative hybrid design. Extensive numerical experiments evaluate the proposed solutions comparing to state-of-the-art strategies, and illustrating design tradeoffs in overhead, complexity, and performance.**

I. Introduction

Multiuser MIMO communication is challenging in frequency selective millimeter wave communication systems. The main difficulties are a byproduct of the hybrid analog-digital array architecture, which splits MIMO beamforming algorithms between analog and digital domains, to reduce overall power consumption [1]. In such a system, the analog combining network prevents direct access to each antenna output, the SNR is low due to the large bandwidth, the channel dimensionality is high, and the analog combining is nominally frequency flat while the channel is frequency selective. This tends to require a lengthy training phase to estimate the channel [2], a process which is aggravated in the multiuser setting, where channels are required for each active user. Precoding and combining is also non-trivial because the analog portion is frequency flat with constrained values, and the number of radio frequency (RF) chains is limited. This requires designing multiple precoders and combiners subject to non-convex constraints, which makes it difficult to design optimal algorithms with low complexity. In this paper, we develop strategies for frequency selective channel estimation and precoder/combiner design in MU-MIMO communication systems.

This work has been partially funded by Xunta de Galicia (ED431C 2016-045, ED341D R2016/012, ED431G/01, ED431G/04), AEI of Spain (TEC2015-69648-REDC, TEC2016-75067-C4-1-R, TEC2016-75103-C2-2-R), ERDF funds (AEI/FEDER, EU) and the National Science Foundation under Grant No. 1702800.

A. Prior Work

In this section, we review prior work on channel estimation and precoder design in millimeter wave MIMO systems, emphasizing contributions that assume a frequency selective channel model.

Most of prior work on channel estimation at mmWave focuses on narrowband channels [2], [3], [4], while the mmWave channel is indeed frequency selective. The limited work assuming a mmWave frequency-selective channel [5], [6], only considers the single user (SU) scenario. In multiuser settings, recent work focuses on narrowband channel estimation [7], [8], [9]. To the best of our knowledge, only [10] and [11] address the problem of frequency-selective multiuser channel estimation at mmWave. In [10], the channels for the different users are estimated on the uplink, but the channels have a large Ricean factor thus only the strongest path for each user is estimated. In [11], the channels are estimated by the MSs on the downlink and then the resulting channel state information (CSI) is fed back to the BS. The main limitations of this algorithm are that it does not exploit reciprocity and further it supposes that the channel directions lie on a specific grid.

The idea of a hybrid analog-digital solution for the precoders and combiners in a MIMO system was first proposed in [12], and developed in [13] for sparse mmWave channels. A common approach to design the hybrid precoders/combiners is to factorize an all-digital solution into the analog and baseband domains. Several methods have been proposed to perform this factorization for the SU and narrowband case, most of them depending on the relationship between the number of streams and the number of RF chains. Perfect factorization of the precoders in the downlink is obtained in [12], but implies a large number of RF chains. A sparse reconstruction problem with hardware constraints is formulated in [13] to find the precoders, but results in high complexity and assumes perfect

CSI at the receiver side. Good approximations of the all-digital solution can be obtained with the greedy algorithm in [14], and fewer RF chains. Other work such as [15] proposes low complexity approaches to compute the precoders, but are also limited to the SU and narrowband case. For the MU large antenna array setup, asymptotically optimal hybrid precoders were analyzed in [16]. A joint solution for precoders and combiners in a MU scenario based on imperfect CSI was proposed in [17] for frequency flat channels. Therein, the RF precoders and combiners are found using beam training algorithms, while inter-user interference is canceled by baseband precoders. A similar approach was presented in [18]. The RF precoder is chosen from a codebook whereas the baseband counterpart is designed employing second-order channel statistics.

The narrowband solutions for the design of the hybrid precoders and combiners cannot be, in general, extended to the wideband scenario. The main reason behind is that the RF precoders and combiners are frequency flat, and have to be jointly designed for all the subcarriers, while the digital precoders/combiners are frequency selective. While a few solutions have been proposed for the wideband scenario in the SU case [16], [19], all of them focus only on precoding without estimation. Considering imperfect CSI and a single user, a precoder design was proposed in [20]. Different search methods to find the RF precoder from a codebook were proposed, while the baseband counterpart maximizes the mutual information. The precoder design in [20] though is only useful for the single user setting.

Different precoding algorithms were proposed in [19] for the frequency selective single user case, where a search on the Riemannian manifold was performed to find the RF precoders. The complexity of the alternating minimization (AM) algorithm proposed there is prohibitive in our scenario, since the space search scales with the number of streams and subcarriers, which is large in the MU frequency selective scenario. In addition, authors in [19] present a low complexity alternative by restricting the columns of the baseband precoders to be mutually orthogonal. The motivation behind this constraint does not apply for some metrics as, e.g. mean square error (MSE) as considered in our paper.

Two solutions have been proposed for the frequency selective MU case [21], [22]. The first work assumes perfect CSI, only addresses the problem of precoder design, and evaluates the solution only for the high SNR regime. The algorithm for the MU frequency selective scenario developed in [22] exhibits a good tradeoff performance-complexity, but it only works with one RF chain at both the transmitter and the receiver.

To the best of our knowledge, a solution for the the joint precoder and combiner design for the mmWave frequency selective MU scenario under imperfect CSI is missing except for [23]. In that work, the autocorrelation of the received signal and the cross-correlation between the transmitted and received signal are estimated to compute precoders and combiners based on MSE metric. The iterative design in [23] employs several transmissions in the downlink and the uplink to set precoders and combiners. The digital solutions are next factorized via [15, HD-LSR Algorithm] at each step. This results in accumulated performance losses. Furthermore, the large training overhead might be unrealistic when the time coherence of the channel is short, for example with high mobility MSs.

B. Contributions

In this paper, we propose solutions to the problems of channel estimation and hybrid design of the precoders and combiners for a multiuser multistream MIMO system operating in a frequency selective mmWave channel. We assume orthogonal frequency-division multiplexing (OFDM) modulation, which is commonly employed to simplify equalization in frequency selective channels. Next, we summarize the specific contributions of our work:

- We propose an approach for joint estimation of the MU frequency selective channels, leveraging the jointly-sparse structure of the channels for different MSs and exploiting reciprocity. Compared to MIMO DL channel estimation, the consequences of this choice are twofold: i) it unnecessarily increases feedback overhead for the BS to design hybrid precoders and combiners, and ii) since the number of antennas at a MS is expected to be significantly smaller than at the BS, a more lengthy training stage is required to obtain channel estimates. Our design overcomes these limitations. In addition, previous work on MU mmWave frequency selective channel estimation only considers on-grid angular parameters [11].
- We show theoretically and numerically that with a hybrid architecture at the receiver and transmitter sides, MU uplink (UL) channel estimation at the BS, exploiting reciprocity of time-division duplexing (TDD) systems, outperforms DL channel estimation at the MSs for a fixed overhead, even neglecting feedback. This is well known in massive MIMO with fully digital and hybrid architectures [24], but it is not obvious with compressive estimation. Further, we show that when the AoAs and AoDs fall within a quantized spatial grid, the estimation error lies very close to the Cramér-Rao lower bound (CRLB).
- We design UL precoders at each MS that maximize their individual mutual information. The design we propose for the UL combiner at the BS is based on several criteria, exploiting knowledge on the UL precoder used by each MS, such that inter-user interference is mitigated. This scheme only depends on the CSI of each MS, and presents advantages to perform hybrid factorizations.
- We propose two approaches to compute baseband and RF filters: 1) factorization of a digital design into baseband and RF counterparts, as done in [12], [13], [19], [21], [25], [26]; 2) similar to [17], [27], we propose an iterative solution based on AM to directly design the hybrid precoders and combiners.
- We numerically show the connections between the performance of the designed precoders and combiners and the quality of the channel estimates obtained by the proposed algorithm. Extensive experiments show the good performance of our designs, compared to state of the art methods.

The paper is organized as follows: Section II introduces the system model. Section III formulates the problem for the design of precoders and combiners. In section III-B, different approaches to design the hybrid precoders and combiners are introduced. Section IV describes an algorithm for the estimation of multi-user frequency-selective channels. Section V is devoted to the numerical experiments. Finally, Section VI collects the main conclusions derived from this work.

II. System Model

In this section, we introduce the models and assumptions for the different blocks of the communication system considered in this paper.

a) System block diagram: We consider first the DL of a MIMO communication system operating at mmWave, where a BS equipped with N_{BS} antennas transmits information to U MSs, each of them using $N_{\text{MS},u}$ antennas. We assume that the number of RF chains at the BS and the MSs are $L_{\text{BS}} < N_{\text{BS}}$ and $L_{\text{MS},u} < N_{\text{MS},u}$, which is realistic for practical scenarios. We consider a network of phase shifters connecting each antenna with all the RF chains at both the transmitter an the receiver. Moreover, BS allocates several independent data streams $N_{\text{s},u} \leq L_{\text{MS},u}$ to each MS. The total number of data streams is assumed to be smaller than the number of RF chains at the BS, i.e., $N_{\text{s}} = \sum_{u=1}^{U} N_{\text{s},u} \leq L_{\text{BS}}$.

To deal with the frequency selective channel, OFDM modulation is employed with a large enough cyclic prefix to avoid intercarrier interference. We denote by $\boldsymbol{s}_u^{\text{DL}}[k]$ the Gaussian signal data transmitted to MS u and subcarrier k in the DL. We assume the following statistical properties: $\mathbb{E}[\boldsymbol{s}_u^{\text{DL}}[k]] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{s}_u^{\text{DL}}[k]\boldsymbol{s}_u^{\text{DL,H}}[k]] = \mathbf{I}_{N_{\text{s},u}}$ and $\mathbb{E}[\boldsymbol{s}_u^{\text{DL}}[k]\boldsymbol{s}_j^{\text{DL,H}}[k]] = \mathbf{0}$ for $j \neq u$. The symbol vectors are linearly processed using the hybrid precoder $\boldsymbol{P}_u[k] = \boldsymbol{P}_{\text{RF}}\boldsymbol{P}_{\text{BB}}^u[k] \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{s},u}}$ resulting from the product of the baseband and the RF precoders, $\boldsymbol{P}_{\text{BB}}^u[k] \in \mathbb{C}^{L_{\text{BS}} \times N_{\text{s},u}}$ and $\boldsymbol{P}_{\text{RF}} \in \mathbb{C}^{N_{\text{BS}} \times L_{\text{BS}}}$. At the MS end, the received signal is linearly filtered with the hybrid combiner $\boldsymbol{W}_u[k] = \boldsymbol{W}_{\text{RF}}^u\boldsymbol{W}_{\text{BB}}^u[k]$, with $\boldsymbol{W}_{\text{RF}}^u \in \mathbb{C}^{N_{\text{MS},u} \times L_{\text{MS},u}}$ and $\boldsymbol{W}_{\text{BB}}^u[k] \in \mathbb{C}^{L_{\text{MS},u} \times N_{\text{s},u}}$. Since the RF precoders and combiners are implemented using analog phase shifters, their entries are restricted to have constant modulus $|[\boldsymbol{P}_{\text{RF}}]_{i,j}|^2 = 1$, $|[\boldsymbol{W}_{\text{RF}}^u]_{m,n}|^2 = 1$.

Channel estimation is performed at the BS during a training phase prior to data transmission. In particular, the BS simultaneously estimates the channels for all the MS. The assumption of channel reciprocity between the DL and the UL of TDD systems is not impractical when hardware calibration is considered (see strategies proposed in [28], [29] and references therein). We assume the BS can feedforward its channel estimate to the i-th MS. Accordingly, the CSI estimate at each MS matches that in the BS. No further information is shared to design precoders and combiners.

b) Channel model: Let us introduce the D-delay channel model for the link between the BS and the MS u. The d-th delay tap, $\boldsymbol{H}_{u,d} \in \mathbb{C}^{N_{\text{MS},u} \times N_{\text{BS}}}$, follows the expression [16]

$$\boldsymbol{H}_{u,d} = \gamma \sum_{p=1}^{N_{\text{p},u}} \alpha_{u,p} p_{\text{rc},u}(dT_s - \tau_{u,p}) \boldsymbol{a}_{\text{MS},u}(\theta_{u,p}) \boldsymbol{a}_{\text{BS}}^{\text{H}}(\phi_{u,p}), \tag{1}$$

where $N_{\text{p},u}$ is the number of channel paths, $\gamma = \sqrt{N_{\text{BS}} N_{\text{MS},u} / N_{\text{p},u}}$ with the total number of receive antennas $N_{\text{MS}} = \sum_{u=1}^{U} N_{\text{MS},u}$, $\boldsymbol{a}_{\text{MS},u}(\theta_{u,p}) \in \mathbb{C}^{N_{\text{MS},u}}$ and $\boldsymbol{a}_{\text{BS}}(\phi_{u,p}) \in \mathbb{C}^{N_{\text{BS}}}$ are the array response vectors of the receiver and the transmitter, $p_{\text{rc},u}(t)$ denotes the combined effects of pulse shaping and analog filtering and T_s, $\tau_{u,p}$ and $\alpha_{u,p}$ are the sampling interval, the delay, and the gain. In the frequency domain, the channel response is

$$\boldsymbol{H}_u[k] = \sum_{d=1}^{D} \boldsymbol{H}_{u,d} \mathrm{e}^{-\mathrm{j}\frac{2\pi k d}{K}}. \tag{2}$$

An equivalent representation of the uplink channel is given by

$$\boldsymbol{H}_u^{\text{H}}[k] = \boldsymbol{A}_{\text{BS}_u} \boldsymbol{\Delta}_u[k] \boldsymbol{A}_{\text{MS}_u}^{\text{H}}, \tag{3}$$

with the matrices $\boldsymbol{A}_{\text{BS}_u} \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{p},u}}$ and $\boldsymbol{A}_{\text{MS}_u} \in \mathbb{C}^{N_{\text{MS},u} \times N_{\text{p},u}}$ containing the steering vectors for the BS and the MS evaluated at the angles-of-arrival (AoA) and angles-of-departure (AoD), and the diagonal matrix $\boldsymbol{\Delta}_u[k] \in \mathbb{C}^{N_{\text{p},u} \times N_{\text{p},u}}$ containing the gains for each of the $N_{\text{p},u}$ paths.

Using the extended virtual channel model [30], the channel matrix in (3) can be approximated as

$$\boldsymbol{H}_u^{\text{H}}[k] \approx \tilde{\boldsymbol{A}}_{\text{BS}} \boldsymbol{\Delta}_u^{\text{v}}[k] \tilde{\boldsymbol{A}}_{\text{MS},u}^{\text{H}}, \tag{4}$$

where $\boldsymbol{\Delta}_u^{\text{v}}[k] \in \mathbb{C}^{G_{\text{BS}} \times G_{\text{MS}}}$ is a sparse matrix containing the channel gains of the quantized spatial frequencies in its non-zero elements. The dictionary matrices $\tilde{\boldsymbol{A}}_{\text{BS}} \in \mathbb{C}^{N_{\text{BS}} \times G_{\text{MS}}}$, $\tilde{\boldsymbol{A}}_{\text{MS},u} \in \mathbb{C}^{N_{\text{MS},u} \times G_{\text{MS}}}$ contain the array response vectors for the BS and u-th MS evaluated on spatial grids of sizes G_{BS} and G_{MS}.

c) Downlink received signal model: According to our assumption of the cyclic prefix length, the channel in the frequency domain is decomposed into K equivalent narrowband channels. For subcarrier k and MS u, the received signal in the downlink is given by

$$\boldsymbol{x}_u[k] = \boldsymbol{H}_u[k] \sum\nolimits_{i=1}^{U} \boldsymbol{P}_{\text{RF}} \boldsymbol{P}_{\text{BB}}^i[k] \boldsymbol{s}_i^{\text{DL}}[k] + \boldsymbol{n}_u[k]. \tag{5}$$

where $\boldsymbol{n}_u[k] \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_{\text{MS},u}})$ is the additive Gaussian noise (AWGN). It is important to highlight that the RF precoders are frequency flat, i.e., they have to be jointly optimized for all the subcarriers. Analogously, the RF combiners $\boldsymbol{W}_{\text{RF}}^u$ are also frequency flat. The output signal after combining at the MS u and subcarrier k is

$$\boldsymbol{y}_u^{\text{DL}}[k] = \boldsymbol{W}_{\text{BB}}^{u,\text{H}}[k] \boldsymbol{W}_{\text{RF}}^{u,\text{H}} \boldsymbol{x}_u[k]. \tag{6}$$

d) Uplink received signal model: $\boldsymbol{H}_u^{\text{H}}[k]$ is the channel response in the frequency domain for the link between the MS u and the BS. The baseband and RF precoders and combiners are $\boldsymbol{T}_{\text{BB}}^u[k] \in \mathbb{C}^{L_{\text{MS},u} \times N_{\text{s},u}}$, $\boldsymbol{T}_{\text{RF}}^u \in \mathbb{C}^{N_{\text{MS},u} \times L_{\text{MS},u}}$ and $\boldsymbol{F}_{\text{RF}} \in \mathbb{C}^{N_{\text{BS}} \times L_{\text{BS}}}$, $\boldsymbol{F}_{\text{BB}}^u[k] \in \mathbb{C}^{L_{\text{BS}} \times N_{\text{s},u}}$, and the noise is $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_n^2 \mathbf{I}_N)$. Thus, the received signal in the uplink at subcarrier k is given by

$$\boldsymbol{x}[k] = \sum_{i=1}^{U} \boldsymbol{H}_i^{\text{H}}[k] \boldsymbol{T}_{\text{RF}}^i \boldsymbol{T}_{\text{BB}}^i[k] \boldsymbol{s}_i^{\text{UL}}[k] + \boldsymbol{n}[k]. \tag{7}$$

The received signal after hybrid combining is $\boldsymbol{y}_u^{\text{UL}}[k] = \boldsymbol{F}_{\text{BB}}^{u,\text{H}}[k] \boldsymbol{F}_{\text{RF}}^{\text{H}} \boldsymbol{x}[k]$.

Taking into account the signal model in (7), the BS simultaneously estimates the U frequency-selective channels $\boldsymbol{H}_i^{\mathrm{H}}[k]$, $i = 1,\ldots,U,\ k = 0,\ldots,K-1$ following the procedure described in section IV, such that channel estimates $\hat{\boldsymbol{H}}_i^{\mathrm{H}}[k]$ are computed.

e) Performance metrics: Assuming Gaussian signaling and linear precoding/combining, we establish the achievable sum-rate as the system performance metric, yielding [13], [31]

$$\sum_{u=1}^{U} R_u = \frac{1}{K}\sum_{u=1}^{U}\sum_{k=1}^{K} \log_2 \det\left(\mathbf{I}_{N_{\mathrm{s},u}} + \boldsymbol{X}_u^{-1}[k] \times \boldsymbol{W}_u^{\mathrm{H}}[k]\boldsymbol{H}_u[k]\boldsymbol{P}_u[k]\boldsymbol{P}_u^{\mathrm{H}}[k]\boldsymbol{H}_u^{\mathrm{H}}[k]\boldsymbol{W}_u[k]\right), \quad (8)$$

with $\boldsymbol{X}_u[k] = \sum_{i\neq u} \boldsymbol{W}_u^{\mathrm{H}}[k]\boldsymbol{H}_u[k]\boldsymbol{P}_i[k]\boldsymbol{P}_i^{\mathrm{H}}[k]\boldsymbol{H}_u^{\mathrm{H}}[k]\boldsymbol{W}_u[k] + \sigma_n^2\boldsymbol{W}_u^{\mathrm{H}}[k]\boldsymbol{W}_u[k]$. The interference is treated as noise.

To evaluate the performance of the proposed channel estimation strategy, we use the normalized mean-squared error (NMSE), which is defined as

$$\mathrm{NMSE} = \frac{\sum_{u=1}^{U}\sum_{k=1}^{K} ||\hat{\boldsymbol{H}}_u^{\mathrm{H}}[k] - \boldsymbol{H}_u^{\mathrm{H}}[k]||_F^2}{\sum_{u=1}^{U}\sum_{k=1}^{K} ||\boldsymbol{H}_u^{\mathrm{H}}[k]||_F^2}. \quad (9)$$

III. Hybrid Precoder/Combiner Design

To design the hybrid precoders and combiners for the DL, the desired goal would be finding a solution that maximizes the achievable sum-rate in (8) employing the available CSI, i.e.,

$$\max_{\boldsymbol{P}_{\mathrm{RF}},\boldsymbol{P}_{\mathrm{BB}}^u[k],\boldsymbol{W}_{\mathrm{RF}}^u,\boldsymbol{W}_{\mathrm{BB}}^u[k]} \sum_{u=1}^{U} R_u \text{ s.t. } \|\boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}^u[k]\|_{\mathrm{F}}^2 = \frac{P_{\mathrm{tx}}}{UN_{\mathrm{s},u}}, \forall u,k \text{ and } |[\boldsymbol{P}_{\mathrm{RF}}]_{i,j}|^2 = 1, \forall i,j \ \ |[\boldsymbol{W}_{\mathrm{RF}}^u]_{m,n}|^2 = 1, \forall m,n,u, \quad (10)$$

with a power constraint for each MS and subcarrier, such that the total transmit power is P_{tx}. We also define the signal-to-noise ratio as $\mathrm{SNR} = P_{\mathrm{tx}}/(U\sigma_n^2)$ for $\mathbb{E}[\alpha_{u,p}] = 1$ in (1).

The problem formulation in (10) is however difficult to tackle, even neglecting the constraints imposed by the phase shifters, due to the coupling between the precoders. In the following section we propose alternative optimization problems and solve them for an all-digital filter as a previous step to obtain the hybrid solution.

A. Digital Design

A number of solutions have been proposed for the design of precoders and combiners for the MU DL, with the goal of removing the inter-user interference, e.g. [32]–[34]. Another family of solutions is based on MSE dualities between the DL and the UL [35]–[37]. That is, to calculate the combiners in the dual UL, and finding scaling factors to obtain the DL precoders. Leveraging the duality, the UL precoders and combiners can be designed in closed form to minimize the sum-MSE. The sum-MSE establishes a lower bound for the achievable sum-rate, and it gets tighter when MSs have similar power constrains and average channel gains [38]. Hence, the MSE criterion is a common choice when trying to maximize the achievable sum-rate [39], [40].

We propose to design the UL precoders in a selfish way using the available CSI at the MS, similar to coordinated transmit-receive processing in [32], i.e.,

$$\boldsymbol{T}_u[k] = [\boldsymbol{V}_u[k]]_{1:N_{\mathrm{MS},u},1:N_{\mathrm{s},u}}\boldsymbol{\Lambda}_u[k], \quad (11)$$

where $\hat{\boldsymbol{H}}_u^{\mathrm{H}}[k] = \boldsymbol{U}_u[k]\boldsymbol{\Sigma}_u[k]\boldsymbol{V}_u^{\mathrm{H}}[k]$, and $\boldsymbol{\Lambda}_u[k]$ is the usual water-filling solution [41].

Observe that we have neglected the inter-user interference in the design of the uplink precoders. This is motivated by the lack of knowledge of the channels of the rest of the MSs and the UL combiners employed at the BS. On the contrary, the BS has information of all MS' channels and can unilaterally determine the precoders at the MSs (cf. (11)). Thus, for the uplink combiners we focus on minimizing the MSE. The MSE for MS u and subcarrier k in the uplink is given by $\mathrm{MSE}_u^{\mathrm{UL}}[k] = \mathbb{E}[\|\boldsymbol{s}_u^{\mathrm{UL}}[k] - \boldsymbol{y}_u^{\mathrm{UL}}[k]\|_2^2]$. Expanding the expectation and simplifying

$$\mathrm{MSE}_u^{\mathrm{UL}}[k] = N_{\mathrm{s},u} - 2\mathfrak{Re}\left\{\mathrm{tr}\left(\boldsymbol{F}_u^{\mathrm{H}}[k]\boldsymbol{H}_u^{\mathrm{H}}[k]\boldsymbol{T}_u[k]\right)\right\} + \mathrm{tr}\left(\boldsymbol{F}_u^{\mathrm{H}}[k]\left(\sum_{i=1}^{U}\boldsymbol{H}_i^{\mathrm{H}}[k]\boldsymbol{T}_i[k]\boldsymbol{T}_i^{\mathrm{H}}[k]\boldsymbol{H}_i[k] + \sigma_n^2\mathbf{I}\right)\boldsymbol{F}_u[k]\right). \quad (12)$$

Note from (12) that the combiners are decoupled and can be independently designed for each MS and subcarrier. The minimum mean square error (MMSE) combiner [42], [43] is

$$\boldsymbol{F}_{\mathrm{MMSE}}^u[k] = \left(\sum_{i=1}^{U}\hat{\boldsymbol{H}}_i^{\mathrm{H}}[k]\boldsymbol{T}_i[k]\boldsymbol{T}_i^{\mathrm{H}}[k]\hat{\boldsymbol{H}}_i[k] + \sigma_n^2\mathbf{I}\right)^{-1} \times \hat{\boldsymbol{H}}_u^{\mathrm{H}}[k]\boldsymbol{T}_u[k]. \quad (13)$$

Provided that $\hat{\boldsymbol{H}}_u[k]$ and consequently $\boldsymbol{T}_u[k]$ are known at the BS, we obtain (13) for the UL combiners.

Taking into account the characteristics of mmWave channels, it is intriguing to evaluate the performance of other schemes for lower frequencies. In particular, maximum ratio combiner (MRC) strategy presents low complexity and good performance in the low SNR regime; coordinated beamforming (CB) removes the inter-user interference, and it is asymptotically optimal for the achievable sum rate [26], [32]. MRC maximizing the SNR of the received signal is then [44]

$$\boldsymbol{F}_{\mathrm{MRC}}^u[k] = \hat{\boldsymbol{H}}_u^{\mathrm{H}}[k]\boldsymbol{T}_u[k]. \quad (14)$$

According to the equivalent channel model (3), the inter-user interference matrix is

$$\boldsymbol{R}_{\bar{u}}^{\mathrm{UL}}[k] = \sum_{i\neq u}\boldsymbol{A}_{\mathrm{BS}_i}\boldsymbol{\Delta}_i[k]\boldsymbol{A}_{\mathrm{MS}_i}^{\mathrm{H}}\boldsymbol{T}_i[k]\boldsymbol{T}_i^{\mathrm{H}}[k]\boldsymbol{A}_{\mathrm{MS}_i}\boldsymbol{\Delta}_i^{\mathrm{H}}[k]\boldsymbol{A}_{\mathrm{BS}_i}^{\mathrm{H}}.$$

Observe that the rank of this matrix is bounded, that is $\mathrm{rank}(\boldsymbol{R}_{\bar{u}}^{\mathrm{UL}}[k]) \leq \sum_{i\neq u}\min(N_{\mathrm{p},i}, N_{\mathrm{s},i}) \leq L_{\mathrm{BS}} - N_{\mathrm{s},u}$, according to the assumptions in our setup. Similar to [32], we propose to point the combiners in the direction of the nullspace of $\hat{\boldsymbol{R}}_{\bar{u}}^{\mathrm{UL}}[k]$. Therefore, we obtain the CB combiner as

$$\boldsymbol{F}_{\mathrm{CB}}^u[k] = \boldsymbol{U}_u[k]\boldsymbol{U}_u[k]^{\mathrm{H}}\hat{\boldsymbol{H}}_u^{\mathrm{H}}[k]\boldsymbol{T}_u[k]. \quad (15)$$

with $v = \mathrm{rank}(\hat{\boldsymbol{R}}_{\bar{u}}^{\mathrm{UL}}[k])$ and $\boldsymbol{U}_u[k] \in \mathbb{C}^{N_{\mathrm{BS}}\times N_{\mathrm{BS}}-v}$ being the basis of $\hat{\boldsymbol{R}}_{\bar{u}}^{\mathrm{UL}}[k]$ nullspace.

B. Hybrid Design

In this section, we propose different approaches to design hybrid precoders and combiners. In Sec. III-B1 we present an algorithm to solve the hybrid factorization of digital schemes. Moreover, in Sec. III-B2 we directly compute the hybrid precoders and combiners by means of an alternating minimization (AM).

1) Factorization of Digital Designs: In this section, we consider the factorization of the digital solution obtained in Sec. III-A. As in prior work, we use the Frobenius norm as the means of computing the distance between the unconstrained and the hybrid solution. This was shown in [13] to be meaningful for precoding with a unitary constraint; low complexity approaches to the computation were devised in [15]. Minimizing the Frobenius norm is also meaningful for the MSE, as shown in Appendix B. Unlike [19], [45] where a single user is considered, the MSs and the subcarriers are coupled in $\boldsymbol{F}_{\text{RF}}$. By introducing the matrices $\boldsymbol{F}_{\text{MMSE}} = [\boldsymbol{F}^1_{\text{MMSE}}[1], \ldots, \boldsymbol{F}^1_{\text{MMSE}}[K], \ldots, \boldsymbol{F}^U_{\text{MMSE}}[K]] \in \mathbb{C}^{N_{\text{BS}} \times KN_{\text{s}}}$ and $\boldsymbol{F}_{\text{BB}} = [\boldsymbol{F}^1_{\text{BB}}[1], \ldots, \boldsymbol{F}^1_{\text{BB}}[K], \ldots, \boldsymbol{F}^U_{\text{BB}}[K]] \in \mathbb{C}^{L_{\text{BS}} \times KN_{\text{s}}}$, the factorization problem becomes

$$\min_{\boldsymbol{F}_{\text{RF}}, \boldsymbol{F}_{\text{BB}}} \|\boldsymbol{F}_{\text{MMSE}} - \boldsymbol{F}_{\text{RF}}\boldsymbol{F}_{\text{BB}}\|^2_{\text{F}} \text{ s.t. } |[\boldsymbol{F}_{\text{RF}}]_{i,j}|^2 = 1, \, \forall i, j \\ \text{and } \|\boldsymbol{F}_{\text{RF}}\boldsymbol{F}^u_{\text{BB}}[k]\|^2_{\text{F}} = \|\boldsymbol{F}^u_{\text{MMSE}}[k]\|^2_{\text{F}}, \, \forall u, k. \quad (16)$$

An analogous formulation is obtained for the UL precoders, but the factorization is individually performed for each MS. Since the ideas for the UL combiners apply for the UL precoders, we focus on the hybrid design of the UL combiners.

From the proposed methods in the literature, only those allowing $L_{\text{BS}} \ll KN_{\text{s}}$ apply for the problem formulation in (16). The convex relaxation over the set of complex matrices with unitary entries presented in [15, HD-LSR Alg.] can be extended to the frequency selective scenario [23], and its performance will be evaluated in our experiments. It consist of finding the RF precoders performing a search over a ball of radius $1 + \beta$, where β establishes a compromise between the search space size and the relaxation precision. Then, the baseband precoders for such RF update are calculated using least squares (LS). After the AM procedure, a final power normalization is performed to fulfill the power constraints.

a) Projected Gradient Factorization: To design the UL baseband combiners for given UL RF combiners, the main approach in the literature is to use the LS solution. That is, the UL baseband combiner $\boldsymbol{F}_{\text{BB}}$ is the solution of the minimization problem

$$\min_{\boldsymbol{F}_{\text{BB}}} \|\boldsymbol{F}_{\text{MMSE}} - \boldsymbol{F}_{\text{RF}}\boldsymbol{F}_{\text{BB}}\|^2_{\text{F}}, \quad (17)$$

which is given by the closed-form expression $\boldsymbol{F}_{\text{BB}} = \boldsymbol{F}^{\dagger}_{\text{RF}}\boldsymbol{F}_{\text{MMSE}}$. This solution, however, has to be properly scaled to ensure that the power constraints are satisfied. Therefore, we compute the UL baseband combiner as

$$\boldsymbol{F}_{\text{BB}} = \boldsymbol{F}^{\dagger}_{\text{RF}}\boldsymbol{F}_{\text{MMSE}}\tilde{\boldsymbol{\Theta}}, \quad (18)$$

$\tilde{\boldsymbol{\Theta}} = \text{blockdiag}(\theta_1[k]\mathbf{I}_{N_{\text{s},1}}, \ldots, \theta_1[K]\mathbf{I}_{N_{\text{s},1}}, \ldots, \theta_U[K]\mathbf{I}_{N_{\text{s},U}})$ contains the factors resulting from the power normalization $\theta_u[k] = \sqrt{P_{\text{tx}}/(UN_{\text{s},u})}/\|\boldsymbol{F}_{\text{RF}}\boldsymbol{F}^u_{\text{BB}}[k]\|_{\text{F}}$. Substituting (18) in (17) we obtain the distortion metric d

$$d = \|\boldsymbol{F}_{\text{MMSE}}\|^2_{\text{F}} - \text{tr}\left(\boldsymbol{\Theta}\boldsymbol{F}^{\text{H}}_{\text{MMSE}}\boldsymbol{F}_{\text{RF}}(\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{RF}})^{-1}\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{MMSE}}\right), \quad (19)$$

with $\boldsymbol{\Theta} = 2\tilde{\boldsymbol{\Theta}} - \tilde{\boldsymbol{\Theta}}^2$. Note that $\boldsymbol{F}_{\text{RF}}$ is a tall matrix, and $\boldsymbol{\Pi} = \boldsymbol{F}_{\text{RF}}(\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{RF}})^{-1}\boldsymbol{F}^{\text{H}}_{\text{RF}}$ is an orthogonal projector. It is apparent that the accuracy of the LS solution depends on the relationship between the rank of the digital solution and the number of RF chains L_{BS}. For example, in low-frequency architectures deploying an RF chain per antenna we have $\text{rank}(\boldsymbol{F}_{\text{MMSE}}) \leq L_{\text{BS}} = N_{\text{BS}}$. Hence, for any full rank $\boldsymbol{F}_{\text{RF}}$ the projector becomes $\boldsymbol{\Pi} = \mathbf{I}_{N_{\text{BS}}}$ and perfect factorization is ensured.

Taking into account the distortion d introduced by the LS baseband filters and the power normalization (19), we propose an alternative to the methods studied in [15] which does not rely on AM to design the RF and baseband combiners. Indeed, the baseband combiners are implicitly updated in (19), and the expression depends only on the RF combiners.

Now, we propose to design the RF combiners by means of a projected gradient method. The objective function is d in (19), and $|[\boldsymbol{F}_{\text{RF}}]_{i,j}|^2 = 1$ are the constraints defining the feasible set of solutions. First, let us split the distortion up for the MSs and subcarriers as $d = \sum_{u=1}^{U}\sum_{k=1}^{K} d_{u,k}$,

$$d_{u,k} = \|\boldsymbol{F}^u_{\text{MMSE}}[k]\|^2_{\text{F}} - 2\sqrt{\frac{P_{\text{tx}}}{UN_{\text{s},u}}}\|\boldsymbol{A}_u[k]\|_{\text{F}} + \frac{P_{\text{tx}}}{UN_{\text{s},u}} \quad (20)$$

where $\boldsymbol{A}_u[k] = (\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{RF}})^{-1/2}\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}^u_{\text{MMSE}}[k]$. Thus, we pose the optimization problem as follows

$$\min_{\boldsymbol{F}_{\text{RF}}} \sum_{u=1}^{U}\sum_{k=1}^{K} d_{u,k} \quad \text{s.t.} \quad |[\boldsymbol{F}_{\text{RF}}]_{i,j}|^2 = 1, \forall i, j. \quad (21)$$

This problem is non-convex because of the analog restrictions. However, it is possible to reach a local optimum following the direction of the gradient, and projecting the solution to the feasible set afterwards [46]. The gradient step obeys

$$\bar{\boldsymbol{F}}_{\text{RF}} = \boldsymbol{F}_{\text{RF}} + s\frac{\partial d}{\partial \boldsymbol{F}^*_{\text{RF}}}, \quad (22)$$

with the diminishing step size s [46]. The gradient in (22) is computed in Appendix A as

$$\frac{\partial d}{\partial \boldsymbol{F}^*_{\text{RF}}} = \sum_{u=1}^{U}\sum_{k=1}^{K} \frac{\sqrt{P_{\text{tx}}/(UN_{\text{s},u})}}{\|\boldsymbol{A}_u[k]\|_{\text{F}}}\left(\boldsymbol{F}_{\text{RF}}(\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{RF}})^{-1}\boldsymbol{F}^{\text{H}}_{\text{RF}} - \mathbf{I}_{N_{\text{BS}}}\right) \\ \times \boldsymbol{F}^u_{\text{MMSE}}[k]\boldsymbol{F}^{u,\text{H}}_{\text{MMSE}}[k]\boldsymbol{F}_{\text{RF}}(\boldsymbol{F}^{\text{H}}_{\text{RF}}\boldsymbol{F}_{\text{RF}})^{-1}. \quad (23)$$

Note that for $L_{\text{BS}} = N_{\text{BS}}$ the gradient is zero for full rank $\boldsymbol{F}_{\text{RF}}$. Alternatively, the necessary condition for optimality is reached if all $\boldsymbol{F}^u_{\text{MMSE}}[k]$ are contained in the span of the RF combiner.

As aforementioned, the RF precoders do not satisfy the unitary entry constraint after the update. Hence, at each iteration $\bar{\boldsymbol{F}}_{\text{RF}}$ is projected to the feasible set by simply taking the phases

$$\boldsymbol{F}_{\text{RF}} = \arg(\bar{\boldsymbol{F}}_{\text{RF}}). \quad (24)$$

Algorithm 1 Hybrid Design by Projected Gradient (HD-PG)

Initialize: $\ell \leftarrow 0$, random $\boldsymbol{F}_{\text{RF}}(\ell)$
repeat
  $\ell \leftarrow \ell + 1$
  $s \leftarrow s_0$
  repeat
    $\bar{\boldsymbol{F}}_{\text{RF}} \leftarrow \boldsymbol{F}_{\text{RF}}(\ell-1) + s\frac{\partial d}{\partial \boldsymbol{F}_{\text{RF}}^*}$
    $\boldsymbol{F}_{\text{RF}}(\ell) \leftarrow \arg(\bar{\boldsymbol{F}}_{\text{RF}})$
    $s \leftarrow \frac{s}{2}$
  until $d(\boldsymbol{F}_{\text{RF}}(\ell-1)) \geq d(\boldsymbol{F}_{\text{RF}}(\ell))$
until $d(\boldsymbol{F}_{\text{RF}}(\ell-1)) - d(\boldsymbol{F}_{\text{RF}}(\ell)) < \delta$ or $\ell \geq \varepsilon$
$\boldsymbol{F} \leftarrow \boldsymbol{F}^{\text{RF},\dagger}(\ell)\boldsymbol{F}_{\text{MMSE}}$
for all u, l **do**
  $\boldsymbol{F}_{\text{BB}}^u[k] \leftarrow \boldsymbol{F}_{\text{BB}}^u[k] \times \sqrt{P_{\text{tx}}/(UN_{\text{s},u})}/\|\boldsymbol{F}_{\text{RF}}(\ell)\boldsymbol{F}_{\text{BB}}^u[k]\|_{\text{F}}$
end for

Algorithm 1 summarizes the proposed method. Since every step of the algorithm reduces the error metric, which is bounded below by zero, the decreasing step size guarantees convergence to a local optimum [46, Prop. 2.2.1]. It is remarkable that, contrary to methods previously proposed where the power is merely normalized after algorithm's convergence, the power constraint is considered to compute the gradient. In other words, while the power constraints are transparent in Algorithm 1, their performance impact has not been determined for previous approaches.

Algorithm 1 can also be employed to find the UL RF precoders. However, they are individually designed for each MS. Although the number of columns of the digital UL precoding matrix is smaller than that of the UL combining matrix, the number of subcarriers times the number of streams is still large. This fact makes the strategies assuming $L_{\text{MS},u} \geq KN_{\text{s},u}$ hardly applicable in the frequency selective scenario.

Observe that the flexibility of the HD-PG method makes it possible its utilization to factorize the precoder and combiners for any number of MSs, streams, or subcarriers.

2) Hybrid Design by Alternating Minimization: In this section we propose ad hoc solutions directly incorporating the hybrid constraint in the design, as in [15], [19]. The main advantage is that the computationally complex factorization is avoided. This family of solutions consider a fixed RF precoder/combiner already designed according to some criterion. Next, the baseband counterpart is computed to optimize the same or a different criterion taking into account the RF part previously obtained. Now, since the RF optimization usually depends on the baseband matrices, the RF and baseband precoders/combiners are alternatively updated until convergence.

We start designing the UL hybrid precoders by maximizing the mutual information, as in the digital solution in Sec. III-A. Since the RF precoder is common to all subcarriers, it has to be jointly designed for all of them. Let us define $\boldsymbol{H}_u = [\boldsymbol{H}_u^*[1], \ldots, \boldsymbol{H}_u^*[K]]^{\text{T}}$ containing all the channels for all the subcarriers. According to recent work, the common support of the frequency selective mmWave channel model in (1) can be exploited to design the precoders/combiners [47]. Likewise, we propose to use the right singular vectors corresponding to the largest $L_{\text{MS},u}$ singular values as columns of the UL precoder, with the singular value decomposition (SVD) $\boldsymbol{H}_u = \boldsymbol{U}_{\boldsymbol{H}_u}\boldsymbol{\Sigma}_{\boldsymbol{H}_u}\boldsymbol{V}_{\boldsymbol{H}_u}^{\text{H}}$. The motivation behind this choice is to simultaneously increase the SNR for all the subcarriers. These vectors, however, have to be entry-wise normalized to satisfy the analog restrictions,

$$\boldsymbol{T}_{\text{RF}}^u = \arg\left([\boldsymbol{V}_{\boldsymbol{H}_u}]_{1:N_{\text{MS},u},1:L_{\text{MS},u}}\right). \tag{25}$$

For the RF precoders in (25), we compute the equivalent channels $\tilde{\boldsymbol{H}}_u[k] = \boldsymbol{H}_u^{\text{H}}[k]\boldsymbol{T}_{\text{RF}}^u$. Next, the baseband precoding matrices $\boldsymbol{T}_{\text{BB}}^u[k]$ are computed as the mutual information maximizers of a MIMO point-to-point link with the equivalent channel $\tilde{\boldsymbol{H}}_u[k]$ [41] [cf. (11)]. Again, this UL hybrid precoding matrix can be determined at the BS using only the channel estimates, meaning that the UL precoders can be determined at the BS for computing the UL combiners.

Let us define the interference-plus-noise matrix in the uplink, $\boldsymbol{R}_{\text{I}}^{\text{UL}}[k]$, and the desired signal matrix, $\boldsymbol{R}_{\text{D},u}^{\text{UL}}[k]$, as

$$\begin{aligned}\boldsymbol{R}_{\text{I}}^{\text{UL}}[k] &= \sum_{i=1}^{U} \boldsymbol{H}_i^{\text{H}}[k]\boldsymbol{T}_{\text{RF}}^i\boldsymbol{T}_{\text{BB}}^i[k]\boldsymbol{T}_{\text{BB}}^{i,\text{H}}[k]\boldsymbol{T}_{\text{RF}}^{i,\text{H}}\boldsymbol{H}_i[k] + \sigma_n^2\mathbf{I}_{N_{\text{BS}}},\\ \boldsymbol{R}_{\text{D},u}^{\text{UL}}[k] &= \boldsymbol{H}_u^{\text{H}}[k]\boldsymbol{T}_{\text{RF}}^u\boldsymbol{T}_{\text{BB}}^u[k].\end{aligned} \tag{26}$$

Using this compact notation, the MSE considering the RF and the baseband combiners is

$$\begin{aligned}\text{MSE}_u^{\text{UL}}[k] = N_{\text{s},u} &- 2\mathfrak{Re}\left\{\text{tr}(\boldsymbol{F}_{\text{BB}}^{u,\text{H}}[k]\boldsymbol{F}_{\text{RF}}^{\text{H}}\boldsymbol{R}_{\text{D},u}^{\text{UL}}[k])\right\}\\ &+ \text{tr}\left(\boldsymbol{F}_{\text{BB}}^{u,\text{H}}[k]\boldsymbol{F}_{\text{RF}}^{\text{H}}\boldsymbol{R}_{\text{I}}^{\text{UL}}[k]\boldsymbol{F}_{\text{RF}}\boldsymbol{F}_{\text{BB}}^u[k]\right).\end{aligned} \tag{27}$$

For the combiner in the uplink we propose a different criterion design as that of the precoder. In particular, the RF combiner aims at maximizing the second term of (27) for all MSs and subcarriers, i.e., maximizing the SNR of the received signal

$$\boldsymbol{F}_{\text{RF}} = \arg\left(\sum_{k=1}^{K}\sum_{u=1}^{U}\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k]\boldsymbol{F}_{\text{BB}}^{u,\text{H}}[k]\right), \tag{28}$$

where $\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k]$ is the desired signal matrix in (26) with the channel estimate $\hat{\boldsymbol{H}}_u^{\text{H}}[k]$. This way, we have a closed-form expression for the RF combiner for which the baseband combiners are considered to be given. The baseband combiners $\boldsymbol{F}_{\text{BB}}^u[k]$ are independently designed for each MS and subcarrier for given RF combiners, according to the MSE criterion

$$\boldsymbol{F}_{\text{BB}}^{u,\text{MMSE}}[k] = \left(\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\text{I}}^{\text{UL}}[k]\boldsymbol{F}_{\text{RF}}\right)^{-1}\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k], \tag{29}$$

with $\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k]$ being the interference-plus-noise matrix (26) for the estimated channel $\hat{\boldsymbol{H}}_u^{\text{H}}[k]$. Equations (29) and (28) motivate the AM process. Thereby, the baseband and the RF combiners are alternatively updated until convergence. Unfortunately, convergence cannot be theoretically guaranteed. Since the updates of the RF and baseband combiners obey different criterion, it is possible that the cost function does not reduce at every step. However, we observe good results from numerical experiments.

In Section III-A we posed alternatives to the digital MMSE UL combiner. Likewise, we propose different baseband combiner designs to include in the AM procedure. That is, for

the RF combiner in (28) we employ MRC-like and CB-like baseband combiners instead of (29)

$$\boldsymbol{F}_{\text{BB}}^{u,\text{MRC}}[k] = \left(\boldsymbol{F}_{\text{RF}}^{\text{H}}\boldsymbol{C}_{\boldsymbol{n}}\boldsymbol{F}_{\text{RF}}\right)^{\dagger}\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k], \text{ and}$$
$$\boldsymbol{F}_{\text{BB}}^{u,\text{CR}}[k] = \boldsymbol{U}_u[k]\boldsymbol{U}_u^{\text{H}}[k]\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\text{D},u}^{\text{UL}}[k]. \tag{30}$$

where $\boldsymbol{U}_u[k]$ is the basis of null space of $\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\bar{u}}^{\text{UL}}[k]\boldsymbol{F}_{\text{RF}}$ (cf. (15)). Note that the bound for the rank in Sec. III-A holds also here $\text{rank}(\boldsymbol{F}_{\text{RF}}^{\text{H}}\hat{\boldsymbol{R}}_{\bar{u}}^{\text{UL}}[k]\boldsymbol{F}_{\text{RF}}) \leq \text{rank}(\hat{\boldsymbol{R}}_{\bar{u}}^{\text{UL}}[k])$.

IV. Compressive Channel Estimation

This section is devoted to provide a solution for the problem of estimating the multi-user uplink mmWave frequency-selective MIMO channel. In this work, unlike in [11], we propose an alternative to perform channel estimation at the BS in the uplink. We provide the mathematical model for simultaneous estimation of the UK channel matrices $\boldsymbol{H}_u^{\text{H}}[k]$, $\forall u, k$, and use the simultaneous weighted - orthogonal matching pursuit (SW-OMP) algorithm in [5] to perform this task. Furthermore, the CRLB for the estimation of the multi-user channel is provided. Finally, we propose a reasonably fair criterion to compare uplink and downlink channel estimation. In the simulations section, we will show that the estimation performance in the BS significantly outperforms downlink channel estimation, and will also compare the CRLB for both scenarios.

A. Uplink Compressive Channel Estimation

We assume that the BS and the U MSs undertake a training stage in which the different MSs simultaneously send M training frames to the BS using OFDM signaling. For the transmission of the m-th training frame at subcarrier k, each MS uses a precoder $\boldsymbol{T}_{\text{tr},u}^{(m)}[k] \in \mathbb{C}^{N_{\text{MS},u}\times L_{\text{MS},u}}$ to transmit a training symbol $\boldsymbol{s}_u^{(m)}[k]$, and the BS process the received MIMO signal with a combiner $\boldsymbol{F}_{\text{tr}}^{(m)}[k] \in \mathbb{C}^{N_{\text{BS}}\times L_{\text{BS}}}$. Accordingly, the multi-user received signal at the BS can be written as

$$\boldsymbol{y}^{(m)}[k] = \boldsymbol{F}_{\text{tr}}^{(m)\text{H}}[k]\sum_{u=1}^{U}\boldsymbol{H}_u^{\text{H}}[k]\boldsymbol{T}_{\text{tr},u}^{(m)}[k]\boldsymbol{s}_u^{(m)}[k] + \boldsymbol{n}^{(m)}[k], \tag{31}$$

where $\boldsymbol{y}^{(m)}[k]$ is the $L_{\text{BS}} \times 1$ received signal during the m-th training step, and $\boldsymbol{n}^{(m)}[k] \in \mathbb{C}^{L_{\text{BS}}\times 1}$ is the received post-combining additive colored Gaussian noise vector with distribution $\boldsymbol{n}^{(m)}[k] \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{F}_{\text{tr}}^{(m)\text{H}}[k]\boldsymbol{F}_{\text{tr}}^{(m)}[k])$. The received signal in (31) can be further expressed as

$$\boldsymbol{y}^{(m)}[k] = \boldsymbol{F}_{\text{tr}}^{(m)\text{H}}[k]\underbrace{\left[\begin{array}{ccc}\boldsymbol{H}_1^{\text{H}}[k] & \dots & \boldsymbol{H}_U^{\text{H}}[k]\end{array}\right]}_{\boldsymbol{H}^{\text{H}}[k]}\times \tag{32}$$
$$\times\underbrace{\left[\begin{array}{ccc}\boldsymbol{s}_1^{(m)\text{T}}[k]\boldsymbol{T}_{\text{tr},1}^{(m)\text{T}}[k] & \dots & \boldsymbol{s}_U^{(m)\text{T}}[k]\boldsymbol{T}_{\text{tr},U}^{(m)\text{T}}[k]\end{array}\right]^{\text{T}}}_{\boldsymbol{x}[k]} + \boldsymbol{n}^{(m)}[k],$$

which will be used as baseline to derive the relation between the received signal and the vectorized channel matrices for each MS. If we apply the $\text{vec}\{\cdot\}$ operator to the previous equation, it is possible to write

$$\boldsymbol{y}^{(m)}[k] = \left(\boldsymbol{x}^{\text{T}}[k]\otimes\boldsymbol{F}_{\text{tr}}^{(m)\text{H}}[k]\right)\text{vec}\left\{\boldsymbol{H}^{\text{H}}[k]\right\} + \boldsymbol{n}^{(m)}[k]. \tag{33}$$

Now, using the extended virtual channel model in (4), the vectorized channel matrix $\boldsymbol{H}_u^{\text{H}}[k]$ for the u-th MS can be approximated as

$$\text{vec}\{\boldsymbol{H}_u^{\text{H}}[k]\} \approx \overbrace{\left(\tilde{\boldsymbol{A}}_{\text{MS},u}^{*}\otimes\tilde{\boldsymbol{A}}_{\text{BS}}\right)}^{\tilde{\boldsymbol{\Psi}}_u}\text{vec}\{\tilde{\boldsymbol{\Delta}}_u[k]\}, \tag{34}$$

with $\tilde{\boldsymbol{\Psi}}_u \in \mathbb{C}^{N_{\text{BS}}N_{\text{MS},u}\times G_{\text{BS}}G_{\text{MS}}}$ the dictionary matrix for the u-th user. The expression in (34) approximately holds if the grid sizes of the dictionary matrices are large enough [30]. Typically, the choices $G_{\text{BS}} \geq 2N_{\text{BS}}$ and $G_{\text{MS}} \geq 2N_{\text{MS},u}$ are considered to estimate the channel [5], [6]. Thereby, the approximation in (34) can be plugged into (33) to obtain

$$\boldsymbol{y}^{(m)}[k] \approx \overbrace{\left(\boldsymbol{x}^{\text{T}}[k]\otimes\boldsymbol{F}_{\text{tr}}^{(m)\text{H}}[k]\right)}^{\boldsymbol{\Phi}^{(m)}[k]}\left(\bigoplus_{u=1}^{U}\tilde{\boldsymbol{\Psi}}_u\right)\boldsymbol{h}_{\text{MU}}^{\text{v}}[k] + \boldsymbol{n}^{(m)}[k], \tag{35}$$

with the vector $\boldsymbol{h}_{\text{MU}}^{\text{v}}[k] = [\text{vec}\{\boldsymbol{\Delta}_1^{\text{v}}[k]\}^{\text{T}}\dots\text{vec}\{\boldsymbol{\Delta}_U^{\text{v}}[k]\}^{\text{T}}]^{\text{T}}$ containing the concatenation of the vectorized U sparse matrices $\boldsymbol{\Delta}_u^{\text{v}}[k]$, and the matrix $\boldsymbol{\Phi}^{(m)}[k] \in \mathbb{C}^{L_{\text{BS}}\times N_{\text{BS}}\sum_{u=1}^{U}N_{\text{MS},u}}$ is termed the measurement matrix for channel reconstruction.

The signal model in (35) considers the general use of frequency-selective hybrid precoders and combiners to estimate the channel. However, the task of compressive channel estimation has computational complexity growing with the number of subcarriers K [5]. To reduce this complexity, we consider the use of frequency-flat training precoders and combiners. Hence, $\boldsymbol{T}_{\text{tr},u}^{(m)}[k] = \boldsymbol{T}_{\text{tr},u}^{(m)}$, and $\boldsymbol{F}_{\text{tr}}^{(m)}[k] = \boldsymbol{F}_{\text{tr}}^{(m)}$, $\forall k$, and the channel vectors $\boldsymbol{h}_{\text{MU}}^{\text{v}}[k]$, $k = 1,\dots,K$, can be recovered by using only a single subcarrier-independent measurement matrix. Observe, however, that the L_{BS} and $L_{\text{MS},u}$ available degrees of freedom at the transmitter and receiver are exploited. Since OFDM training pilots are forwarded through the MIMO channel, the frequency-selective nature of the transmitted signal cannot be avoided. However, we can express the measurement matrix $\boldsymbol{\Phi}^{(m)}[k]$ as $\boldsymbol{\Phi}^{(m)}[k] = s^{(m)}[k]\boldsymbol{\Phi}^{(m)}$ if the OFDM symbols sent by the different MSs are rotated versions of a baseline OFDM symbol. This can be written as $\boldsymbol{s}_u^{(m)}[k] = \boldsymbol{q}_u^{(m)}s^{(m)}[k]$, $\forall u$, with $\boldsymbol{q}_u^{(m)} \in \mathbb{C}^{L_{\text{MS},u}\times 1}$ being a spatial modulation vector. With this choice, we can simultaneously exploit the degrees of freedom coming from having several RF chains at each MS and have a frequency-independent measurement matrix, as shown hereafter.

Taking these considerations into account, the measurement matrix at the m-th training step can be written as

$$\boldsymbol{\Phi}^{(m)} = \left[\begin{array}{ccc}\boldsymbol{q}_1^{(m)T}\boldsymbol{T}_{\text{tr},1}^{(m)\text{T}} & \dots & \boldsymbol{q}_U^{(m)T}\boldsymbol{T}_{\text{tr},U}^{(m)\text{T}}\end{array}\right]\otimes\boldsymbol{F}_{\text{tr}}^{(m)\text{H}}, \tag{36}$$

in which the pilot $s^{(m)}[k]$ is a subcarrier-dependent scalar whose effect can be eliminated at the receiver by simply multiplying the received vector by $(s^{(m)})^{-1}[k]$ or by $s^{(m)*}[k]$, where the latter would only apply for constellations with

symbols having the same energy. Thereby, the subcarrier dependence of the matrix $\boldsymbol{\Phi}^{(m)}[k]$ in (35) is avoided. It is important to remark that the different MSs must use different precoders to transmit each training frame. The interpretation of such fact relies on the ability of the BS to obtains enough information and discriminate the information of the channel for each MS. In terms of sparse recovery, this means that users' using different precoders for a given training step translates into reduced correlation between the columns in $\boldsymbol{\Phi}^{(m)}$, thereby enabling support recovery for subsequent estimation of the channel gains and the channel matrices themselves.

If we consider the transmission of M consecutive training frames, the received measurement vectors for each subcarrier can be written as an extension of (36) as

$$\underbrace{\begin{bmatrix} \boldsymbol{y}^{(1)\mathrm{T}}[k] \\ \vdots \\ \boldsymbol{y}^{(M)\mathrm{T}}[k] \end{bmatrix}^{\mathrm{T}}}_{\boldsymbol{y}[k]} \approx \underbrace{\begin{bmatrix} \boldsymbol{\Phi}^{(1)} \\ \vdots \\ \boldsymbol{\Phi}^{(M)} \end{bmatrix}}_{\boldsymbol{\Phi}} \underbrace{\left(\bigoplus_{u=1}^{U} \tilde{\boldsymbol{\Psi}}_u \right)}_{\tilde{\boldsymbol{\Psi}}_{\mathrm{MU}}} \tilde{\boldsymbol{h}}_{\mathrm{MU}}[k] + \quad (37)$$

$$+ \quad \underbrace{\begin{bmatrix} \boldsymbol{n}^{(1)\mathrm{T}}[k] & \dots & \boldsymbol{n}^{(M)\mathrm{T}}[k] \end{bmatrix}^{\mathrm{T}}}_{\boldsymbol{n}[k]},$$

where $\tilde{\boldsymbol{\Psi}}_u \in \mathbb{C}^{N_{\mathrm{BS}} N_{\mathrm{MS},u} \times G_{\mathrm{BS}} G_{\mathrm{MS}}}$ is the dictionary matrix defined as $\tilde{\boldsymbol{\Psi}}_u \triangleq \left(\tilde{\boldsymbol{A}}^{*}_{\mathrm{MS},u} \otimes \tilde{\boldsymbol{A}}_{\mathrm{BS}} \right)$. The matrix $\tilde{\boldsymbol{\Psi}}_{\mathrm{MU}} \in \mathbb{C}^{N_{\mathrm{BS}} \sum_{u=1}^{U} N_{\mathrm{MS},u} \times U G_{\mathrm{BS}} G_{\mathrm{MS}}}$ is the multi-user dictionary. The $ML_{\mathrm{BS}} \times 1$ vector $\boldsymbol{n}[k]$ is distributed according to $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{C}_{\mathrm{w}})$, where $\boldsymbol{C}_{\mathrm{w}}$ is the positive definite covariance matrix given by $\boldsymbol{C}_{\mathrm{w}} = \mathrm{blkdiag}\left\{ \boldsymbol{F}_{\mathrm{tr}}^{(1)\mathrm{H}} \boldsymbol{F}_{\mathrm{tr}}^{(1)}, \dots, \boldsymbol{F}_{\mathrm{tr}}^{(M)\mathrm{H}} \boldsymbol{F}_{\mathrm{tr}}^{(M)} \right\}$. Likewise, we also define the Cholesky factor of $\boldsymbol{C}_{\mathrm{w}}$ as $\boldsymbol{D}_{\mathrm{w}} \in \mathbb{C}^{ML_{\mathrm{r}} \times ML_{\mathrm{r}}}$, $\boldsymbol{C}_{\mathrm{w}} = \boldsymbol{D}_{\mathrm{w}}^{\mathrm{H}} \boldsymbol{D}_{\mathrm{w}}$, which will be used in the proposed multi-user wideband channel estimation algorithm.

Therefore, the sparse vectors $\tilde{\boldsymbol{h}}_{\mathrm{MU}}[k]$ can be recovered by solving the relaxed optimization problem

$$\tilde{\boldsymbol{h}}_{\mathrm{MU}}[k] = \underset{\boldsymbol{h}[k], k=1,\dots,K}{\arg\min} \sum_{k=1}^{K} ||\boldsymbol{h}[k]||_1 \ \text{ s.t. } \ \sum_{k=1}^{K} L(\boldsymbol{h}[k]) \leq \epsilon,$$

where

$$L(\boldsymbol{h}[k]) = \left(\boldsymbol{y}[k] - \boldsymbol{\Phi} \tilde{\boldsymbol{\Psi}}_{\mathrm{MU}} \boldsymbol{h}^{\mathrm{v}}[k] \right)^{\mathrm{H}} \boldsymbol{C}_{\mathrm{w}}^{-1} \left(\boldsymbol{y}[k] - \boldsymbol{\Phi} \tilde{\boldsymbol{\Psi}}_{\mathrm{MU}} \boldsymbol{h}^{\mathrm{v}}[k] \right) \quad (38)$$

is the cost function associated to the log-likelihood function (LLF) of the received Gaussian signal [48]. The solution to this problem can be obtained by using either the SW-OMP or the subcarrier-selection simultaneous weighted - orthogonal matching pursuit + thresholding (SS-SW-OMP+Th) algorithms in [5], which have been shown to obtain estimation performance lying within the CRLB when the AoA and AoD fall within the quantized spatial grid. We propose to use the SW-OMP algorithm for illustration. In the next subsection, the CRLB for the estimation of the channel matrices $\boldsymbol{H}_u[k]$, for all MSs and subcarriers is provided. The channel estimation algorithm is provided in Algorithm 2.

Algorithm 2 Simultaneous Weighted Orthogonal Matching Pursuit (SW-OMP)

1: Initialize: $\boldsymbol{\Upsilon}_{\mathrm{w}} \leftarrow \boldsymbol{D}_{\mathrm{w}}^{-\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\Psi}$, $\boldsymbol{y}_{\mathrm{w}}[k] \leftarrow \boldsymbol{D}_{\mathrm{w}}^{-\mathrm{H}} \boldsymbol{y}[k]$, and $\boldsymbol{r}[k] \leftarrow \boldsymbol{y}_{\mathrm{w}}[k]$, $k = 1, \dots, K$
2: **repeat**
3: $\quad \boldsymbol{c}[k] \leftarrow \boldsymbol{\Upsilon}_{\mathrm{w}}^{\mathrm{H}} \boldsymbol{r}[k], \quad k = 1, \dots, K$
4: $\quad p^{\star} \leftarrow \arg\max_{p} \sum_{k=1}^{K} |[\boldsymbol{c}[k]]_p|$
5: $\quad \hat{\mathcal{T}} \leftarrow \hat{\mathcal{T}} \cup p^{\star}$ common support candidate
6: $\quad$ **for all** k **do**
7: $\quad\quad \boldsymbol{x}_{\hat{\mathcal{T}}}[k] \leftarrow \left([\boldsymbol{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}} \right)^{\dagger} \boldsymbol{y}_{\mathrm{w}}[k]$ support's subspace proj.
8: $\quad\quad \boldsymbol{r}[k] \leftarrow \boldsymbol{y}_{\mathrm{w}}[k] - [\boldsymbol{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}} \boldsymbol{x}_{\hat{\mathcal{T}}}[k]$ update residual
9: $\quad$ **end for**
10: $\quad$ MSE $\leftarrow \frac{1}{KML_{\mathrm{r}}} \sum_{k=1}^{K} \boldsymbol{r}^{\mathrm{H}}[k] \boldsymbol{r}[k]$
11: **until** MSE $< \epsilon$

B. Cramér-Rao Lower Bound

This section is devoted to obtain the theoretical expression of the CRLB for the estimation of the different subchannels for each MS. We will focus on obtaining the bound assuming perfect estimation of the support of the multi-user channel vector. The motivation in doing so is to assess how well the estimator of the support performs, since the estimator of the channel gains derived in [5] is efficient if the support is correctly estimated. Assuming that the AoA and AoD have been found, the signal model in (38) reduces to

$$\boldsymbol{y}[k] = \overbrace{\boldsymbol{\Phi} \left(\bigoplus_{u=1}^{U} \underbrace{\left(\boldsymbol{A}^{*}_{\mathrm{MS}_u} \circ \boldsymbol{A}_{\mathrm{BS}_u} \right)}_{\boldsymbol{\Psi}_u} \right)}^{\boldsymbol{\Upsilon}} \boldsymbol{\xi}[k] + \boldsymbol{n}[k], \quad (39)$$

with $\boldsymbol{A}_{\mathrm{BS}_u}$ and $\boldsymbol{A}_{\mathrm{MS}_u}$ the array matrices in (3) for the BS and the u-th MS. The $\sum_{u=1}^{U} N_{\mathrm{p},u} \times 1$ vector $\boldsymbol{\xi}[k]$ contains the actual channel gains for the multi-user channel and is defined as

$$\boldsymbol{\xi}[k] \triangleq \mathrm{vec}\left\{ \begin{bmatrix} \mathrm{diag}\{\boldsymbol{\Delta}_1[k]\} & \dots & \mathrm{diag}\{\boldsymbol{\Delta}_U[k]\} \end{bmatrix} \right\}, \quad (40)$$

where $\boldsymbol{\Delta}_u[k]$ are the $N_{\mathrm{p},u} \times N_{\mathrm{p},u}$ diagonal matrices in (3). The Fisher Information Matrix for $\boldsymbol{\xi}[k]$ is given by [48], [5]

$$\boldsymbol{I}(\boldsymbol{\xi}[k]) = \frac{\boldsymbol{\Upsilon}^{\mathrm{H}} \boldsymbol{C}_{\mathrm{w}}^{-1} \boldsymbol{\Upsilon}}{\sigma^2}. \quad (41)$$

Thereby, the CRLB matrix for the estimation of the multi-user channel matrix $\boldsymbol{H}^{\mathrm{H}}[k]$ is given by the *CRLB Theorem for Transformed Parameters* [48] as

$$\boldsymbol{I}^{-1}(\mathrm{vec}\{\boldsymbol{H}^{\mathrm{H}}[k]\}) = \left(\bigoplus_{u=1}^{U} \boldsymbol{\Psi}_u \right) \boldsymbol{I}^{-1}(\boldsymbol{\xi}[k]) \left(\bigoplus_{u=1}^{U} \boldsymbol{\Psi}_u^{\mathrm{H}} \right). \quad (42)$$

Finally, assuming that no prior information concerning the relation between the different channels $\boldsymbol{H}_u[k], \forall k, u$ is available, the overall variance for the estimation of the UK different channel matrices is denoted by $\gamma(\boldsymbol{h}_{\mathrm{total}})$ and given by

$$\gamma(\boldsymbol{h}_{\mathrm{total}}) = \sum_{k=1}^{K} \mathrm{tr}\left\{ \boldsymbol{I}^{-1}(\mathrm{vec}\{\boldsymbol{H}_u[k]\}) \right\}, \quad (43)$$

where $\boldsymbol{h}_{\text{total}} \in \mathbb{C}^{KN_{\text{BS}}\sum_{u=1}^{U} N_{\text{MS},u}\times 1}$ is given by

$$\boldsymbol{h}_{\text{total}} = \text{vec}\left\{\left[\ \boldsymbol{H}^{\text{H}}[1]\ldots,\boldsymbol{H}^{\text{H}}[K]\ \right]\right\}, \tag{44}$$

such that the variance for the estimation of the different subchannels is added up owing to the lack of knowledge concerning possible correlation between the different channel matrices.

In the next subsection, we will introduce the different considerations to be made for a proper and fair comparison between uplink and downlink multi-user wideband channel estimation.

C. Uplink vs. Downlink Multi-user Wideband Channel Estimation

This subsection aims at providing a proper comparison between uplink and downlink mmWave multi-user wideband channel estimation. The main problem is how to perform a comparison between both approaches comes from the different SNR measured at the MSs and the BS. Provided that the different signals transmitted by each MS are uncorrelated, the SNR measured at the BS will be the sum of the different SNRs concerning the individual signals transmitted by each MS. This can be proven by considering a multi-user scenario with U MS.

Let $\boldsymbol{H}_u^{\text{H}}[k]$, $\boldsymbol{T}_{\text{tr},u}^{(m)}$, and $\boldsymbol{s}_u^{(m)}[k]$ denote the uplink channel matrix for the u-th user at subcarrier k, the training precoder used by the u-th user during the m-th training step, and the vector of training symbols sent by the u-th user during the m-th training step at k-th subcarrier, respectively. Let us also denote the received noise vector for the m-th training step and at the k-th subcarrier as $\boldsymbol{n}^{(m)}[k]$. Then, the received signal at the BS may be expressed as

$$\boldsymbol{y}^{(m)}[k] = \underbrace{\sum_{u=1}^{U} \boldsymbol{H}_u^{H}[k]\boldsymbol{T}_{\text{tr},u}^{(m)}\boldsymbol{s}_u^{(m)}[k]}_{\boldsymbol{x}^{(m)}[k]} + \boldsymbol{n}^{(m)}[k]. \tag{45}$$

Next, let the receive SNR be defined as

$$\text{SNR} = \frac{\text{tr}\{\mathbb{E}\{\boldsymbol{x}^{(m)}[k]\boldsymbol{x}^{(m)H}[k]\}\}}{\text{tr}\{\mathbb{E}\{\boldsymbol{n}^{(m)}[k]\boldsymbol{n}^{(m)H}[k]\}\}}. \tag{46}$$

Now, let us recall that the training symbols are assumed to be uncorrelated for the different users, i.e., $\mathbb{E}\{\boldsymbol{s}_u^{(m)}[k]\boldsymbol{s}_j^{(m)H}[k]\} = \boldsymbol{0}$, for $u \neq j$. Accordingly, the cross-product terms in the numerator of (46) are zero-valued. Then, it turns out that the expression in (46) yields

$$\text{SNR} = \sum_{u=1}^{U} \frac{P_u\,\text{tr}\{\|\boldsymbol{H}_u^{H}[k]\boldsymbol{F}_{\text{tr},u}\|_F^2\}}{UN_{\text{s}}N_{\text{r}}\sigma^2}, \tag{47}$$

which is the summation of the SNR received by the different users, i.e., the summation of the SNR corresponding to the downlink for every user.

Uncorrelation between the different signals can be claimed provided that the training precoders are pseudorandomly built with a network of phase-shifters having N_{Q} quantization bits. Thereby, the comparison between UL and DL channel estimation will be made taking into account the SNR for a single user. This decision is made based on the fact that each MS measures a SNR U times smaller than that of the BS. Yet, the MSs separately estimate K $N_{\text{p},u}$-sparse vectors each, while the BS estimates $K\sum_{u=1}^{U} N_{\text{p},u}$-sparse vectors, such that the proposed SNR criterion seems reasonable for comparison. In fact, the proposed criterion is the expected ratio between the SNR and the number of multipath components to be estimated, for either uplink or downlink scenarios.

V. Numerical Results

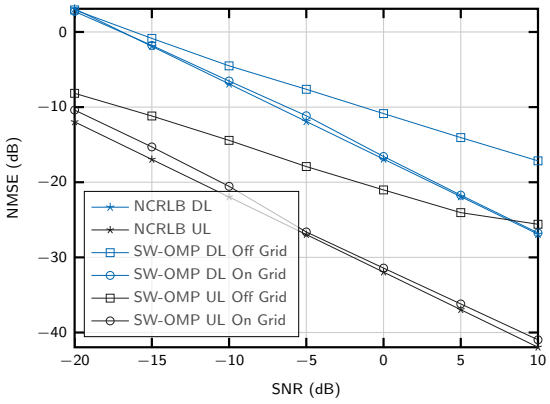
In this section, we compute several numerical experiments to obtain further insight of the proposed methods. First, we present a performance comparison for the channel estimation in the uplink and the downlink in Sec. V-A. In Sec. V-B, we evaluate the sum achievable rates obtained with digital and hybrid precoder/combiner designs considering perfect CSI. Finally, Sec. V-C is devoted to experiments taking into account the available CSI.

The setup considered for our experiments consist of a BS with $N_{\text{BS}} = 128$ antennas transmitting to $U = 4$ MSs deploying $N_{\text{MS},u} = 16$ antennas each, which individually allocate $N_{\text{s},u} = 2$ streams. The number of RF chains are $L_{\text{BS}} = 8$ and $L_{\text{MS},u} = 2$, while number of subcarriers considered is $K = 32$. The channels are generated according to the D delay model in (1), with delay $D = 8$ and $N_{\text{p}} = 4$ paths. The results are averaged by performing Monte Carlo simulations over 100 channel realizations. We consider spatially withe noise for the uplink and the downlink, and its power is fixed according to $\text{SNR} = P_{\text{tx}}/(U\sigma_n^2)$. The maximum number of iterations ε is set to 200 for Alg. 1, 100 for HD-LSR and 50 for the AM methods.

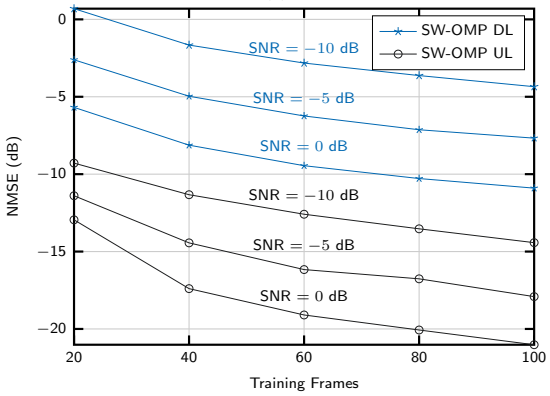
A. Channel Estimation Performance

This subsection is devoted to show simulation results concerning channel estimation performance for both UL and DL channel estimation. The NMSE expression in (9) will be used to compare the achievable performance when estimation is performed at the BS and each of the U MSs. Besides the simulation parameters introduced at the beginning of this section, we take $G_{\text{BS}} = 2N_{\text{BS}} = 256$ and $G_{\text{MS}} = 2N_{\text{MS}} = 32$. The phase-shifters used in both the BS and MS are assumed to have N_{Q} quantization bits, so that the entries of the training precoders and combiners $\boldsymbol{F}_{\text{tr}}^{(m)}$ and $\boldsymbol{W}_{\text{tr}}^{(m)}$, $m = 1, 2, \ldots, M$ are drawn from a set $\mathcal{A} = \left\{0, \frac{2\pi}{2^{N_{\text{Q}}}}, \ldots, \frac{2\pi(2^{N_{\text{Q}}}-1)}{2^{N_{\text{Q}}}}\right\}$. The number of quantization bits is set to $N_{\text{Q}} = 4$, and the number of training frames is set to $M = 100$.

We show in Fig. 1 the sample average NMSE versus both SNR and training overhead (number of training frames M). We first observe a large difference between channel estimation performance at the BS and the MSs. This is due to the higher amount of information the BS can manage to estimate the channel. This comes from the larger dictionary used at the BS, which makes it possible to perform a better compressive estimation. Furthermore, the BS uses $L_{\text{BS}} = 8$ RF chains for channel estimation at the, while each MS is restricted to use $L_{\text{MS},u} = 2$ RF chains. This fact, jointly with the use of a larger



(a)



(b)

Fig. 1: Estimation performance (NMSE) for both UL and DL multiuser channel estimation considering both on-grid and off-grid channel realizations: (a) NMSE versus SNR. (b) NMSE versus number of training frames M for SNR ranging $\{-10, -5, 0\}$ dB.

dictionary, are the main reason why uplink channel estimation greatly outperforms its downlink counterpart. We also see that the gap between both approaches is reflected in the CRLB for both cases. Both the estimation at the BS and the MSs exhibit performance lying very close to the CRLB, although there is a small yet non-negligible gap between uplink channel estimation and the CRLB, when the AoA/AoD fall within the quantized spatial grid. This gap can be further reduced by considering a larger number of subcarriers. This is because the sparsity level the BS must estimate is approximately U times larger (if the grid sizes are large enough), such that the estimation error of the channel gains is approximately U larger as well. Therefore, according to the Law of Large Numbers, either a larger number of training frames or a larger number of subcarriers must be used to obtain better estimates of the noise variance. Thereby, the SW-OMP algorithm can be halted more efficiently. This effect does not appear in the case of channel estimation in the DL, since the sparsity level of each subchannel is lower. Nonetheless, it should be clear that, even with this small gap, channel estimation at the BS greatly outperforms estimation at the MSs. In the event of off-grid AoA/AoD, there is a considerable loss in performance for both UL and DL channel estimation, although the performance gap between them is approximately maintained. Further, for SNR ranging from -10 up to 0 dB, the NMSE performance with off-grid parameters is lower than -10 dB, which is a clear indicator that the proposed estimation approach is suitable for multi-user UL channel estimation with very-low SNR conditions.

B. Performance Evaluation for Perfect CSI

Fig. 2a compares the different digital designs in Sec. III-A and the hybrid approaches in sections III-B1 and III-B2 for perfect CSI. The performance of digital CB, and MMSE is very similar, whereas the gap for MRC is increasing with the SNR. Iterative methods in Sec. III-B2 present better performance, specially for MMSE combiner. Nevertheless, the performance of Alg. 1 can be greatly improved using a smarter initialization, as stated in Sec. V-C. Fig. 2b shows that the performance of the digital solution can be achieved by the factorizations when the number of RF chains is large enough, whereas the gap with respect to AM is always greater than zero. Furthermore, less RF chains are needed to obtain an accurate factorization for MRC, compared to that for MMSE.

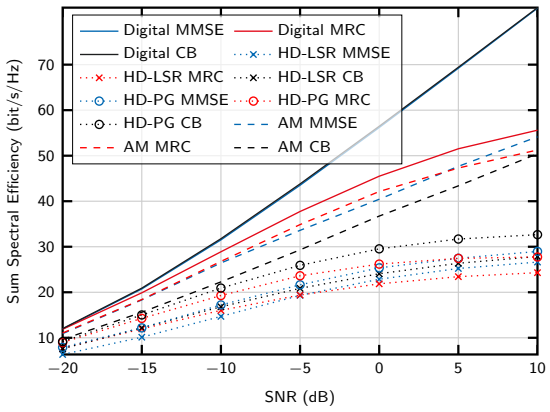
C. Performance Evaluation for Imperfect CSI

Now we evaluate the performance loss resulting from the imperfect CSI. Let us start by studying the purely digital case to determine the robustness of the different combiner designs against the channel uncertainties.

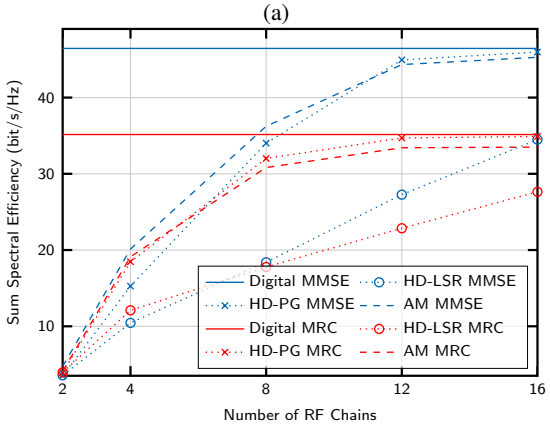
As can be observed in Fig. 3a, the gap introduced because of the estimation error is noticeable for MMSE, CB combiners. Contrarily, the robustness of MRC leads to very similar performance for perfect and imperfect CSI.

The gaps in Fig. 3b come from the combination of several impairments that individually contribute to worsening the achievable performance. From the channel model in (2), it is clear that the rank of the resulting frequency-domain channel matrices increases with the number of paths N_{p}. Moreover, the multi-user uplink channel $\boldsymbol{H}[k] = [\boldsymbol{H}_1^{\mathrm{H}}[k] \dots \boldsymbol{H}_U^{\mathrm{H}}[k]]$ can be vectorized and expressed in terms of the array matrices employed by the BS and MSs as $\mathrm{vec}\{\boldsymbol{H}[k]\} = \bigoplus_{u=1}^{U} \left(\boldsymbol{A}_{\mathrm{MS},u}^{*} \circ \boldsymbol{A}_{\mathrm{BS},u}\right) \mathrm{vec}\{\mathrm{diag}\{\boldsymbol{\Delta}_u[k]\}\}$. Then, if the AoA and AoD are i.i.d. and follow a continuous uniform distribution, and $\sum_{u=1}^{U} N_{\mathrm{p},u} \leq \min(N_{\mathrm{BS}}, N_{\mathrm{MS},u})$, the probability that transmit or receive steering vectors of the channels for different MSs are linearly dependent is zero. Accordingly, $\mathrm{vec}\{\boldsymbol{H}[k]\}$ lies in a $N_{\mathrm{BS}} \sum_{u=1}^{U} N_{\mathrm{MS},u}$-dimensional subspace spanned by $\sum_{u=1}^{U} N_{\mathrm{p},u}$ non-orthogonal linearly independent vectors. Therefore, we can conclude that the rank of the multi-user channel grows linearly with the number of MSs U such that $\mathrm{rank}\{\boldsymbol{H}[k]\} = \min(N_{\mathrm{BS}}, \sum_{u=1}^{U} N_{\mathrm{p},u})$.

Observe that the lack of orthogonality of the steering vectors leads to inter-user interference. Moreover, the proposed compressive approach aims at simultaneously estimating the different UK subchannels without prior knowledge on the sparsity level. This brings about additional difficulty to accurately estimate the actual number of paths (i.e., the rank) of $\boldsymbol{H}[k]$. The estimation of a UN_{p}-sparse vector employing finite-resolution dictionaries leads to a larger grid quantization error

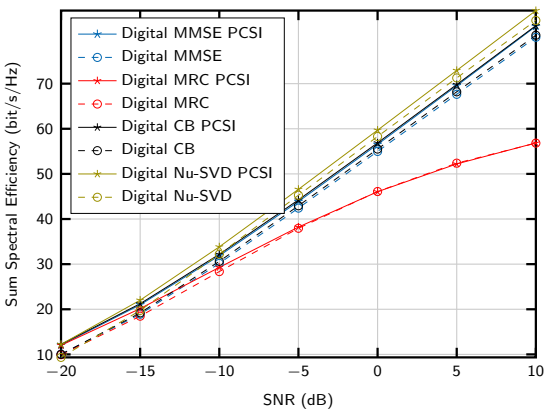


(a)

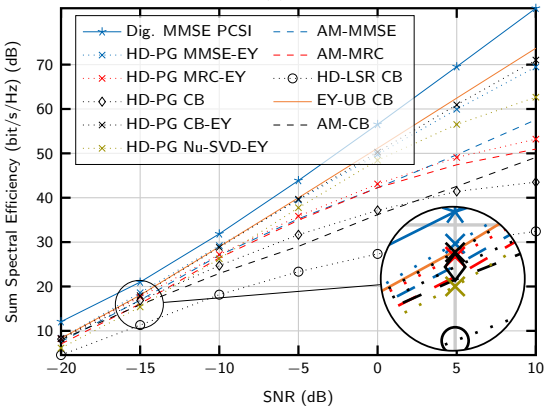


(b)

Fig. 2: (a) Achievable Sum-Rate vs. SNR for error-free channel estimation. Digital MMSE, Digital MRC, and Digital CB represent the digital benchmarks, HD-LSR and HD-PG are the factorizations using [15, Least Squares relaxation] and Alg. 1. Finally, AM MMSE, AM CB, and AM MRC are the methods from Sec. III-B2. (b) Achievable Sum-Rate vs. N# of RF Chains with $L_{\text{BS}} = 2, 4, 8, 12, 16$. We consider $N_{\text{BS}} = 64$, $L_{\text{MS},u} = 4$, and a fixed SNR of 0dB.



(a)



(b)

Fig. 3: (a) Achievable Sum-Rate vs. SNR for perfect and imperfect CSI using digital solutions with off-grid channel model. We include precoders and combiners jointly obtained with the algorithm [33, Nu-SVD] as benchmark. (b) Achievable Sum-Rate vs. SNR with hybrid precoders and combiners. EY-UB CB represents the Eckart-Younglow rank approximation of CB. HD-PG and HD-LSR are the hybrid designs using Alg. 1 and [15, HD-LSR]. MMSE-EY, CB-EY, MRC-EY, Nu-SVD-EY correspond to Eckart-Young low rank approximations for the methods MMSE, CB, maximum ratio combiner MRC, and Nu-SVD. AM are the alternating minimization methods.

compared to the single MS case. Therefore, the estimation of the channel gains and, consequently, the noise variance, is more challenging. As a result, this creates a non-negligible performance gap between perfect and imperfect CSI.

Figure 3b depicts the results obtained with the hybrid approaches. Regarding the hybrid factorization problem, the main difficulty is that the analog UL combiner (DL precoder) must be designed to accommodate the different KN_{s} data streams. The number of available degrees of freedom of this analog precoder is $L_{\text{BS}} \ll KN_{\text{s}}$, whereby it is very hard to design a low-rank hybrid precoder to simultaneously approximate KU different rank-N_{s} matrices. Together with the stringent constraints of the RF matrix, it makes hybrid factorization the system bottleneck.

After stacking the different combiners to perform the hybrid factorization, the rank of the overall combiner grows rapidly with MMSE and Nu-SVD approaches. Since $N_{\text{p}} < \min(N_{\text{BS}}, N_{\text{MS},u})$, $\forall u$, for a given MS, the column spaces of digital MRCs at the different subcarriers are equal to each other, up to a unitary rotation which is invariant in the Grassmannian manifold [47]. Therefore, for MRC the rank of the overall combiner does not sharply grow with the number of subcarriers K, yet it does with the number of users U (see (14)). A similar effect is observed for CB combiners in (15). Due to the limited rank of the inter-user interference matrix, the projection to the nullspace is approximately the identity matrix, i.e., $\boldsymbol{U}_u[k]\boldsymbol{U}_u[k]^{\text{H}} \approx \mathbf{I}_{N_{\text{BS}}}$. By sharp contrast, provided that the inverse of the received covariance matrix in (13) is different for each subcarrier, MMSE combiners destroy the relationship between column spaces of different subcarriers. Therefore, the rank of the overall UL digital combiner grows both with U and K, leading to the noticeable performance gap between the digital and the hybrid MMSE combiners. This effect is even more accentuated in the case of Nu-SVD since

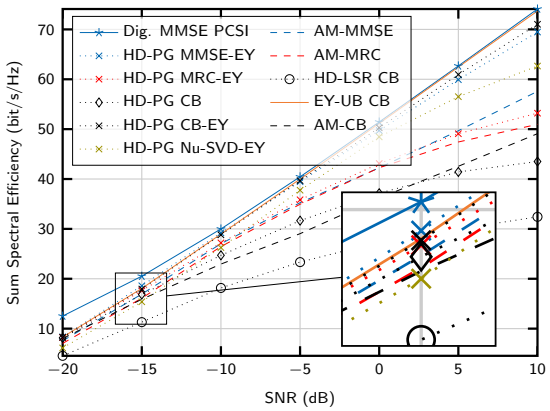


Fig. 4: Achievable Sum-Rate vs. SNR with hybrid precoders and combiners. Compare with Fig. 3b.

the UL precoders are not chosen as the right singular vectors of the channel estimate, as in the proposed digital designs.

The latter discussion indicates the suitability of MRC for the low SNR regime expected at mmWave. However, at low SNR, the diagonal elements of the matrix in (13) are dominated by the noise term, whilst at high SNR the signal term is dominant. Thereby, the lower the SNR, the less significant the alteration of the column space of the MMSE combiner is.

Recall that equivalent overall DL precoders and combiners have to be approximated by matrix products of ranks $L_{\text{BS}} = 8$ and $L_{\text{MS},u} = 2$, respectively. Moreover, minimizing the Frobenius norm leads to near-optimal hybrid approximations [13]. Therefore, by means of the Eckart-Young theorem [49], it is possible to set an upper bound for the performance of solutions based on factorization of digital designs. According to [49], for the digital precoder $\boldsymbol{F}_{\text{d}}$ the error is upper bounded as $\|\boldsymbol{F}_{\text{d}} - \boldsymbol{F}_{\text{EY}}\|_{\text{F}} \leq \|\boldsymbol{F}_{\text{d}} - \boldsymbol{F}_{\text{RF}}\boldsymbol{F}_{\text{BB}}\|_{\text{F}}$ with the SVD $\boldsymbol{F}_{\text{d}} = \boldsymbol{U}_{\boldsymbol{F}}\boldsymbol{\Sigma}_{\boldsymbol{F}}\boldsymbol{V}_{\boldsymbol{F}}^{\text{H}}$, $\boldsymbol{\Sigma}_{\text{EY}}$ containing the largest L_{BS} values in the diagonal of $\boldsymbol{\Sigma}_{\boldsymbol{F}}$, and the optimal L_{BS}-rank approximation $\boldsymbol{F}_{\text{EY}} = \boldsymbol{U}_{\boldsymbol{F}}\boldsymbol{\Sigma}_{\text{EY}}\boldsymbol{V}_{\boldsymbol{F}}^{\text{H}}$. Remarkably, by factorizing $\boldsymbol{F}_{\text{EY}}$ instead of $\boldsymbol{F}_{\text{d}}$, and initializing Alg. 1 to $\boldsymbol{F}_{\text{RF}} = \arg([\boldsymbol{U}_{\boldsymbol{F}}]_{1:N_{\text{BS}},1:L_{\text{BS}}})$, the performance loss of the hybrid solution is mainly caused by the rank limitation and not by the factorization itself. Remember that the initialization is very important due to the local optimality of Alg. 1. The AM approaches provide a good tradeoff among computational complexity and performance. It is very interesting for MRC due to the good throughputs and computational efficiency.

Contrarily to the off-grid channel model, considering on-grid parameters does not result in the aforementioned quantization error, thereby simplifying channel estimation and hybrid precoder and combiner design. This follows from the fact that the probability of several paths sharing the same AoA and AoD in the multi-user channel is non-zero, being one if $UN_{\text{p}} \geq \min(G_{\text{MS}}, G_{\text{BS}})$. Thereby, and jointly with on-grid delay paths, the rank of the multi-user channel does not necessarily increase with U and N_{p}. Consequently, the difference in degrees of freedom between hybrid and digital solutions is reduced. Figure 4 depicts the achievable rates for the case of on-grid AoA/AoD with hybrid precoders and combiners. Observe that both the estimation and factorization errors are negligible, and the overall performance is near-optimal under on-grid AoA/AoD.

VI. Conclusions

This work tackles channel estimation and the design of precoders and combiners in multi-user mmWave MIMO systems. The channel is estimated on the uplink, performed simultaneously for all the MSs and subcarriers, and has been shown to greatly outperform its downlink counterpart for both on-grid and off-grid AoA/AoD. The precoders and combiners are designed according to different optimization criteria, presenting advantages depending on the SNR regime as shown in Table I. Specifically, although the MMSE and CB combiners perform better, the low complexity MRC appears as an interesting alternative for low and mid SNRs. Simulation results exhibit the good performance of the proposed hybrid solutions, and the rank of the digital solutions is shown to be the performance-limiting factor. While the AM approaches offers lower computational complexity, the digital factorization methods are more flexible and present better performance.

Appendix

A. Gradient Computation

To compute the gradient in (22), we use the separation in (20), which yields

$$\frac{\partial d}{\partial \boldsymbol{F}_{\text{RF}}^{*}} = \sum_{u=1}^{U}\sum_{k=1}^{K}\frac{\partial d_{k,l}}{\partial \boldsymbol{F}_{\text{RF}}^{*}}, \tag{48}$$

allowing us to compute the derivative for each MS and subcarrier, as follows

$$\begin{aligned}\frac{\partial d_{k,l}}{\partial \boldsymbol{F}_{\text{RF}}^{*}} &= -\frac{\sqrt{P_{\text{tx}}/(UN_{\text{s},u})}}{\|\boldsymbol{A}_u[k]\|_{\text{F}}}\frac{\partial\|\boldsymbol{A}_u[k]\|_{\text{F}}^2}{\partial \boldsymbol{F}_{\text{RF}}^{*}} \\ &= \frac{\sqrt{P_{\text{tx}}/(UN_{\text{s},u})}}{\|\boldsymbol{A}_u[k]\|_{\text{F}}}\left(\boldsymbol{F}_{\text{RF}}(\boldsymbol{F}_{\text{RF}}^{\text{H}}\boldsymbol{F}_{\text{RF}})^{-1}\boldsymbol{F}_{\text{RF}}^{\text{H}} - \mathbf{I}_{N_{\text{BS}}}\right) \\ &\quad\times \boldsymbol{F}_{\text{MMSE}}^{u}[k]\boldsymbol{F}_{\text{MMSE}}^{u,\text{H}}[k]\boldsymbol{F}_{\text{RF}}(\boldsymbol{F}_{\text{RF}}^{\text{H}}\boldsymbol{F}_{\text{RF}})^{-1}.\end{aligned}$$

Adding the individual contributions for each MS and subcarrier we arrive at (23).

B. Frobenius Norm for MSE

Let us introduce the matrix $\boldsymbol{E}_u[k] = \boldsymbol{F}_{\text{MMSE}}^{u}[k] - \boldsymbol{F}_{\text{RF}}\boldsymbol{F}_{\text{BB}}^{u}[k]$ containing the difference between the unconstrained and the hybrid combiners. Now, we show that minimizing $\|\boldsymbol{E}_u[k]\|_{\text{F}}^2$ leads to a reduction in a lower bound for the MSE in (27)

$$\begin{aligned}\text{MSE}_u^{\text{UL}}[k] &= N_{\text{s},u} - 2\mathfrak{Re}\{\text{tr}\left((\boldsymbol{F}_{\text{MMSE}}^{u,\text{H}}[k] - \boldsymbol{E}_u[k]^{\text{H}})\boldsymbol{R}_D^{\text{UL}}[k]\right)\} \\ &\quad + \text{tr}\left((\boldsymbol{F}_{\text{MMSE}}^{u,\text{H}}[k] - \boldsymbol{E}_u^{\text{H}}[k])\boldsymbol{R}_I^{\text{UL}}[k](\boldsymbol{F}_{\text{MMSE}}^{u}[k] - \boldsymbol{E}_u[k])\right) \\ &= 2\mathfrak{Re}\{\text{tr}\left(\boldsymbol{E}_u^{\text{H}}[k]\boldsymbol{R}_D^{\text{UL}}[k]\right)\} + \text{tr}\left(\boldsymbol{E}_u^{\text{H}}[k]\boldsymbol{R}_I^{\text{UL}}[k]\boldsymbol{E}_u[k]\right) \\ &\quad + \text{MMSE}_u^{\text{UL}}[k] - 2\mathfrak{Re}\{\text{tr}\left(\boldsymbol{E}_u^{\text{H}}[k]\boldsymbol{R}_I^{\text{UL}}[k]\boldsymbol{F}_{\text{MMSE}}^{u}[k]\right)\} \\ &= \text{MMSE}_u^{\text{UL}}[k] + \text{tr}\left(\boldsymbol{E}_u^{\text{H}}[k]\boldsymbol{R}_I^{\text{UL}}[k]\boldsymbol{E}_u[k]\right) \\ &\leq \text{MMSE}_u^{\text{UL}}[k] + \|\boldsymbol{E}_u[k]\|_{\text{F}}^2\|\boldsymbol{R}_I^{\text{UL},1/2}[k]\|_{\text{F}}^2.\end{aligned} \tag{49}$$

TABLE I: Performance of different approaches

	Digital			HD-PG			AM		
	MMSE	CB	MRC	MMSE-EY	CB-EY	MRC-EY	MMSE	CB	MRC
Low-Mid SNR	Good	Good	Good	Good	Good	Good	Good	Poor	Good
Mid-High SNR	Good	Good	Mid	Good	Good	Mid	Mid	Poor	Mid

REFERENCES

[1] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, "MIMO Precoding and Combining Solutions for Millimeter-Wave Systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, December 2014.

[2] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[3] J. Lee, G. T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 3326–3331.

[4] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[5] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath Jr., "Frequency-domain Compressive Channel Estimation for Frequency-selective Hybrid MIMO Systems at Millimeter Wave," *submitted to IEEE Transactions on Wireless Communications, available at arxiv*.

[6] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath Jr., "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, September 2017.

[7] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel Estimation for Millimeter-Wave Multiuser MIMO Systems via PARAFAC Decomposition," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7501–7516, Nov 2016.

[8] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-User Precoding and Channel Estimation for Hybrid Millimeter Wave Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1576–1590, July 2017.

[9] Q. Wang, Z. Zhou, J. Fang, and Z. Chen, "Compressive channel estimation for millimeter wave multiuser MIMO systems via pilot reuse," in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, Oct 2016, pp. 1–6.

[10] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding Over Frequency-Selective Fading Channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, June 2016.

[11] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, and L. Castedo, "Channel Estimation and Hybrid Precoding/Combining for Frequency Selective Multiuser mmWave Systems," *Proc. Global Communications Conference (GLOBECOM)*, December 2017.

[12] X. Zhang, A. F. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, November 2005.

[13] 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.

[14] R. Méndez-Rial, C. Rusu, N. González-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmWave MIMO systems," in *Proc. Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 151–155.

[15] C. Rusu, R. Méndez-Rial, N. González-Prelcic, and R. W. Heath, "Low Complexity Hybrid Precoding Strategies for Millimeter Wave MIMO Communication Systems," *IEEE Transactions on Wireless Communications*, 2016.

[16] S. Park, A. Alkhateeb, and R. W. Heath Jr., "Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems," *arXiv:1606.08405*, vol. abs/1606.08405, 2016.

[17] 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, November 2015.

[18] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical csit," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7025–7038, November 2017.

[19] 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.

[20] 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, May 2016.

[21] T. E. Bogale, L. B. Le, A. Haghighat, 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.

[22] X. Cheng, M. Wang, and S. Li, "Compressive Sensing-Based Beamforming for Millimeter-Wave OFDM Systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 371–386, Jan 2017.

[23] J. P. González-Coma, N. González-Prelcic, L. Castedo, and R. W. Heath, "Frequency selective multiuser hybrid precoding for mmWave systems with imperfect channel knowledge," in *Asilomar Conference on Signals, Systems and Computers*, November 2016, pp. 291–295.

[24] Jose Flordelis, Fredrik Rusek, Fredrik Tufvesson, Erik G. Larsson, and Ove Edfors, "Massive MIMO Performance - TDD Versus FDD: What Do Measurements Say?," *CoRR*, vol. abs/1704.00623, 2017.

[25] I. Ahmed, H. Khammari, and A. Shahid, "Resource Allocation for Transmit Hybrid Beamforming in Decoupled Millimeter Wave Multiuser-MIMO Downlink," *IEEE Access*, vol. 5, pp. 170–182, 2017.

[26] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid Beamforming for Large Antenna Arrays With Phase Shifter Selection," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7258–7271, November 2016.

[27] R. A. Stirling-Gallacher and M. S. Rahman, "Multi-user MIMO strategies for a millimeter wave communication system using hybrid beam-forming," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 2437–2443.

[28] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[29] S. Haghighatshoar and G. Caire, "Massive MIMO Channel Subspace Estimation From Low-Dimensional Projections," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, January 2017.

[30] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.

[31] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[32] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.

[33] Z. Pan, K. Wong, and T. Ng, "Generalized multiuser orthogonal space-division multiplexing," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1969–1973, November 2004.

[34] R. Chen, Z. Shen, J. G. Andrews, and R. W. Heath, "Multimode Transmission for Multiuser MIMO Systems With Block Diagonalization," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3294–3302, July 2008.

[35] T.E. Bogale and L. Vandendorpe, "Robust Sum MSE Optimization for Downlink Multiuser MIMO Systems With Arbitrary Power Constraint: Generalized Duality Approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1862–1875, April 2012.

[36] J. González-Coma, M. Joham, P. Castro, and L. Castedo, "QoS constrained power minimization in the MISO broadcast channel with imperfect CSI," *Signal Processing*, vol. 131, pp. 447 – 455, February 2017.

[37] J. P. González-Coma, A. Gründinger, M. Joham, and L. Castedo, "MSE Balancing in the MIMO BC: Unequal Targets and Probabilistic Interference Constraints," *IEEE Transactions on Signal Processing*, vol. 65, no. 12, pp. 3293–3305, June 2017.

[38] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing Sum Rate and Minimizing MSE on Multiuser Downlink: Optimality, Fast Algorithms and Equivalence via Max-min SINR," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6127–6143, December 2011.

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

[40] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020–4030, August 2008.

[41] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[42] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, Technology press books in science and engineering. Technology Press of the Massachusetts Institute of Technology, 1964.

[43] V. Venkateswaran and A. J. van der Veen, "Analog Beamforming in MIMO Communications With Phase Shift Networks and Online Channel Estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, August 2010.

[44] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal Multiuser Transmit Beamforming: A Difficult Problem with a Simple Solution Structure [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, July 2014.

[45] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *International Conference on Communications (ICC)*, June 2014, pp. 5838–5842.

[46] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, Massachusetts, 1999.

[47] K. Venugopal, N. González-Prelcic, and R. W. Heath Jr., "Optimality of Frequency Flat Precoding in Frequency Selective Millimeter Wave Channels," *IEEE Wireless Communications Letters*, vol. PP, no. 99, pp. 1–1, 2017.

[48] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, 1993.

[49] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.