Received 27 January 2020; accepted 6 February 2020. Date of publication 11 February 2020; date of current version 26 February 2020.

Digital Object Identifier 10.1109/OJCOMS.2020.2973185

Compressed Sensing Based Non-Orthogonal Pilot Designs for Multi-Antenna Grant-Free Access Systems

AYON QUAYUM (Student Member, IEEE), AND HLAING MINN (Fellow, IEEE)

Electrical and Computer Engineering Department, University of Texas at Dallas, Richardson, TX 75080, USA

CORRESPONDING AUTHOR: A. QUAYUM (e-mail: ayonqm@gmail.com)

ABSTRACT Grant-free access is an attractive approach to enable spectrum-efficient low-latency access for systems with massive number of users. Pilot design plays a crucial role for grant-free access as it needs to provide a large number of access codes with fast collision detection capability and good channel estimation performance. Recently, pilot designs with fast collision detection capability have been proposed for compressed sensing (CS) based channel estimation. But the existing designs are not optimized for the estimation of highly sparse and block-sparse channels. In this paper, we present several propositions related to the performances of the CS based sparse and block-sparse channel estimation. Utilizing these propositions, we develop a novel non-orthogonal pilot design with fast collision detection capability for grant-free access in block-sparse channels. We also propose two methods to optimize the Peak-to-Average Power Ratio (PAPR) of the proposed non-orthogonal pilot codes. The simulation results illustrate that the proposed design provides similar or better channel estimation and collision detection performances with much better pilot resource efficiency when compared to the existing designs. Finally, we investigate the trade-offs among different design parameters and the channel estimation performances to facilitate better design choices.

INDEX TERMS Compressed sensing, grant-free access, MIMO, non-orthogonal pilot, sparse channel.

I. INTRODUCTION

THE EMERGENCE of new/future wireless applications in various sectors including self-driving cars, video broadcasting on social media and smart cities are driving the demand for higher uplink capacity in terms of data throughput and user access with low latency in wireless communication systems. The current uplink access techniques are not designed to support the high density of devices, which will be orders of magnitude higher, with low latency for fifth generation (5G) and other future standards [1]. In current wireless standards, grant-based access is the most common uplink access technique. It needs initial signaling protocols to establish connection between user and the base station (BS). This requires valuable time and spectrum resources for signaling which will become excessive for supporting high density of devices. Grant-free access is an attractive alternative access technique which can reduce the latency and signaling load by eliminating random access and resource allocation procedures [2]. In grant-free access systems, a user can directly send data and embedded pilot signals without waiting for resource allocation.

The conventional orthogonal pilot codes are not suitable for grant-free access systems with large number of users because of two reasons. First, they can support only a limited number of pilot codes which is insufficient for large number of grant-free uplink access users. Second, the conventional orthogonal pilot codes are not designed to facilitate fast collision detection at the receiver which is important for grant-free access systems. In grant-based access systems, the uplink pilot collision is resolved during random access procedure. As the grant-free access systems bypass the random access and resource allocation procedures and there are no dedicated access grant procedures among the users, any pilot collision needs to be detected by the BS from the initial transmission. Compared to the orthogonal pilot codes, non-orthogonal pilot codes can be designed to support much larger number of users, thus increasing the efficiency of the pilot resource utilization [3]. They can be used without

pre-assignment and be designed to facilitate fast collision detection at the receiver. When designed with these qualities, non-orthogonal pilot codes make a better choice for grant-free access systems compared to the orthogonal pilot codes.

An uplink grant-free transmission based on sparse code multiple access (SCMA) with non-orthogonal pilot codes was discussed in [4]. But the SCMA pilot codes were not designed to enable fast collision detection at the receiver. Non-orthogonal pilot designs were also developed based on Zadoff-Chu sequence in [5], [6] and Reed-Muller sequence in [7]. But they were designed for non-sparse channels in a single transmit antenna system without collision detection capability. Another non-orthogonal pilot design was proposed in [8] using the assumption that the channel correlation matrix was known. Such assumption is not applicable for grant-free access systems.

Grant-free access was also discussed in [9], [10] within the context of massive connectivity in Internet of Things (IoT) and massive machine type communication (mMTC) with sporadic traffic. Joint activity detection and channel estimation was proposed by exploiting the sparsity of the device activity pattern. Pilot entries were generated from independent and identically distributed (i.i.d.) complex Gaussian distribution for these schemes. In [11], the pilot codes were generated by sampling an i.i.d. symmetric Bernoulli distribution for grant-free access in the same context. They also assumed single transmit antenna and did not offer collision detection capability.

A pilot design based on on-off type non-orthogonal pilot codes with collision-detection capability has been proposed in [12]. The non-orthogonal pilot codes (grant-free access codes) are created by selecting P non-zero pilot tones and P' null pilot tones from a set of $P_{\text{tot}} = P + P'$ pilot tones. Different combinations of non-zero and null pilot tones create different non-orthogonal pilot codes. At the receiver, more than P non-zero pilot tones indicate a pilot collision and could be quickly detected. A deployment of these non-orthogonal pilot codes within densely populated remote radio heads (RRH) [13] can support a higher number of users compared to the orthogonal pilot codes [14]. A more advanced non-orthogonal pilot design has been proposed in [15] which addresses both fast collision detection capability and optimized channel estimation performances.

While designing grant-free access codes, especially for multiple input multiple output (MIMO) systems with multiple antennas, large pilot overhead becomes the limiting factor for supporting massive connectivity. A promising way to optimize the pilot codes in terms of pilot resource usage is by taking advantage of the time domain sparsity of the MIMO channels which occurs in several important wideband systems [16]–[18]. Clustered multipath components in delay domain have been observed by Saleh and Valenzuela [19]. The sparse cluster positions which are common across different antennas in the MIMO systems make it possible to approximate the MIMO channels by only a few non-zero time domain channel coefficients [20]. It has been shown that the approximated sparse channel could be estimated successfully using compressed sensing (CS) based techniques with a smaller number of pilot tones compared to traditional approaches such as the least-square (LS) based technique [20], [21]. Among different CS algorithms, linear programming based Dantzig selector was proposed in [22], orthogonal matching pursuit (OMP) was used in [23], and subspace pursuit was developed in [24]. Several orthogonal pilot designs were proposed for CS based channel estimation in [25]–[31]. But they offer neither non-orthogonal pilot designs nor pilot collision detection capabilities. Using Dantzig selector based CS algorithm, non-orthogonal pilot codes with collision detection property were proposed in [32]. But the authors did not consider channel estimation performances in their pilot design. Better non-orthogonal pilot designs for multi-users CS based channel estimation, with both optimized channel estimation performances and collision detection property, have been proposed in [15].

The CS based non-orthogonal pilot designs in [15] provide optimized channel estimation performance but with some limitations. First, in this design, only a single configuration of the pilot codes is possible for a specific system where the number of pilot tones in each pilot code is equal to the half of the maximum channel length. When the number of non-zero channel taps is small compared to the maximum channel length, this design results in longer than necessary pilot codes (inefficient pilot resource utilization). Second, although the existing non-orthogonal pilot designs (e.g., [15]) can be extended for MIMO channels, they are not optimized to take advantage of the block-sparsity property of the MIMO channels where the locations of the non-zero channel taps are the same across all the channels [33]. Third, the Peak-to-Average Power Ratio (PAPR) is not considered in the existing non-orthogonal pilot designs, thus these pilot codes may experience nonlinear distortions due to high PAPR.

By taking advantage of the block-sparsity property of MIMO channels, a block optimized orthogonal matching pursuit (BOOMP) algorithm has been proposed for estimation of block-sparse channels in [33]. In [34], the authors have proposed optimized orthogonal pilot design for block-sparse channels using genetic algorithm (GA). But the orthogonal pilot design in [34] is optimized for single user scenario and does not support large number of pilot codes with similar channel estimation performances as needed for grant-free access in multiuser scenario.

In this paper, we propose a non-orthogonal pilot design for uplink grant-free access which overcomes the above-mentioned limitations of the existing pilot designs. The major contributions of this paper are listed below.

1) We first prove several propositions which establish the equivalencies of the block-sparse channel estimation performances among the pilot codes with certain different configurations. We also analyze the upper bound of the channel estimation performance metric

in terms of the cyclic differences of the pilot tone indexes within structured pilot codes for sparse and block-sparse channels.

2) We propose a novel non-orthogonal pilot design for block-sparse channels with three critical characteristics which are essential for grant-free access but not fully supported by any of the existing pilot designs. First, using the propositions, we develop a structured multi-user pilot design that ensures similar channel estimation performances for all the users while optimizing the pilot resource usage. Second, the proposed pilot design provides large number of non-orthogonal pilot codes with collision detection capabilities for grant-free access. Finally, the proposed pilot design is optimized for both sparse and block-sparse channels in terms of channel estimation performances and pilot resource usage.
3) We present two methods to optimize the PAPR of the proposed non-orthogonal pilot codes. The proposed methods offer large gains in terms of PAPR reduction and low memory requirement for large number of access codes.
4) Analytical expressions for performance of pilot code detection/collision are also presented.
5) Using simulations, we show the superiority of our pilot design in terms of pilot resource utilization compared to existing non-orthogonal pilot designs for both sparse and block-sparse channels while maintaining similar/better channel estimation performance. Through simulation results, we also illustrate the trade-offs among different pilot configurations in terms of the total number of pilot resources, the number of available grant-free access codes and the channel estimation performances. Such trade-off flexibility is not available for the existing pilot design in [15] where only a single configuration is possible for any specific system.

The remainder of this paper is organized as follows. Section II describes the system model. In Section III, we present several propositions which characterize the channel estimation performances of certain structured pilot codes for block-sparse channels. Next, Section IV develops the novel non-orthogonal pilot design for grant-free access. Section V provides performance comparisons between the proposed design and the existing ones in terms of pilot resource usage, complexity, channel estimation and pilot collision detection performances. Finally, Section VI concludes the paper.

We will use the following notations in this paper. $\boldsymbol{A}$ is a matrix, $\boldsymbol{a}$ is a vector, $\{\boldsymbol{A}\}$ is a set and a is a scalar. a_{ij} represents the element in i-th column and j-th row of the matrix $\boldsymbol{A}$ and $\boldsymbol{a}_k$ represents the k-th column of the matrix $\boldsymbol{A}$. $||\boldsymbol{a}||$ represents the Euclidean norm of $\boldsymbol{a}$ and $|a|$ is the absolute value of a. diag$\{\boldsymbol{a}\}$ is a square diagonal matrix with the vector $\boldsymbol{a}$ as the diagonal elements. $(.)^T$ and $(.)^H$ represent the transpose and hermitian operators, $\setminus$ indicates the set difference, $\lceil\cdot\rceil$ denotes the ceiling operation, $\lfloor\cdot\rfloor$ denotes the floor operation and $\lfloor\cdot\rceil$ denotes the rounding operation to the nearest integer.

II. SYSTEM AND SIGNAL MODEL

We consider an uplink MIMO system with a high density of devices. To enable spectrum-efficient low-latency access for massive number of devices, we focus on a grant-free access system. In such a system, a user can directly send its data and embedded pilot code without needing to go through random access and resource allocation procedures. The immediate advantages are substantial reductions of access latency and signaling overhead, which will become more prominent for systems with massive number of devices. A crucial component for grant-free access is the set of pilot codes which serve for dual purposes of multiple access and channel knowledge acquisition. To handle massive number of users, traditionally used orthogonal pilot codes are insufficient but non-orthogonal pilot codes offer an efficient solution. Thus, we focus on non-orthogonal pilot designs. However, in the first stage of our design, we develop orthogonal pilot codes. Then, building up on them, we develop non-orthogonal pilot codes. We note that our orthogonal pilot codes can also be used in the grant-based access of the current cellular wireless systems to offer improved performance in terms of pilot resource utilization and channel estimation. In this regard, our technical developments cover both grant-based and grant-free access systems.

For grant-free access, we consider a densely deployed small cells system. A user will randomly select a non-orthogonal pilot code with collision detection capability, and send its pilot-embedded data packet to the BS. Fast collision detection capability enables the receivers to quickly determine the viability of the received packet. If there is no collision detected, channel estimation and user data detection take place. To accommodate easy collision detection for large number of devices with a large system bandwidth, we consider orthogonal frequency division multiplexing (OFDM) transmission. Its fine resource granularity in frequency domain offers numerous orthogonal pilot codes which are disjoint in subcarrier domain, thus yielding easy decoupling among orthogonal users. Collision detection of non-orthogonal pilot codes can also be easily accomplished by comparing the received energies of null tones of the pilot codes with a threshold (e.g., as in [15]).

For presentation convenience, we will first present a signal model for the single input single output (SISO) system and then extend it for the MIMO system. Descriptions and dimensions of some important symbols used in the signal model are listed in Table 1 for easy reference. For the OFDM system with discrete Fourier transform (DFT) of size N, define the N-point unitary DFT matrix $\boldsymbol{F} = [\boldsymbol{f}_0, \ldots, \boldsymbol{f}_{N-1}]$ where $\boldsymbol{f}_k = \frac{1}{\sqrt{N}}[1, \ldots, e^{-j\frac{2\pi k}{N}(N-1)}]^T$. The cyclic prefix length of OFDM is chosen to be not smaller than the maximum channel delay to avoid inter-symbol interference. Let us consider the channel impulse response (CIR) vector $\boldsymbol{h} = [h_0, h_1, \ldots, h_{(L_h-1)}]^T$ to be sample-spaced and sparse

TABLE 1. Description and dimension of different symbols.

Symbol	Dimension	Description
N	1	DFT size
L_h	1	Channel length
τ	1	Number of non-zero channel taps
M	1	Number of transmit antennas
R	1	Number of receive antennas
P	1	Number of non-zero pilot tones
P'	1	Number of null pilot tones
$\boldsymbol{F}$	$N \times N$	N-point unitary DFT matrix
$\boldsymbol{y}$	$P \times 1$	Frequency domain received pilot vector
$\boldsymbol{h}$	$L_h \times 1$	Channel impulse response vector
$\boldsymbol{q}$	$P \times 1$	Subcarrier index vector of pilot tones
$\boldsymbol{c_q}$	$P \times 1$	Frequency domain pilot symbol vector
$\boldsymbol{C}$	$P \times P$	Pilot symbol matrix (diag $\{\boldsymbol{c_q}\}$)
$\boldsymbol{F_q}$	$P \times L_h$	$\boldsymbol{F_q}$ is selected by taking P rows, corresponding to $\boldsymbol{q}$, from the first L_h columns of $\boldsymbol{F}$
$\boldsymbol{A}(\boldsymbol{q})$	$P \times L_h$	Dictionary matrix ($\sqrt{N}\boldsymbol{C}\boldsymbol{F_q}$)
$\bar{\boldsymbol{h}}$	$L_h MR \times 1$	Block-sparse channel vector, $\bar{\boldsymbol{h}}_i = [h_i^{(0,0)}, \ldots, h_i^{(M-1,R-1)}]^T$
$\bar{\boldsymbol{y}}$	$PMR \times 1$	Regrouped frequency domain received pilot vector for block-sparse channel, $\bar{\boldsymbol{y}}_i = [y_i^{(0,0)}, \ldots, y_i^{(M-1,R-1)}]^T$
$\boldsymbol{D}(\boldsymbol{q})$	$PMR \times L_h MR$	$\boldsymbol{D}_{i,j}(\boldsymbol{q}) =$ diag$(a_{ij}^0(\boldsymbol{q}^{(0)}), \ldots, a_{ij}^{(R-1)}(\boldsymbol{q}^{(M-1)}))$

with τ non-zero channel taps where $\tau \ll L_h$. We assume the effects of timing errors and non-sample-spaced channel delays have been incorporated in $\boldsymbol{h}$. The average power delay profile of $\boldsymbol{h}$ is with a $-20/(L_h - 1)$ dB per tap decay factor so that the average power ratio between h_0 and $h_{(L_h-1)}$ is 20 dB. The locations of the non-zero channel taps are uniformly distributed within $\boldsymbol{h}$ and their average powers are according to the channel power delay profile.

To estimate $\boldsymbol{h}$ of length L_h, we will use a pilot code consisting of P ($< L_h$) non-zero pilot tones. Let $\boldsymbol{q} = [q_0, q_1, \ldots, q_{(P-1)}]^T$ be the subcarrier index vector of P pilot tones and $\boldsymbol{c_q} = [c_{q_0}, c_{q_1}, \ldots, c_{q_{(P-1)}}]^T$ be the corresponding frequency domain pilot symbol vector with c_{q_k} being the pilot symbol transmitted on subcarrier q_k. Define $\boldsymbol{F_q}$ to be a $P \times L_h$ matrix constructed by selecting P rows defined by $\boldsymbol{q}$ from the first L_h columns of $\boldsymbol{F}$. Then, the frequency domain received pilot vector $\boldsymbol{y} = [y_0, y_1, \ldots, y_{P-1}]^T$ extracted from the pilot tone locations is

$$\boldsymbol{y} = \sqrt{N}\boldsymbol{C}\boldsymbol{F_q}\ \boldsymbol{h} + \boldsymbol{n} = \boldsymbol{A}(\boldsymbol{q})\ \boldsymbol{h} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{C} = \text{diag}\{\boldsymbol{c_q}\}$ is the pilot symbol matrix, $\boldsymbol{n} = [n_0, n_1, \ldots, n_{P-1}]^T$ is the zero-mean complex Gaussian noise vector with covariance matrix $\sigma_n^2\boldsymbol{I}$ where $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{A}(\boldsymbol{q}) = \sqrt{N}\boldsymbol{C}\boldsymbol{F_q}$ is the *dictionary matrix* for the pilot tone index set $\boldsymbol{q}$. The performance of the compressed sensing based channel estimation depends on the properties of the dictionary matrix. We will discuss these properties in Section III in more details.

Now let us consider a MIMO system with M transmit and R receive antennas. Let $\boldsymbol{q}^{(m)} = [q_0^{(m)}, q_1^{(m)}, \ldots, q_{(P-1)}^{(m)}]^T$ be the index set of the P pilot tones used for transmit antenna $m \in \{0, 1, \ldots, (M-1)\}$. We impose the pilot design condition that $\boldsymbol{q}^{(m)} \cap \boldsymbol{q}^{(k)} = \emptyset\ \forall m \neq k$ to simply maintain orthogonality of the pilot codes for different transmit antennas. Let $\boldsymbol{h}^{(m,r)} = [h_0^{(m,r)}, h_1^{(m,r)}, \ldots, h_{(L_h-1)}^{(m,r)}]^T$ be the CIR vector from transmit antenna m to the receive antenna r. Using (1), we can express the received pilot vector from transmit antenna m to receive antenna r as

$$\boldsymbol{y}^{(m,r)} = \boldsymbol{A}^r\left(\boldsymbol{q}^{(m)}\right)\ \boldsymbol{h}^{(m,r)} + \boldsymbol{n}^{(m,r)} \tag{2}$$

where $\boldsymbol{n}^{(m,r)}$ is the corresponding noise vector.

We assume that MIMO channels share the same sparsity property that each channel has τ ($\ll L_h$) non-zero taps and their positions are the same across all the channel vectors $\boldsymbol{h}^{(0,0)}, \ldots, \boldsymbol{h}^{(M-1,R-1)}$ [35], [36]. To exploit this property in MIMO channel estimation, we rearrange our signal model in the following. By regrouping the received pilot vectors at the receiver from all the transmit antennas, we define the vector $\bar{\boldsymbol{y}} = [\bar{\boldsymbol{y}}_0^T, \ldots, \bar{\boldsymbol{y}}_{P-1}^T]^T$, $\bar{\boldsymbol{h}} = [\bar{\boldsymbol{h}}_0^T, \bar{\boldsymbol{h}}_1^T, \ldots, \bar{\boldsymbol{h}}_{(L_h-1)}^T]^T$, and $\bar{\boldsymbol{n}} = [\bar{\boldsymbol{n}}_0^T, \bar{\boldsymbol{n}}_1^T, \ldots, \bar{\boldsymbol{n}}_{P-1}^T]^T$ where $\bar{\boldsymbol{y}}_i = [y_i^{(0,0)}, \ldots, y_i^{(M-1,R-1)}]^T$, $\bar{\boldsymbol{h}}_i = [h_i^{(0,0)}, \ldots, h_i^{(M-1,R-1)}]^T$, and $\bar{\boldsymbol{n}}_i = [n_i^{(0,0)}, \ldots, n_i^{(M-1,R-1)}]^T$. Because of the common sparsity property of the channels, $\bar{\boldsymbol{h}}$ is a block-sparse vector where the blocks are given by $\bar{\boldsymbol{h}}_i$, $i = 0, 1, \ldots, L_h - 1$.

Now, using the elements from the dictionary matrices of all the transmit antennas, we define $\boldsymbol{D}_{i,j}(\boldsymbol{q}) = \text{diag}(a_{ij}^0(\boldsymbol{q}^{(0)}), \ldots, a_{ij}^{(R-1)}(\boldsymbol{q}^{(M-1)}))$ and then the $PMR \times L_h MR$ block-diagonal matrix

$$\boldsymbol{D}(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{D}_{0,0}(\boldsymbol{q}) & \boldsymbol{D}_{0,1}(\boldsymbol{q}) & \ldots & \boldsymbol{D}_{0,(L_h-1)}(\boldsymbol{q}) \\ \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{D}_{(P-1),0}(\boldsymbol{q}) & \boldsymbol{D}_{(P-1),1}(\boldsymbol{q}) & \ldots & \boldsymbol{D}_{(P-1),(L_h-1)}(\boldsymbol{q}) \end{bmatrix}. \tag{3}$$

Then, using (2), we can express $\bar{\boldsymbol{y}}$ as

$$\bar{\boldsymbol{y}} = \boldsymbol{D}(\boldsymbol{q})\ \bar{\boldsymbol{h}} + \bar{\boldsymbol{n}} \tag{4}$$

where $\bar{\boldsymbol{y}}_i = \sum_{l=0}^{L_h-1} \boldsymbol{D}_{i,l}(\boldsymbol{q})\bar{\boldsymbol{h}}_l + \bar{\boldsymbol{n}}_l$, $i = 1, 2, \ldots, P-1$.

It is more efficient to estimate the locations of non-zero blocks from the block-sparse vector $\bar{\boldsymbol{h}}$ using a CS technique [37]. Similar to the dictionary matrix $\boldsymbol{A}(\boldsymbol{q})$ in the SISO scenario, the properties of the matrix $\boldsymbol{D}(\boldsymbol{q})$ determine the performance of the CS based estimation of block-sparse channels. We will discuss the properties of $\boldsymbol{D}(\boldsymbol{q})$ in more details in Section III.

III. PROPERTIES OF THE PILOT CODES IN SPARSE AND BLOCK-SPARSE CHANNELS

In this section, we prove several propositions on the properties of pilot codes as they are important to characterize the channel estimation performances for the block-sparse channels. We start with several definitions.

Definition 1: Define *coherence* $g(\boldsymbol{q})$ of the pilot code with the tones index set $\boldsymbol{q} = [q_0, \ldots, q_{(P-1)}]$ as the maximum

absolute correlation between two columns of the dictionary matrix $\boldsymbol{A}(\boldsymbol{q})$,

$$g(\boldsymbol{q}) = \max_{0\le m<n\le L_h-1} |\langle \boldsymbol{a}_m, \boldsymbol{a}_n\rangle|, \tag{5}$$

where $\boldsymbol{a}_i$ is the ith column of the dictionary matrix $\boldsymbol{A}(\boldsymbol{q})$. Using the definition of dictionary matrix $\boldsymbol{A}(\boldsymbol{q})$ and defining $d = n - m$, the coherence is given by

$$g(\boldsymbol{q}) = \max_{1\le d\le L_h-1} \left| \sum_{k=0}^{P-1} \left|c_{q_k}\right|^2 e^{-j\frac{2\pi}{N} q_k d} \right|. \tag{6}$$

Coherence is a good metric of the CS based channel estimation performance [38], [39]. Lower coherence is associated with better channel estimation performance.

Definition 2: For a set of orthogonal pilot codes defined by the pilot tones index sets $\boldsymbol{Q} = \{\boldsymbol{q}^{(0)}, \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(M-1)}\}$, we define the *block-coherence* of $\boldsymbol{Q}$ according to [37], [40] as

$$\boldsymbol{G}(\boldsymbol{Q}) = \frac{1}{M} \max_{k\ne l} \rho\left(\boldsymbol{D}_k^H \boldsymbol{D}_l\right) \tag{7}$$

where $\rho(\boldsymbol{\Theta}) = \lambda_{\max}^{1/2}(\boldsymbol{\Theta}^H\boldsymbol{\Theta})$ is the spectrum norm of matrix $\boldsymbol{\Theta}$ and $\lambda_{\max}(\boldsymbol{\Theta}^H\boldsymbol{\Theta})$ is the largest eigenvalue of the positive semidefinite matrix $\boldsymbol{\Theta}^H\boldsymbol{\Theta}$. Similar to coherence, a lower block-coherence is associated with better channel estimation performances [34]. We will use block-coherence as a metric to design optimized pilot codes for block-sparse channels.

Definition 3: Let the set $\{\boldsymbol{s}\} = \{s_0, s_1, \ldots, s_{M-1}\}$ be a subset of $\{z_N\} = \{0, 1, 2, \ldots, N-1\}$ of cardinality N. Then we define *cyclic difference (CD)* as the cyclical difference between two members of $\boldsymbol{s}$ given by

$$\Delta_{ij} = \left(s_i - s_j\right) \text{ modulo } N, i \ne j. \tag{8}$$

We make the observation that within the members of a set of cardinality M, there are a total of $M(M-1)$ CDs.

Definition 4: Let the set $\{\boldsymbol{w}\} = \{w_0, w_1, \ldots, w_{M-1}\}$ of cardinality M be a subset of $\{z_N\} = \{0, 1, \ldots, N-1\}$ of cardinality N. Then we define *Groupwise Cyclic Difference Set* (GCDS) as a subset $\{\boldsymbol{a}(\boldsymbol{w}, N, K, \Lambda)\} = [a_0, \ldots, a_{K-1}]$ of $\{\boldsymbol{w}\}$ where the cyclic differences within the members of $\{\boldsymbol{a}\}$ take each one of all the possible non-zero values of the members of $\{\boldsymbol{w}\}$ exactly Λ times. There are a total of $K(K-1)$ cyclic differences within the members of $\{\boldsymbol{a}\}$ given by

$$\left(a_i - a_j\right) \text{ modulo } N, \ \forall i \ne j. \tag{9}$$

In the special case when $\{\boldsymbol{w}\} = \{z_N\}$, the subset $\{\boldsymbol{a}(\boldsymbol{w}, N, K, \Lambda)\}$ is commonly defined in literature as *Cyclic Difference Set* (CDS).

Definition 5: For a pilot tones index set $\{\boldsymbol{q}\} = [q_0, q_1, \ldots, q_{(P-1)}]$, define mirror indexed pilot set (MIPS) as $\{\boldsymbol{q}_{\mathrm{R}}\} \triangleq \{N - \boldsymbol{q}\}$ modulo $N = [(N - q_0), \ldots, (N - q_{(P-1)})]$ modulo N and a shifted mirror indexed pilot set (SMIPS) as $\{\boldsymbol{q}_{\mathrm{R}} + b\}$ modulo N for any integer b.

Using [15, Proposition 5], for equal energy pilot tones the coherence $g(\boldsymbol{q}) = g(\boldsymbol{q}_{\mathrm{R}}) = g(\boldsymbol{q}_{\mathrm{R}} + b)$.

Now we define the set $\{\boldsymbol{u} : u_l = \frac{N}{L} l;\ l = 0, 1, 2, \ldots, L-1\}$ where N/L is an integer and $L \ge L_h$. Then, the [15, Proposition 6] states that a pilot tones index set $\{\boldsymbol{q}(\boldsymbol{u}, N, P, \Lambda)\}$ taken from $\{\boldsymbol{u}\}$ with P equal energy pilot tones will have the lowest possible coherence when it is GCDS where $\Lambda = \frac{P(P-1)}{L-1}$. But for most practical scenarios such a set does not exist [41]. In the following proposition we provide an upper bound for the coherence of any pilot tone set chosen from the set $\{\boldsymbol{u}\}$.

Proposition 1: For an OFDM system with DFT size N and channel length L_h, let the pilot tone index set $\{\boldsymbol{q}\}$ consist of the $P < L_h$ pilot tones taken from the subcarriers index set $\{\boldsymbol{u} : u_l = \frac{N}{L} l;\ l = 0, 1, 2, \ldots, L-1\}$. Let the CDs of the value u_l occur with the frequency $n_{u_l}^{(\boldsymbol{q})}$ within $\{\boldsymbol{q}\}$. Then the coherence $g(\boldsymbol{q})$ is bounded by

$$E_p\sqrt{K} \le g(\boldsymbol{q}) \le E_p \sqrt{K + \sum_{m=1}^{L-1} \left|\tilde{n}_{u_l}^{(\boldsymbol{q})}\right|} \tag{10}$$

where $\tilde{n}_{u_l}^{(\boldsymbol{q})} = n_{u_l}^{(\boldsymbol{q})} - \Lambda$, $\Lambda = \frac{P(P-1)}{L-1}$, $K = \frac{P(L-P)}{L-1}$ and E_P is the energy of a pilot tone.

Proof: From the definition of coherence,

$$g(\boldsymbol{q}) = E_p \max_{1\le d\le L_h-1} \left| \sum_{k=0}^{P-1} e^{-j\frac{2\pi}{N} q_k d} \right|. \tag{11}$$

Let us define $\psi_{\boldsymbol{q}}(d) = |\sum_{k=0}^{P-1} e^{-j\frac{2\pi}{N} q_k d}|$, $d \in [1, \ldots, L_h - 1]$. Now we can express $\psi_{\boldsymbol{q}}^2(d)$ as

$$\begin{aligned} \psi_{\boldsymbol{q}}^2(d) &= \sum_{k=0}^{P-1}\sum_{r=0}^{P-1} e^{-j\frac{2\pi}{N}(q_k - q_r)d} \\ &= P + \sum_{k=0}^{P-1}\sum_{r=0, r\ne k}^{P-1} e^{-j\frac{2\pi}{N}(q_k - q_r)d}. \end{aligned} \tag{12}$$

Let CDs among pilot tone indexes be denoted by $\{v\} = (q_k - q_r)$ modulo $N, k \ne r$ and n_v be the number of times v appears in the last term of (12). Then $\psi_{\boldsymbol{q}}^2(d)$ could be written in terms of n_v and v as

$$\psi_{\boldsymbol{q}}^2(d) = P + \sum_{v=1}^{N-1} n_v e^{-j\frac{2\pi}{N} v d}. \tag{13}$$

We make two observations here. First, because of $\boldsymbol{q} \subset \boldsymbol{u}$, the values of v are limited to the possible values of the CDs within index set $\boldsymbol{u}$. The second observation is that, from the definition of $\{\boldsymbol{u}\}$, the CDs within index set $\boldsymbol{u}$ can only have the values from the set $\{u_1, u_2, \ldots, u_{L-1}\}$. Using these two observations, we replace v with u_l and n_v with $n_{u_l}^{(\boldsymbol{q})}$ in (13) where $l = 1, 2, \ldots, L-1$ as

$$\psi_{\boldsymbol{q}}^2(d) = P + \sum_{l=1}^{L-1} n_{u_l}^{(\boldsymbol{q})} e^{-j\frac{2\pi}{N} u_l d}. \tag{14}$$

A total of $P(P-1)$ CDs are possible within the set $\{\boldsymbol{q}\}$ with each one of them taking one of the possible $(L-1)$ CD

values. So, the average frequency of the occurrence of CDs with the value of u_l is $\Lambda = \frac{P(P-1)}{L-1}$. We can express (14) as

$$\psi_{\boldsymbol{q}}^2(d) = \sum_{l=1}^{L-1}\Big(n_{u_l}^{(\boldsymbol{q})} - \Lambda\Big)e^{-j\frac{2\pi}{N}u_l d} + \Lambda\sum_{l=0}^{L-1} e^{-j\frac{2\pi}{N}u_l d} + (P-\Lambda). \tag{15}$$

With $u_l = \frac{N}{L}l$, the second term in (15) vanishes and we have

$$\psi_{\boldsymbol{q}}^2(d) = (P-\Lambda) + \sum_{l=1}^{L-1}\Big(n_{u_l}^{(\boldsymbol{q})} - \Lambda\Big)e^{-j\frac{2\pi}{N}u_l d} = K + \sum_{l=1}^{L-1}\tilde{n}_{u_l}^{(\boldsymbol{q})}e^{-j\frac{2\pi}{N}u_l d}, \tag{16}$$

where $\tilde{n}_{u_l}^{(\boldsymbol{q})} = n_{u_l}^{(\boldsymbol{q})} - \Lambda$ and $K = \frac{P(L-P)}{L-1}$. Now using the fact that by definition $\psi_{\boldsymbol{q}}(d)$ represents an absolute value and applying triangle inequality, we have

$$\psi_{\boldsymbol{q}}^2(d) = K + \sum_{l=1}^{L-1}\tilde{n}_{u_l}^{(\boldsymbol{q})}e^{-j\frac{2\pi}{N}u_l d} \le K + \sum_{l=1}^{L-1}\Big|\tilde{n}_{u_l}^{(\boldsymbol{q})}\Big|. \tag{17}$$

So, maximum of $\psi_{\boldsymbol{q}}(d)$, $d = 1, 2\ldots, L_h-1$, is bounded by

$$\max_{1\le d\le L_h-1}\psi_{\boldsymbol{q}}(d) \le \sqrt{K + \sum_{l=1}^{L-1}\Big|\tilde{n}_{u_l}^{(\boldsymbol{q})}\Big|}. \tag{18}$$

Combining (18) and the lower bound $E_p\sqrt{K} \le g(\boldsymbol{q})$ from [15], we obtain the coherence bounded as given in (10). ■

Proposition 2: For an OFDM system with DFT size N and channel length L_h, let us define a group of M orthogonal pilot tones index sets $\boldsymbol{Q} = \{\boldsymbol{q}^{(m)} : m = 0, 1, \ldots, M-1\}$ where each set consists of $P < L_h$ pilot tones taken from the subcarriers index set $\{\boldsymbol{u} : u_l = \frac{N}{L}l;\ l = 0, 1, \ldots, L-1\}$ where N/L is an integer (L is a power of 2), $L \ge L_h$, and $L/2 \ge MP$. Let $n_{u_l}^{(m)}$ be the total number of CDs within $\boldsymbol{q}^{(m)}$ with the value of u_l, $\Lambda = \frac{P(P-1)}{L-1}$ be the mean frequency of CDs and $K = \frac{P(L-P)}{L-1}$. Then, the block coherence $G(\boldsymbol{Q})$ is bounded by

$$G(\boldsymbol{Q}) \le \frac{E_p}{M}\sqrt{MK + \sum_{m=1}^{M-1}\sum_{l=1}^{L-1}\Big|n_{u_l}^{(m)} - \Lambda\Big|}. \tag{19}$$

Proof: From [33 and 34, Th. 1], we can define the block-coherence $G(\boldsymbol{Q})$ in terms of the coherences of pilot codes with the tones index sets $\{g(\boldsymbol{q}^{(m)})\}$ as

$$G(\boldsymbol{Q}) = \frac{1}{M}\max_{m\in\{0,\ldots,M-1\}} g\Big(\boldsymbol{q}^{(m)}\Big). \tag{20}$$

From Proposition III, the coherence $g(\boldsymbol{q}^m)$ for pilot tones index set $\boldsymbol{q}^{(m)}$ is bounded by

$$g\big(\boldsymbol{q}^m\big) \le E_p\sqrt{K + \sum_{l=1}^{L-1}\Big|n_{u_l}^{(m)} - \Lambda\Big|}. \tag{21}$$

For a group of non-negative values $x_0, x_1, \ldots, x_{M-1}$, we have

$$\max(x_0, x_1, \ldots, x_{M-1}) \le \sum_{m=0}^{M-1} x_m. \tag{22}$$

Using the relationship in (22) and squaring both sides in inequality (21), we obtain

$$\max_{m\in\{1,\ldots,M\}} g^2\Big(\boldsymbol{q}^{(m)}\Big) \le E_p^2\left(MK + \sum_{m=1}^{M-1}\sum_{l=1}^{L-1}\Big|n_{u_l}^{(m)} - \Lambda\Big|\right). \tag{23}$$

As $g(\boldsymbol{q}^m)$ is non-negative, $\max_m(g^2(\boldsymbol{q}^m)) = (\max_m g(\boldsymbol{q}^m))^2$. So from (20) and (23), we arrive at (19). ■

Proposition 3: For equal energy pilot tones, block-coherence is not affected by the constant cyclical shift of the pilot tones, i.e., block-coherence $G(\boldsymbol{Q}) = G(\boldsymbol{Q}_b)$ where $\boldsymbol{Q} = \{\boldsymbol{q}^{(m)} : m = 0, \ldots, M-1\}$ and $\boldsymbol{Q}_b = \{(\boldsymbol{q}^{(m)} + b) \text{ modulo } N : m = 0, 1, \ldots, M-1\}$ for any integer b and DFT size N.

Proof: According to [15, Proposition 4], coherence is not affected by the constant cyclical shift of the pilot tones. So, we have the following coherence equality

$$g\Big(\boldsymbol{q}^{(m)}\Big) = g\Big(\boldsymbol{q}_b^{(m)}\Big) \tag{24}$$

where $\boldsymbol{q}_b^{(m)} = \{(\boldsymbol{q}^{(m)} + b) \text{ modulo } N\}$. Now combining (20) and (24), we can express the block-coherence for $\boldsymbol{Q}_b$ as

$$G(\boldsymbol{Q}_b) = \frac{1}{M}\max_{m\in\{1,\ldots,M\}} g\Big(\boldsymbol{q}_b^{(m)}\Big) = G(\boldsymbol{Q}). \tag{25}$$

■

Proposition 4: For pilot tones index sets $\{\boldsymbol{q}^{(m)}\}$, define mirror indexed pilot set (MIPS) as $\{\boldsymbol{q}_{\mathrm{R}}^{(m)}\} \triangleq \{N - \boldsymbol{q}^{(m)}\}$ modulo N and shifted mirror indexed pilot set (SMIPS) as $\{(\boldsymbol{q}_{\mathrm{R}}^{(m)} + b) \text{ modulo } N\}$ for any integer b. Let $\{\boldsymbol{Q}\} = \{\boldsymbol{q}^{(m)} : m = 0, 1, \ldots, M-1\}$, $\{\boldsymbol{Q}_{\mathrm{R}}\} = \{\boldsymbol{q}_{\mathrm{R}}^{(m)} : m = 0, 1, \ldots, M-1\}$ and $\{\boldsymbol{Q}_{\mathrm{R}} + b\} = \{(\boldsymbol{q}_{\mathrm{R}}^{(m)} + b) \text{ modulo } N : m = 0, 1, \ldots, M-1\}$ for any integer b. Then for equal energy pilot tones, the following block-coherence equality holds: $G(\boldsymbol{Q}) = G(\boldsymbol{Q}_{\mathrm{R}}) = G(\boldsymbol{Q}_{\mathrm{R}} + b)$.

Proof: Reference [15, Proposition 5] states that the coherences for two pilot tones sets are the same if they are MIPS of each other. So, we have

$$g\big(\boldsymbol{q}^m\big) = g\big(\boldsymbol{q}_{\mathrm{R}}^m\big). \tag{26}$$

From (20) and (26), we obtain

$$G\big(\boldsymbol{Q}_{\mathrm{R}}\big) = \frac{1}{M}\max_{m\in\{1,\ldots,M\}} g\big(\boldsymbol{q}_{\mathrm{R}}^m\big) = G(\boldsymbol{Q}). \tag{27}$$

Now, using Proposition III and (27), we have

$$G(\boldsymbol{Q}) = G\big(\boldsymbol{Q}_{\mathrm{R}}\big) = G\big(\boldsymbol{Q}_{\mathrm{R}} + b\big). \tag{28}$$

■

IV. NON-ORTHOGONAL PILOT DESIGN FOR GRANT-FREE ACCESS IN SPARSE AND BLOCK-SPARSE CHANNELS

In this section, we develop a novel non-orthogonal pilot design with fast collision detection capability for grant-free access which is optimized for block-sparse channels. While CS based pilot design for a sparse channel has been discussed in [15], CS based estimation of block-sparse MIMO channels have not been considered for the pilot design. Our proposed pilot design have following new contributions compared to the reference design in [15]:

1) The reference pilot design in [15] did not consider MIMO channels. In the proposed pilot design, we have leveraged the block-sparsity property of the MIMO channels to reduce the number of pilot resources needed for the multiple antenna system (up to 62.5% reduction for a system with 4 transmit antennas).
2) The reference pilot design provides a single configuration for each system requiring fixed length pilot codes without considering any system requirements. Overcoming this limitation, the proposed pilot design provides flexible configurations with adjustable parameters such as the pilot code length and the number of grant-free access codes. Using the proposed pilot design, a better pilot configuration could be chosen considering the number of users, the availability of the pilot resources, and other system requirements.
3) The PAPR was not considered in the reference pilot design. The proposed pilot design offers large reduction of PAPR with low memory requirements.
4) The proposed pilot design provides better pilot detection performances compared to the reference pilot design.

Our pilot design is based on the following criteria:

- The non-orthogonal pilot design should be modular for easy pilot code generation.
- The receiver should be able to detect a collision easily when more than one non-orthogonal pilot codes with some common subcarriers are received.
- Pilot codes should be optimized for block-sparse channel estimation performances.
- The channel estimation performances of different non-orthogonal pilot codes should be similar to ensure fairness across users.
- The proposed non-orthogonal pilot design should optimize the pilot resource utilization while providing a large number of non-orthogonal pilot codes.

We develop our pilot design in two steps.

1) In the first step, we design modular orthogonal root pilot codes, each with P_{tot} pilot tones.
2) In the second step, we design several non-orthogonal pilot codes from each root pilot code.

Specifically, we generate non-orthogonal pilot codes from each root pilot code by choosing different combinations of the positions of P' null pilot tones within the allocated $P_{\text{tot}} = P + P'$ pilot tones with some constraints. Thus, each final pilot code has P non-zero pilot tones and P' null pilot tones. Our design yields several groups of non-orthogonal pilot codes where the pilot codes within each group are non-orthogonal but those from different groups are orthogonal.

Note that we assign the pilot codes for different transmit antennas of each user to be orthogonal (i.e., they are from different root pilot codes). However, pilot codes of different users are mostly non-orthogonal. At the receiver, detection of more than P non-zero pilot tones within the corresponding P_{tot} pilot tones of a pilot code indicates a pilot code collision.

Details of the two design steps are described in Sections IV-A and IV-B. We also show that the non-orthogonal pilot design for sparse channels could be obtained as a special case of the proposed design in Section IV-C. Sections IV-D and IV-E discuss the PAPR optimization and threshold based pilot detection for the proposed pilot design.

A. STEP 1: DESIGN OF ROOT PILOT CODES

Possessing a modular structure among pilot codes is much desirable as it simplifies pilot code generation. Thus, we consider orthogonal modular structure to develop root pilot codes with P_{tot} pilot tones. The root pilot codes are designed based on following criteria:

- Different pilot codes should be orthogonal to each other with some modularity for easy signal generation.
- Pilot design should ensure similar channel estimation performances for different pilot codes.
- Block-sparse channel estimation performances of the pilot codes should be optimized.

We first consider the length of pilot code P_{tot} which includes P non-zero and P' null pilot tones. Pilot codes with more non-zero pilot tones can provide better channel estimation performances but may reduce the total number of available pilot codes. The appropriate number for P will depend on the system performance requirements and resource availability. Note that when we turn a root pilot code into several non-orthogonal pilot codes in Step 2, the channel estimation performance of the non-orthogonal pilot codes will be slightly degraded from that of the root pilot code due to changing some non-zero pilot tones into null tones. To account for such degradation, we choose a number for P which provides slightly better channel estimation performance (for the root code) than the desired performance (for the non-orthogonal code).

Next, we decide the number of null pilot tones per antenna P'. An increasing number of null pilot tones will generate more non-orthogonal pilot codes. But a greater number of null pilot tones will also use more pilot resources. The best choice for P' while fixing P will depend on the total number of users and the availability of the pilot resources.

To design the pilot codes with $P_{\text{tot}} = P + P'$ pilot tones for MIMO block-sparse channels, we consider a MIMO system

with DFT size N, M transmit antennas per user, block-sparse channels of length L_h and the number of non-zero channel taps τ where $\tau \ll L_h$. Let each transmit antenna be equipped with a pilot code of length P_{tot}.

In our orthogonal design, subcarriers of different pilot codes are disjoint to guarantee orthogonality of the pilot codes in frequency. We first divide all the subcarriers into several orthogonal groups of length L. We define L as a power of 2 with $L \geq L_h$, N/L being an integer, and an additional condition of $L/2 \geq MP_{\text{tot}}$. Note that for scenarios with $L/2 < MP_{\text{tot}}$, we can split M antennas into several antenna sub-groups (e.g., two sub-groups, the first with M_1 antennas and the second with M_2 antennas where $M_1 + M_2 = M$) and design the pilots of different antenna sub-groups using different OFDM symbols. Thus, without loss of generality, in the following, we will present the design assuming $L/2 \geq MP_{\text{tot}}$.

We define the above-mentioned orthogonal subcarrier groups as $\{\boldsymbol{U}_i : i = 0, 1, \ldots, \frac{N}{L} - 1\}$ where $\boldsymbol{U}_i = \{(N/L)l + i : l = 0, 1, \ldots, L - 1\}$ using the same considerations and steps that have been used in [15]. In fact, $\{\boldsymbol{U}_i\}$ corresponds to the frequency division multiplexing pilot design in [42].

Next, using Proposition III, we design optimized tone indexes of pilot codes for two users from the first subcarrier group $\boldsymbol{U}_0$ with similar channel estimation performances. To achieve this, we design the two sets of pilot codes (one set per user) to be SMIPS of each other. According to Proposition III, the block-coherence of the two sets of pilot codes, selected with SMIPS property, will be the same, hence ensuring similar channel estimation performances. Each user needs M pilot codes to be used by M transmit antennas. For two users, tone indexes for a total of $2M$ pilot codes with P_{tot} pilot tones in each code are selected from $\boldsymbol{U}_0$. The pilot design can easily be extended to select the pilot codes for $2S$ users from $\boldsymbol{U}_0$ by selecting two sets of SMIPS pilot codes with a total of SM pilot codes in each one with the condition $L/2 \geq SMP_{\text{tot}}$. Without loss of generality, we will assume $S = 1$ in following discussion.

Next, we design the tone indexes of the pilot codes from other subcarrier groups $\{\boldsymbol{U}_i\}$, $i = 1, 2, \ldots, \frac{N}{L} - 1$, by shifting the tone indexes of the pilot codes of $\boldsymbol{U}_0$ by i subcarriers. Proposition III ensures that all resulting sets of pilot codes from different subcarrier groups will have the same block-coherence which in turn guarantees similar channel estimation performances for all users. This modularity of the pilot codes also facilitates easy pilot code generation. Now our design problem just needs to focus on selecting two sets of tone indexes of pilot codes from $\boldsymbol{U}_0$ that are SMIPS of each other and will yield optimized channel estimation performances for block-sparse channels. Recall that a lower value of the block-coherence indicates a better channel estimation performance. Thus, our design aims to minimize the block-coherence of the pilot codes. As the complexity of using the exact block-coherence metric is prohibitively large, we resort to minimizing the upper bound of the block-coherence.

To solve this pilot design problem, we use a game theoretic approach. We observe that the design problem could be formulated as an M persons dynamic game [43]. We consider M players representing M antennas of a user. Each player starts with an empty pilot tone index set and chooses P_{tot} pilot tone indexes during the game. The goal of the game is to minimize the cost function. The cost function is based on the upper bound of the block-coherence from Proposition III. As the cost function is calculated by computing the cyclic differences for each player separately, it has very low computational complexity of $\mathcal{O}(P^2)$ compared to the complexity of $\mathcal{O}(M^5 L_h^2 P)$ for computing the block-coherence directly using Definition III. There are a total of P_{tot} rounds of play. In each round each player takes a turn to choose one pilot tone from the available pilot resources. After the turn, the chosen pilot tone and its shifted mirror indexed pilot tone are removed from the available pilot resource set.

In each turn, the player uses a strategy of directly minimizing the cost function when choosing the pilot tone. Each player considers his own previously selected pilot tones indexes while choosing the next pilot tone index from the available pilot resources. Thus, the cost function for the j-th player is defined as the term from the upper limit of the blcok-coherence in Proposition III which is related to $\{\boldsymbol{q}^{(j)}\}$ given by

$$\boldsymbol{\beta}_j\left(\boldsymbol{q}^{(j)}\right) = \sum_{l=1}^{L-1} \left| n_{ul}^{(j)} - \Lambda \right|. \tag{29}$$

In the first round, all the choices for the first pilot tone selection are equal with respect to cost function minimization as there is zero cyclic difference within a set of cardinality one. Thus for simplicity, in our strategy, all players select the first available pilot tone with the minimum index value in the first round. For other rounds, player j first calculates the cost resulting from adding an available pilot tone according to (29) for all possible choices of pilot tone from the available pilot resources and then simply chooses the one which yields the minimum cost.

Algorithm **1** describes these steps for choosing optimized tone indexes of pilot codes. The algorithm takes the number of pilot tones in each pilot code P_{tot}, the number of transmit antennas per user M, the subcarrier group length L of $\boldsymbol{U}_0$, and DFT size N as inputs. Note that $\boldsymbol{U}_0$ has been determined earlier in this section. The outputs of the algorithm are two groups of pilot tones index sets $\{\boldsymbol{Q}_i\}$ and $\{\boldsymbol{Q}_{\text{R}i}\}$, $i = 0, 1, \ldots, \frac{N}{L} - 1$, for $2N/L$ users with M antennas each. In line 1 of Algorithm **1**, we initialize pilot resource set $\boldsymbol{\mathcal{A}} = \boldsymbol{U}_0$ and $\Lambda = \frac{P_{\text{tot}}(P_{\text{tot}}-1)}{L-1}$. Next, we initialize (line 2) the first group of M pilot tones index sets $\boldsymbol{Q}_0 = \{\boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(M-1)}\}$ and their corresponding mirror indexed pilot sets $\boldsymbol{Q}_{\text{R}0} = \{\boldsymbol{q}_{\text{R}}^{(0)}, \ldots, \boldsymbol{q}_{\text{R}}^{(M-1)}\}$ as null.

Then the pilot tones index sets $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_{\text{R}0}$ are chosen from $\boldsymbol{\mathcal{A}}$ which are SMIPS of each other. There are a total of P_{tot} iterations (line 3-14 in Algorithm **1**), corresponding to the P_{tot} rounds of game play, for selection of P_{tot} pilot tone

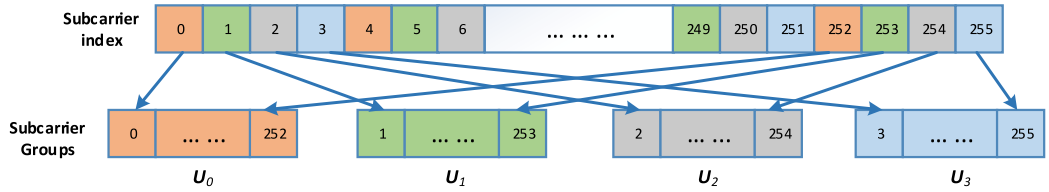

FIGURE 1. Subcarrier groups $\{U_j\}$ for the system with $N = 256$ and $L = 64$.

Algorithm 1 Orthogonal Pilot Design for Block-Sparse Channels.

Input: M, P_{tot}, L, N

Output: $\{\boldsymbol{Q}_i\}$, $\{\boldsymbol{Q}_{\text{R}i}\}$ where $i = 0, 1, \ldots, \frac{N}{L} - 1$

1: **Initialization**: $\boldsymbol{\mathcal{A}} = \boldsymbol{U}_0 = \{(N/L)l : l = 0, 1, \ldots, L-1\}$, $\Lambda = \frac{P_{\text{tot}}(P_{\text{tot}}-1)}{L-1}$,

2: $\boldsymbol{Q}_0 = \{\boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(M-1)}\} = \{\emptyset, \ldots, \emptyset\}$, $\boldsymbol{Q}_{\text{R}0} = \{\boldsymbol{q}_{\text{R}}^{(0)}, \ldots, \boldsymbol{q}_{\text{R}}^{(M-1)}\} = \{\emptyset, \ldots, \emptyset\}$.

3: **for** $k = 0$ to $P_{\text{tot}} - 1$ **do**

4: **for** $m = 0$ to $M - 1$ **do**

5: **if** $k = 0$ **then**

6: $q_0 = a_0$

7: **else**

8: Choose $q = \min_a\left(\sum_{l=1}^{L-1}\left|n'_{u_l}(\boldsymbol{q}_a^{(m)})\right|\right)$ where $\boldsymbol{q}_a^{(m)} = \boldsymbol{q}^{(m)} \cup a$ and $a \in \boldsymbol{\mathcal{A}}$

9: **end if**

10: $q_{\text{Rk}} = N - q_k - \frac{N}{L}$

11: $\boldsymbol{q}^{(m)} = \boldsymbol{q}^{(m)} \cup q_k$, $\boldsymbol{q}_{\text{R}}^{(m)} = \boldsymbol{q}_{\text{R}}^{(m)} \cup q_{\text{Rk}}$

12: $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}} \setminus \{q_k, q_{\text{Rk}}\}$

13: **end for**

14: **end for**

15: $\boldsymbol{Q}_i = \boldsymbol{Q}_0 + i$, $\boldsymbol{Q}_{\text{R}i} = \boldsymbol{Q}_{\text{R}0} + i$ where $i = 1, 2, \ldots, \frac{N}{L} - 1$.

indexes of each pilot code. Within each of these iterations there are M turns (line 4-13), one for each of the M players, to choose one pilot tone. In first of the P_{tot} iterations (line 6), each player chooses the first available pilot tone index a_0 from the set $\boldsymbol{\mathcal{A}}$. In subsequent iterations (line 8), the player j chooses the pilot tone index from the available subcarrier indexes that minimizes the absolute deviations of CDs within $\{\boldsymbol{q}^{(j)}\}$ as defined in the cost function (29).

Let the chosen pilot tone index for player j in k-th turn be q_k. Then, we choose the shifted mirror indexed pilot tone index q_{Rk} from subcarrier group $\{\boldsymbol{U}_0\}$ in line 10 of Algorithm **1** as

$$q_{\text{Rk}} = N - q_k - N/L. \tag{30}$$

We first add q_k to the pilot tone index set $\{\boldsymbol{q}^{(j)}\} = \{\boldsymbol{q}^{(j)}\} \cup q_k$ and q_{Rk} to pilot tone index set $\{\boldsymbol{q}_{\text{R}}^{(j)}\} = \{\boldsymbol{q}_{\text{R}}^{(j)}\} \cup q_{\text{Rk}}$ in line 11 of Algorithm **1**. Then q_k and q_{R_k} are removed from the available pilot resources $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}} \setminus \{q_k, q_{\text{Rk}}\}$ before the next turn (line 12).

After selecting $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_{\text{R}0}$, we select the pilot codes from other subcarrier groups $\{\boldsymbol{U}_i\}$ as $\boldsymbol{Q}_i = \boldsymbol{Q}_0 + i$ and $\boldsymbol{Q}_{\text{R}i} = \boldsymbol{Q}_{\text{R}0} + i$ where $i = 1, 2, \ldots, \frac{N}{L} - 1$ (line 15).

Now, we provide an example of the pilot design using Algorithm **1** for block-sparse channels. For illustration purposes, we will use the system with DFT size $N = 256$, $M = 2$ antennas per user, $P_{\text{tot}} = 14$ pilot tones per antenna, the block-sparse channel length $L_h = 64$, subcarrier groups $\{\boldsymbol{U}_i\}$ of length $L = 64$ each and $\tau = 4$ non-zero taps per channel. We start by dividing the subcarriers into $N/L = 4$ orthogonal groups. The groups are given by $\{\boldsymbol{U}_i\} = \{(4k + i) : k = 0, 1, \ldots, 63;\ i = 0, \ldots, 3\}$. Fig. 1 illustrates the process of configuring these subcarrier groups.

We use Algorithm **1** to select $M = 2$ optimized pilot codes and their corresponding $M = 2$ SMIPS codes from $\boldsymbol{U}_0$. We set up the dynamic game with two players and the initial available pilot resource set $\boldsymbol{\mathcal{A}} = \boldsymbol{U}_0 = \{0, 4, \ldots, 252\}$. Each player starts with a null set for the pilot tone index set. After the first iteration, the pilot tone indexes for two players are $\boldsymbol{q}^{(0)} = \{0\}$ and $\boldsymbol{q}^{(1)} = \{4\}$. For next 13 iterations, each player selects the pilot tone index that minimizes the cost in (29). The complete selected pilot tone index sets for the two antennas using Algorithm **1** are shown in Fig. 2(A) which are $\boldsymbol{q}^{(0)} = \{0, 8, 16, 28, 32, 40, 68, 84, 88, 112, 148, 160, 188, 200\}$ and $\boldsymbol{q}^{(1)} = \{4, 12, 20, 24, 36, 56, 60, 80, 116, 132, 144, 176, 180, 208\}$. We select the SMIPSs from $\boldsymbol{U}_0$ using (30) as $\boldsymbol{q}_{\text{R}}^{(0)} = 252 - \boldsymbol{q}^{(0)}$ and $\boldsymbol{q}_{\text{R}}^{(1)} = 252 - \boldsymbol{q}^{(1)}$ which yield pilot codes of the two transmit antennas for another user.

The complexity of Algorithm **1** is $\mathcal{O}(MLP_{\text{tot}})$. The complexity is linear in terms of both the total number of transmit antennas and the total number of pilot tones per pilot code. This ensures reasonable complexity cost when Algorithm **1** is used in a system with large number of transmit antennas. Also we note that the complexity remains constant with respect to the DFT size N which is usually much larger than L. One of the advantages of our modular design is that all of the orthogonal root codes can be generated from a much smaller subset of the root codes. Algorithm **1** can be used once offline to generate the pilot codes $\{\boldsymbol{q}^{(0)}, \ldots, \boldsymbol{q}^{(M-1)}\}$ and then they can be stored in the user devices, thus eliminating the need of using Algorithm **1** by the user devices with limited computing powers. All other orthogonal root pilot codes can be generated directly by selecting MIPS and shifted version of the stored pilot codes. For our example mentioned above, only two pilot codes $\boldsymbol{q}^{(0)}$ and $\boldsymbol{q}^{(1)}$ need to be stored in the user devices requiring very low memory.

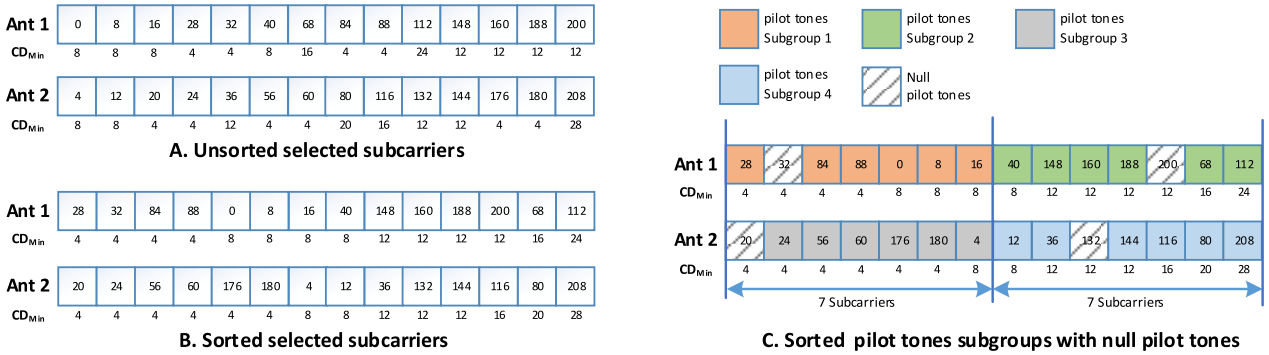

FIGURE 2. An illustration of the proposed non-orthogonal pilot design with $P = 12$ and $P' = 2$ for block-sparse channels in a MIMO system with $N = 256$, $M = 2$, $L_h = 64$, $L = 64$, and $\tau = 4$.

B. STEP 2: DESIGN OF GRANT-FREE ACCESS CODES

In our next step, we choose the locations of P' pilot tones within P_{tot} tones of the root pilot code to be used as null pilots. Recall that Algorithm **1** selects P_{tot} pilot tones indexes so that the frequencies of all possible CDs within P_{tot} tones are close to $\Lambda_{\text{tot}} = \frac{P_{\text{tot}}(P_{\text{tot}}-1)}{L-1}$. For the non-orthogonal pilot codes, we want the frequencies of all CDs to be close to $\Lambda = \frac{P(P-1)}{L-1}$. So we want to choose the null pilot tone positions so that the frequencies of all distinct CDs decrease by the same amount.

Now each pilot tone is associated with $2(P_{\text{tot}} - 1)$ CDs and removing the pilot tone affects the frequencies of all associated CDs. To simplify the selection process, we will consider only the smallest CD (CD_{min}) that is associated with a pilot tone. Our strategy is to distribute the null pilot tones so that they affect the frequencies of all the CD_{min} approximately equally.

For this purpose, we first sort the pilot tone indexes in an increasing order according to the smallest CDs associated with them which is the cyclical distance to the nearest pilot tone index. Let $\{\tilde{q}_0, \tilde{q}_1, \ldots, \tilde{q}_{P_{\text{tot}}-1}\}$ be the sorted pilot tone indexes. Now we divide the sorted indexes into P' subgroups of approximately equal sizes. For each subgroup, we sequentially choose $\lfloor \frac{P_{\text{tot}}}{P'} \rceil$ indexes from the sorted pilot tone indexes. For the last subgroup, we choose all of the remaining subcarriers of the P_{tot} sorted indexes. Then we select one of the pilot tones indexes from each subgroup as null tone. Each combination of null pilot tones together with the corresponding non-zero pilot tones will create a non-orthogonal pilot code. Note that due to the above sorting and grouping, different choices of null tone locations would approximately yield similar effects on the frequencies of all the CD_{min}. The number of non-orthogonal pilot codes available from an orthogonal root code is given by

$$\lfloor P_{\text{tot}}/P' \rceil^{(P'-1)} \times \big(P_{\text{tot}} - \lfloor P_{\text{tot}}/P' \rceil (P'-1)\big). \quad (31)$$

As there are $2MN/L$ root codes, the total number of available non-orthogonal pilot codes is $2MN/L$ times the number in (31).

The complexity of step 2 of designing non-orthogonal pilot codes is $\mathcal{O}(P_{\text{tot}} \log(P_{\text{tot}}))$ which includes the sorting of root pilot codes. Thus, the complexity of both the steps together is $\mathcal{O}(MLP_{\text{tot}} \log(P_{\text{tot}}))$ as we can directly select the grant-free access codes by choosing null pilot tones within the sorted pilot tones index sets of root pilot codes. As discussed in Section IV-A, we can eliminate the pilot design complexity by storing the sorted orthogonal root pilot codes at the user devices. This will allow the users to directly access all the grant-free access codes using the stored root pilot codes removing the design complexity for the user.

Now, we give an example of the proposed non-orthogonal pilot design for block-sparse channels. We will use the same example from the previous section with $N = 256$, $M = 2$, $L_h = 64$, $L = 64$, and $\tau = 4$. For this example, we choose $P = 12$ non-zero pilot tones and $P' = 2$ null pilot tones per antenna. Then we design two orthogonal root pilot codes with $P_{\text{tot}} = 14$ pilot tones per antenna using Algorithm **1**. Fig. 2(a) shows the selected root pilot code and the minimum CD associated with each pilot tone index. Then the pilot tones are sorted according to the minimum CDs. Fig. 2(b) shows the sorted pilot tones. Next, we divide pilot tone indexes of each pilot code into $P' = 2$ subgroups with 7 pilot tone indexes in each group. We select one pilot tone as null pilot in each subgroup. Fig. 2(c) shows different subgroups. We will have 49 non-orthogonal pilot codes (see Eq.(31)) from each of the 16 orthogonal root codes, thus a total of 784 non-orthogonal pilot codes.

We summarize the design process of non-orthogonal pilot codes for block-sparse channels here.

1) Define the number of non-zero pilot tones per antenna considering the system performance requirements and resource availability.
2) Define the number of null pilot tones per antenna considering the number of available pilot resources and expected number of users.
3) Design orthogonal root pilot codes with $P_{\text{tot}} = P + P'$ pilot tones each by using Algorithm 1. Select M orthogonal pilot codes for the M transmit antennas with $P_{\text{tot}} = P + P'$ pilot tones per antenna.
4) Sort the selected pilot tones index sets in an increasing order by the minimum CD associated with the indexes.

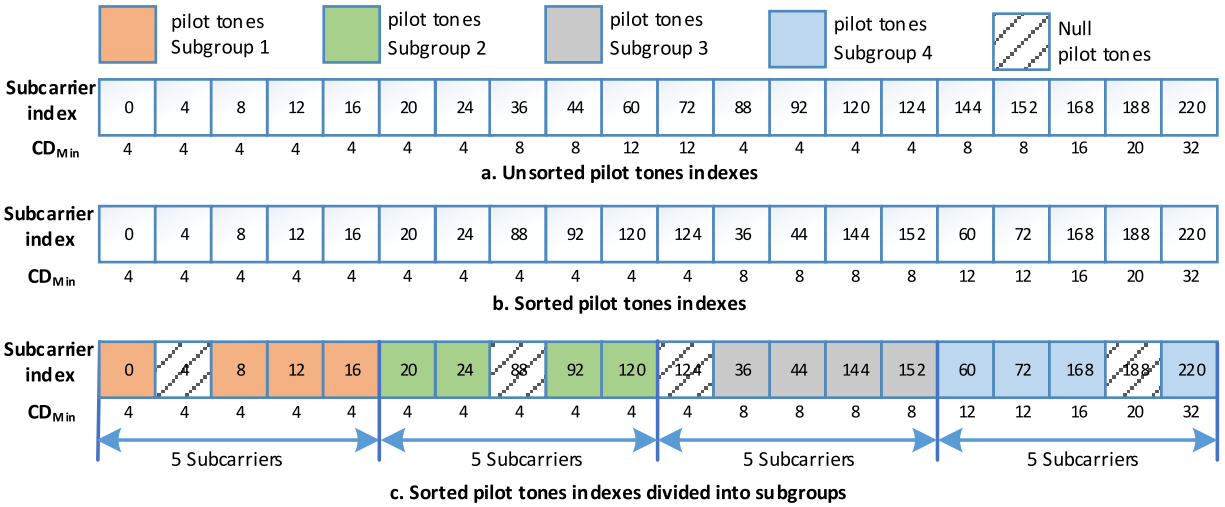

FIGURE 3. Construction of non-orthogonal pilot codes from an orthogonal pilot code for an OFDM system with $N = 256$, $M = 1$, $L_h = L = 64$, and $\tau = 4$.

5) Divide each of the sorted pilot tones index sets into P' subgroups of approximately equal sizes by selecting the subcarrier indexes sequentially.
6) Use one of the pilot tone indexes within each of the subgroups as a null pilot tone. Different combinations of the null pilot tone together with the corresponding non-zero pilot tones will create different non-orthogonal pilot codes.

C. NON-ORTHOGONAL PILOT DESIGN FOR SPARSE CHANNELS IN A SINGLE-ANTENNA SYSTEM

We note that the proposed non-orthogonal pilot design for block-sparse channels could also be used to design optimized non-orthogonal pilot codes for sparse channels as a special case with $M = 1$ transmit antenna per user. In single-antenna systems with sparse channels, the proposed pilot design has several advantages compared to the existing pilot design in [15]. The proposed pilot design uses a flexible pilot length which could be chosen according to system requirements whereas the pilot length is always half of the maximum channel length in the existing design. Due to this flexibility of pilot length together with the exploitation of block sparsity, the proposed design requires less number of pilot resources compared to the existing design, especially when the number of non-zero channel taps is much smaller than the channel length. Also the proposed pilot design enables us to incorporate optimum trade-offs for different design parameters, i.e., the total number of pilot resources used, total number of available grant-free access codes and the channel estimation performances, according to system requirements. For the existing design, they could not be optimized as there is only a single configuration available for a specific system. We will discuss these advantages in more details in Section V.

To design the non-orthogonal pilot codes for sparse channels, we follow the same steps as defined in Section IV-B. Note that, for single-antenna systems the cost function $\boldsymbol{\beta}(\boldsymbol{q}) = \sum_{l=1}^{L-1} |n_{u_l}^{(\boldsymbol{q})} - \Lambda|$ is the term from the upper bound of coherence for the sparse channel in Proposition III. The lower bound for coherence is reached when the cost function is zero.

Now, we give an example of non-orthogonal pilot design in single-antenna sparse channels. Let us consider the system with $N = 256$, $L_h = L = 64$, $M = 1$ and $\tau = 4$. We first consider the length of the pilot code. Suppose the desired channel estimation performance is well achieved by 16 non-zero pilot tones per pilot code. Then, we use $P = 16$ non-zero pilot tones. Next, based on the targeted number of users to be supported in the system, the desired number of non-orthogonal pilot codes per orthogonal subcarrier group can be obtained. Then, from (31), we can obtain the required number of null tones P' per pilot code which should be checked against the system resources. Suppose system requirements and available resources suggest $P' = 4$. Next, we use Algorithm **1** to design tone indexes $\{\boldsymbol{Q}_i, \boldsymbol{Q}_{\mathrm{R}\,i}\}$ of orthogonal root pilot codes with $P + P' = 20$ pilot tones per code using the steps described in Section IV-A. The selected pilot tone indexes for $\boldsymbol{Q}_0$ with their corresponding minimum CDs are depicted in Fig. 3a. Then, we sort the pilot tone indexes according to their $\mathrm{CD}_{\min}$ in an increasing order. The sorted indexes are shown in Fig. 3b. Finally, we divide the sorted pilot tone indexes into $P' = 4$ subgroups of $\frac{P_{\mathrm{tot}}}{P'} = 5$ subcarrier indexes. Then we select one pilot tone from each subgroup as null pilot. Pilot tone subgroups and possible null pilot tones are shown in Fig. 3c. In this way, a total of 625 non-orthogonal pilot codes (see (31)) are possible from this set of 20 pilot tones. We can select a total of 8 orthogonal pilot tones index sets $\{\boldsymbol{Q}_0, \ldots, \boldsymbol{Q}_3, \boldsymbol{Q}_{\mathrm{R}\,0}, \ldots, \boldsymbol{Q}_{\mathrm{R}\,3}\}$ each with 20 (non-zero+null) pilot tones using Algorithm **1** from the full bandwidth. So a total of $625 \times 8 = 5000$ grant-free access codes could be generated. Also we use a total of $20 \times 8 = 160$ pilot tones indexes out of 256 tones and the

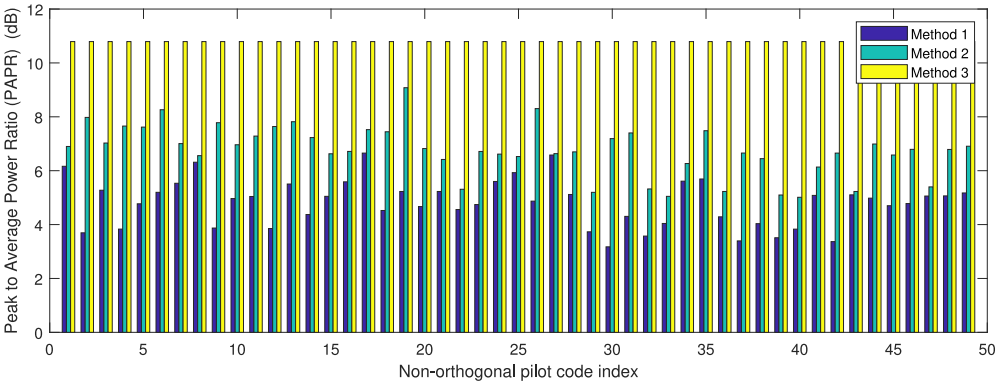

FIGURE 4. PAPR of different non-orthogonal pilot codes using different PAPR optimization methods in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, $P = 12$, and $P' = 2$.

remaining tones can be used for data transmission. So the proposed design uses 37.5% less resources compared to the existing design [15] which uses 32 pilot tones for each pilot code.

D. PEAK-TO-AVERAGE POWER RATIO (PAPR) OPTIMIZATION FOR NON-ORTHOGONAL PILOT CODES

PAPR optimization is important for OFDM systems as the high peaks require more expensive and larger linear power amplifier to be used at lower efficiency [44]. As the phases of the non-zero pilot tones do not affect our non-orthogonal pilot designs presented in the previous sections, we can optimize the PAPR of the time-domain signals of the pilot codes by designing the phases of the non-zero pilot tones.

To illustrate our design strategies, let us consider Quadrature Phase Shift Keying (QPSK) symbols for the pilot codes. To avoid large storage for the access codes, we propose two different approaches to optimize the PAPR where only one optimized QPSK code needs to be stored. In Method 1, we choose the QPSK symbols for the orthogonal pilot code of length $P + P'$. We use an exhaustive search among all possible pilot codes with different QPSK symbols and choose the one with lowest PAPR. Then we puncture the selected QPSK symbols at the P' null tones locations to generate non-orthogonal pilot codes. The same PAPR could be achieved for the codes at SMIPS by choosing the conjugate of the QPSK symbols at the corresponding shifted mirror indexes. Because of Fourier transform property, other codes obtained by the simple tone index shift will also maintain the same PAPR.

In Method 2, we choose the P QPSK symbols for the non-zero pilot tones within a non-orthogonal pilot code of length $P+P'$ with lowest PAPR using exhaustive search. We use the obtained QPSK symbols in the same order across tones for other non-orthogonal codes obtained from the same orthogonal code. Similar to the first method, for the SMIPS codes we use conjugate of the QPSK symbols at the corresponding shifted mirror indexes.

To illustrate performance of these methods, let us consider the same example from Section IV-B for the system with $N = 256$, $L_h = L = 64$, $M = 2$, $\tau = 4$, $P = 12$ and $P' = 2$. Fig. 4 compares Method 1 and Method 2 with the reference (Method 3) where the same QPSK symbol is used at all the pilot tones without any optimization. PAPR values for different methods are shown for all 49 non-orthogonal pilot codes (see example in Section IV-B). We observe that both Method 1 and Method 2 offer much lower PAPR for the access codes compared to the reference while Method 1 provides slightly lower PAPR compared to Method 2.

E. THRESHOLD BASED PILOT DETECTION

In this section, we define two important pilot detection performance metrics for the proposed pilot design, namely probability of single user detection (P_{SUD}) and probability of two codes collision detection (P_{CD}). P_{SUD} is defined as the probability of successfully detecting a pilot code when there are no other non-orthogonal pilot codes present. P_{CD} is defined as the probability of detecting a collision when two pilot codes are present which are non-orthogonal to each other. If more than two non-orthogonal codes are present, the collision detection performance will be better and hence the above P_{CD} represents the worst-case collision detection performance.

We use a threshold based pilot detection where the receiver detects a non-zero pilot tone when the average received energy across all received antennas is greater than the threshold θ. The probability of detecting a single grant-free access code in absence of any collision is given by [15]

$$\left[1 - F\left(\frac{2\theta}{E_{p_1} + \sigma_n^2}\right)\right]^P \left[F\left(\frac{2\theta}{\sigma_n^2}\right)\right]^{\gamma P} \tag{32}$$

where E_p is the energy of a non-zero pilot tone, σ_n^2 is the noise power per tone, $\gamma = P'/P$, and $F(x)$ is a chi-square cumulative distribution function with $2R$ degrees of freedom given by

$$F(x) = 1 - e^{-\frac{x}{2}} \sum_{k=0}^{R-1} \frac{1}{k!} \left(\frac{x}{2}\right)^k. \tag{33}$$

An optimum threshold θ_{opt} which maximizes (32) could be found by numerical method and used for threshold based pilot detection.

Now, let us consider two non-orthogonal pilot codes $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ with non-zero pilot tone energies E_{p_1} and E_{p_2}. When both the pilot codes are received simultaneously, four different combinations of the received energies are possible. A pilot tone may contain non-zero pilot energies from both the pilot codes, from pilot code $\boldsymbol{c}_1$, from pilot code $\boldsymbol{c}_2$ or from none of the pilot codes. Let us assume t_1, t_2, t_3 and t_4 represent the total numbers of received pilot tones with these combinations accordingly. A collision is detected when more than P non-zero pilot tones are detected at the receiver. This allows for fast collision detection at the receiver. Then the probability of detecting a collision for the considered pilot code pair is given by

$$
\begin{aligned}
P_{\mathrm{CD}} \approx & \sum_{t_1'+t_2'+t_3'+t_4'>P} \\
& \times \left[\binom{t_1}{t_1'}\left[1-F\left(\frac{2\theta}{E_{p_1}+E_{p_2}+\sigma_n^2}\right)\right]^{t_1'}\right. \\
& \qquad \left.\times\left[F\left(\frac{2\theta}{E_{p_1}+E_{p_2}+\sigma_n^2}\right)\right]^{t_1-t_1'}\right. \\
& \times \binom{t_2}{t_2'}\left[1-F\left(\frac{2\theta}{E_{p_1}+\sigma_n^2}\right)\right]^{t_2'}\left[F\left(\frac{2\theta}{E_{p_1}+\sigma_n^2}\right)\right]^{t_2-t_2'} \\
& \times \binom{t_3}{t_3'}\left[1-F\left(\frac{2\theta}{E_{p_2}+\sigma_n^2}\right)\right]^{t_3'}\left[F\left(\frac{2\theta}{E_{p_2}+\sigma_n^2}\right)\right]^{t_3-t_3'} \\
& \times \binom{t_4}{t_4'}\left[1-F\left(\frac{2\theta}{\sigma_n^2}\right)\right]^{t_4'}\left[F\left(\frac{2\theta}{\sigma_n^2}\right)\right]^{t_4-t_4'}\Bigg]. \qquad (34)
\end{aligned}
$$

The terms t_1', t_2', t_3' and t_4' represent the total numbers of pilot tones with the aforementioned different combinations of pilot energies for which the detected pilot energies are greater than θ. If the average probability of collision detection is desired, P_{CD} in (34) can be averaged over all pairs of non-orthogonal pilot codes.

V. PERFORMANCE EVALUATION AND COMPARISON

In this section, we evaluate several performance metrics of the proposed non-orthogonal pilot design for various configurations and compare them with the existing pilot designs. In Sections V-A and V-B, we compare different aspects of the proposed pilot designs (i.e., the channel estimation performances, pilot resource usage, pilot detection complexity, and pilot collision detection performances) with the existing pilot designs. In Section V-C, we will compare the channel estimation performances and trade-offs for different pilot codes configurations of the proposed pilot design. In Section V-D, we will explore the effects of the increasing number of transmit and receive antennas.

A. CHANNEL ESTIMATION PERFORMANCE AND PILOT RESOURCE USAGE COMPARISONS WITH EXISTING DESIGNS

We will first compare the proposed non-orthogonal pilot design with the non-orthogonal pilot codes based on Zadoff-chu (ZC) sequence. Several pilot designs based on ZC sequence have been proposed in [5], [6]. There are three major limitations when using the ZC sequence based reference pilot designs to estimate block sparse channels in a system with multiple transmit antennas.

First, the reference pilot designs do not consider a multiple transmit antenna system where each antenna needs orthogonal pilot code to avoid pilot contamination. There are limited number of orthogonal pilot codes available when using ZC sequence based pilot design as a ZC sequence of length P can support only $N_u = \lfloor P/L_h \rfloor$ orthogonal pilot codes [6]. For example, let us consider a system with channel length $L_h = 64$ using ZC sequence of length $P_{ZC} = 139$ as defined in the standard for long term evaluation (LTE) [45]. For such a system, there are only 2 orthogonal pilot codes available which could be used by 2 transmit antennas. To generate more orthogonal pilot codes for the larger antenna systems, a longer ZC sequence will be needed, resulting inefficient pilot resource usage.

Second, the pilot codes used by different users will be non-orthogonal among themselves with high probability when most of the orthogonal pilot codes are used by multiple transmit antennas. As the ZC sequence based reference pilot designs do not have collision detection capability, the non-orthogonal pilot codes will result in severe pilot contamination and poor channel estimation performance when the number of users grows.

Third, while the pilot design in [6] uses CS based user detection, none of the ZC sequence based reference pilot designs use CS based channel estimation. As a result, the proposed pilot design outperforms the reference pilot designs in terms of channel estimation performances and pilot resource usage for sparse and block-sparse channels.

To illustrate this using simulation, let us consider the same example from the previous section with $N = 256$, $L_h = L = 64$, $M = 1$, $R = 1$, and $\tau = 4$. For sparse channel estimation, we use the OMP algorithm from [15] and stop the algorithm when the residual is less than $P\sigma_n^2$ where P is the number of non-zero pilot tones and σ_n^2 is the noise variance. The performance results are averaged over 10^5 simulation trials. In each trial, the sparse channels are generated independently using random positions for the non-zero channel taps.

Fig. 5 shows the channel estimation performance of the proposed pilot design with $P_{\mathrm{tot}} = 24$ [config. 1, Table 2] and the ZC sequence based pilot design from reference [6] with pilot sequence length $P_{\mathrm{ZC}} = 139$. For the ZC sequence based pilot codes, we have used oracle channel estimation where the user activity is known to the receiver. We have simulated two different scenarios using reference [6] pilot codes with $K = 1, 2$, where K represents the total number of active users transmitting non-orthogonal pilot codes. We observe that the channel estimation performance for the proposed pilot design is better compared to reference [6] pilot design even without any pilot collision when $K = 1$. The performance of [6] pilot design deteriorates further when $K = 2$ non-orthogonal users are present. Note that the pilot code length of the proposed

TABLE 2. Different configurations for the proposed pilot design with optimized pilot code length.

Design	Total # of antennas per user	Total # of tones per pilot code	# of non-zero tones per pilot code	# of orthogonal codes	multiplicity factor per orthogonal code	total # of access codes	Total pilot resources used
Reference [15]	1	32	28	8	4096	32768	100%
Proposed (Config. 1)	1	24	21	8	512	4096	75%
Proposed (Config. 2)	1	21	18	8	343	2744	65.63%
Proposed (Config. 3)	2	16	14	8	64	512	50%
Proposed (Config. 4)	4	12	10	8	36	288	37.5%
Proposed (Config. 5)	1	25	21	8	1512	24192	78.13%
Proposed (Config. 6)	2	18	14	8	256	2048	56.25%
Proposed (Config. 7)	4	14	10	8	128	1024	43.75%

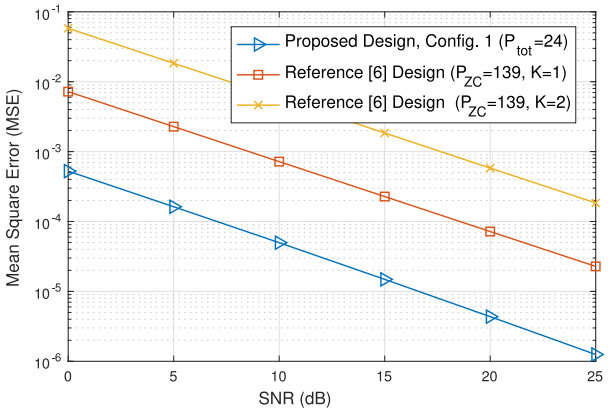

FIGURE 5. Average channel estimation performance of the proposed pilot design and the reference [6] non-orthogonal pilot design in a system with $N = 256$, $L_h = L = 64$, $M = 1$, $R = 1$, and $\tau = 4$.

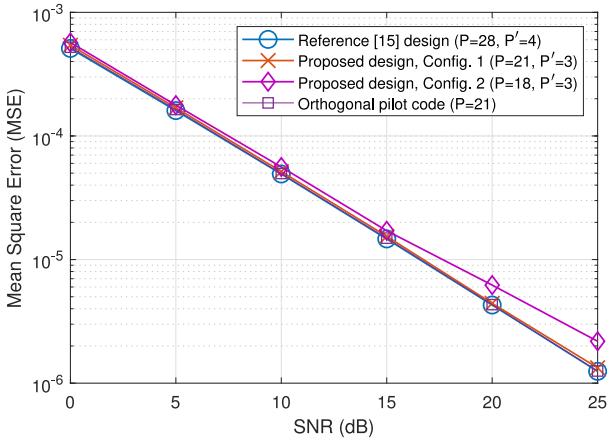

FIGURE 6. Average channel estimation performance of the proposed pilot design, the reference non-orthogonal pilot design, and orthogonal pilot design [15] in a system with $N = 256$, $L_h = L = 64$, $M = 1$, $R = 1$, and $\tau = 4$.

pilot design is only 24 compared to 139 in the reference design resulting in much more efficient pilot resource usage.

Next, we will compare the proposed pilot design in more details with the reference design in [15] which uses CS based channel estimation and has collision detection capability. The proposed pilot design offers large savings in terms of pilot resource usage by leveraging optimized pilot code length and block-sparsity property of the MIMO channels compared to reference [15] design. Through simulations, we will first evaluate the reduction of pilot resource usage for the proposed pilot design in sparse channels compared to the existing reference [15] non-orthogonal pilot design while keeping similar channel estimation performances.

Fig. 6 compares the average channel estimation performances of the proposed non-orthogonal pilot design with that of the orthogonal and reference [15] non-orthogonal pilot designs in a single-antenna system. The reference [15] pilot design uses a total of $L/2 = 32$ pilot tones with 28 non-zero pilot tones and 4 null tones for non-orthogonal pilot codes. While the length of the non-orthogonal pilot codes in the reference [15] design is required to be fixed, the length of the pilot codes in the proposed design is a flexible design parameter. From Fig. 6, we observe that the proposed non-orthogonal pilot design (configuration 1) with 21 non-zero pilot tones and 3 null pilot tones has similar channel estimation performance as the existing design. Configuration 1 generates 512 grant-free access codes (given by (31)) from 24 pilot resources and a total of 4096 grant-free access codes using the full bandwidth. The details of different configurations are given in Table 2.

This configuration accounts for a total of 25% reduction in pilot resource usage compared to the existing design without any loss of channel estimation performances. We can further reduce the number of pilot resources needed by using 18 non-zero pilot tones and 3 null tones (configuration 2, Table 2) with only slightly degraded average channel estimation performance. This configuration uses 34.37% less pilot resources compared to the reference design while still generating 2744 grant-free access codes. We also compare these channel estimation performances to the orthogonal pilot codes with $P = 21$, generated by using Algorithm **1**. The channel estimation performance of the orthogonal pilot code is similar to that of the configuration 1 consisting of same number of non-zero pilot tones. Note that the orthogonal pilot design can support only 8 orthogonal pilot codes compared to 4096 non-orthogonal pilot codes for configuration 1.

For the proposed pilot design, the pilot resource usage could be further reduced by exploiting block-sparsity property of the MIMO channels in multiple antenna systems. Fig. 7 showcases the reduction of pilot resource usage for the proposed pilot designs compared to the reference [15]

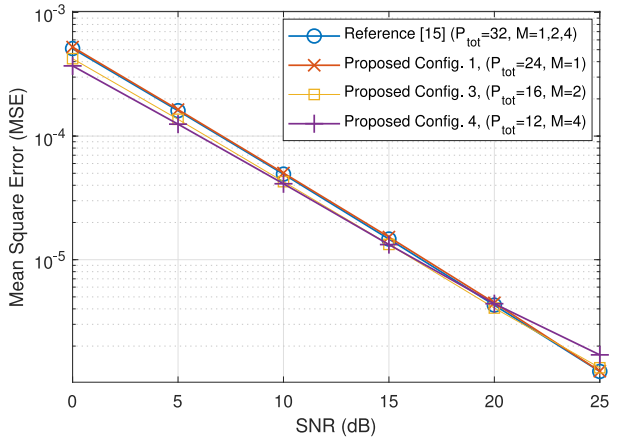

FIGURE 7. Average channel estimation performances of the proposed pilot design and the reference pilot design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 1, 2, 4$.

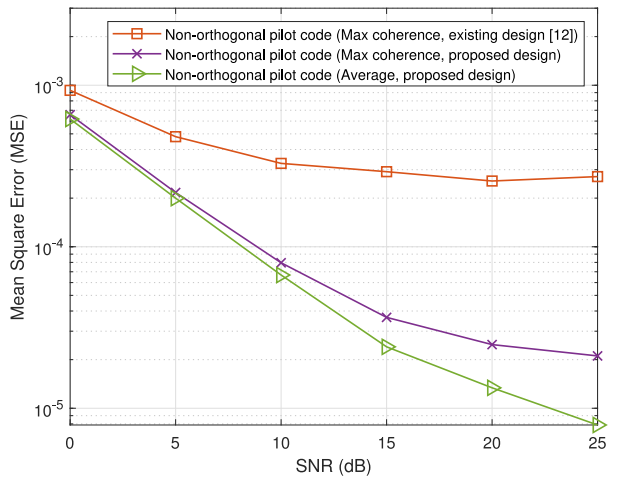

FIGURE 8. Average channel estimation performance of the proposed non-orthogonal pilot design and the reference non-orthogonal pilot design [12], both with 16 non-zero pilot tones plus 4 null pilot tones, in a system with $N = 256$, $L_h = L = 64$, $M = 1$, $R = 1$, and $\tau = 4$.

design for several values of M, the number of user's transmit antennas. As the reference [15] design is not optimized for the block-sparse MIMO channel, it uses $L/2 = 32$ pilot tones for all the antennas. In our simulations, we have used *block optimized orthogonal matching pursuit* (BOOMP) algorithm to estimate the block-sparse channels [33]. From Fig. 7, for similar channel estimation performances the proposed design uses a total of 24, 16 and 12 pilot tones per antenna for the systems with $M = 1, 2$, and 4, respectively compared to the 32 pilot tones per antenna in the reference [15] design. They result in a total of 25%, 50%, and 62.5% reduction of pilot resource usage for the proposed pilot design with 1, 2, and 4 transmit antennas respectively.

Finally, we will compare the channel estimation performances for different non-orthogonal pilot codes generated from the same root pilot code. The non-orthogonal pilot design in [15] uses $P_{\text{tot}} = L/2 = 32$ pilot tones for each pilot code which is much higher than the proposed design.

TABLE 3. Complexity of non-orthogonal pilot detection schemes for different pilot designs.

Pilot detection schemes	Complexity of the pilot detection schemes
Shuffling based detection [7]	$\mathcal{O}(\log_2 P(\log_2 P + 1)P)$
SIC-based detection [7]	$\mathcal{O}(k \log_2 P(\log_2 P + 1)P)$
Approximate Message Passing (AMP) based detection [11]	$\mathcal{O}(M_b N_u L_h)$ per iteration
Vector AMP based detection [9]	$\mathcal{O}(M_b N_u P)$ per iteration
(OMP-NS) algorithm [6]	$2PN_pK_u \log(P) + \mathcal{O}(P)$
Threshold based detection [15]	$\mathcal{O}(P)$
Proposed Threshold based detection	$\mathcal{O}(P)$

To compare with an existing method using the same pilot resource amount, we use the existing design in [12] for the placements of the null pilot tones where the P' null pilot tones are selected randomly from the P_{tot} pilot tones. For these non-orthogonal pilot codes, the coherences of different codes are different. Fig. 8 shows the channel estimation performances of the non-orthogonal pilot codes with highest coherence for the proposed design and the existing design [12] under the same pilot resources and the same number of null pilot tones (with $P_{\text{tot}} = 20$ and $P' = 4$). The proposed pilot design provides much better channel estimation performance for the worst pilot code than the existing design, thus ensuring much better fairness among the users.

B. PILOT DETECTION PERFORMANCE AND COMPLEXITY COMPARISONS WITH EXISTING DESIGNS

Threshold based pilot detection for the proposed pilot design, as discussed in Section IV-E, is very low in complexity requiring only P comparisons when the pilot code length is P. Most of the existing pilot detection techniques for non-orthogonal pilot designs require algorithms of much higher computation complexity. Table 3 compares the complexity of different pilot detection schemes for existing and proposed non-orthogonal pilot designs. Here N_u, M_b, K_u, and N_p refer to the total number of users in the system, total number of base station antennas, total number of active users, and total number of pilot resources used. From the table, the threshold based detection approaches for reference [15] and the proposed pilot design have the lowest complexity of $\mathcal{O}(P)$. Also only these two schemes support fast collision detection. While they both have the same pilot detection and collision detection complexity for the pilot codes of same length, the proposed pilot design requires much shorter pilot code length compared to the reference design in [15] for equivalent channel estimation performances. So, the complexity of the pilot detection and collision detection for the proposed pilot design will be less compared to [15].

To compare the pilot detection and collision detection performances of the proposed pilot design with the reference design of [15] in more details, we will explore the pilot detection performance metrics P_{SUD} and P_{CD} for the pilot designs. We use 3 different configurations (config 5, 6, and 7, Table 2) of the proposed pilot design for the next

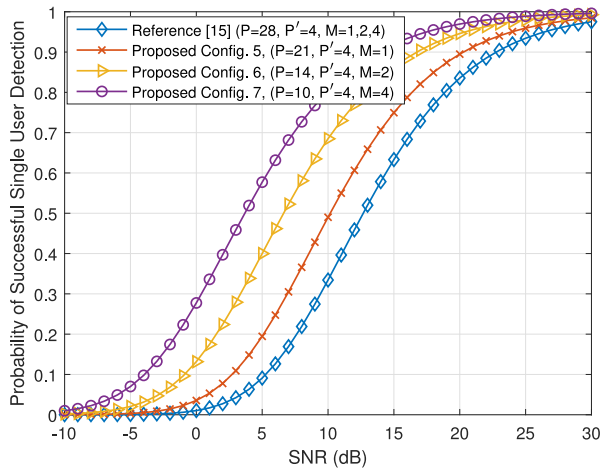

FIGURE 9. Comparison of P_{SUD} (equation (32)) for the proposed pilot design configurations 5, 6, 7 and the reference design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 1, 2, 4$.

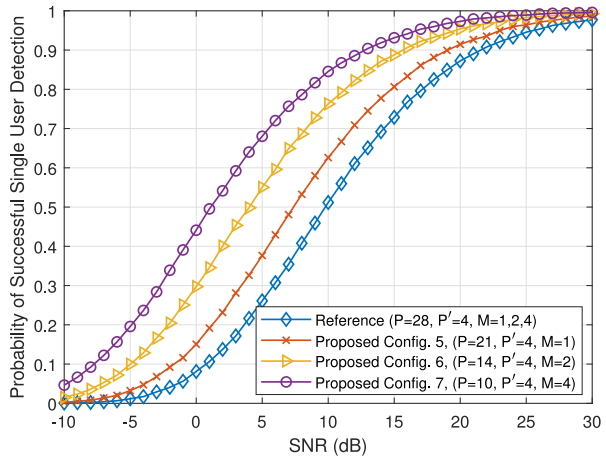

FIGURE 10. Comparison of P_{SUD} (Monte Carlo simulation) for the proposed pilot design configurations 5, 6, 7 and the reference design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 1, 2, 4$.

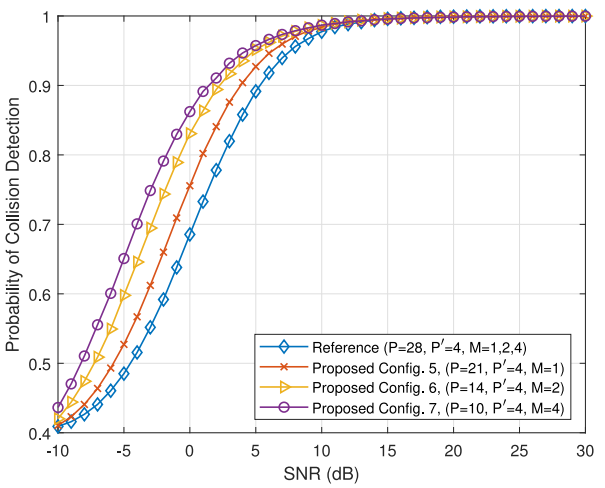

FIGURE 11. Comparison of P_{CD} (equation (34)) for the proposed pilot design configurations 5, 6, 7 and the reference design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 1, 2, 4$.

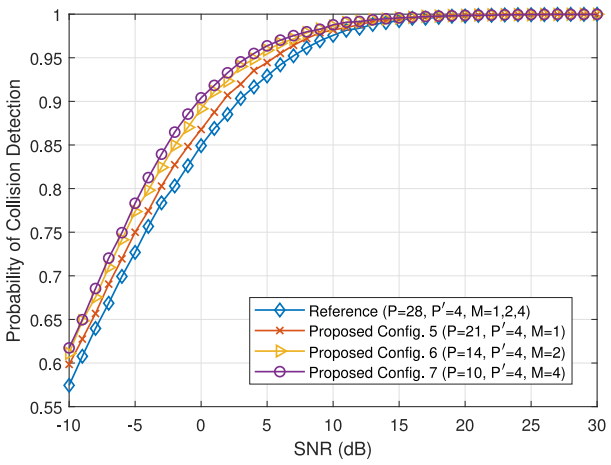

FIGURE 12. Comparison of P_{CD} (Monte Carlo simulation) for the proposed pilot design configurations 5, 6, 7 and the reference design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 1, 2, 4$.

performance comparisons. They use a total of 21, 14, and 10 non-zero pilot tones per antenna. These configurations are chosen because they provide similar channel estimation performances for the systems with $R = 1$, $M = 1, 2$ and 4 when compared to the reference design [15] with 28 non-zero pilot tones per antenna. For fair comparison, all the configurations use 4 null pilot tones per non-orthogonal pilot code similar to the reference design.

Fig. 9 and Fig. 10 compares P_{SUD} of the proposed pilot designs with that of the reference [15] pilot design. Fig. 9 uses the analytical result in equation (32) while Fig. 10 uses Monte Carlo simulation for performance comparison. The analytical and simulation results show the same trend with slight differences in value. These differences are expected due to the assumption of independent channel gains at the pilot subcarriers when deriving (33). We observe that less number of pilot tones provide better single user detection performance due to greater pilot energy concentration. This is reflected in the higher P_{SUD} values of the proposed pilot designs compared to the reference [15] pilot design.

P_{CD} performances for different configurations of the proposed pilot design are compared with the reference design [15] in Fig. 11. Here, we consider the scenario where two randomly selected non-orthogonal pilot codes with same pilot energy are colliding at the receiver as described by the analytical result in (34). Fig. 12 uses Monte Carlo simulation to compare P_{CD} performances for different configurations under same scenario. The analytical and simulation results show similar trends with slight differences in value, and the differences are due to the assumption of independence channel gains at the pilot subcarriers when deriving (35). We observe that the probability of detecting a collision is higher with shorter pilot code length due to greater pilot energy concentration and the proposed pilot design outperforms the reference [15] design.

TABLE 4. Different configurations for the proposed pilot design in block-sparse channels.

Design	Total # of antennas per user	Total # of tones per pilot code	# of non-zero tones per pilot code	# of orthogonal codes	multiplicity factor per orthogonal code	total # of access codes
Proposed (Config. 8)	2	14	12	16	49	784
Proposed (Config. 9)	2	15	12	16	125	2000
Proposed (Config. 10)	2	16	12	16	256	4096
Proposed (Config. 11)	2	16	13	16	150	2400
Proposed (Config. 12)	2	15	13	16	56	896

C. CHANNEL ESTIMATION PERFORMANCES AND TRADE-OFFS FOR DIFFERENT PILOT CODES CONFIGURATIONS

In this section, we will investigate the trade-offs among the pilot resource usage, total number of available grant-free access codes, and the average channel estimation performances for different configurations of the proposed pilot design. The proposed pilot design is for uplink systems and hence the number of transmit antennas per user (device) will be typically small due to power and size constraints. Hence, we will use $M = 2$ for most of the configurations used in this and following sections to compare different design choices. The details of the additional configurations that have not been discussed before are listed in Table 4.

First, we will explore the trade-offs between pilot resource usage and total number of available grant-free access codes. We can increase the total number of available grant-free access codes without degrading the channel estimation performances by adding more pilot resources as the null pilot tones. Fig. 13 compares the channel estimation performances of configuration 8, 9 and 10 with $M = 2$, $R = 1$. The configurations have 12 non-zero pilot tones with 2, 3 and 4 null pilot tones for each antenna. Fig. 13 shows that the channel estimation performances for all three configurations are similar while providing 784, 2000 and 4096 grant-free access codes. These results indicate that when needed according to system requirements, a higher number of grant-free access codes could easily be generated by the proposed design using more pilot resources without degrading the average channel estimation performances. Although the complexity of the pilot design increases with the addition of more pilot tones, this could be avoided by storing the sorted root pilot codes at the user devices as discussed in Sections IV-A and IV-B.

We can also increase the number of total available grant-free access codes while using a fixed number of pilot resources by using some of the non-zero pilot tones as null tones. To illustrate this, Fig. 14 compares the performances of the pilot configurations 10, 11 and 3 for block-sparse channels with $M = 2$ and $R = 1$. Each configuration has a total of 16 pilot tones including 12, 13 and 14 non-zero pilot tones with 4, 3 and 2 null tones. The configurations provide a total of 4096, 2400 and 1024 grant-free access codes. Fig. 14 shows that a greater number of grant-free access codes could be supported by the proposed design with only a slight degradation of the average channel estimation performances while using the same amount of pilot resources and same pilot design complexity. This offers a good trade-off when a system is supporting a large number of users with limited resources.

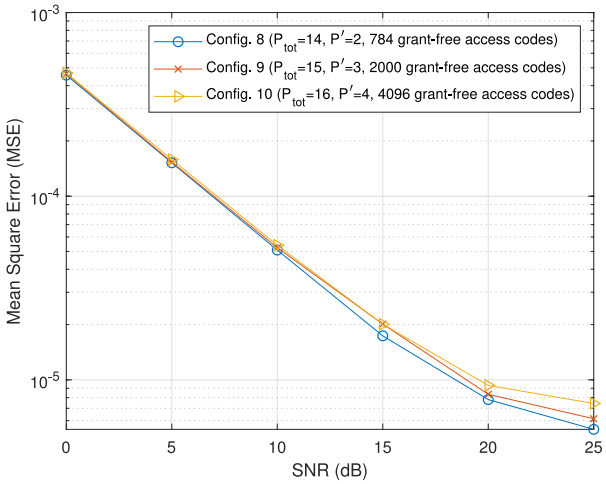

FIGURE 13. Average block-sparse channel estimation performances of the proposed pilot design configurations 8, 9, 10 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1$.

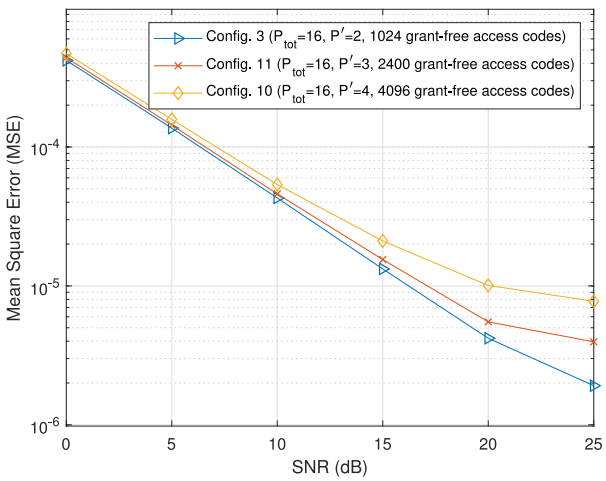

FIGURE 14. Average block-sparse channel estimation performances of the proposed pilot design configurations 10, 11, 3 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1$.

Next, we will discuss the effect of increasing the number of non-zero pilot tones P in a pilot code. Fig. 15 compares the channel estimation performances of the pilot configurations 8, 12 and 3 for block-sparse channels. These configurations use a total of 12, 13 and 14 non-zero pilot tones and 2

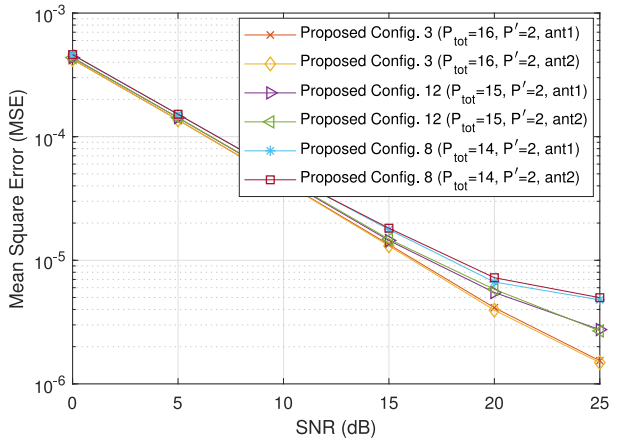

FIGURE 15. Average block-sparse channel estimation performances of the proposed pilot design configurations 8, 12, 3 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1$.

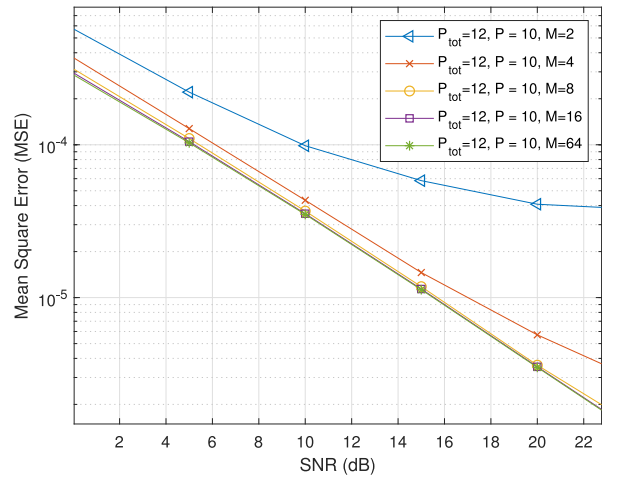

FIGURE 16. Channel estimation performances of the proposed pilot design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 2, 4, 8, 16, 64$.

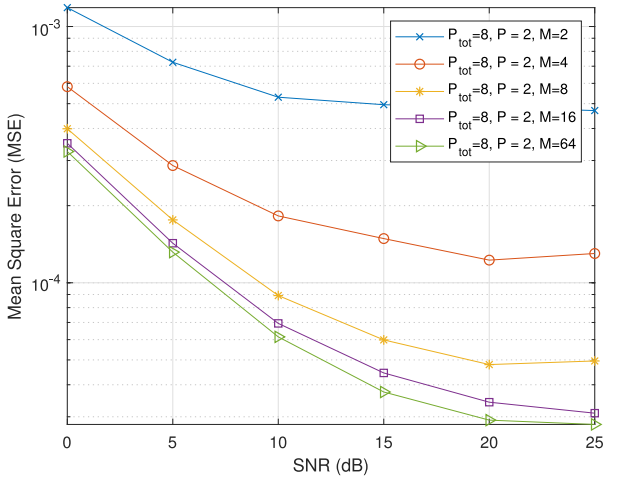

FIGURE 17. Channel estimation performances of the proposed pilot design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $R = 1$, and $M = 2, 4, 8, 16, 64$.

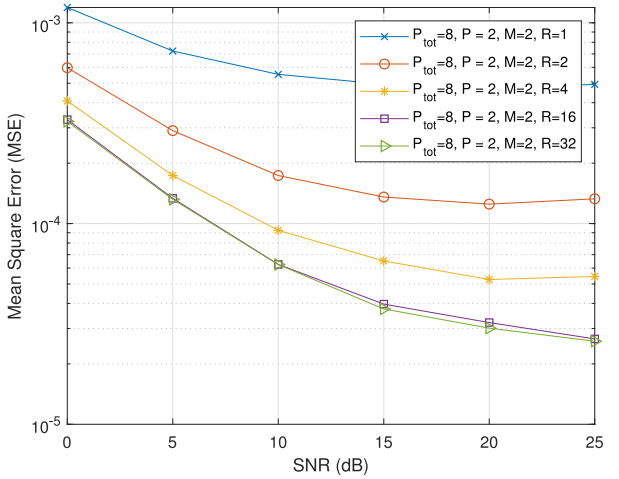

FIGURE 18. Channel estimation performances of the proposed pilot design in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1, 2, 4, 16, 32$.

null pilot tones for each antenna. Fig. 15 shows that we can improve the channel estimation performances by allocating more pilot resources as non-zero pilot tones in our pilot design. The pilot design complexity increases slightly with additional pilot tones when the root pilot codes are not stored at the user devices. We also note that the proposed pilot design has similar channel estimation performances for both of the antennas.

D. PERFORMANCE EVALUATION FOR INCREASING NUMBERS OF TRANSMIT AND RECEIVE ANTENNAS

Next, we will explore the effect of increasing M, the number of transmit antennas per user. When using a fixed number of pilot tones per pilot code, the channel estimation performance could be improved by increasing M. To illustrate this, we use a pilot code with $P_{\text{tot}} = 12$ and $P = 10$, similar to configuration 4 (Table 2), for different values of M. The number of all possible combinations of the non-orthogonal pilot codes for all antennas becomes prohibitively large for simulation when M is large (e.g., 36^{16} combinations for $M = 16$). So, we use the non-orthogonal pilot code with the highest coherence for each antenna which represents the worst case scenario for $M = 2, 4, 8, 16,$ and 64. Fig. 16 shows that the channel estimation performance improves when using a higher number of transmit antennas. The reason is due to the block-sparsity property where a larger M could yield better channel path delay identification. For this configuration, the performance gain diminishes when M is greater than 8 as very good performance is already achieved by $M = 8$. The performance gain of increasing M could be larger for a shorter pilot code length. Fig. 17 shows that the channel estimation performance improves more with increasing values of M when using a pilot code with $P_{\text{tot}} = 10$ and $P = 8$. It could be ascribed to the smaller pilot resource which leaves more room for performance improvement by increasing M.

Next, we investigate the effect of increasing R, the number of receive antennas, on channel estimation and collision detection performances. When the channels for all received

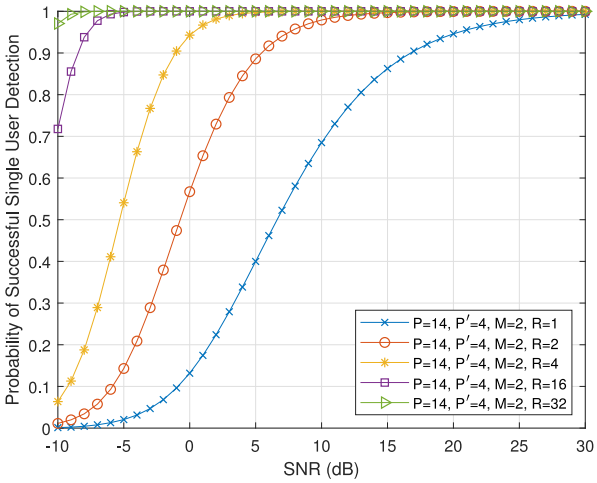

FIGURE 19. Comparison of P_{SUD} (equation (32)) for the proposed pilot design configuration 6 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1, 2, 4, 16, 32$.

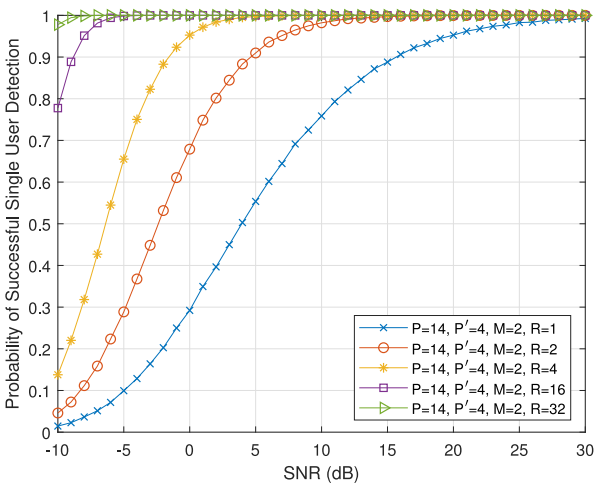

FIGURE 20. Comparison of P_{SUD} (Monte Carlo simulation) for the proposed pilot design configuration 6 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1, 2, 4, 16, 32$.

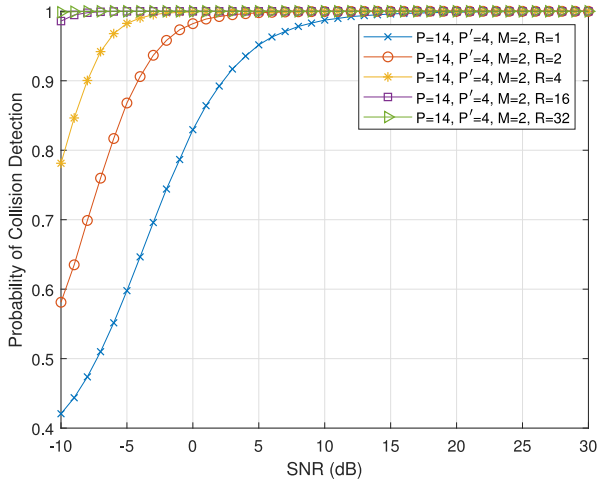

FIGURE 21. Comparison of P_{CD} (equation (34)) for the proposed pilot design configuration 6 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1, 2, 4, 16, 32$.

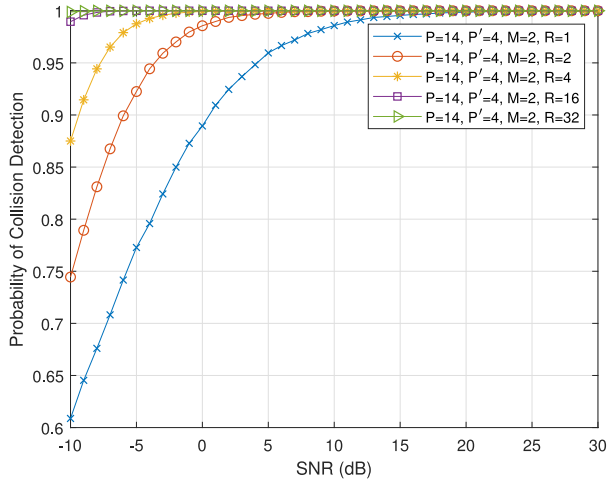

FIGURE 22. Comparison of P_{CD} (Monte Carlo simulation) for the proposed pilot design configuration 6 in a system with $N = 256$, $L_h = L = 64$, $\tau = 4$, $M = 2$, and $R = 1, 2, 4, 16, 32$.

antennas exhibit same block-sparsity property, the estimation of non-zero channel tap positions could be improved by including the signals from all received antennas in BOOMP algorithm, thus resulting in better channel estimation. Fig. 18 shows the channel estimation performances for different values of R when using a pilot code with $P_{\text{tot}} = 8$ and $P = 2$. We observe that the higher number of receive antennas improves channel estimation performance because of the improved estimation of the positions of non-zero channel taps. From Fig. 17 and Fig. 18, we can note that the channel estimation performance depends on the value of $M \times R$ which determines the block size of the block-sparse system.

Finally, we evaluate P_{SUD} and P_{CD} for increasing values of R. The average received pilot energy across all receive antennas is used for threshold based pilot detection. Fig. 19 and Fig. 20 compares the P_{SUD} for increasing number of received antennas using the analytical result in (32) and Monte Carlo simulation. The P_{SUD} improves with the increasing values of R. Fig. 21 compares the P_{CD} for different values of R using the analytical result in (34) while Fig. 22 does the same using Monte Carlo simulation. The analytical and simulation results, except with some differences in value due to the assumption of independence channel gains at pilot subcarriers in the analytical development, show the same trends. From the figures, we observe that higher numbers of receive antennas provide much improved P_{CD}. The improvements in P_{SUD} and P_{CD} with increasing R values are due to the diminishing of fluctuations of the received pilot energies (channel hardening effect) when combined from multiple receive antennas.

VI. CONCLUSION

In this paper, considering grant-free access to support a large number of users, we have proposed compressed sensing based pilot designs for novel non-orthogonal pilot codes with fast collision detection capability for highly sparse and

block-sparse channels. The proposed non-orthogonal pilot designs use only a fraction of the pilot resources compared to the existing pilot designs while not compromising the channel estimation performances. They provide much better performance fairness of grant-free access codes across users and/or more pilot resource savings than the existing non-orthogonal pilot designs. They also offer flexibility through the choice of design parameters in satisfying different system requirements and constraints, as well as very low memory requirement for generating low PAPR access codes.The results also show that the channel estimation performance improves with the increasing number of the product of the numbers of transmit and receive antennas due to the corresponding improvement in channel block sparsity identification. Similarly, the pilot code/collision detection performance improves with the increasing number of the receive antennas due to channel hardening effect of the receive antenna combining.

REFERENCES

[1] NTT DOCOMO, *New Sid Proposal: Study on New Radio Access Technology*, document RAN 1 RP-160671, 3GPP, Sophia Antipolis, France, 2016.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[3] Z. Li, N. Rupasinghe, O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "Directional training and fast sector-based processing schemes for mmWave channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–7.

[4] J. Zhang *et al.*, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1353–1362, Mar. 2017.

[5] T. Kim and S. H. Chae, "A channel estimator via non-orthogonal pilot signals for uplink cellular IoT," *IEEE Access*, vol. 7, pp. 53419–53428, 2019.

[6] A. T. Abebe and C. G. Kang, "Joint channel estimation and MUD for scalable grant-free random access," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2229–2233, Dec. 2019.

[7] H. Zhang, R. Li, J. Wang, Y. Chen, and Z. Zhang, "Reed–Muller sequences for 5G grant-free massive access," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2017, pp. 1–7.

[8] R. Deng, S. Zhou, and Z. Niu, "Scalable non-orthogonal pilot design for massive MIMO systems with massive connectivity," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2016, pp. 1–6.

[9] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[10] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, "Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6212–6225, Aug. 2019.

[11] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

[12] O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "RRH based massive MIMO with 'on the FL' pilot contamination control," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2016, pp. 1–7.

[13] M. Peng, S. Yan, and H. V. Poor, "Ergodic capacity analysis of remote radio head associations in cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 365–368, Aug. 2014.

[14] O. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "A novel alternative to cloud RAN for throughput densification: Coded pilots and fast user-packet scheduling at remote radio heads," in *Proc. IEEE 15th Asilomar Conf. Signals Syst. Comput.*, 2016, pp. 3–10.

[15] A. Quayum, H. Minn, and Y. Kakishima, "Non-orthogonal pilot designs for joint channel estimation and collision detection in grant-free access systems," *IEEE Access*, vol. 6, pp. 55186–55201, 2018.

[16] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Dept. ECE, Univ. Wisconsin-Madison, Madison, WI, USA, Rep. 5, pp. 10–15, 2007.

[17] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. H. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1465–1475, Apr. 2007.

[18] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 4–27, Jan. 2000.

[19] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 2, pp. 128–137, Feb. 1987.

[20] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[21] W. U. Bajwa, A. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Proc. IEEE 13th Digit. Signal Process. Workshop 5th IEEE Signal Process. Educ. Workshop*, 2009, pp. 320–325.

[22] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[23] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jun. 2011.

[24] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, Aug. 2009.

[25] X. He, R. Song, and W.-P. Zhu, "Optimal pilot pattern design for compressed sensing-based sparse channel estimation in OFDM systems," *Circuits Syst. Signal Process.*, vol. 31, no. 4, pp. 1379–1395, 2012.

[26] R. Mohammadian, A. Amini, and B. H. Khalaj, "Compressive sensing-based pilot design for sparse channel estimation in OFDM systems," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 4–7, Jan. 2017.

[27] X. He, R. Song, and W.-P. Zhu, "Pilot allocation for sparse channel estimation in MIMO-OFDM systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 60, no. 9, pp. 612–616, Sep. 2013.

[28] X. Ren, W. Chen, and M. Tao, "Position-based compressed channel estimation and pilot design for high-mobility OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 1918–1929, May 2015.

[29] J.-C. Chen, C.-K. Wen, and P. Ting, "An efficient pilot design scheme for sparse channel estimation in OFDM systems," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1352–1355, Jul. 2013.

[30] C. Qi and L. Wu, "A study of deterministic pilot allocation for sparse channel estimation in OFDM systems," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 742–744, May 2012.

[31] C. Qi, G. Yue, L. Wu, Y. Huang, and A. Nallanathan, "Pilot design schemes for sparse channel estimation in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1493–1505, Apr. 2015.

[32] W. C. Ao, C. Wang, O. Y. Bursalioglu, and H. Papadopoulos, "Compressed sensing-based pilot assignment and reuse for mobile UEs in mmWave cellular systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2016, pp. 1–7.

[33] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 4558–4563.

[34] X. He, R. Song, and W.-P. Zhu, "Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2990–3004, May 2016.

[35] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2012.

[36] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, 2014.

[37] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[38] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[39] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[40] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
[41] P. Xia, S. Zhou, and G. B. Giannakis, "Achieving the Welch bound with difference sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1900–1907, May 2005.
[42] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
[43] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, vol. 23. Philadelphia, PA, USA: SIAM, 1999.
[44] S. H. Han and J. H. Lee, "PAPR reduction of OFDM signals using a reduced complexity PTS technique," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 887–890, Nov. 2004.
[45] *Physical Channels and Modulation, V15.7.0*, 3GPP Standard TS 36.211, 2019.

AYON QUAYUM (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2005, the M.Sc. degree in electrical and computer engineering from North Carolina State University, Raleigh, NC, USA, in 2008. He is currently pursuing the Ph.D. degree in electrical engineering with the University of Texas at Dallas, Richardson, TX, USA. From 2008 to 2016, he was a Staff Engineer with Qualcomm, Inc., where he was involved in research and development of WCDMA and 4G systems. His research interests include signal processing, wireless communication, and system design for next-generation wireless standards.

HLAING MINN (Fellow, IEEE) received the B.E. degree in electrical electronic engineering from Yangon Institute of Technology, Yangon, Myanmar, in 1995, the M.E. degree in telecommunications from the Asian Institute of Technology, Thailand, in 1997, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2001.

He was a Postdoctoral Fellow with the University of Victoria from January to August 2002. He has been with the University of Texas at Dallas, USA, since September 2002, where he is currently a Full Professor. His research interests include wireless communications, signal processing, signal design, dynamic spectrum access and sharing, next-generation wireless technologies, and biomedical signal processing. He has been serving as an Editor-at-Large since 2016 and served as an Editor from 2005 to 2016 for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was also an Editor of the *International Journal of Communications and Networks* from 2008 to 2015. He served as the Technical Program Co-Chair for the Wireless Communications Symposium of the IEEE GLOBECOM 2014 and the Wireless Access Track of the IEEE VTC, Fall 2009, as well as the Technical Program Committee Member for over 30 IEEE conferences.