IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2021, accepted March 6, 2021, date of publication March 10, 2021, date of current version March 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065581

Wideband User Grouping for Uplink Multiuser mmWave MIMO Systems With Hybrid Combining

DARIAN PÉREZ-ADÁN[1], ÓSCAR FRESNEDO[1], (Member, IEEE), JOSÉ P. GONZÁLEZ-COMA[2], (Member, IEEE), AND LUIS CASTEDO[1], (Senior Member, IEEE)

[1]Department of Computer Engineering & Center for Research in Information and Communication Technologies (CITIC), University of A Coruña, 15071 A Coruña, Spain

[2]Defense University Center, Spanish Naval Academy, 36920 Marín, Spain

Corresponding author: Darian Pérez-Adán (d.adan@udc.es)

This work was supported in part by the Xunta de Galicia under Grant ED431C 2020/15, in part by the Centro de Investigación de Galicia CITIC under Grant ED431G2019/01, in part by the Agencia Estatal de Investigación of Spain under Grant RED2018-102668-T and Grant PID2019-104958RB-C42, in part by the European Regional Development Funds (ERDF) of the EU (ERDF Galicia 2014-2020 & AEI/ERDF programs, UE), and in part by the Predoctoral Grant BES-2017-081955.

ABSTRACT Analog-digital hybrid precoding and combining schemes constitute an interesting approach to millimeter-wave (mmWave) multiple-input multiple-output (MIMO) systems due to the low hardware complexity and/or low power required for its deployment. However, the design of the hybrid precoders and combiners of a wideband multiuser (MU) mmWave MIMO system is challenging because the signal processing in the analog domain is constrained to be frequency flat. Furthermore, the number of radio frequency (RF) chains limits the number of individual streams that a common base station (BS) can simultaneously serve. This work jointly addresses the user scheduling, the user precoder design, and the BS hybrid combining design for the uplink of wideband MU mmWave MIMO systems. On the one hand, user precoding and BS hybrid combining are jointly designed to minimize the impact of having frequency-flat RF components. On the other hand, a number of users larger than the number of RF chains are served at the BS by employing a distributed quantizer linear coding (DQLC)-based non-orthogonal multiple access (NOMA) scheme. The use of this encoding strategy also allows exploiting the spatial correlation between the source information. Simulation results show remarkable performance gains of the proposed approaches for wideband mmWave MIMO hardware-constrained systems.

INDEX TERMS User scheduling, wideband mmWave, multiuser communications, non-orthogonal multiple access, joint source-channel coding.

I. INTRODUCTION

It is expected that mmWave MIMO systems will accomplish the required capacity increase for the next generations of wireless communications systems [1]–[3]. Hybrid analog-digital architectures for precoding and combining have been proposed to provide a suitable trade-off between the power consumption and the flexibility of fully digital solutions while providing high beamforming gains [4]. Hybrid transceiver architectures are based on decoupling the signal processing for precoding and combining into analog and baseband domains to reduce the number of RF chains [3], [5].

Designing hybrid analog-digital schemes is particularly critical for the MU uplink where a common combiner is employed at the BS to receive all the K_s served user streams. In addition, MU communication systems nowadays demand the feasibility of handling simultaneously a huge amount of nodes (e.g., wireless sensor networks (WSNs) or Internet of things (IoT) systems) and this would lead to the need of using a large number of RF chains at the BS (N_{RF}^{r}), since at least one RF chain is necessary to receive each data stream [3]. For this reason, we have explored in [6] a new user grouping approach and hybrid system design for practical narrowband mmWave MIMO systems having fewer RF chains than user streams (i.e., $N_{RF}^{r} \leq K_s$). This scheme, by properly canceling the inter-group interference, is able to deal with the hardware and power limitations by leveraging a NOMA strategy based on the use of DQLC. DQLC mappings are also able to exploit the spatial correlation between the sources [7], [8], which is relevant in scenarios like WSNs or IoT where the sources usually produce correlated information. However,

The associate editor coordinating the review of this manuscript and approving it for publication was Zesong Fei.

DQLC-based schemes are particularly sensitive to interferences and for this reason, it is fundamental to properly design precoders and combiners in the uplink to successfully eliminate the MU interference. Several works have been developed by assuming single user schemes in wideband scenarios [9]–[13]. However, in MU wideband scenarios, the design of hybrid precoders and combiners to manage the MU interference becomes more challenging since the common RF combiner is frequency flat and, hence, the analog combiner has to be jointly designed for all users and subcarriers. Moreover, the large channel bandwidth leads to the beam direction of the steering vectors, an eventuality known as beam squint effect [14]. This effect also needs to be considered in order to effectively remove the MU interference.

Limited work has been carried out for strongly hardware-constrained hybrid precoding and combining in MU wideband scenarios [15]–[19]. The authors in [15] and [16] considered MU setups having more than one phase shifter (PS) for each RF chain connection to a single antenna, thus leading to more complex architectures. The authors in [17] develop a projected gradient (PG) algorithm to leverage the common structure of the channel response matrices among different subcarriers in wideband scenarios. In [18] a hybrid MU equalizer with dynamic subarray antennas for the uplink of single-carrier frequency-division multiple access (SC-FDMA) systems is proposed. The equalizer design is based on two steps by minimizing the mean square error (MSE) of all the subcarriers: the digital equalizers are iteratively designed on a per subcarrier basis, whereas the analog equalizer is fixed over the subcarriers and iterations. A dynamic antenna mapping was derived for the analog combiner in order to connect the best set of antennas and the phase shifter to each RF chain. The authors in [19] focus on the hybrid precoding design in a downlink mmWave MU scheme. A two-step scheme was developed to cancel the inter-user interference by considering infinite angular resolution and a non-uniformly spaced quantization codebook based analog precoder with finite angular resolution. Furthermore, the authors proposed a phase compensation operation to alleviate the beam squint effect.

All these works assume that the number of RF chains is at least equal to the number of individual data streams to be collected, i.e., $N_{\mathrm{RF}} \geq K_{\mathrm{s}}$, and therefore the developed algorithms are not suitable for NOMA schemes [20] with user grouping where $N_{\mathrm{RF}} < K_{\mathrm{s}}$. Moreover, most of these works do not consider the beam squint effect and its critical impact on the cancellation of interferences.

A. CONTRIBUTIONS

In this work, we propose a novel wideband hybrid combining design, with its corresponding user grouping and scheduling scheme, for the uplink of mmWave MIMO systems where a strong hardware constraint at the BS is considered ($N^{\mathrm{r}}_{\mathrm{RF}} < K_{\mathrm{s}}$). This corresponds to a scenario where more streams than available RF chains at the BS could be allocated for each subcarrier. The proposed transmission scheme is specifically designed to deal with the difficulties described above when considering wideband scenarios. In particular, the main contributions of this work are the following:

- The design of a novel approach for the interference cancellation at both ends of the communication link considering the hardware constraints imposed by the hybrid architecture at the BS in wideband scenarios. The proposed algorithm jointly configures the user precoders and the BS hybrid combiner to minimize the inter-group interference in the user grouping NOMA scheme.
- The development of an extended version of the factorization algorithm proposed in [6] to decouple the BS filter which considers the hardware constraints of hybrid architectures in wideband mmWave MIMO systems.
- The design of two user grouping and scheduling strategies for wideband scenarios that consider the beam squint effect present in wideband systems. One of them also considers the hardware constraints related to hybrid architectures in wideband mmWave MIMO systems.

B. ORGANIZATION

The remainder of the paper is structured as follows. The system model is detailed in Section II. The joint design of the hybrid combining and the scheduling algorithm to handle the user grouping deployment in wideband mmWave MIMO systems is described in Section III. A novel algorithm for the joint computation of the user precoders and the BS combiner is approached in Section IV. The computational complexity of the proposed algorithms is addressed in Section V. Finally, we present simulation results and comparisons in Section VI, while Section VII is devoted to present the conclusions.

C. NOTATION

The following notation is employed throughout this paper: a is a scalar, $\mathbf{a}$ stands for a vector, $[\mathbf{A}]_{i,j}$ is the entry on the i-th row and the j-th column of the matrix $\mathbf{A}$ and $[\mathbf{A}]_{i,:}$ represents the i-th row of $\mathbf{A}$. Transpose and conjugate transpose of $\mathbf{A}$ are $\mathbf{A}^T$ and $\mathbf{A}^*$, respectively. $\|\mathbf{A}\|_F$ denotes the Frobenius norm of $\mathbf{A}$ and $\mathbf{A}^\dagger$ its pseudoinverse. The operator blkdiag $(\cdot)$ constructs a block diagonal matrix from the input matrices and span $(\cdot)$ is the operator that generates an orthonormal basis for the range of the input matrix. Calligraphic letters are used to denote sets and sequences being $|\mathcal{A}|$ the cardinality of $\mathcal{A}$. Finally, $\otimes$ represents the Kronecker product and $\mathbb{E}[\cdot]$ stands for the expectation operator.

II. SYSTEM MODEL

Let us consider an uplink MU wireless scenario with K users communicating with a common BS equipped with $N^{\mathrm{r}}_{\mathrm{RF}}$ RF chains and N_{r} antennas as shown in Figure 1. To fully exploit the large bandwidths available at mmWave frequencies, users send wideband signals. Consequently, these signals are affected by a frequency selective channel that is decomposed into L parallel subchannels thanks to the use of the orthogonal frequency-division multiplexing (OFDM) modulation. The resulting L symbols associated

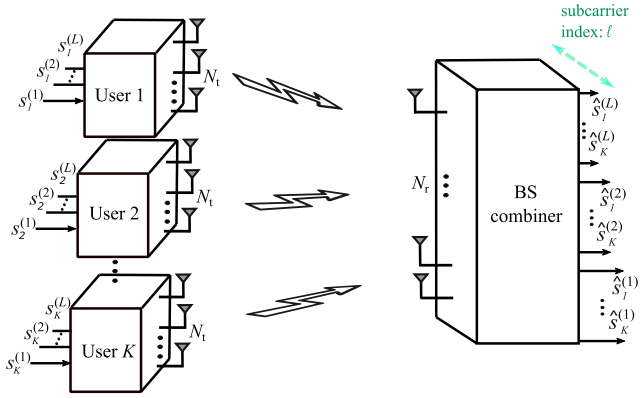

FIGURE 1. System model of an uplink MU wideband mmWave MIMO system with K users and L subcarriers.

to the L subcarriers are transmitted by the users employing a cyclic prefix long enough to avoid inter-carrier interference (ICI) and inter-symbol interference (ISI). Although single-carrier schemes present low peak-to-average-power-ratio (PAPR), OFDM provides greater flexibility for scheduling than single-carrier schemes. Moreover, the PAPR can also be reduced substantially by employing several strategies with a limited impact on the system performance [21].

We also consider that the users send a single data stream to the common BS per channel use and subcarrier. At each time instant, $K_{\mathrm{s}}^{(\ell)} < K$ users are served while $K_{\mathrm{I}}^{(\ell)}$ users remain idle at the ℓ-th subcarrier. The set $\mathcal{K}$ contains the subset of active users $\mathcal{K}_{\mathrm{s}}^{(\ell)}$ and the subset of idle users $\mathcal{K}_{\mathrm{I}}^{(\ell)}$ such that $\mathcal{K} = \mathcal{K}_{\mathrm{s}}^{(\ell)} \cup \mathcal{K}_{\mathrm{I}}^{(\ell)}$, $\forall\ \ell$. The number of served and idle users is given by the cardinality of the sets, i.e., $K_{\mathrm{s}}^{(\ell)} = |\mathcal{K}_{\mathrm{s}}^{(\ell)}|$ and $K_{\mathrm{I}}^{(\ell)} = |\mathcal{K}_{\mathrm{I}}^{(\ell)}|$, respectively. Finally, the set $\mathcal{K}$ will contain all the available users in the system, i.e., $|\mathcal{K}| = K = K_{\mathrm{s}}^{(\ell)} + K_{\mathrm{I}}^{(\ell)}, \forall\ \ell$.

We follow an approximation where the users are gathered in $G^{(\ell)}$ groups by means of an appropriate scheduling algorithm such that $G_{\mathrm{s}} \leq G^{(\ell)}$ groups contain the served users, whereas $G_{\mathrm{I}}^{(\ell)}$ is the number of groups with idle users at the ℓ-th subcarrier. We establish that the number of served groups is the same for the whole bandwidth, i.e., $G_{\mathrm{s}} = N_{\mathrm{RF}}^{\mathrm{r}}$, whereas the total number of groups $G^{(\ell)}$ is subcarrier dependent and changes according to the considered scheduling algorithm. Denoting $\mathcal{G}_i^{(\ell)}$ as the i-th group at ℓ-th subcarrier, equalities $\mathcal{K}_{\mathrm{s}}^{(\ell)} = \cup_{i=1}^{G_{\mathrm{s}}} \mathcal{G}_i^{(\ell)}$ and $\mathcal{K}_{\mathrm{I}}^{(\ell)} = \cup_{i=1}^{G_{\mathrm{I}}^{(\ell)}} \mathcal{G}_i^{(\ell)}$ hold $\forall\ \ell = 1, \ldots, L$.

For convenience, we also introduce the vector $\mathbf{g}^{(\ell)}$ which contains the number of users per group, where each component $g_i^{(\ell)} = |\mathcal{G}_i^{(\ell)}|$ represents the number of users in the i-th group such that $\sum_{i=1}^{G_{\mathrm{s}}} g_i^{(\ell)} = K_{\mathrm{s}}^{(\ell)}$. Note that, although the number of served groups is fixed throughout the whole bandwidth, the number of collected user symbols per subcarrier at the BS, $K_{\mathrm{s}}^{(\ell)}$, is variable since it depends on the specific composition of the served groups.

We assume that each user has N_{t} antennas to send a single data stream of discrete-time continuous-amplitude symbols to the BS. The vector of the K user symbols at subcarrier ℓ is $\mathbf{s}^{(\ell)} = \left[s_1^{(\ell)}, s_2^{(\ell)}, \ldots, s_K^{(\ell)}\right]^T$ which follows a zero-mean spatially correlated multivariate complex-valued Gaussian distribution with covariance matrix $\mathbf{C}_{\mathrm{s}}^{(\ell)} = \mathbb{E}[\mathbf{s}^{(\ell)}\mathbf{s}^{*(\ell)}]$, such that $[\mathbf{C}_{\mathrm{s}}^{(\ell)}]_{k,k} = 1,\ \forall\ k$, and $[\mathbf{C}_{\mathrm{s}}^{(\ell)}]_{i,j} = \rho_{i,j}, 0 \leq \rho_{i,j} \leq 1$, $\forall\ i, j$ with $i \neq j$.

The source symbols corresponding to the served users are individually encoded by means of the mapping functions $f_{i,j}^{(\ell)}(s_{i,j}^{(\ell)})$, $\forall\ i = 1, \ldots, G_{\mathrm{s}}$, $\forall\ j = 1, \ldots, g_i^{(\ell)}$, $\forall\ \ell = 1, \ldots, L$, where $f_{i,j}^{(\ell)}(\cdot)$ corresponds to the mapping function applied to the symbols of the j-th user in the i-th group at the ℓ-th subcarrier. We assume that the encoded symbols satisfy the inequality constraint $\mathbb{E}\left[\left|f_{i,j}^{(\ell)}\left(s_{i,j}^{(\ell)}\right)\right|^2\right] \leq 1$, $\forall\ i, j, \ell$. All the encoded users symbols at subcarrier ℓ are stacked in the following vector

$$\mathbf{f}^{(\ell)}(\mathbf{s}^{(\ell)}) = \left[f_{1,1}^{(\ell)}\left(s_{1,1}^{(\ell)}\right), \ldots, f_{1,g_1^{(\ell)}}^{(\ell)}\left(s_{1,g_1^{(\ell)}}^{(\ell)}\right), \ldots, f_{G_{\mathrm{s}},g_{G_{\mathrm{s}}}^{(\ell)}}^{(\ell)}\left(s_{G_{\mathrm{s}},g_{G_{\mathrm{s}}}^{(\ell)}}^{(\ell)}\right)\right]^T, \quad (1)$$

Next, the encoded symbols are transformed by the linear precoder $\mathbf{p}_{i,j}^{(\ell)} \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ and then transmitted over the uplink channel. An individual per-subcarrier power constraint is considered at each user as in [17], such that $\|\mathbf{p}_{i,j}^{(\ell)}\|_F^2 \leq T_{i,j},\ \forall\ i, j, \ell$ with $T_{i,j} = T'_{i,j}/L$, being $T'_{i,j}$ the total power constraint per user. In this paper, digital precoding is assumed at the users as in [22]. Therefore, we can consider that the number of RF chains required at the users is equal to the number of transmit antennas, i.e., $N_{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{t}}$ since N_{t} is often a small number.

The received signal from the active users at the ℓ-th subcarrier reads as

$$\mathbf{y}^{(\ell)} = \sum_{i=1}^{G_{\mathrm{s}}} \sum_{j=1}^{g_i^{(\ell)}} \mathbf{H}_{i,j}^{(\ell)} \mathbf{p}_{i,j}^{(\ell)} f_{i,j}^{(\ell)}(s_{i,j}^{(\ell)}) + \mathbf{n}^{(\ell)}, \quad (2)$$

where $\mathbf{H}_{i,j}^{(\ell)} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ is the mmWave channel response corresponding to subcarrier ℓ of the j-th user in the i-th group, and $\mathbf{n}^{(\ell)} = \left[n_1, n_2, \ldots, n_{N_{\mathrm{r}}}\right]^T$ represents the complex-valued additive white Gaussian noise (AWGN) such that $\mathbf{n}^{(\ell)} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{n}}^2 \mathbf{I})$. The received signal at subcarrier ℓ can also be rewritten in a more compact way as

$$\mathbf{y}^{(\ell)} = \mathbf{H}^{(\ell)} \mathbf{P}^{(\ell)} \mathbf{f}^{(\ell)}(\mathbf{s}^{(\ell)}) + \mathbf{n}^{(\ell)}, \quad (3)$$

where

$$\mathbf{H}^{(\ell)} = \left[\mathbf{H}_{1,1}^{(\ell)}, \ldots, \mathbf{H}_{G_{\mathrm{s}},g_{G_{\mathrm{s}}}^{(\ell)}}^{(\ell)}\right] \quad (4)$$

and

$$\mathbf{P}^{(\ell)} = \mathrm{blkdiag}\left(\mathbf{p}_{1,1}^{(\ell)}, \ldots, \mathbf{p}_{G_{\mathrm{s}},g_{G_{\mathrm{s}}}^{(\ell)}}^{(\ell)}\right). \quad (5)$$

At the receiver, a BS hybrid combiner $\mathbf{W}_{\mathrm{H}}^{(\ell)}\ \forall\ \ell = 1, \ldots, L$, is implemented with $N_{\mathrm{RF}}^{\mathrm{r}}$ RF chains to fulfill the task of

decoupling the uplink received signal $\mathbf{y}^{(\ell)}$ as

$$\mathbf{z}^{(\ell)} = \mathbf{W}_{\mathrm{H}}^{*(\ell)}\mathbf{y}^{(\ell)}, \tag{6}$$

where $\mathbf{W}_{\mathrm{H}}^{*(\ell)} = \mathbf{W}_{\mathrm{BB}}^{*(\ell)}\mathbf{W}_{\mathrm{RF}}^{*}$, with $\mathbf{W}_{\mathrm{BB}}^{(\ell)} \in \mathbb{C}^{N_{\mathrm{RF}}^{\mathrm{r}} \times G_{\mathrm{s}}}$ the baseband combiner for the ℓ-th subcarrier and $\mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}$ the frequency-flat RF combiner. The term $\mathcal{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}^{\mathrm{r}}}$ represents the set of feasible RF combiners imposing a modulus constraint at each entry. The hybrid combiner aims at canceling the interference between the different user groups such that each entry of the vector $\mathbf{z}^{(\ell)}$ only contains the combination of the symbols transmitted by the users in the corresponding served group. Note that the dimension of the vectors of filtered symbols at each subcarrier is actually $\mathbf{z}^{(\ell)} \in \mathbb{C}^{\,G_{\mathrm{s}} \times 1}$, $\forall\, \ell$.

Then, the set of decoding functions $\mathbf{q}^{(\ell)}(\mathbf{z}^{(\ell)}) = \left[q_1^{(\ell)}(z_1^{(\ell)}), \ldots, q_{G_{\mathrm{s}}}^{(\ell)}(z_{G_{\mathrm{s}}}^{(\ell)})\right]^T$ provides an estimate of the source symbols for the served users at subcarrier ℓ as

$$\hat{\mathbf{s}^{(\ell)}}_i = q_i^{(\ell)}(z_i^{(\ell)}) = \left[\hat{s}_{i,1}^{(\ell)}, \ldots, \hat{s}_{i,g_i^{(\ell)}}^{(\ell)}\right]^T, \quad \forall\, i = 1, \ldots, G_{\mathrm{s}}.$$

Since we are considering the transmission of analog complex-valued source symbols, the performance of the communication system is measured in terms of the mean square error (MSE) between the source and the estimated symbols. Table 1 summarizes the main system model parameters considered in this paper.

A. CHANNEL MODEL

The channel response matrix which describes the link between the BS and the j-th user in the i-th group at the m-th delay tap with $m \in \{0, \ldots, L_{\mathrm{D}} - 1\}$, being L_{D} the maximum number of delay taps, is defined by [23]–[25]

$$\mathbf{H}_{i,j}^{(m)} = \gamma \sum_{n=1}^{N_{\mathrm{p}_{i,j}}} \beta_{i,j,n} p_{\mathrm{rc}}\left(mT_s - \tau_{i,j,n}\right) \mathbf{a}_{\mathrm{BS}}^{(\ell)}(\phi_{i,j,n}^{\mathrm{BS}}) \mathbf{a}_{\mathrm{t}_{i,j}}^{*(\ell)}(\phi_{i,j,n}^{\mathrm{t}}), \tag{7}$$

where $p_{\mathrm{rc}}(t)$ stands for the raised cosine pulse-shaping filter, $\tau_{i,j,n}$ is the relative delay for the n-th path, T_s is the sampling period, $\gamma = \sqrt{N_{\mathrm{t}} N_{\mathrm{r}} / N_{\mathrm{p}_{i,j}}}$ is a power normalization factor and $\beta_{i,j,n}$ represents the complex path gain for the n-th path. The term ϕ^{t} stands for the azimuth angles of departure (AoD) at each transmitter and ϕ^{BS} are the azimuth angles of arrival (AoA) at the BS.

In the frequency domain, the channel response (7) for the link between the BS and the j-th user in the i-th group can be represented as [3]

$$\begin{aligned} \mathbf{H}_{i,j}^{(\ell)} &= \sum_{m=0}^{L_{\mathrm{D}}-1} \mathbf{H}_{i,j}^{(m)} e^{j2\pi m(\ell-1)/L} \\ &= \sum_{n=1}^{N_{\mathrm{p}_{i,j}}} \beta_{i,j,n}^{(\ell)} \mathbf{a}_{\mathrm{BS}}^{(\ell)}(\phi_{i,j,n}^{\mathrm{BS}}) \mathbf{a}_{\mathrm{t}_{i,j}}^{*(\ell)}(\phi_{i,j,n}^{\mathrm{t}}), \end{aligned} \tag{8}$$

where $\ell \in \{1, \ldots, L\}$. The path gain $\beta_{i,j,n}^{(\ell)}$ is obtained as $\beta_{i,j,n}^{(\ell)} = \gamma \beta_{i,j,n} \sum_{m=0}^{L_{\mathrm{D}}-1} p_{\mathrm{rc}}(mT_{\mathrm{s}} - \tau_{i,j,n}) e^{j2\pi m(\ell-1)/L}$. The notation $\mathbf{a}_{\mathrm{t}_{i,j}}^{*(\ell)}(\phi_{i,j,n}^{\mathrm{t}})$ and $\mathbf{a}_{\mathrm{BS}}^{(\ell)}(\phi_{i,j,n}^{\mathrm{BS}})$ is introduced to highlight that the array response vectors are affected by the beam squint effect present when considering large bandwidths [25]–[27]. Since we are assuming uniform linear arrays (ULAs) at both ends, the array response vectors $\mathbf{a}_{\mathrm{t}}^{(\ell)}(\phi^{\mathrm{t}})$ and $\mathbf{a}_{\mathrm{BS}}^{(\ell)}(\phi^{\mathrm{BS}})$ are given by [28], [29]

$$\mathbf{a}^{(\ell)}(\phi) = \frac{1}{\sqrt{N}}\left[1, e^{j\frac{2\pi}{f_\ell} c\, d\, \sin\phi}, \ldots, e^{j\frac{2\pi}{f_\ell} c\, d(N-1)\, \sin\phi}\right]^T, \tag{9}$$

where N is the number of antennas in the array, f_ℓ is the frequency at the ℓ-th subcarrier, c is the speed of light, and d is the inter antenna spacing which is often equal to $\lambda/2$, being $\lambda = \frac{c}{f_{\mathrm{c}}}$ the wavelength and f_{c} the carrier frequency.

TABLE 1. System model parameters.

Parameter	Setting
Number of users	K
Number of antennas per users	N_{t}
Number of subcarriers	L
Vector of the K user symbols at subcarrier ℓ	$\mathbf{s}^{(\ell)}$
Correlation matrix (user symbols at subcarrier ℓ)	$\mathbf{C}_{\mathrm{s}}^{(\ell)}$
Mapping function: j-th user in the i-th group at subcarrier ℓ	$f_{i,j}^{(\ell)}(\cdot)$
Precoder: j-th user in the i-th group at subcarrier ℓ	$\mathbf{p}_{i,j}^{(\ell)}$
Number of RF chains at the users	$N_{\mathrm{RF}}^{\mathrm{t}}$
Total power constraint per user	$T'_{i,j}$
Vector of AWGN at subcarrier ℓ	$\mathbf{n}^{(\ell)}$
Subset of active users at subcarrier ℓ	$\mathcal{K}_{\mathrm{s}}^{(\ell)}$
Subset of idle users at subcarrier ℓ	$\mathcal{K}_{\mathrm{I}}^{(\ell)}$
Number of served user groups for all the subcarriers	G_{s}
Served users at subcarrier ℓ	$K_{\mathrm{s}}^{(\ell)}$
Idled users at subcarrier ℓ	$K_{\mathrm{I}}^{(\ell)}$
i-th group of users at ℓ-th subcarrier	$\mathcal{G}_i^{(\ell)}$
Vector with the number of users per group at subcarrier ℓ	$\mathbf{g}^{(\ell)}$
Number of antennas at the BS	N_{r}
Number of RF chains at the BS	$N_{\mathrm{RF}}^{\mathrm{r}}$
BS hybrid combiner at subcarrier ℓ	$\mathbf{W}_{\mathrm{H}}^{(\ell)}$
Baseband combiner at subcarrier ℓ	$\mathbf{W}_{\mathrm{BB}}^{(\ell)}$
Frequency-flat RF combiner	$\mathbf{W}_{\mathrm{RF}}$
Post-combining group symbols at subcarrier ℓ	$\mathbf{z}^{(\ell)}$
Decoding functions at subcarrier ℓ	$\mathbf{q}^{(\ell)}(\mathbf{z}^{(\ell)})$
Estimated served user symbols at subcarrier ℓ	$\hat{\mathbf{s}}^{(\ell)}$

III. USER GROUPING

In [6] we proposed a NOMA scheme based on the use of DQLC mappings which enables to serve more than one user over the same RF chain for narrowband scenarios. This approach requires the design of an adequate user grouping strategy to balance the trade-off between the system performance and the number of served users. We aim at extending this idea to wideband scenarios which leads to significant changes in the design of the user precoder and the BS hybrid combiner, as well as in the scheduling algorithm.

A. NON-ORTHOGONAL MAPPINGS AND DIGITAL DESIGN

We assume that non-orthogonal DQLC mappings are individually applied into the $G_{\mathrm{s}} \leq G^{(\ell)}$ served groups. The implementation of DQLC per group establishes that $g_i^{(\ell)} - 1$ users of the i-th group at the ℓ-th subcarrier transmit a quantized version of their symbols while the symbol of the remaining user is just scaled by a power factor ($\alpha_{i,j}^{(\ell)}$) so that it can be placed between two quantization steps of the encoded user symbols. Therefore, the DQLC mapping function for the i-th group at the ℓ-th subcarrier is mathematically stated as [7]

$$f_{i,j}^{(\ell)}\left(s_{i,j}^{(\ell)}\right) = \begin{cases} \alpha_{i,j}^{(\ell)} \left\lfloor \dfrac{s_{i,j}^{(\ell)}}{\Delta_{i,j}^{(\ell)}} - \dfrac{1}{2} \right\rceil + \dfrac{1}{2}, & j < g_i^{(\ell)} \\ \alpha_{i,j}^{(\ell)} s_{i,j}^{(\ell)}, & j = g_i^{(\ell)} \end{cases}, \quad (10)$$

where $\alpha_{i,j}^{(\ell)}$ represents the scale factor and $\Delta_{i,j}^{(\ell)}$ represents the quantization step employed by the j-th user in the i-th group at the ℓ-th frequency sub-band. The DQLC parameters are adjusted according to the signal-to-interference-plus-noise ratio (SINR) in a way similar to the narrowband scenario approached in [6]. Here, we consider that the channel symbols at the BS for the ℓ-th subcarrier are given by

$$\mathbf{z}^{(\ell)} = \mathbf{R}^{(\ell)}\mathbf{f}^{(\ell)}(\mathbf{s}^{(\ell)}) + \tilde{\mathbf{n}}^{(\ell)}, \quad (11)$$

where

$$\mathbf{R}^{(\ell)} = \mathbf{W}_{\mathrm{H}}^{*(\ell)}\mathbf{H}^{(\ell)}\mathbf{P}^{(\ell)} \quad (12)$$

represents the equivalent channel response for the encoded symbols after the combining process and $\tilde{\mathbf{n}}^{(\ell)} = \mathbf{W}_{\mathrm{H}}^{*(\ell)}\mathbf{n}^{(\ell)}$ is the equivalent noise. Note that the SINR corresponding to the ℓ-th subcarrier of the uplink signal can be obtained directly from the equivalent channel matrix $\mathbf{R}^{(\ell)} \in \mathbb{C}^{G_{\mathrm{s}} \times K_{\mathrm{s}}^{(\ell)}}$ and the noise power. The vector of filtered symbols can be decomposed into its individual components as

$$\begin{aligned} z_i^{(\ell)} &= \sum_{j=1}^{g_i^{(\ell)}} [\mathbf{R}^{(\ell)}]_{i,[b(i)+j]} f_{i,j}^{(\ell)}(s_{i,j}^{(\ell)}) \\ &\quad + \sum_{r \neq i}^{G_{\mathrm{s}}} \sum_{j}^{g_r^{(\ell)}} [\mathbf{R}^{(\ell)}]_{i,[b(r)+j]} f_{r,j}^{(\ell)}(s_{r,j}^{(\ell)}) + \tilde{\mathbf{n}}_i^{(\ell)}, \ \forall\, i = 1, \ldots, G_{\mathrm{s}}, \end{aligned} \quad (13)$$

where $b(i) = \sum_{t=1}^{i-1} g_t^{(\ell)} + 1$ is an auxiliary index that determines the first component of the equivalent channel matrix corresponding to the i-th group. The first term in (13) is the desired signal for the i-th group, the second term is the inter-group interference, and the third term is the i-th component of the equivalent noise after filtering. It is worth noting that the quality of the DQLC demapping process is determined by the SINR level at the input of the demapping functions assuming that the mapping parameters are properly adjusted according to such a value. In addition, SINR values will be maximum when the second term in (13) vanishes which happens when the equivalent response matrices have the following structure

$$\mathbf{R}^{(\ell)} = \begin{bmatrix} r_{1,1}^{(\ell)} & \ldots & r_{1,g_1^{(\ell)}}^{(\ell)} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & r_{2,1}^{(\ell)} & \ldots & r_{2,g_2^{(\ell)}}^{(\ell)} & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & r_{G_{\mathrm{s}},1}^{(\ell)} & \ldots & r_{G_{\mathrm{s}},g_{G_{\mathrm{s}}}^{(\ell)}}^{(\ell)} \end{bmatrix}, \quad (14)$$

where the zero entries represent the inter-group interference at subcarrier ℓ, while the non-zero elements are the desirable post-combining equivalent channel gains. This structure can be obtained by implementing a fully digital combiner at the BS by imposing the condition

$$\begin{aligned} &[\mathbf{W}^{*(\ell)}]_{k,:}\mathbf{H}_{i,j}^{(\ell)}\mathbf{p}_{i,j}^{(\ell)} = 0, \\ &\quad \forall\, i \neq k, \ \text{with} \ k, i = 1, \ldots, G_{\mathrm{s}}, \ \text{and} \ j = 1, \ldots, g_i^{(\ell)}. \end{aligned} \quad (15)$$

Thus, the fully digital combiner is designed in the context of DQLC-based systems to cancel the interference caused by all the users from the other groups. The digital filter implementation that guarantees the fulfilment of the conditions in (15) was derived in [6, Algorithm 3]. This implementation considers digital maximum ratio transmitter (MRT) precoding in the users and a common digital combiner for the BS to cancel the inter-group interference per subcarrier. Throughout this paper, this strategy will be labeled as DCc-DMRT (Digital Combiner cancellation - Digital MRT). The same result can also be reached by considering a digital maximum ratio combiner (MRC) at the BS and canceling the inter-group interference with the digital precoders (DMRC-DPc). This approach is similar to [30, Section 3] and invokes the MSE duality principle [31].

After the filtering step, the DQLC demapping functions $\mathbf{q}^{(\ell)}(\mathbf{z}^{(\ell)})$ produce the estimated symbols $\hat{\mathbf{s}}_i^{(\ell)}$ $\forall\ i, \ell$. DQLC decoding methods can be found in [32, Section III] and [33, Section III].

B. HYBRID COMBINING FOR USER GROUPING IN WIDEBAND SYSTEMS

In this section we extend the algorithm proposed for hybrid combining in [6, Algorithm 4] for hybrid mmWave wideband scenarios assuming the condition $N_{\mathrm{RF}}^{\mathrm{r}} = G_{\mathrm{s}} \leq K_{\mathrm{s}}^{(\ell)}$ $\forall$ ℓ. In this case, we design a PG algorithm for hybrid combining considering the structure imposed in (15) where the analog part is frequency-flat, as shown in Figure 2.

Let us define the matrices $\mathbf{W}^* = \left[\mathbf{W}^{*(1)}, \ldots, \mathbf{W}^{*(L)}\right]$ and $\mathbf{W}_{\mathrm{H}}^* = \mathbf{W}_{\mathrm{BB}}^*(\mathbf{I}_L \otimes \mathbf{W}_{\mathrm{RF}}^*) = \left[\mathbf{W}_{\mathrm{BB}}^{*(1)}, \ldots, \mathbf{W}_{\mathrm{BB}}^{*(L)}\right](\mathbf{I}_L \otimes \mathbf{W}_{\mathrm{RF}}^*)$ such that $\mathbf{W}^* \in \mathbb{C}^{G_{\mathrm{s}} \times N_{\mathrm{r}} L}$, $\mathbf{W}_{\mathrm{BB}}^* \in \mathbb{C}^{G_{\mathrm{s}} \times G_{\mathrm{s}} L}$ and $\mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}$. Most works in the literature (e.g., [5], [17], [34]–[36]) address the optimization of $\mathbf{W}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{RF}}$ as the minimization of the difference between the digital combiner and the hybrid one

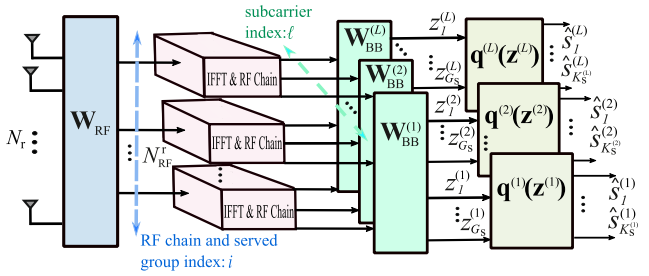

FIGURE 2. Hybrid combiner architecture in a wideband (L subcarriers) mmWave MIMO system with DQLC NOMA.

in the euclidean space, i.e.,

$$\min_{\mathbf{W}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}}} \|\mathbf{W}-(\mathbf{I}\otimes\mathbf{W}_{\mathrm{RF}})\mathbf{W}_{\mathrm{BB}}\|_F^2 \quad \text{s.t.} \quad \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}. \tag{16}$$

We instead focus on diminishing the distance between the digital combiner response and the hybrid combiner response, therefore considering the following distortion metric [6]

$$d\left(\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}\right)=\|\mathbf{R}-\mathbf{W}_{\mathrm{BB}}^{*}(\mathbf{I}_L\otimes\mathbf{W}_{\mathrm{RF}}^{*})\mathbf{H}\mathbf{P}\|_F^2, \tag{17}$$

and posing the following minimization problem

$$\min_{\mathbf{W}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}}} d\left(\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}\right) \quad \text{s.t.} \quad \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}, \tag{18}$$

where $\mathbf{R}=[\mathbf{R}^{(1)},\ldots,\mathbf{R}^{(L)}]$, $\mathbf{H}=\mathrm{blkdiag}(\mathbf{H}^{(1)},\ldots,\mathbf{H}^{(L)})$ and $\mathbf{P}=\mathrm{blkdiag}(\mathbf{P}^{(1)},\ldots,\mathbf{P}^{(L)})$. Note that $\mathbf{R}\in\mathbb{C}^{G_{\mathrm{s}}\times\sum_{l=1}^{L}K_{\mathrm{s}}^{(\ell)}}$, $\mathbf{W}_{\mathrm{BB}}^{*}\in\mathbb{C}^{G_{\mathrm{s}}\times LG_{\mathrm{s}}}$, $\mathbf{W}_{\mathrm{RF}}\in\mathcal{W}_{\mathrm{RF}}$, $\mathbf{H}\in\mathbb{C}^{N_{\mathrm{r}}L\times N_{\mathrm{t}}\sum_{l=1}^{L}K_{\mathrm{s}}^{(\ell)}}$ and $\mathbf{P}\in\mathbb{C}^{N_{\mathrm{t}}\sum_{l=1}^{L}K_{\mathrm{s}}^{(\ell)}\times\sum_{l=1}^{L}K_{\mathrm{s}}^{(\ell)}}$.

Note that this approach is especially suitable for the DQLC-based system because we are rather interested in preserving the structure of the equivalent channel response $\mathbf{R}$ than in preserving the entries of the digital combiner. In the following, we derive a PG algorithm (see Algorithm 1) to solve the non-convex optimization problem in (18) under the assumption $N_{\mathrm{RF}}^{\mathrm{r}}=G_{\mathrm{s}}$. We start determining the gradient of the cost function (17) which leads to

$$\frac{\partial d}{\partial\mathbf{W}_{\mathrm{RF}}^{*}}=\sum_{\ell=1}^{L}\mathbf{W}_{\mathrm{BB}}^{(\ell)}\mathbf{R}^{(\ell)}\mathbf{P}^{*(\ell)}\mathbf{H}^{*(\ell)} + \mathbf{W}_{\mathrm{BB}}^{(\ell)}\mathbf{W}_{\mathrm{BB}}^{*(\ell)}\mathbf{W}_{\mathrm{RF}}^{*(\ell)}\mathbf{H}^{(\ell)}\mathbf{P}^{(\ell)}\mathbf{P}^{*(\ell)}\mathbf{H}^{*(\ell)}. \tag{19}$$

The analog combiner in the first iteration ($\mathbf{W}_{\mathrm{RF}}(0)$) is calculated as the projection of the digital combiner corresponding to the central subcarrier onto the set $\mathcal{W}_{\mathrm{RF}}$ (step 2, Algorithm 1) whereas the step size μ is diminished in order to reach a local optimum. Next, at each algorithm iteration, the unconstrained solution for the analog RF component is given by

$$\tilde{\mathbf{W}}_{\mathrm{RF}}^{*}=\mathbf{W}_{\mathrm{RF}}^{*}-\mu\frac{\partial d}{\partial\mathbf{W}_{\mathrm{RF}}^{*}}, \tag{20}$$

which is then projected onto the set of feasible solutions $\mathcal{W}_{\mathrm{RF}}$ (step 7, Algorithm 1). The least squares (LS) solution is next employed to update the baseband combiner such that

Algorithm 1 PG

Input: $\{\mathbf{H}^{(\ell)}\}_{\ell=1}^{L}\in\mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}K_{\mathrm{s}}^{(\ell)}}$, $\{\mathbf{P}^{(\ell)}\}_{\ell=1}^{L}\in\mathbb{C}^{N_{\mathrm{t}}K_{\mathrm{s}}^{(\ell)}\times K_{\mathrm{s}}^{(\ell)}}$, $\{\mathbf{W}^{(\ell)}\}_{\ell=1}^{L}\in\mathbb{C}^{N_{\mathrm{r}}\times G_{\mathrm{s}}}$, $\{\mathbf{R}^{(\ell)}\}_{\ell=1}^{L}\in\mathbb{C}^{G_{\mathrm{s}}\times K_{\mathrm{s}}^{(\ell)}}$, μ_0, δ, ϵ

1: **Initialize:** $c\leftarrow 0$
2: $[\mathbf{W}_{\mathrm{RF}}(0)]_{i,j}=\frac{1}{\sqrt{N_{\mathrm{r}}}}\exp\left(j\arg\left([\mathbf{W}^{(L/2)}]_{i,j}\right)\right),\ \forall\, i,j$
3: $\mu\leftarrow\mu_0$
4: **repeat**
5: $\quad c\leftarrow c+1$
6: $\quad\tilde{\mathbf{W}}_{\mathrm{RF}}^{*}\leftarrow\mathbf{W}_{\mathrm{RF}}^{*}(c-1)-\mu\frac{\partial d}{\partial\mathbf{W}_{\mathrm{RF}}^{*}}$
7: $\quad[\mathbf{W}_{\mathrm{RF}}^{*}]_{i,j}(c)=\frac{1}{\sqrt{N_{\mathrm{r}}}}\exp\left(j\arg\left([\tilde{\mathbf{W}}_{\mathrm{RF}}^{*}]_{i,j}\right)\right),\ \forall\, i,j$
8: $\quad\mathbf{W}_{\mathrm{BB}}^{*}(c)=\mathbf{R}\left(\mathbf{P}^{*}\mathbf{H}^{*}(\mathbf{I}_L\otimes\mathbf{W}_{\mathrm{RF}}(c))\right)^{\dagger}$
9: $\quad\mathrm{error}(c)=d\left(\mathbf{W}_{\mathrm{RF}}(c),\mathbf{W}_{\mathrm{BB}}(c)\right)$
10: $\quad$**if** $\mathrm{error}(c)>\mathrm{error}(c-1)$ **then**
11: $\quad\quad\mu\leftarrow\mu/2$
12: **until** $\mathrm{error}(c)<\delta$ or $c\geq\epsilon$
13: $\mathbf{W}_{\mathrm{H}}^{(\ell)}=\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}^{(\ell)},\ \forall\,\ell$

Output: $\{\mathbf{W}_{\mathrm{H}}^{(\ell)}\}_{\ell=1}^{L}$

$\mathbf{W}_{\mathrm{BB}}^{*}=\mathbf{R}(\mathbf{P}^{*}\mathbf{H}^{*}(\mathbf{I}_L\otimes\mathbf{W}_{\mathrm{RF}}))^{\dagger}$. Algorithm 1 shows the sequence of steps corresponding to the extended PG-based approach to compute the hybrid combiner from the digital version. This iterative algorithm is stopped when the distortion d falls below a certain threshold value or when the maximum number of iterations ϵ is reached.

C. SCHEDULING FOR WIDEBAND USER GROUPING

The scheduling procedure aims at determining the number of user groups, the composition of each group and which groups are served at the BS per subcarrier, i.e., it must determine $\mathbf{g}^{(\ell)}$ and $\mathcal{G}_i^{(\ell)}\forall\, i,\ell$. It hence comprises two intertwined steps: the grouping strategy and the allocation policy. Both steps have been approached in [6, Algorithm 1 and Algorithm 2] for a narrowband scenario. For wideband systems, we will follow a strategy that considers the limitations imposed by the use of DQLC mappings but with the novelty of incorporating the intrinsic limitations of working with a common RF combiner for all the subcarriers, and considering the per-subcarrier dependent impact of the beam squint effect. Notice that the two main factors which will determine the performance of the proposed wideband system are the source correlation and the similarity between the user and the subcarrier channels.

On the one hand, the performance of DQLC schemes is definitively conditioned by the spatial correlation among the source symbols and the channel similarity for those users which are in the same group and for a specific subcarrier [6], [7]. Hence, it is essential to consider these two factors, intra-group correlation and intra-group channel similarity, to gather users in the same group. On the other hand, the requirement of canceling inter-group interference with a hybrid combiner where the analog part is common for all subcarriers necessarily leads to a scheduling procedure where it is preferable to select users with similar channels

for the same group index through all the subcarriers. For example, the composition of the first user group for the L subcarriers, i.e., $\mathcal{G}_1^{(\ell)}\ \forall\ \ell$, should not only consider the channel similarity among the users individually at each subcarrier, but also the inter-subcarrier channel similarity. Thus, it is desirable that the inter-subcarrier channels for the users in $\mathcal{G}_1^{(1)}, \mathcal{G}_1^{(2)}, \ldots, \mathcal{G}_1^{(L)}$ are also as similar as possible. A user allocation based on this premise will enable a better overall inter-group interference cancellation by using the common RF filter.

In addition, the channel similarity threshold γ_s and the cross-correlation threshold γ_ρ are considered to discard users that could degrade the system performance when including them into the groups. These thresholds allow us to manage the system requirements: either incorporating as many users as possible to the detriment of the quality of service or introducing a smaller number of users while guaranteeing a high level of the received signal quality (see [6, Subsection B, Section III]). Also, we set the maximum number of users per group to $g_{\max} = 7$ and focus on the allocation by also considering the different subcarrier channels.

As was mentioned, two important factors must be taken into account when considering DQLC mappings to perform the allocation:

1) The spatial correlation between the users per group (intra-group correlation).
2) The similarity of the user channels per group (intra-group channel similarity).

These two factors are jointly integrated in [6, Eq. 18] by using the metric

$$m_k = \delta_\rho \tilde{\rho}_k + \delta_\mathrm{s} C_{\mathrm{sim}k}, \tag{21}$$

where the factors δ_ρ and δ_s define the weight of the correlation and similarity criteria, respectively, $\tilde{\rho}_k$ is the mean of the users cross-correlation in a group $\mathcal{G}_i$, whereas the parameter $C_{\mathrm{sim}k}$ measures the convenience of including a user in the group $\mathcal{G}_i$ according to its channel response similarity w.r.t. the already allocated users in $\mathcal{G}_i$. Note that this part of the allocation procedure should be incorporated into the new allocation policies for wideband scenarios as it is a consequence of using DQLC mappings.

When considering wideband communications, two different situations can arise. First, in wideband scenarios where the beam squint effect is negligible, the user channels across all the subcarriers have rotated row and column spaces [13], i.e., the channels are very similar, since the actual steering vectors are the same at each subcarrier channel ($\mathbf{a}_{\mathrm{t}_{i,j}}^{(\ell)} = \mathbf{a}_{\mathrm{t}_{i,j}}$ and $\mathbf{a}_{\mathrm{BS}_{i,j}}^{(\ell)} = \mathbf{a}_{\mathrm{BS}_{i,j}}, \forall\ i, j, \ell$). In these scenarios, an individual allocation per subcarrier following a straightforward approximation of [6, Algorithm 2] can be performed.

Nevertheless, when considering larger bandwidths, channel similarity is lost and the per subcarrier AoA and AoD can be spatially distinguished due to the beam squint effect [13]. Hence, the design of the user allocation also needs to consider the inter-subcarrier channel similarity. In the following, we describe three possible allocation strategies.

1) INDIVIDUAL ALLOCATION

In this strategy, an individual allocation is performed per subcarrier and thus the user groups, the composition of each group, the served group indexes and the number of served users $K_\mathrm{s}^{(\ell)}$ will be different per subcarrier. This strategy is a straightforward extension of [6, Algorithm 2] and it does not consider the beam squint effect.

2) COMMON ALLOCATION

In this strategy, the user distribution per group is obtained by considering a similar metric to (21), $m_{\mathrm{c}k} = \delta_\rho \tilde{\rho}_k + \delta_\mathrm{s} C_{\mathrm{csim}k}$, and this distribution will be the same for all the subcarriers, i.e., $\mathcal{G}_i^{(\ell)} = \mathcal{G}_i,\ \forall\ \ell$ and thus $K_\mathrm{s}^{(\ell)} = K_\mathrm{s},\ \forall\ \ell$. Therefore, we start including in the group the user with the best channels i.e., with the largest singular values across the subcarrier channels. Next, we use the metric $C_{\mathrm{csim}k}$ to measure the convenience of including a new user $\mathcal{K}(k)$ in $\mathcal{G}_i$ by jointly computing the channel similarity of all the user channels in $\mathcal{G}_i$ and considering all the subcarriers. With this aim, we define the matrix

$$\mathbf{H}_\mathrm{c} = \left[\mathbf{H}_{\mathcal{G}_i(1)}^{T(1)}, \ldots, \mathbf{H}_{\mathcal{G}_i(1)}^{T(L)}, \ldots, \mathbf{H}_{\mathcal{G}_i(Z)}^{T(L)}\right]^T, \tag{22}$$

which includes all the subcarrier channels for the Z users already allocated in the sequence $\mathcal{G}_i$, and the matrix

$$\mathbf{H}'_\mathrm{c} = \left[\mathbf{H}_\mathrm{c}^T, \mathbf{H}_{\mathcal{K}(k)}^{T\ (1)}, \ldots, \mathbf{H}_{\mathcal{K}(k)}^{T\ (L)}\right]^T, \tag{23}$$

which also includes the L channels for the candidate user $\mathcal{K}(k)$. We now employ the auxiliary vector $\mathbf{p}_i$, i.e., the right singular vector associated to the largest singular value of $\mathbf{H}_\mathrm{c}$, and similarly $\mathbf{p}'_i$ for $\mathbf{H}'_\mathrm{c}$, to compute the parameter $C_{\mathrm{csim}k}$ as

$$C_{\mathrm{csim}k} = \frac{\sum_{\ell=1}^{L}\sum_{z=1}^{Z} \|\mathbf{H}_{\mathcal{G}_i(z)}^{(\ell)}\mathbf{p}_i\|^2}{\sum_{\ell=1}^{L}\sum_{z=1}^{Z} \|\mathbf{H}_{\mathcal{G}_i(z)}^{(\ell)}\mathbf{p}'_i\|^2 + \sum_{\ell=1}^{L} \|\mathbf{H}_{\mathcal{K}(k)}^{(\ell)}\mathbf{p}'_i\|^2}. \tag{24}$$

Note that this allocation strategy incorporates the inter-subcarrier channel similarity and is thus especially suitable for large bandwidth schemes where the beam squint effect is present. Indeed, by including all the subcarrier channels corresponding to the allocated users as well as the candidate user, the expression (24) will provide an overall similarity across all the subcarrier channels that measures the convenience of including each candidate user.

3) SEQUENTIAL ALLOCATION

This strategy sequentially includes the inter-subcarrier channel similarity to build the groups. We aim at designing an allocation strategy to approximately overlap the inter-group interference subspaces across all the subcarriers.

We start by including in the first group the user with the best channel and by considering the inter-subcarrier channel similarity as in the previous common allocation. In this case, however, the users per group are specifically chosen at each subcarrier. This allocation procedure is performed sequentially across the subcarriers such that we select the

user whose channels are more similar considering the previous subcarriers of the users already allocated in the same group index. Thus, the proposed policy performs a sequential allocation where the best user is always selected according to the source correlation, the intra-group user channel similarity, and the inter-subcarrier channel similarity. Hence, unlike the common allocation, the group composition might be different for different subcarriers. Then, after projecting any calculated fully digital filter onto the set of feasible RF combiners, a common frequency-flat RF filter will be produced to cancel the inter-group interference at each sub-band.

After selecting the first user for each group $\mathcal{G}_i^{(\ell)}$, we next define the metric C_{psim_k} to measure the convenience of including a user $\mathcal{K}(k)$ in $\mathcal{G}_i^{(\ell)}$. Such metric jointly considers the channel similarity of all the subcarrier channels for the users in $\mathcal{G}_i^{(\ell)}$ and all the subcarrier channels of the candidate user $\mathcal{K}(k)$. Once the allocation for the first subcarrier is completed, the group scheduling of the next subcarriers is sequentially determined by taking into account the group composition of the already allocated subcarriers. Hence, the overall channel similarity for the i-th group at the $\tilde{\ell}$-th subcarrier is defined as

$$C_{\text{psim}_k}^{(\tilde{\ell})} = \frac{\sum_{\ell=1}^{L}\sum_{z=1}^{Z}\|\mathbf{H}^{(\ell)}_{\mathcal{G}_i^{(\tilde{\ell})}(z)}\mathbf{p}_i\|^2 + \tilde{C}}{\sum_{\ell=1}^{L}\sum_{z=1}^{Z}\|\mathbf{H}^{(\ell)}_{\mathcal{G}_i^{(\tilde{\ell})}(z)}\mathbf{p}'_i\|^2 + \sum_{\ell=1}^{L}\|\mathbf{H}^{(\ell)}_{\mathcal{K}(k)}\mathbf{p}'_i\|^2 + \tilde{D}}, \tag{25}$$

with

$$\tilde{C} = \sum_{\ell=1}^{L}\sum_{m=1}^{\tilde{\ell}-1}\sum_{\tilde{z}=1}^{g_i^{(m)}}\|\mathbf{H}^{(\ell)}_{\mathcal{G}_i^{(m)}(\tilde{z})}\mathbf{p}_i\|^2 \tag{26}$$

and

$$\tilde{D} = \sum_{\ell=1}^{L}\sum_{m=1}^{\tilde{\ell}-1}\sum_{\tilde{z}=1}^{g_i^{(m)}}\|\mathbf{H}^{(\ell)}_{\mathcal{G}_i^{(m)}(\tilde{z})}\mathbf{p}'_i\|^2. \tag{27}$$

As observed, Z users are assumed to be already allocated to the i-th group at the current $\tilde{\ell}$-th subcarrier, i.e., in $\mathcal{G}_i^{(\tilde{\ell})}$. The terms $\tilde{C}$ and $\tilde{D}$ take into account the channels corresponding to the users which have already been allocated in the same group index i but at the previous subcarriers. Similarly to the common allocation, we also use the auxiliary vectors $\mathbf{p}_i$ and $\mathbf{p}'_i$, but here $\mathbf{H}_c$ results from the concatenation of the L sub-channels for the users allocated in the same group index i at the previous subcarriers and for the Z users already allocated for that group at the current $\tilde{\ell}$-th subcarrier.

It is remarkable to recall that the user groups, their composition, and thus the number of served user symbols $K_s^{(\ell)}$ are subcarrier dependent in this allocation algorithm. Note also that, by taking int account the channel similarity of both the users and the subcarriers, this strategy jointly considers the DQLC singularities and the limitations of working with a common RF combiner at the BS.

IV. GNU-SVD INTER-GROUP INTERFERENCE CANCELLATION ALGORITHM

In this section, we derive an algorithm termed group null-space directed SVD (GNU-SVD) for the joint design of the user precoders and the hybrid combiner at the BS in the uplink MU mmWave MIMO setting under consideration. This algorithm assumes that the composition of each group, the served groups and the number of served users $K_s^{(\ell)}$ per subcarrier provided by the scheduling is known. The inter-user correlation per group is also calculated in this algorithm to compute the DQLC parameter optimization (step 5, Algorithm 2). GNU-SVD is summarized in Algorithm 2 and can be divided into two main steps.

Step 1: the signal-to-noise ratio (SNR) is maximized at transmission by means of a MRT precoder per user and at each subcarrier $\tilde{\mathbf{p}}_{i,j}^{(\ell)}\ \forall\ i,j,\ell$ (step 6, Algorithm 2). Then, the ℓ-th digital unconstrained combiner $\mathbf{W}^{(\ell)}$ is obtained as in [6, Algorithm 3] to cancel the inter-group interference (step 9, Algorithm 2), but using the conditions given in (15). Up to this point, fully digital precoders and combiners are considered and the inter-user interference is just removed among those users in different groups at each subcarrier. Next, we design the hybrid combiner by using Algorithm 1 (PG) in the step 17 of Algorithm 2 and, consequently, a residual inter-user interference will appear since the constraints of the hybrid implementation make infeasible the full cancellation of the inter-group interference. This residual interference can cause large losses in the system performance because it seriously affects the optimization of the DQLC parameters.

Step 2: the residual inter-group interference is removed by means of the digital precoders at the users and the hybrid combiner obtained in the previous step. Notice that the limitations imposed by the hardware, namely the frequency flat response of the analog part of the combiner, demands a refinement of the joint precoder and combiner design in order to lower the interference down to a level that enables the use of DQLC. Hence, we define the equivalent channels $\tilde{\mathbf{H}}_i^{(\ell)}$ for the i-th user group at subcarrier ℓ as

$$\tilde{\mathbf{H}}_i^{(\ell)} = \left[\mathbf{H}_{i,1}^{*\,(\ell)}\mathbf{W}_{\text{H}}^{(\ell)}, \ldots, \mathbf{H}_{i,g_i}^{*\,(\ell)}\mathbf{W}_{\text{H}}^{(\ell)}\right]. \tag{28}$$

Next, we project each MRT user precoder $\tilde{\mathbf{p}}_{i,j}^{(\ell)}\ \forall\ i,j,\ell$ onto the nullspace of the equivalent channels of the users allocated in different groups (step 23, Algorithm 2). We eventually get the user precoders as $\mathbf{p}_{i,j}^{(\ell)} = \left(\mathbf{I} - \mathbf{N}_i^{(\ell)}\mathbf{N}_i^{(\ell)*}\right)\tilde{\mathbf{p}}_{i,j}^{(\ell)}$, where $\mathbf{N}_i^{(\ell)}$ is the basis for the subspace spanned by the columns of the matrix

$$\mathbf{I}_{\mathcal{G}_i^{(\ell)}} = \left[\tilde{\mathbf{H}}_1^{(\ell)}, \ldots, \tilde{\mathbf{H}}_{i-1}^{(\ell)}, \tilde{\mathbf{H}}_{i+1}^{(\ell)}, \ldots, \tilde{\mathbf{H}}_{G_s}^{(\ell)}\right].$$

Algorithm 2 summarizes the GNU-SVD procedure to determine the user precoders and the hybrid combiner at the BS.

V. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, we analyze the computational complexity of the proposed Algorithm 1 for hybrid combining, the proposed

Algorithm 2 GNU-SVD

Input: $\mathbf{H}_k^{(\ell)}\ \forall\, k, l,\ \mathbf{C}_\mathrm{s}, N_\mathrm{RF}^\mathrm{r}, \mathbf{g}$

1: **Initialize:** $\tilde{\mathbf{P}} = [\,], \mathbf{W} = [\,], \mathbf{H} = [\,], \mathbf{R} = [\,],$
2: **for** $\ell = 1, \ldots, L$ **do**
3: **repeat**
4: $i \leftarrow i + 1$
5: $\mathbf{C}_i^{(\ell)} \leftarrow$ Inter-user correlation per group
6: $\tilde{\mathbf{P}}_i^{(\ell)} = \mathrm{blkdiag}\left(\tilde{\mathbf{p}}_{i,1}^{(\ell)}, \ldots, \tilde{\mathbf{p}}_{i,g_i}^{(\ell)}\right) \leftarrow$ MRT
7: $\mathbf{H}_i^{(\ell)} = \left[\mathbf{H}_{i,1}^{(\ell)}, \ldots, \mathbf{H}_{i,g_i}^{(\ell)}\right]$
8: **until** $i = N_\mathrm{RF}^\mathrm{r}$
9: $\mathbf{W}^{(\ell)} \leftarrow$ [6, Algorithm 3] per subcarrier
10: $\tilde{\mathbf{P}}^{(\ell)} = \mathrm{blkdiag}(\tilde{\mathbf{P}}_1^{(\ell)}, \ldots, \tilde{\mathbf{P}}_{G_\mathrm{s}}^{(\ell)})$
11: $\mathbf{H}^{(\ell)} = \left[\mathbf{H}_1{}^{(\ell)}, \ldots, \mathbf{H}_{G_\mathrm{s}}^{(\ell)}\right]$
12: $\mathbf{R}^{(\ell)} = \mathbf{W}^{*(\ell)}\mathbf{H}^{(\ell)}\mathbf{P}^{(\ell)}$
13: $\mathbf{W}^* = \left[\mathbf{W}^{*(1)}, \ldots, \mathbf{W}^{*(L)}\right]$
14: $\tilde{\mathbf{P}} = \mathrm{blkdiag}(\tilde{\mathbf{P}}^{(1)}, \ldots, \tilde{\mathbf{P}}^{(L)})$
15: $\mathbf{H} = \mathrm{blkdiag}(\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(L)})$
16: $\mathbf{R} = [\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(L)}]$
17: $\mathbf{W}_\mathrm{H}^{(\ell)} = \mathbf{W}_\mathrm{RF}\mathbf{W}_\mathrm{BB}^{(\ell)} \leftarrow$ Algorithm 1 $(\mathbf{H}, \mathbf{P}, \mathbf{W}, \mathbf{R})$
18: **for** $\ell = 1, \ldots, L$ **do**
19: **for** $i = 1, \ldots, G_\mathrm{s}$ **do**
20: $\tilde{\mathbf{H}}_i^{(\ell)} = \left[\mathbf{H}_{i,1}^{*\,(\ell)}\mathbf{W}_{\mathrm{H}_i}^{(\ell)}, \ldots, \mathbf{H}_{i,g_i}^{*\,(\ell)}\mathbf{W}_{\mathrm{H}_i}^{(\ell)}\right].$
21: $\mathbf{I}_{\mathcal{G}_i^{(\ell)}} = \left[\tilde{\mathbf{H}}_1^{(\ell)}, \ldots, \tilde{\mathbf{H}}_{i-1}^{(\ell)}, \tilde{\mathbf{H}}_{i+1}^{(\ell)}, \ldots, \tilde{\mathbf{H}}_{G_\mathrm{S}}^{(\ell)}\right]$
22: $\mathbf{N}_i^{(\ell)} \leftarrow$basis for the subspace span $\left(\mathbf{I}_{\mathcal{G}_i^{(\ell)}}\right)$
23: **for** $j = 1, \ldots, g_i$ **do**
24: $\mathbf{p}_{i,j}^{(\ell)} = \left(\mathbf{I} - \mathbf{N}_i^{(\ell)}\mathbf{N}_i^{(\ell)*}\right)\tilde{\mathbf{p}}_{i,j}^{(\ell)}$

Output: $\mathbf{P}, \mathbf{W}_\mathrm{H}$

Algorithm 2 for user precoding and hybrid combining, and the considered allocation policies: individual, common and sequential. The complexity orders for the different algorithms are shown in Table 2.

The main contribution to the overall computational complexity of Algorithm 1 corresponds to the calculation of each candidate for the RF combiner (step 6), which involves the gradient computation. The overall complexity order of this step is $\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}N_\mathrm{RF}^\mathrm{r}K_\mathrm{s}\right)$. It is remarkable that the computational complexity of the whole Algorithm 1 increases with the number of iterations ϵ needed to reach convergence. This leads to an overall complexity order $\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}N_\mathrm{RF}^\mathrm{r}K_\mathrm{s}\epsilon\right)$.

In the GNU-SVD algorithm, the main contribution to the overall complexity is the computation of all the MRT user precoders $\mathbf{p}^{(\ell)}{}_{i,j}$ in step 6. The computational complexity order of this step is $\mathcal{O}\left(N_\mathrm{r}^3L^2N_\mathrm{RF}^\mathrm{r}\right)$ because it is required to carried out the singular value decomposition (SVD) of the user channels.

Regarding the allocation algorithms in Section III-C, the sequential allocation algorithm is the more complex. Its complexity order is bounded by $\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2L^2K^2g_\mathrm{max}^2\right)$, which corresponds to the computation of the overall channel similarity in (25). On the other hand, the individual allocation algorithm is the least complex one since its complexity order is bounded by $\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2Lg_\mathrm{max}^2\right)$. Finally, the common allocation algorithm exhibits an intermediate complexity, mainly determined by the SVD decompositions in (24), with an order of $\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2L^2g_\mathrm{max}^2\right)$.

TABLE 2. Computational complexity of the proposed algorithms and scheduling policies.

Algorithm	Complexity order
Algorithm 1 PG	$\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}N_\mathrm{RF}^\mathrm{r}K_\mathrm{s}L\epsilon\right)$
Algorithm 2 G-Nu-SVD	$\mathcal{O}\left(N_\mathrm{r}^3L^2N_\mathrm{RF}^\mathrm{r}\right)$
Sequential allocation	$\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2L^2K^2g_\mathrm{max}^2\right)$
Common allocation	$\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2L^2g_\mathrm{max}^2\right)$
Individual allocation	$\mathcal{O}\left(N_\mathrm{r}N_\mathrm{t}^2Lg_\mathrm{max}^2\right)$

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed algorithms through different computer simulations where several combining design techniques and scheduling implementations are compared. We consider an exponential correlation model for the user symbols where the source correlation matrix entries are given by $[\mathbf{C}_\mathrm{s}^{(\ell)}]_{i,j} = \rho^{|\mathrm{i-j}|},\ \forall\, i, j$, with ρ the correlation factor. According to this model, the vectors of the K user symbols are generated from a multivariate circularly-symmetric Gaussian distribution with zero mean and covariance matrix $\mathbf{C}_\mathrm{s} = \mathbf{C}_\mathrm{s}^{(\ell)},\ \forall\, \ell$. A MIMO setup with $N_\mathrm{r} = 100$ antennas deployed at the BS and $N_\mathrm{t} = 8$ antennas per user is considered.

Results are averaged over $N = 1000$ channel realizations for each experiment. The parameters for the channel model are set to $L_\mathrm{D} = 8$ delay taps and $N_{\mathrm{p}_{i,j}} = 3$ channel paths $\forall\, i, j$. The AoA's and the AoD's are assumed to be uniformly distributed over the interval $[0, \pi]$ as in [25]. The relative delays $\tau_{i,j,n}$ are also random and assumed to be uniformly distributed over the interval $\tau_{i,j,n} \in [0, (L_\mathrm{D} - 1)T_\mathrm{s}]$, with $T_s = 1/f_s$ and $f_s = 1760$ MHz. The complex-valued channel gains are i.i.d. random variables $\beta_{k,m} \sim \mathcal{N}_\mathbb{C}(0, 1)$. The central carrier frequency is $f_\mathrm{c} = 28$ GHz whereas the signal bandwidth varies from 800 MHz to 3200 MHz. The number of subcarriers is set to $L = 32$.

The system performance is assessed in terms of the average signal-to-distortion ratio (SDR) defined as

$$\mathrm{SDR\ (dB)} = 10\log_{10}\left(1/\hat{\xi}_\mathrm{sum}\right), \tag{29}$$

where

$$\hat{\xi}_\mathrm{sum} = \frac{1}{NL\sum_{l=1}^{L}K_\mathrm{s}^{(\ell)}}\sum_{n=1}^{N}\sum_{\ell=1}^{L}\sum_{i=1}^{G_\mathrm{s}}\sum_{j=1}^{g_i^{(\ell)}}\left|s_{n,i,j}^{(\ell)} - \hat{s}_{n,i,j}^{(\ell)}\right|^2 \tag{30}$$

represents the average MSE between the source symbols and the estimated ones obtained after the demapping operation at the BS.

TABLE 3. Simulation parameter setting.

Parameter	Setting
Number of users	$K = 9$
Spatial correlation model	$[\mathbf{C}_{\mathrm{s}}^{(\ell)}]_{i,j} = \rho^{\lvert i-j\rvert},\ \forall\, i,j$
Correlation per subcarrier	$\mathbf{C}_{\mathrm{s}}^{(\ell)} = \mathbf{C}_{\mathrm{s}},\ \forall\, \ell$
Number of antennas at the BS	$N_{\mathrm{r}} = 100$
Number of RF chains at the BS	$N_{\mathrm{RF}}^{\mathrm{r}} = 3$
Number of antennas per user	$N_{\mathrm{t}} = 8$
Number of taps (channel model)	$L_{\mathrm{D}} = 8$
Channel paths (channel model)	$N_{\mathrm{p}_{i,j}} = 3,\ \forall\, i,j$
AoA's and AoD's (channel model)	$[0, \pi]$
Relative delay (channel model) ($\tau_{i,j,n}$)	$\tau_{i,j,n} \in [0, (L_{\mathrm{D}}-1)T_{\mathrm{s}}]$
Sampling frequency (channel model)	$f_{\mathrm{s}} = 1760$ MHz
Channel gain (channel model)	$\beta_{k,m} \sim \mathcal{N}_{\mathbb{C}}(0,1)$
Carrier frequency (channel model)	$f_{\mathrm{c}} = 28$ GHz
Signal bandwidths	800 MHz $\sim$ 3200 MHz
Number of subcarriers (channel model)	$L = 32$
Channel realizations	$N = 1000$
Number of iterations (PG algorithms)	$\epsilon = 1000$

Without loss of generality, we assume $\sigma_{\mathrm{n}}^2 = 1$ and therefore define the SNR per user as

$$\mathrm{SNR\ (dB)}_{i,j} = 10\log_{10}(T'_{i,j}),\quad \forall\, i = 1,\ldots,G_{\mathrm{s}} \text{ and } j = 1,\ldots,g_i^{(\ell)}. \tag{31}$$

For simplicity, the same power constraint is considered for all the users. Finally, the maximum number of iterations ϵ is set to 1000 in Algorithm 1 as well as in the other iterative algorithms considered in this section for comparison. The main parameters considered in the simulations are summarized in Table 3.

Two experiments have been carried out to evaluate the performance of the different proposed methods. First, we evaluate the behavior of the GNU-SVD interference cancellation method in combination with the extended hybrid PG algorithm and compare it with other possible system implementations. In the second experiment, we evaluate the performance of the proposed user allocation approaches.

Figure 3 shows the SDR (dB) versus SNR (dB) obtained for $K = 9$ users, $B = 3200$ MHz, $N_{\mathrm{RF}}^{\mathrm{r}} = 3$, $\rho = 0.85$ and different strategies for the design of the user precoding and the BS combiner. In particular, we consider the following alternatives: 1) the fully digital combining cancellation with MRT precoders at the users (DCc-DMRT); 2) a digital precoding cancellation with MRC (DMRC-DPc); 3) a hybrid version of 1) by factorizing the combiner with the PG algorithm in [17] labeled as HCc-PG-DMRT; 4) a hybrid version of 1) by factorizing the combiner with the proposed Algorithm 1, labeled as HCc(Alg. 1)-DMRT; 5) a hybrid version of 2) that utilizes the proposed Algorithm 1 to factorize the combiner and digital precoder cancellation (HMRC-DPc); and 6) the proposed GNU-SVD algorithm. Simulations have considered $\gamma_\rho = 0$ and $\gamma_{\mathrm{s}} = 0$ to isolate the impact of the group allocation. We have also chosen $\delta_\rho = 0.8$ and $\delta_{\mathrm{s}} = 0.2$ because those values are a suitable trade-off between the impact of the channel similarity and the correlation for $\rho = 0.85$ [6].

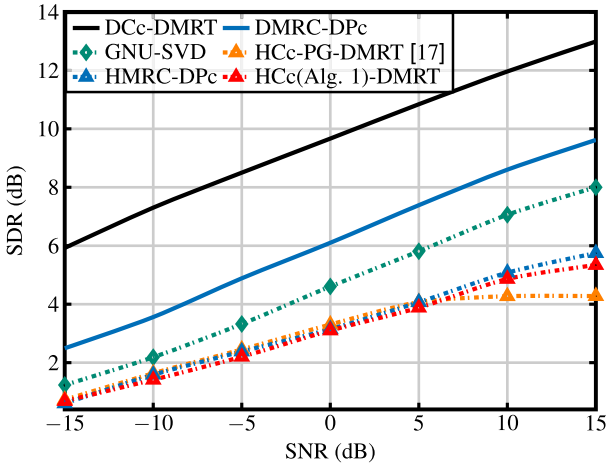

FIGURE 3. SDR (dB) performance for $K = 9$ users, $N_{\mathrm{RF}}^{\mathrm{r}} = 3$, $B = 3200$ MHz and correlation factor $\rho = 0.85$.

As observed, Figure 3 shows a gap (around 2.8 dBs) between the two fully digital implementation strategies (DCc-DMRT and DMRC-DPc) which is related to the available degrees of freedom to implement the inter-group interference cancellation. Note that the DMRC-DPc approach performs the cancellation at the user precoders ($\mathbf{p}_{i,j}^{(\ell)} \in \mathbb{C}^{N_{\mathrm{t}}\times 1}$, $\forall\ i,j,\ell$) with $N_{\mathrm{t}} = 16$ antennas, whereas the DCc-DMRT approach implements the cancellation at the common combiner $\mathbf{W}^{(\ell)} \in \mathbb{C}^{N_{\mathrm{r}}\times G_{\mathrm{s}}},\ \forall\, \ell$ with $N_{\mathrm{r}} = 100$ antennas.

In order to compare the hybrid combiner implementations, we need to take into account two issues: 1) the structure of the common combiner to be factorized, and 2) how the factorization problem is stated. By considering this last issue, we can sense that the factorization algorithm proposed in [17, Section 7]) provides the less appropriate strategy to obtain the hybrid combiner since that algorithm completely neglects the desired structures $\mathbf{R}^{(\ell)}$ in (14), and therefore it is reasonable that the HCc-PG-DMRT strategy exhibits the worst performance, especially for high SNR values. As observed, the implementation strategies HMRC-DPc and HCc(Alg. 1)-DMRT provide a similar behaviour. Although the digital precoding interference cancellation implemented in the HMRC-DPc is performed with less degrees of freedom, the hybrid factorization of the MRCs is more feasible for wideband scenarios than the one performed over the digital combiner implemented according to [6, Algorithm 3]. This is because this latter approach destroys the relationship between the column spaces at the different subcarriers [17]. Finally, by implementing the user precoder cancellation in a second refinement step, the proposed GNU-SVD approach provides the best performance among all the hybrid implementations and also offers a suitable hardware-complexity system

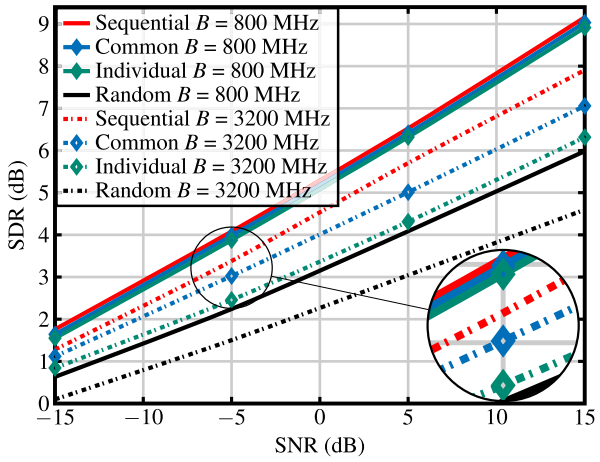

FIGURE 4. SDR (dB) versus SNR (dB) for $K = 9$ users, $N_{\text{RF}}^{\text{r}} = 3$, $B \in \{800\ \text{MHz}, 3200\ \text{MHz}\}$ and correlation factor $\rho = 0.85$.

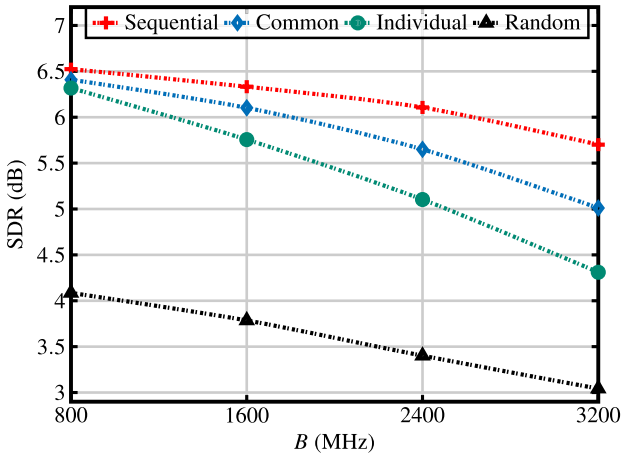

FIGURE 5. SDR (dB) performance for $K = 9$ users, $N_{\text{RF}}^{\text{r}} = 3$, SNR = 5 dB and correlation factor $\rho = 0.85$.

performance trade-off. Note that in all these hybrid strategies we are able to handle 9 user symbols per sub-band ($L = 32$) with only two RF chains ($N_{\text{RF}}^{\text{r}} = 3$).

Figure 4 shows the evaluation results for four different user allocation algorithms that determine the user groups: 1) sequential allocation, 2) common allocation per subcarrier, 3) individual allocation per subcarrier neglecting the beam squint effect, and 4) random allocation. We consider $K = 9$ users, $N_{\text{RF}}^{\text{r}} = 3$, and $\rho = 0.85$. The proposed GNU-SVD cancellation algorithm is employed to configure the user precoders and the hybrid combiner. The comparison is performed for two different bandwidths, i.e., $B \in \{800\ \text{MHz}, 3200\ \text{MHz}\}$. The beam squint effect is more apparent for $B = 3200$ MHz, where the subcarrier offsets $\zeta(\ell) = ((\ell - 1) - \frac{L-1}{2})\frac{B}{L},\ \forall\ \ell$ are larger, while this effect practically vanishes when considering $B = 800$ MHz [25].

As observed in Figure 4, the random allocation provides the worst performance for both bandwidths since it neglects the main factors (cross-correlation and channel similarity) to define an appropriate allocation policy. Figure 4 also shows that, when $B = 800$ MHz, the remaining allocation strategies present a similar behaviour. In this case, the sequential and the common allocation approximately offer the same system performance as the individual allocation because the inter-subcarrier channel similarity does not impact on the allocation. However, the sequential allocation presents a small gain w.r.t. the other strategies because the design of the digital filters takes into account the channel similarity of both the users and the subcarriers, and therefore reduces the losses of working with a frequency-flat analog combiner.

A different behaviour is observed when $B = 3200$ MHz due to the beam squint effect. Figure 4 shows that the sequential and common allocation strategies achieve a higher performance than the individual one since they incorporate the inter-subcarrier channel similarity. Moreover, the sequential allocation strategy provides additional performance gains by considering the limitations of working with a common RF combiner at the BS.

Figure 5 shows the SDR (dB) versus the signal bandwidth (B) for the four considered user allocation strategies: 1) sequential allocation, 2) common allocation per subcarrier, 3) individual allocation per subcarrier, and 4) random allocation. We consider $K = 9$ users, $N_{\text{RF}}^{\text{r}} = 3$ RF chains, $\rho = 0.85$ and a particular SNR value, namely, SNR = 5 dB. The proposed GNU-SVD cancellation algorithm is employed to define the user precoders and the hybrid combiner. The beam squint is practically negligible when considering $B = 800$ MHz but its effect starts to be more apparent while the bandwidth increases because the subcarrier offsets $\zeta(\ell) = ((\ell - 1) - \frac{L-1}{2})\frac{B}{L},\ \forall\ \ell$ become larger.

As observed, Figure 5 clearly illustrates the behavior of the different user allocation strategies under the beam squint effect. Note how the random allocation strategy provides the worst performance in the whole range of bandwidths because it neglects the correlation and the channel similarity factor to determine the allocation. The performance of random allocation also worsens as the bandwidth increases because the probability of clustering users with significantly different channels is higher. Figure 5 also shows that the remaining allocation policies (sequential, common, and individual) present a similar behavior when $B = 800$ MHz since, under a negligible beam squint effect, the inter-subcarrier channel similarity—which is incorporated in the common and the sequential allocation—does not impact on the user allocation. However, their behavior is different as the bandwidth increases because the sequential and the common allocation, by considering the inter-subcarrier channel similarity, offer superior performance than the individual allocation. Although the performance of all allocation policies degrades as the bandwidth increases, the individual allocation is the most sensitive to the beam squint effect as it presents the highest losses. On the other hand, the sequential allocation presents the lowest losses when the channel bandwidth increases since it considers the inter-subcarrier channel similarity as well as the limitations of working with a common RF combiner at the BS.

VII. CONCLUSION

This work focuses on the design of appropriate strategies for wideband user grouping and hybrid combining in the uplink of MU mmWave MIMO systems. A both-ends interference cancellation procedure has been proposed to deal with the hardware constraints of the system and serve the user groups in a DQLC-based NOMA scheme. The proposed hybrid design for the BS combiner leads to large gains over conventional algorithms based on matrix factorization for wideband scenarios. A novel sequential allocation approach that considers the effects derived from beam squint in wideband mmWave systems has also been developed. The proposed sequential scheduling procedure provides gains of about 25% compared to the straightforward extension of the scheduling strategy developed for narrowband scenarios and significant gains over the random allocation policy. Finally, the obtained results show the benefits of the user grouping strategies for wideband scenarios, which allows serving a number of streams per subcarrier larger than the number of available RF chains.

REFERENCES

[1] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Upper Saddle River, NJ, USA: Prentice-Hall, Sep. 2014.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[4] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501–3511, Dec. 2006.

[5] R. Mendez-Rial, C. Rusu, N. Gonzalez-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.

[6] D. Perez-Adan, O. Fresnedo, J. P. Gonzalez-Coma, and L. Castedo, "User grouping for the uplink of multiuser hybrid mmWave MIMO," *IEEE Access*, vol. 8, pp. 55323–55341, 2020.

[7] O. Fresnedo, P. Suarez-Casal, and L. Castedo, "Transmission of spatio-temporal correlated sources over fading multiple access channels with DQLC mappings," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5604–5617, Aug. 2019.

[8] P. A. Floor, A. N. Kim, T. A. Ramstad, and I. Balasingham, "On transmission of multiple Gaussian sources over a Gaussian MAC using a VQLC mapping," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 50–54.

[9] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution PSs for wideband mmWave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2168–2181, Sep. 2020.

[10] Y. Chen, D. Chen, and T. Jiang, "Non-uniform quantization codebook-based hybrid precoding to reduce feedback overhead in millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2779–2791, Apr. 2019.

[11] J. Zhang, Y. Huang, J. Wang, and L. Yang, "Hybrid precoding for wideband millimeter-wave systems with finite resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11285–11290, Nov. 2018.

[12] Y. Chen, D. Chen, Y. Tian, and T. Jiang, "Spatial lobes division-based low complexity hybrid precoding and diversity combining for mmWave IoT systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3228–3239, Apr. 2019.

[13] K. Venugopal, N. Gonzalez-Prelcic, and R. W. Heath, "Optimal frequency-flat precoding for frequency-selective millimeter wave channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5098–5112, Nov. 2019.

[14] M. Cai, K. Gao, D. Nie, B. Hochwald, J. N. Laneman, H. Huang, and K. Liu, "Effect of wideband beam squint on codebook design in phased-array wireless systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[15] X. Yu, J. Zhang, and K. B. Letaief, "Alternating minimization for hybrid precoding in multiuser OFDM mmWave systems," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 281–285.

[16] 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 Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[17] J. P. Gonzalez-Coma, J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, L. Castedo, and R. W. Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[18] R. Magueta, D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Hybrid multi-user equalizer for massive MIMO millimeter-wave dynamic subconnected architecture," *IEEE Access*, vol. 7, pp. 79017–79029, 2019.

[19] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8315–8328, Dec. 2019.

[20] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62367–62414, 2020.

[21] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: Principles, Models and Technology Components*. New York, NY, USA: Academic, 2018.

[22] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[23] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[24] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 273–277.

[25] J. P. Gonzalez-Coma, W. Utschick, and L. Castedo, "Hybrid LISA for wideband multiuser millimeter-wave communication systems under beam squint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1277–1288, Feb. 2019.

[26] Z. Wang, L. Cheng, J. Wang, and G. Yue, "Digital compensation wideband analog beamforming for millimeter-wave communication," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–5.

[27] J. P. Gonzalez-Coma and L. Castedo, "Wideband hybrid precoding using time modulated arrays," *IEEE Access*, vol. 8, pp. 144638–144653, 2020.

[28] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005.

[29] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for WideBand millimeter wave MIMO systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1847–1860, Mar. 2021.

[30] P. Suarez-Casal, J. P. Gonzalez-Coma, O. Fresnedo, and L. Castedo, "Design of linear precoders for correlated sources in MIMO multiple access channels," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6110–6122, Dec. 2018.

[31] T. E. Bogale and L. Vandendorpe, "Robust sum MSE optimization for downlink multiuser MIMO systems with arbitrary power constraint: Generalized duality approach," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1862–1875, Apr. 2012.

[32] P. Suarez-Casal, O. Fresnedo, and L. Castedo, "DQLC optimization for joint source channel coding of correlated sources over fading MAC," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1292–1296.

[33] P. A. Floor, A. N. Kim, N. Wernersson, T. A. Ramstad, M. Skoglund, and I. Balasingham, "Zero-delay joint source-channel coding for a bivariate Gaussian on a Gaussian MAC," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3091–3102, Oct. 2012.

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

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

[36] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.

DARIAN PÉREZ-ADÁN received the B.S. degree in telecommunications and electronics engineering from the Technological University of Havana José Antonio Echeverría (CUJAE), Cuba, in 2017. He is currently pursuing the Ph.D. degree with the University of A Coruña (UDC), Spain. Since 2018, he has been with the Group of Electronic Technology and Communications, UDC. He was awarded a predoctoral scholarship granted by the Spanish Government. His research interests include signal processing for millimeter-wave and multiuser communications.

ÓSCAR FRESNEDO (Member, IEEE) received the computer engineering degree and the Ph.D. degree in computer engineering from the University of A Coruña, Spain, in 2007 and 2014, respectively. Since 2007, he has been with the Group of Electronic Technology and Communications (GTEC), Department of Electronics and Systems, University of A Coruña, where he had the benefit of a FPI scholarship granted by the Spanish Government, from 2008 to 2012. His main research interests include design of coding schemes, analog joint source-channel coding, multiuser communications, and image processing. He has published 13 articles in international technical journals as well as more than 30 papers in relevant international conferences and workshops in the area of communications and signal processing. He has participated as a research member in more than ten research projects and contracts granted by regional, national, and European administrations. He has also received the Best Student Paper Award at the 14th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Darmstadt, in 2013.

JOSÉ P. GONZÁLEZ-COMA (Member, IEEE) was born in Marín, Spain. He received the computer engineering and the Ph.D. degrees from the University of A Coruña, Spain, in 2009 and 2015, respectively. He was with the Group of Electronic Technology and Communications (GTEC), University of A Coruña, in 2009, where he received the FPI Grant from the Ministerio de Ciencia e Innovación, in 2011. He was appointed as a Visiting Researcher with the Associate Institute for Signal Processing, Technische Universität München, Germany, in 2012, and the Signal Processing in Communications Group (UVIGO), Spain, in 2017. Since 2020, he has been with the Defense University Center at the Spanish Naval Academy. His main research interests include channel estimation and precoding in massive multi-in multi-out systems, and millimeter-wave communications.

LUIS CASTEDO (Senior Member, IEEE) received the Ph.D. degree in telecommunications engineering from the Technical University of Madrid, Spain, in 1993. Since 1994, he has been a Faculty Member with the Department of Computer Engineering, University of A Coruña (UDC), Spain, where he became a Professor, in 2001, and acted as the Chairman, from 2003 to 2009. He had previously held several research appointments at the University of Southern California (USC) and École supérieure d'électricité (SUPELEC). From 2014 to 2018, he has been a Manager of the Communications and Electronic Technologies (TEC) program, State Research Agency, Spain. He has also been the Principal Researcher of more than 50 research projects funded by public organisms and private companies. He has coauthored more than 300 papers in peer-reviewed international journals and conferences. His research interests include signal processing for wireless communications and prototyping of digital communication equipments. His papers have received three Best Student Paper Awards at the IEEE/ITG Workshop on Smart Antennas, in 2007, the IEEE International Workshop on Signal Processing Advances in Wireless Communications, in 2013, and the IEEE International Conference on Internet of Things (iThings), in 2017. He has been a General Co-Chair of the 8th IEEE Sensor Array and Multichannel Signal Processing Workshop, in 2014, and the 27th European Signal Processing Conference, in 2019.