IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 May 2022, accepted 20 June 2022, date of publication 23 June 2022, date of current version 28 June 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3185746

Design Framework of Fourier-Based Non-Orthogonal Sequences for Sparse Activity Detection in Multicell Massive Connectivity

NAM YUL YU, (Senior Member, IEEE)
School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea
e-mail: nyyu@gist.ac.kr

This work was supported by the National Research Foundation of Korea (NRF) Grant through the Korea Government [Ministry of Science and ICT (MSIT)] under Grant NRF-2021R1F1A1046282.

ABSTRACT In massive machine-type communications (mMTC), grant-free access is a key enabler for a massive number of users to be connected to a base station with low signaling overhead and low latency. In this paper, new Fourier-based sequence sets are proposed for user-specific, non-orthogonal, unimodular sequences for uplink grant-free access in multi-cell massive machine-type communications (mMTC). A design framework based on a partial Fourier matrix with masking operations is presented for multiple sets of non-orthogonal sequences, where the correlation matrix is defined to analyze the intra- and inter-cell correlations simultaneously. We use algorithms to find a subsampling index set for the partial Fourier matrix, in an effort to reduce the correlations of the resulting sequences of arbitrary length. Simulation results demonstrate that the Fourier-based non-orthogonal sequences achieve the excellent performance of sparse activity detection for uplink grant-free access in multi-cell mMTC. Compared to the Zadoff-Chu sequences, this design framework presents more sets of non-orthogonal sequences of arbitrary length, which can supply unique signatures for devices in more mMTC cells.

INDEX TERMS Compressed sensing, genetic algorithm, grant-free access, massive machine-type communications, non-orthogonal sequences.

I. INTRODUCTION

Massive connectivity of wireless devices is essential for industrial, commercial, and critical applications of massive machine-type communications (mMTC) [1], [2], which provides a concrete platform for the Internet of Things (IoT). Unlike human-type communications (HTC), mMTC is characterized by small size data, infrequent transmission, low cost devices, low mobility, and so on [3]. In practice, mMTC systems need to support a massive number of devices with low control overhead, low latency, and low power consumption for delay-sensitive and energy efficient communications.

Non-orthogonal multiple access (NOMA) [4], [5] has received a great deal of attention for massive connectivity in 5G wireless systems. In code-domain NOMA, user-specific and non-orthogonal spreading sequences are assigned to users for their non-orthogonal multiplexing through common resources. In sparse code multiple access (SCMA) [6], sparse spreading sequences are assigned to users, where a message passing algorithm (MPA) [7] and a list sphere decoding based MPA decoder [8] can be deployed for reliable multiuser detection with low complexity. Complex-valued spreading sequences are employed for multi-user shared access (MUSA) [9], where the successive interference cancellation (SIC) can be performed for multiuser detection. Also, pattern division multiple access (PDMA) [10] attempted to enable massive connectivity with low complexity through an efficient pattern matrix design and a recursive approach of multiuser detection [11]. For a survey on existing works of code-domain NOMA, readers are referred to [5]. Recently, the state-of-the-art technique of deep learning [12] has been applied for multiuser detection in uplink code-domain NOMA systems [13]– [15]. Also, the end-to-end NOMA transceiver designed by [16], [17] utilizes a

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

model-driven deep learning method to solve the problem of NOMA signature classification.

Grant-free access is of great interest to connect a massive number of devices to mMTC systems with low latency and low signaling overhead [18]. In uplink grant-free access, active devices send their data with no access-grant procedure. Then, a base station (BS) receiver has to identify active devices with no aid of a grant procedure and detect each active device's data from the superimposed signal. The principle of compressed sensing (CS) [19] can be applied for multiuser detection in uplink grant-free access, exploiting the sparse activity that many devices are present in a cell, but only a small fraction of them are active at a time. Many research articles [20]− [26] demonstrated that a CS-based detector can be successfully deployed at BS for sparse activity detection in uplink grant-free access.

To guarantee reliable activity detection for uplink grant-free NOMA, it is essential to design good non-orthogonal sequences with low correlation. In literature, a variety of non-orthogonal sequences, e.g., quasi-orthogonal [27], pseudo-random noise [21], random Gaussian [22]–[25], sinusoidal [28] sequences, etc., have been considered. Also, complex-valued non-orthogonal sequences with finite elements have been employed for multi-user shared access (MUSA) [9]. Golay-based non-orthogonal sequences and sets [29], [30] have been also studied for low correlation and low peak power in OFDM-based NOMA. In practice, the Zadoff-Chu (ZC) sequences [31] have been adopted as preambles for random access in 3GPP-LTE [32].

Besides the constructive methods, one can find other classes of non-orthogonal sequences from the efforts of algorithmic optimization of CS matrices [33]− [38], where each matrix is considered as a collection of non-orthogonal column sequences. Each algorithmic design produces non-orthogonal sequences with low correlation by constructing an optimized CS matrix with low coherence. For example, Elad [33] launched an algorithmic design for a sensing matrix by minimizing the average measure of coherence iteratively. In [34]− [37], each algorithm attempted to optimize a sensing matrix by approximating its Gram matrix to that of an equiangular tight frame (ETF) [39]. The authors of [38] presented several algorithms to construct incoherent frames under various design constraints. Recently, the genetic algorithm (GA) [40] has been applied for optimizing a partial Fourier matrix in radar imaging [41], finding a frozen bit position for polar code design [42], maximizing the energy efficiency of wireless sensor networks [43]. Inspired by these works, a two-stage GA [44] has been proposed for non-orthogonal sequence design for multicarrier transmission in single-cell mMTC.

In this paper, we construct multiple sets of Fourier-based non-orthogonal sequences for uplink grant-free access in multi-cell mMTC. We present a design framework based on a partial Fourier matrix with masking operations and define the correlation matrix, which gives a novel approach to analyze the intra- and inter-cell correlations of multiple sequence sets simultaneously. A subsampling index set for the partial Fourier matrix can be found to reduce the intra- and inter-cell correlations of the resulting sequences. In this work, we use the incoherent harmonic design [38] and the genetic algorithms (GA) [40], respectively, to find the index set of an arbitrary set size. We do not focus on the computational complexity of each algorithm, since each one can be executed off-line only once to find the index set, with no need to be implemented in real-time applications. Finally, we obtain multiple partial Fourier matrices having different masks, where each matrix presents a set of Fourier-based non-orthogonal sequences to be uniquely assigned to devices in an mMTC cell.

In CS-based activity detection, simulation results demonstrate that the Fourier-based non-orthogonal sequences outperform random Gaussian, MUSA, and tight-frame (TF) based sequences [37], while comparable to ZC sequences of prime length. Our design framework can provide more sets of the Fourier-based sequences than the ZC sequences, supplying unique signatures of arbitrary length for devices in more mMTC cells. To sum up, this paper presents the main contribution of the design framework for multiple sets of Fourier-based non-orthogonal sequences of arbitrary length, which show the outstanding performance of sparse activity detection and the practical advantages for multi-cell massive connectivity.

The rest of this paper is organized as follows. Section II describes a system model of uplink grant-free access in multi-cell mMTC, where a CS problem is formulated for sparse activity detection. Section III presents a design framework for multiple sets of Fourier-based non-orthogonal sequences and defines the correlation matrix to analyze the intra- and inter-cell correlations of the sequence sets simultaneously. In Section IV, we introduce the algorithms used in this paper to find a subsampling index set for a partial Fourier matrix. Section V presents simulation results of the Fourier-based non-orthogonal sequences to demonstrate the performance of CS-based activity detection for uplink grant-free access in multi-cell mMTC. Finally, concluding remarks will be given in Section VI.

Notations: In this paper, $\mathbb{Z}_M = \{0, \cdots, M-1\}$. A matrix (or a vector) is represented by a bold-face upper (or a lower) case letter. $\mathbf{F}^T$ and $\mathbf{F}^*$ denote the transpose and the conjugate transpose of a matrix $\mathbf{F}$, respectively. For an $M \times N$ matrix $\mathbf{F}$, $\mathbf{F}(k, l)$ is its entry in the kth row and the lth column, where $0 \le k \le M-1$ and $0 \le l \le N-1$. $\mathbf{F}(a{:}b,\ :)$ is a sub-row matrix of $\mathbf{F}$ extracting the row indices from a to b, while $\mathbf{F}(:,\ c{:}d)$ is a sub-column matrix of $\mathbf{F}$ extracting the column indices from c to d. Also, $\mathbf{F}(k,\ c{:}d)$ is a vector of $\mathbf{F}$ extracting the column indices from c to d at the kth row. $\text{abs}(\mathbf{F}) = [|\mathbf{F}(k, l)|]$ is a matrix taking the magnitude of each element of $\mathbf{F}$. A matrix whose elements are all 1 is denoted by $\mathbf{1}$, where the dimension is determined in the context. $\text{diag}(\mathbf{h})$ is a diagonal matrix whose diagonal entries are from a vector $\mathbf{h}$.

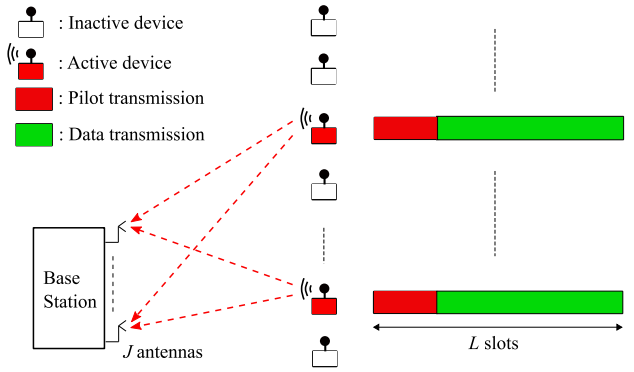

FIGURE 1. Two-phase grant-free access scheme for a single mMTC cell.

The inner product of vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$. The l_p-norm of $\mathbf{x}$ is denoted by $||\mathbf{x}||_p = \left(\sum_{k=1}^{N} |x_k|^p\right)^{\frac{1}{p}}$ for $1 \le p < \infty$. Also, $\|\mathbf{x}\|_\infty$ denotes the maximum magnitude of the elements of $\mathbf{x}$. The Frobenius norm of a matrix $\mathbf{F}$ is $\|\mathbf{F}\|_F = \sqrt{\sum_{k,l} |\mathbf{F}(k,l)|^2}$. The coherence of a matrix $\mathbf{X}$ is defined by $\mu(\mathbf{X}) = \max_{k \neq l} \frac{|\langle \mathbf{x}_k, \mathbf{x}_l \rangle|}{\|\mathbf{x}_k\|_2 \|\mathbf{x}_l\|_2}$, where $\mathbf{x}_k$ and $\mathbf{x}_l$ are the kth and the lth columns of $\mathbf{X}$, respectively. Finally, $\mathbf{h} \sim \mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$ is a circularly symmetric complex Gaussian random vector with mean $\mathbf{m}$ and covariance Σ.

II. SYSTEM MODEL

In this paper, we consider a cellular network of B cells, where each mMTC cell is denoted by $\mathcal{C}_b$, $b = 1, \cdots, B$. Each cell has a base station (BS) equipped with J antennas at the center, to accommodate total N devices each of which transmits with a single antenna. For a fully grant-free access, we assume that each BS serves a fixed set of static devices, which are uniformly distributed in a cell.

In each mMTC cell, we consider a two-phase grant-free access scheme [20], [22] for uplink massive random access. In the first phase, each active device transmits its unique signature as a dedicated pilot, and the BS receiver then tries to identify active devices and estimate their channel profiles from the superimposed signatures. Data can be directly transmitted in the second-phase from active devices with no grant from BS. In this two-phase scheme, we assume that the channels and the device activity remain unchanged during L slots for pilot and data transmissions. Fig. 1 illustrates this access scheme for a single mMTC cell.

At each access, an activity indicator vector of cell $\mathcal{C}_b$ is defined by $\boldsymbol{\alpha}^{(b)} = (\alpha_1^{(b)}, \cdots, \alpha_N^{(b)})^T$, where $\alpha_n^{(b)} = 1$ if device n is active in the cell, and 0 otherwise. Then, $\mathcal{S}^{(b)} = \{n \mid \alpha_n^{(b)} = 1,\ n = 1, \cdots, N\}$ is a set of active devices of $\mathcal{C}_b$, which are synchronized. Each device of $\mathcal{C}_b$ is active with probability $p_a^{(b)}$ in an i.i.d. manner, i.e., $\Pr[\alpha_n^{(b)} = 1] = p_a^{(b)}$ for $n = 1, \cdots, N$. When device n of $\mathcal{C}_b$ is active, it transmits its unique *signature* $\mathbf{s}_n^{(b)} = (s_{1,n}^{(b)}, \cdots, s_{M,n}^{(b)})^T$ for grant-free access, where $M < N$. In this paper, we assume that each active device transmits its signature with the same transmit power ρ.

Without loss of generality, we consider sparse activity detection at a BS receiver of $\mathcal{C}_1$. Assuming the flat Rayleigh fading channel, let $h_{t,n}^{(b)}$ be the channel gain between the BS receiver's antenna t of $\mathcal{C}_1$ and device n of $\mathcal{C}_b$, which remains unchanged during the coherence time interval. Then, the channel vector for the device n is defined by $\mathbf{h}_n^{(b)} = (h_{1,n}^{(b)}, \cdots, h_{J,n}^{(b)})^T$, where $\mathbf{h}_n^{(b)} \sim \mathcal{CN}(\mathbf{0}, \beta_n^{(b)}\mathbf{I})$. The path-loss component $\beta_n^{(b)}$ is determined by the device location from the BS of $\mathcal{C}_1$.

Let $\mathbf{S}^{(b)} = [\mathbf{s}_1^{(b)}, \cdots, \mathbf{s}_N^{(b)}] \in \mathbb{C}^{M \times N}$ be a signature sequence matrix of $\mathcal{C}_b$. At the BS receiver of $\mathcal{C}_1$, the received signal at antenna t is given by

$$\mathbf{y}_t = \sqrt{\xi}\mathbf{S}^{(1)}\mathbf{x}_t^{(1)} + \sqrt{\xi}\sum_{b=2}^{B}\mathbf{S}^{(b)}\mathbf{x}_t^{(b)} + \mathbf{w}_t, \tag{1}$$

where $\xi = \rho M$ and $\mathbf{x}_t^{(b)} = (\alpha_1^{(b)} h_{t,1}^{(b)}, \cdots, \alpha_N^{(b)} h_{t,N}^{(b)})^T$ for $t = 1, \cdots, J$. In (1), $\mathbf{w}_t \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I})$ is the complex Gaussian noise vector at the antenna t. Collecting the received signal of (1) from the J antennas, the BS receiver has a multiple measurement vector (MMV) problem of

$$\mathbf{Y} = \sqrt{\xi}\mathbf{S}^{(1)}\mathbf{X}^{(1)} + \sqrt{\xi}\sum_{b=2}^{B}\mathbf{S}^{(b)}\mathbf{X}^{(b)} + \mathbf{W}, \tag{2}$$

where $\mathbf{X}^{(b)} = [\mathbf{x}_1^{(b)}, \cdots, \mathbf{x}_J^{(b)}] \in \mathbb{C}^{N \times J}$, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_J] \in \mathbb{C}^{M \times J}$, and $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_J] \in \mathbb{C}^{M \times J}$, respectively. In (2), the second-term of the right-hand side is the *inter-cell* interference from active devices in $\mathcal{C}_2, \cdots, \mathcal{C}_B$. Obviously, $\mathbf{X}^{(1)}$ has the row-wise sparsity from the activity indicator $\boldsymbol{\alpha}^{(1)}$. Thus, the BS receiver of $\mathcal{C}_1$ can apply a joint sparse recovery algorithm to identify the nonzero rows of $\mathbf{X}^{(1)}$ from $\mathbf{Y}$ for CS-based detection of active devices in $\mathcal{C}_1$. In specific, if the nonzero rows of $\mathbf{X}^{(1)}$ are estimated, the row indices mean a detected index set of active devices at $\mathcal{C}_1$, denoted by $\widehat{\mathcal{S}}^{(1)}$, while the coefficients of each nonzero row give an estimated channel vector $\widehat{\mathbf{h}}_n^{(1)}$ for device $n \in \widehat{\mathcal{S}}^{(1)}$. The CS-based sparse activity detection and channel estimation complete the first phase of uplink grant-free random access at cell $\mathcal{C}_1$. In the second phase, the BS receiver detects data from active devices with the knowledge of device identity and channel profiles obtained from the first phase [20], [22]. In this paper, we restrict our attention to sparse activity detection in the first phase by identifying $\widehat{\mathcal{S}}^{(1)}$ via joint sparse recovery under the CS MMV model.

As in massive MIMO regime of [23], the inter-cell interference in (2) is treated as noise in CS-based activity detection at $\mathcal{C}_1$. In practice, some advanced techniques such as beamforming, cooperative MIMO [23], multi-point coordination [45], etc., can be deployed for mitigating the inter-cell interference. However, it is noteworthy that the main result of this paper, i.e., multiple sets of non-orthogonal sequences, will not be so relevant to whether or not the interference mitigation techniques are considered, since the design effort relies on the correlation properties of non-orthogonal sequences.

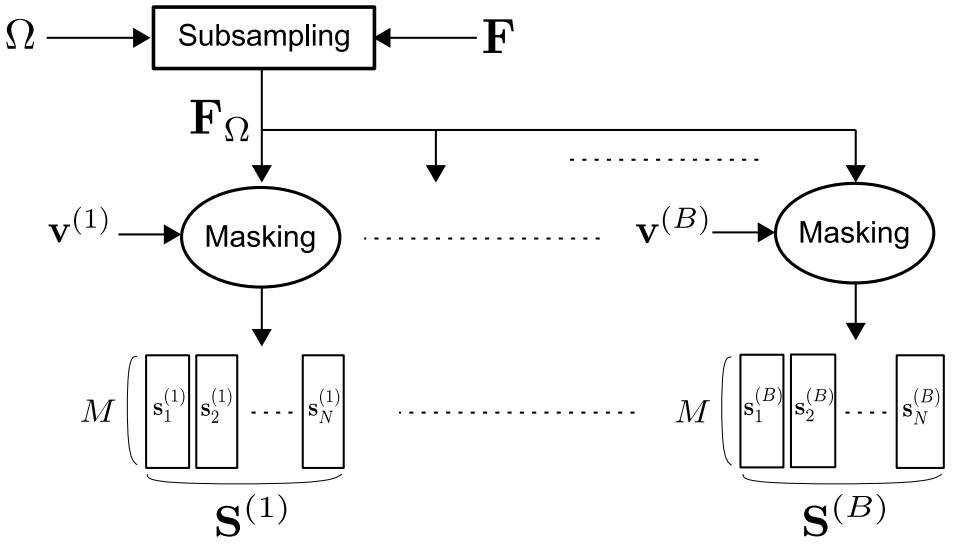

FIGURE 2. Design framework of Construction 1. Each set $\mathbf{S}^{(b)}$, $b = 1, \cdots, B$, has N non-orthogonal sequences of length M, where $M < N$.

III. DESIGN FRAMEWORK OF FOURIER-BASED NON-ORTHOGONAL SEQUENCES

This section presents a design framework based on a partial Fourier matrix with masking operations for multiple sets of non-orthogonal sequences, which is the main contribution. Then, we define the correlation matrix to analyze the intra- and inter-cell correlations simultaneously.

A. DESIGN FRAMEWORK

Let $\mathbf{F}$ be the N-point discrete Fourier matrix of $\mathbf{F}(k, l) = \exp\left(-\frac{j2\pi kl}{N}\right)$ for $0 \le k, l \le N-1$, where $j = \sqrt{-1}$. Then, we define a partial Fourier matrix $\mathbf{F}_\Omega$ by selecting M rows out of $\mathbf{F}$, where the selected row indices are specified by $\Omega = \{r_1, \cdots, r_M\} \subset \{0, \cdots, N-1\}$ in unsorted order. Using $\mathbf{F}_\Omega$, we present a framework to construct B ($\le M$) sequence sets, each of which has N non-orthogonal sequences of length M.

Construction 1: To construct B ($\le M$) sequence sets, define a masking sequence $\mathbf{v}^{(b)} = (v_1^{(b)}, \cdots, v_M^{(b)})^T$ by $v_m^{(b)} = \exp\left(\frac{j2\pi a_m^{(b)}}{M}\right)$ for $b = 1, \cdots, B$, where

$$a_m^{(b)} \equiv a_m^{(1)} + (m-1)(b-1) \pmod{M}, \tag{3}$$

for $m = 1, \cdots, M$. In (3), $a_m^{(1)}$ is randomly taken from $\mathbb{Z}_M$ for each m. Then, a signature sequence matrix $\mathbf{S}^{(b)}$ is

$$\mathbf{S}^{(b)} = \frac{1}{\sqrt{M}}\mathrm{diag}(\mathbf{v}^{(b)}) \cdot \mathbf{F}_\Omega = \left[\mathbf{s}_1^{(b)}, \cdots, \mathbf{s}_N^{(b)}\right], \tag{4}$$

for $b = 1, \cdots, B$, where the subsampling index set Ω can be found algorithmically or algebraically. Each matrix $\mathbf{S}^{(b)}$ is a set of N sequences of length M for cell $\mathcal{C}_b$, where device n is assigned the unique signature $\mathbf{s}_n^{(b)}$ for $n = 1, \cdots, N$.

Construction 1 presents total B ($\le M$) sets of the *Fourier-based* sequences of length M to supply unique signatures for B cells, where each one has N devices. From (4), note that Construction 1 utilizes a common subsampling index set Ω (or equivalently $\mathbf{F}_\Omega$) for all signature matrices $\mathbf{S}^{(1)}, \cdots, \mathbf{S}^{(B)}$, while each matrix $\mathbf{S}^{(b)}$ has its own masking sequence $\mathbf{v}^{(b)}$ for $b = 1, \cdots, M$.

Fig. 2 illustrates the design framework of Construction 1. In this paper, we consider to employ an algorithm to find the subsampling index set Ω of an arbitrary set size, which yields the Fourier-based sequences of arbitrary length. Note that the algorithm can be executed offline only once to find Ω, with no need to be implemented in real-time applications. Also, the masking sequence $\mathbf{v}^{(b)}$ is obtained by (3), not by an algorithm. Thus, the Fourier-based sequences can be efficiently generated in practice by storing the subsampling index set Ω and the masking sequence $\mathbf{v}^{(b)}$.

B. CORRELATION MATRIX

We introduce a correlation measure for the sequences in $\mathbf{S}^{(1)}, \cdots, \mathbf{S}^{(B)}$, which can be used by an algorithm to find a subsampling index set Ω. For $1 \le b_1, b_2 \le B$ and $1 \le n_1, n_2 \le N$, the correlation[1] of $\mathbf{s}_{n_1}^{(b_1)} \in \mathbf{S}^{(b_1)}$ and $\mathbf{s}_{n_2}^{(b_2)} \in \mathbf{S}^{(b_2)}$ is defined by

$$C_{b_1,b_2}(n_1, n_2) = \frac{\left\langle \mathbf{s}_{n_1}^{(b_1)}, \mathbf{s}_{n_2}^{(b_2)} \right\rangle}{\left\| \mathbf{s}_{n_1}^{(b_1)} \right\|_2 \left\| \mathbf{s}_{n_2}^{(b_2)} \right\|_2}, \tag{5}$$

which denotes the *inter-cell* correlation if $b_1 \ne b_2$, and the *intra-cell* correlation if $b_1 = b_2$ and $n_1 \ne n_2$. In $C_{b_1,b_2}(n_1, n_2)$, we assume $b_1 \ge b_2$ and $n_1 \ge n_2$ without loss of generality. From Construction 1, (5) yields

$$\begin{aligned} &C_{b_1,b_2}(n_1, n_2) \\ &= \frac{1}{M}\sum_{m=1}^{M} e^{-j\frac{2\pi r_m(n_1-1)}{N}} e^{j\frac{2\pi r_m(n_2-1)}{N}} \cdot e^{j\frac{2\pi\left(a_m^{(b_1)} - a_m^{(b_2)}\right)}{M}} \\ &= \frac{1}{M}\sum_{m=1}^{M} e^{-j\frac{2\pi r_m l}{N}} \cdot e^{j\frac{2\pi(m-1)k}{M}}, \end{aligned} \tag{6}$$

[1] In fact, it is the (normalized) *inner product* of $\mathbf{s}_{n_1}^{(b_1)}$ and $\mathbf{s}_{n_2}^{(b_2)}$. We call it the correlation, following the convention of synchronous multiple access.

for $0 \le k = b_1 - b_2 \le B-1$ and $0 \le l = n_1 - n_2 \le N-1$, where $l \ne 0$ if $k = 0$. Clearly, (6) shows that the correlation is determined by the M-point inverse discrete Fourier transform (IDFT) of each column of $\mathbf{F}_\Omega$. Let $\mathbf{D} = \mathbf{F}_M(:,\ 0{:}B-1)$, where $\mathbf{F}_M$ is the M-point discrete Fourier matrix. Then, we define a *correlation matrix* by

$$\mathbf{C}_\Omega = \frac{1}{M}\text{abs}\left(\mathbf{D}^*\mathbf{F}_\Omega\right) \in \mathbb{C}^{B\times N}, \tag{7}$$

where the element $\mathbf{C}_\Omega(k,l)$ means the correlation magnitude $\left|C_{b_1,b_2}(n_1,n_2)\right|$, where $k = b_1 - b_2$ and $l = n_1 - n_2$. From (6), the first row of $\mathbf{C}_\Omega$, excluding $\mathbf{C}_\Omega(0,0) = 1$, has the intra-cell correlation magnitudes, while the remaining submatrix contains the inter-cell correlation magnitudes, denoted by

$$\begin{aligned}\mathbf{C}_\Omega^{\text{intra}} &= \mathbf{C}_\Omega(0,\ 1{:}N-1),\\ \mathbf{C}_\Omega^{\text{inter}} &= \mathbf{C}_\Omega(1{:}B-1,\ :).\end{aligned} \tag{8}$$

Clearly, (8) shows that the correlation matrix of (7) has the intra- and inter-cell correlations that need to be reduced for sequence design. In other words, each element of $\mathbf{C}_\Omega(0,\ 1{:}N-1)$ is the correlation magnitude of signatures $\mathbf{s}_{n_1}^{(b)}$ and $\mathbf{s}_{n_2}^{(b)}$, where $k = b-b = 0$ and $l = n_1 - n_2 \ne 0$, which indicates the intra-cell correlation of signatures in cell $\mathcal{C}_b$. On the other hand, each element of $\mathbf{C}_\Omega(1{:}B-1,\ :)$ is the correlation magnitude of signatures $\mathbf{s}_{n_1}^{(b_1)}$ and $\mathbf{s}_{n_2}^{(b_2)}$, where $k = b_1 - b_2 \ne 0$ and $l = n_1 - n_2$, which shows the inter-cell correlation of signatures in cells $\mathcal{C}_{b_1}$ and $\mathcal{C}_{b_2}$.

In general, one has to investigate $\mathcal{O}(B^2N^2)$ correlations to construct B sets of N sequences with low intra- and inter-cell correlations. However, we need to deal with only the BN correlations of $\mathbf{C}_\Omega$ in Construction 1, which are taken from the IDFT of $\mathbf{F}_\Omega$ by (7). Thus, the correlation matrix gives a novel approach to analyze the intra- and inter-cell correlations simultaneously for multiple sequence sets under our design framework.

In what follows, we investigate the average magnitudes of intra- and inter-cell correlations, respectively.

Theorem 1: For given b_1 and b_2, $1 \le b_1 \ne b_2 \le B$, the average (rms-sense) magnitude of inter-cell correlation of $\mathbf{S}^{(b_1)}$ and $\mathbf{S}^{(b_2)}$ from Construction 1 is

$$A_{\text{inter}} = \sqrt{\frac{1}{N^2}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\left|C_{b_1,b_2}(n_1,n_2)\right|^2} = \frac{1}{\sqrt{M}}, \tag{9}$$

regardless of Ω.

For given b, $1 \le b \le B$, the average (rms-sense) magnitude of intra-cell correlation of $\mathbf{S}^{(b)}$ is

$$\begin{aligned}A_{\text{intra}} &= \sqrt{\frac{1}{N(N-1)}\sum_{n_1=1}^{N}\sum_{\substack{n_2=1\\ n_2\ne n_1}}^{N}\left|C_{b,b}(n_1,n_2)\right|^2}\\ &= \sqrt{\frac{N-M}{M(N-1)}},\end{aligned} \tag{10}$$

regardless of Ω.

Proof: From (6),

$$\begin{aligned}&M^2\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\left|C_{b_1,b_2}(n_1,n_2)\right|^2\\ &= \sum_{m=1}^{M}\sum_{s=1}^{M} e^{j\frac{2\pi(m-s)(b_1-b_2)}{M}}\cdot\sum_{n_1=1}^{N}\sum_{n_2=1}^{N} e^{-j\frac{2\pi(r_m-r_s)(n_1-n_2)}{N}}.\end{aligned} \tag{11}$$

If we set $k = b_1 - b_2$ and $l = n_1 - n_2$, (11) becomes

$$\begin{aligned}&\sum_{m=1}^{M}\sum_{s=1}^{M} e^{j\frac{2\pi(m-s)k}{M}}\cdot\left(N\sum_{l=0}^{N-1} e^{-j\frac{2\pi(r_m-r_s)l}{N}}\right)\\ &= \sum_{m=1}^{M} N^2 + N\sum_{m=1}^{M}\sum_{\substack{s=1\\ s\ne m}}^{M} e^{j\frac{2\pi(m-s)k}{M}}\cdot\left(\sum_{l=0}^{N-1} e^{-j\frac{2\pi(r_m-r_s)l}{N}}\right),\\ &= N^2 M,\end{aligned} \tag{12}$$

which is obvious from $\sum_{l=0}^{N-1} e^{-j\frac{2\pi(r_m-r_s)l}{N}} = 0$ for $m \ne s$. Therefore, the average magnitude of inter-cell correlation of (9) is immediate from (11) and (12).

Similarly, if $b_1 = b_2 = b$,

$$\begin{aligned}&\sum_{n_1=1}^{N}\sum_{\substack{n_2=1\\ n_2\ne n_1}}^{N}\left|C_{b,b}(n_1,n_2)\right|^2\\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\left|C_{b,b}(n_1,n_2)\right|^2 - \sum_{n_1=1}^{N}\left|C_{b,b}(n_1,n_1)\right|^2,\\ &= \frac{N^2}{M} - N,\end{aligned} \tag{13}$$

where the first-term is from (11) and (12) with $k = 0$. Then, (10) is obvious from (13). □

IV. ALGORITHMS FOR SUBSAMPLING INDEX SET

Section III presented a design framework for Fourier-based non-orthogonal sequences, where a subsampling index set Ω is necessary for Construction 1. In this section, we describe two different algorithms to find Ω of an arbitrary set size.

A. INCOHERENT HARMONIC DESIGN ALGORITHM

In [38], several algorithms for incoherent frames with various design constraints have been proposed. In particular, Algorithm 2 of [38] returns a subsampling row index set for an incoherent harmonic frame, which is equivalent to a partial Fourier matrix with low coherence. In this paper, we give a brief sketch of the algorithm, and readers are referred to Section 5 of [38] for more details.

Let $\mathbf{g} = (g_1,\cdots,g_N)^T \in \{0,1\}^N$ be a binary indicator vector, where a subsampling index set is determined by $\Omega = \{k-1 \mid g_k = 1, k = 1,\cdots,N\}$. From $\mathbf{F}_\Omega^*\mathbf{F}_\Omega = \mathbf{F}^*\text{diag}(\mathbf{g})\mathbf{F}$, it is clear that the Gram matrix is a circulant matrix in which the first column is $\widehat{\mathbf{g}} = \frac{1}{M}\mathbf{F}^*\mathbf{g} = (\widehat{g}_1,\cdots,\widehat{g}_N)^T$. Thus, each element of $\widehat{\mathbf{g}} \setminus \{\widehat{g}_1\}$ corresponds to the correlation of a distinct column pair of $\mathbf{F}_\Omega$. The essence of the algorithm is to find $\mathbf{g}$ to minimize the maximum magnitude of $\widehat{\mathbf{g}} \setminus \{\widehat{g}_1\}$,

which ultimately minimizes the coherence of a partial Fourier matrix $\mathbf{F}_\Omega$.

In [38], an algorithmic approach has been used to find $\mathbf{g}$ for incoherent harmonic design. Algorithm 2 of [38] starts by randomly generating an initial subsampling index set Ω_0 of size $\lceil \zeta(N-M) \rceil$. Then, it tries to find a solution $\mathbf{x} = (x_1, \cdots, x_N)^T \in \mathbb{R}^N$ to a convex optimization problem [38] formulated by

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \frac{1}{M}\|\mathbf{U}\mathbf{x}\|_\infty + \lambda\|\text{diag}(\mathbf{w})\mathbf{x}\|_1 \\ &\text{subject to } x_1 = 1,\ \|\mathbf{x}\|_1 = M,\ x_k = 0 \text{ for } k-1 \in \Omega_0, \\ &\qquad 0 \le x_k \le 1 \text{ for } 2 \le k \le N, \end{aligned} \tag{14}$$

where $\mathbf{U} = \mathbf{F}^*(1{:}\lfloor N/2 \rfloor, :)$ is a sub-row matrix of $\mathbf{F}^*$. In (14), $\|\text{diag}(\mathbf{w})\mathbf{x}\|_1$ is a regularization term to promote the sparsity for the solution $\mathbf{x}$, where each element of $\mathbf{w} = (w_1, \cdots, w_N)^T$ is $w_k = (x_k + \epsilon)^{-1}$ for a small $\epsilon > 0$.

If (14) finds a solution $\mathbf{x}$, the algorithm determines a subsampling index set by $\Omega = \{k-1 \mid |x_k| > \epsilon,\ k = 1, \cdots, N\}$. If $|\Omega| > M$, it eliminates some indices one by one until only M indices remain in Ω, in such a way that each elimination causes a minimal increase in the coherence [38]. In addition, the algorithm conducts a local search for updating the index set, where some l indices of Ω can be replaced by new ones in $\{0, \cdots, N-1\} \setminus \Omega$ for further reduction of coherence. The entire steps of this algorithm are repeated by a predefined number of iterations I_h, which yields I_h index sets. Finally, we select a subsampling index set Ω with the lowest coherence among them.

In this paper, we use Algorithm 2 of [38] to find the index set Ω for Construction 1, where the optimization problem of (14) is solved via CVX [46]. We modify the algorithm slightly to obtain a subsampling index set Ω that has the indices in an unsorted order, by taking into account the magnitudes of the elements of $\mathbf{x}$ in (14). While its minimizing the coherence of a partial Fourier matrix ultimately guarantees low intra-cell correlation for the Fourier-based sequences, the algorithm in itself makes no effort to reduce the inter-cell correlation. Under our design framework, we found that it is a non-trivial task to modify the algorithm for reducing the intra- and inter-cell correlations simultaneously using the correlation matrix of (7), which will remain as a future research issue.

B. GENETIC ALGORITHM (GA)

The genetic algorithm (GA) [40] is an effective evolutionary technique to find a subsampling index set for a good partial Fourier matrix [41], which motivates us to use GA to find Ω in Construction 1. While the GA of [44] is used for a single sequence set, the GA of this paper is to design multiple sequence sets considering the intra- and inter-cell correlations. Also, the GA of [44] can be applied to an arbitrary unitary matrix, but this GA is only for the Fourier matrix, which facilitates our correlation analysis through the correlation matrix of (7).

We now describe the evolutionary steps of GA used in this paper to find Ω.

1) INITIALIZATION

A population is initialized by $\mathcal{P} = (\Omega_1, \cdots, \Omega_T)$, where $\Omega_t \subset \{0, \cdots, N-1\}$ is a randomly selected (unsorted) index set with $|\Omega_t| = M$ for $t = 1, \cdots, T$.

2) COST FUNCTION

We propose a cost function of

$$f(\mathbf{F}_{\Omega_t}) = \sqrt{d_{\text{intra}}^2 + d_{\text{inter}}^2} \tag{15}$$

for each Ω_t in $\mathcal{P}$, where $d_{\text{intra}} = \frac{1}{\sqrt{N-1}} \left\| \mathbf{C}_{\Omega_t}^{\text{intra}} - A_{\text{intra}}\mathbf{1} \right\|_F$ and $d_{\text{inter}} = \frac{1}{\sqrt{N(B-1)}} \left\| \mathbf{C}_{\Omega_t}^{\text{inter}} - A_{\text{inter}}\mathbf{1} \right\|_F$ are the average distances from intra- and inter-cell correlation magnitudes to their averages, respectively. In (15), $\mathbf{C}_{\Omega_t}^{\text{intra}}$ and $\mathbf{C}_{\Omega_t}^{\text{inter}}$ are computed for $\mathbf{F}_{\Omega_t}$ by (7) and (8). Attempting to minimize the cost function of (15), this GA seeks a subsampling index set that makes the intra- and inter-cell correlations closer to their averages, which will contribute to constructing good Fourier-based sequences by trying to reduce the correlations simultaneously.

3) CROSSOVER

In population $\mathcal{P}$, let us consider a pair of index sets Ω_{t_1} and Ω_{t_2}, $1 \le t_1 \ne t_2 \le T$. Then, $d_1 = \lceil 0.5 \cdot M \rceil$ and $d_2 = M - d_1$ indices are randomly selected from Ω_{t_1} and Ω_{t_2}, respectively. Finally, the selected indices, which should be all distinct, are combined to generate a new index set through *crossover*. Applying the crossover for every pair of index sets in $\mathcal{P}$, we obtain a new population $\mathcal{C}$.

4) MUTATION

In this step, μ indices are randomly selected from each index set of $\mathcal{P}$, and then replaced by new (random) ones through *mutation*. We apply the mutation to all index sets in $\mathcal{P}$, which yields a new population $\mathcal{M}$.

5) POPULATION UPDATE

Through crossover and mutation, we have a new, intermediate population $\mathcal{I} = \mathcal{P} \cup \mathcal{C} \cup \mathcal{M}$, which contains at most $T + \binom{T}{2} + T = \frac{T(T+3)}{2}$ index sets. From $\mathcal{I}$, we select the T index sets with the T lowest cost functions of (15). Then, $\mathcal{P}$ is updated by the T selected index sets.

6) ITERATION AND SELECTION

The evolution steps of crossover, mutation, and population update are repeated by a predefined number of iterations, denoted by $I_{\max}$. In the end, the fittest index set of $\mathcal{P}$, which has the lowest cost function of (15), will be selected as a subsampling index set Ω. Unlike the incoherent harmonic design algorithm, the GA using the cost function (15) makes an effort to find Ω for reducing the intra- and inter-cell correlations simultaneously.

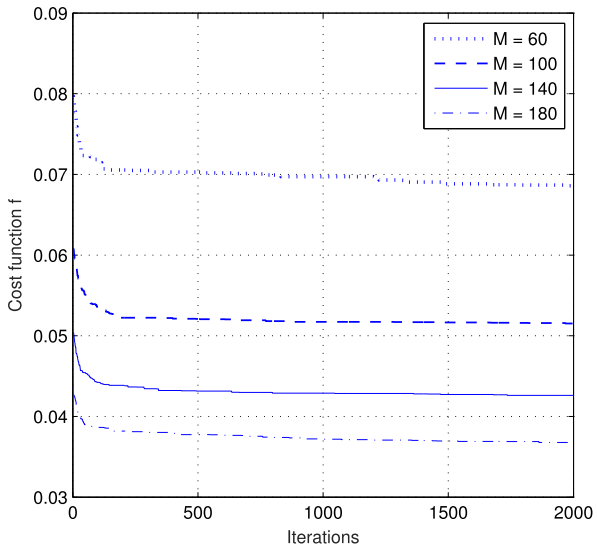

FIGURE 3. Evolution of the cost function of GA for Fourier-based sequences of various lengths, where $T = 20$ and $\mu = 1$.

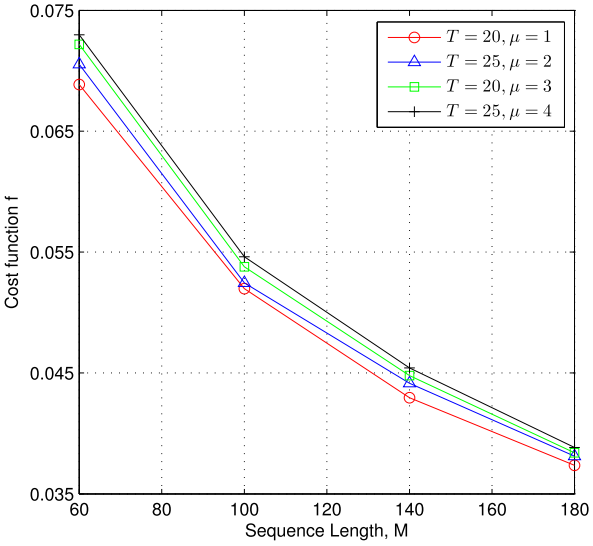

FIGURE 4. Lowest cost function of GA for Fourier-based sequences over various T and μ, where $I_{\max} = 2000$.

With the iterative procedures, the incoherent harmonic design and the genetic algorithms may find the subsampling index set slowly with high computational complexity. Thus, it would be inefficient to implement the algorithms in real-time applications. As discussed in Section III.A, each algorithm can be executed offline only once to find the index set. Then, the Fourier-based sequences can be generated efficiently in practice by storing the index set and the masking sequence.

V. SIMULATION RESULTS

This section presents simulation results of the Fourier-based non-orthogonal sequences to demonstrate the performance for uplink grant-free access in multi-cell mMTC.

A. TESTED NON-ORTHOGONAL SEQUENCES

In simulations, we construct $B = 19$ sets of the Fourier-based sequences, where each set has $N = 500$ sequences. We use three different methods to find the subsampling index set Ω for Construction 1. First, we use a random subsampling by taking a subset Ω randomly from $\{1, \cdots, N\}$. Second, we employ the incoherent harmonic design algorithm introduced in Section IV.A, where $\lambda = \frac{1}{M}$, $\epsilon = 10^{-7}$, $\zeta = 0.1$, $l = 1$, and $I_h = 50$.

Third, the GA of Section IV.B is used to find Ω, where the parameters are empirically determined for $N = 500$ and $B = 19$. Fig. 3 displays the evolution of the cost function of (15) for several sequence lengths $M \in [60, 100, 140, 180]$, where $T = 20$ and $\mu = 1$. From the figure, we set $I_{\max} = 2000$, which confirmed the convergence of the GA for the sequence lengths through numerical experiments. Given $I_{\max}$, we then examined the cost function of the subsampling index set Ω found by the GA, with the parameters of $T \in [5, 10, 15, 20, 25]$ and $\mu \in [1, 2, 3, 4]$. Fig. 4 sketches the lowest cost function among all the tested T for each μ. As can be seen from the figure, $T = 20$ and $\mu = 1$ show the lowest cost function for all the tested M. Consequently, we construct the Fourier-based sequences of Construction 1 by finding the subsampling index set Ω using the GA with $I_{\max} = 2000$, $T = 20$, and $\mu = 1$.

Table 1 gives some examples of the Fourier-based sequences of length $M = 100$, where the subsampling index set Ω has been found by GA. Recall that a set of the Fourier-based sequences for cell $\mathcal{C}_b$ is denoted by $\mathbf{S}^{(b)} = [\mathbf{s}_1^{(b)}, \cdots, \mathbf{s}_N^{(b)}]$, where $N = 500$. Then, each element of $\mathbf{s}_n^{(b)}$ can be represented by $s_{m,n}^{(b)} = \frac{1}{\sqrt{M}} e^{j\frac{2\pi u_{m,n}^{(b)}}{N}}$ with $u_{m,n}^{(b)} \in \mathbb{Z}_N$ for $1 \le n \le N$ and $1 \le m \le M$. Table 1 gives $\mathbf{u}_2^{(1)} = (u_{1,2}^{(1)}, \cdots, u_{M,2}^{(1)})^T$ for $\mathbf{s}_2^{(1)}$ at cell $\mathcal{C}_1$, and $\mathbf{u}_3^{(2)} = (u_{1,3}^{(2)}, \cdots, u_{M,3}^{(2)})^T$ for $\mathbf{s}_3^{(2)}$ at cell $\mathcal{C}_2$, respectively. Note that each sequence may have duplicate elements, due to the operation of random masking in the design framework.

For comparison, we consider some known sequences of flexible lengths. First, we generate random Gaussian sequences of length M, where each element is drawn from the i.i.d. complex Gaussian distribution with zero mean and variance $1/M$. We also use the MUSA sequences of length M, where each element is randomly taken from the 3-level signal constellation [9]. Generating $420,020$ sets[2] of N random Gaussian (or MUSA) sequences each, we choose B sets with the B lowest maximum intra-cell correlations for simulation. In addition, we construct B sets of the tight-frame (TF) based sequences of length M, where each set is obtained from an $M \times N$ optimized sensing matrix by running the iterative algorithm of [37].

[2]For a fair comparison, this number is the sum of total population sizes of our GA, i.e., $T + I_{\max} \cdot T(T+1)/2$ with $T = 20$ and $I_{\max} = 2000$.

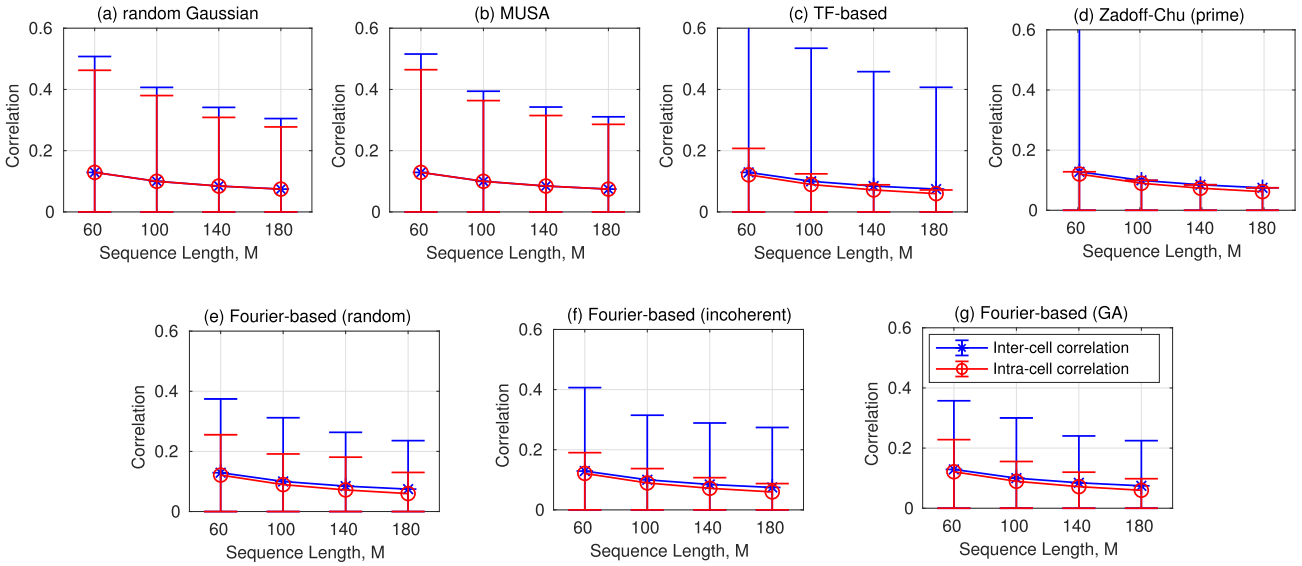

FIGURE 5. Intra- and inter-cell correlation distributions of tested non-orthogonal sequences over sequence length, where $N = 500$ and $B = 19$.

TABLE 1. Examples of Fourier-based sequences obtained from GA, where $N = 500$, $M = 100$, and $B = 19$.

$\mathbf{u}_2^{(1)}$	118, 21, 211, 367, 278, 495, 94, 15, 118, 77, 340, 309, 455, 222, 355, 205, 75, 203, 249, 273, 335, 217, 455, 325, 48, 180, 5, 343, 359, 270, 310, 353, 444, 104, 482, 113, 16, 345, 368, 161, 109, 298, 328, 495, 278, 417, 460, 189, 242, 343, 371, 242, 477, 40, 241, 82, 334, 282, 399, 150, 340, 464, 307, 181, 87, 489, 140, 472, 435, 229, 396, 396, 91, 443, 123, 177, 311, 375, 419, 321, 172, 139, 395, 84, 37, 197, 414, 198, 229, 100, 224, 68, 318, 119, 239, 361, 307, 281, 29, 172.
$\mathbf{u}_3^{(2)}$	231, 282, 7, 19, 431, 495, 323, 70, 41, 464, 475, 353, 215, 314, 420, 355, 205, 1, 408, 91, 240, 414, 135, 250, 91, 50, 210, 206, 178, 305, 395, 256, 98, 73, 84, 221, 197, 280, 101, 102, 198, 76, 396, 350, 186, 159, 355, 3, 409, 256, 77, 139, 284, 185, 97, 44, 158, 104, 148, 170, 425, 48, 174, 182, 64, 288, 160, 459, 290, 233, 307, 492, 142, 86, 481, 474, 167, 160, 173, 447, 29, 288, 410, 13, 29, 464, 58, 316, 358, 10, 38, 246, 401, 343, 298, 442, 159, 442, 463, 489.

The last one for comparison is the ZC-based non-orthogonal sequences of length M_{ZC}, where M_{ZC} is an odd prime number closest to M. We begin with an $M_{\mathrm{ZC}} \times M_{\mathrm{ZC}}$ unitary matrix $\mathbf{Z}_u$ whose columns are all cyclic shifts of a uth root Zadoff-Chu (ZC) sequence [31] of length M_{ZC}, where the kth element is $\exp\left(\frac{j\pi uk(k+1)}{M_{\mathrm{ZC}}}\right)$ and u is a root number between 1 and $M_{\mathrm{ZC}} - 1$. Due to the perfect autocorrelation of the ZC sequence, each column pair of $\mathbf{Z}_u$ are mutually orthogonal for any u. Then, we construct a matrix $\mathbf{Z} \in \mathbb{C}^{M_{\mathrm{ZC}} \times N_r M_{\mathrm{ZC}}}$ by concatenating the matrices $\mathbf{Z}_v$, where $v = (u-1) \pmod{M_{\mathrm{ZC}} - 1} + 1$ for $u = 1, \cdots, N_r$ with $N_r = \lceil \frac{BN}{M_{\mathrm{ZC}}} \rceil$. If $N_r \geq M_{\mathrm{ZC}}$, some submatrices $\mathbf{Z}_v$ are reused in $\mathbf{Z}$, due to the shortage of available roots. Finally, writing by $\mathbf{Z} = [\mathbf{Z}'_1, \mathbf{Z}'_2, \cdots]$, we obtain a set of N sequences of length M_{ZC}, or $\mathbf{Z}'_b \in \mathbb{C}^{M_{\mathrm{ZC}} \times N}$, $b = 1, \cdots, B$, to be assigned to cell $\mathcal{C}_b$. With $M_{\mathrm{ZC}} < N$, the maximum magnitude of intra-cell correlations of $\mathbf{Z}'_b$ is $\frac{1}{\sqrt{M_{\mathrm{ZC}}}}$, due to the cross-correlation of ZC sequences with distinct roots [47]. Also, the maximum magnitude of inter-cell correlations of $\mathbf{Z}'_{b_1}$ and $\mathbf{Z}'_{b_2}$ for $b_1 \neq b_2$ is $\frac{1}{\sqrt{M_{\mathrm{ZC}}}}$ if $N_r < M_{\mathrm{ZC}}$, and 1 otherwise, since some ZC sequences are reused if $N_r \geq M_{\mathrm{ZC}}$.

Remark 1: All cyclic shifts of the ZC sequences with $M_{\mathrm{ZC}} - 1$ roots can present at most $\mathcal{N}_{\mathrm{ZC}} = \lfloor \frac{M_{\mathrm{ZC}}(M_{\mathrm{ZC}}-1)}{N} \rfloor$ sequence sets, i.e., $\mathbf{Z}'_1, \cdots, \mathbf{Z}'_{\mathcal{N}_{\mathrm{ZC}}}$, where each set is assigned to a cell of N devices. If $N = LM_{\mathrm{ZC}}$, then $\mathcal{N}_{\mathrm{ZC}} < M_{\mathrm{ZC}}/L \approx M/L$, where $L > 1$. Thus, the Fourier-based sequences of length M, offering up to M distinct sets, can support more cells than the ZC ones. Also, if sequences are to be assigned to B cells with no reuse, each cell can accommodate at most $\lfloor \frac{M_{\mathrm{ZC}}(M_{\mathrm{ZC}}-1)}{B} \rfloor$ devices with the ZC sequences, whereas the Fourier-based sequences have no such limit, as long as $B \leq M$. To support more devices in a cell, we can increase the number of Fourier-based sequences of a set by choosing $\mathbf{F}$ with a larger N, until the performance degradation due to high overloading (N/M) can be tolerated for activity detection in uplink grant-free access.

Table 2 shows the maximum number of cells that can be supported by the ZC and the Fourier-based sequences with no reuse, denoted by B_{ZC} and B_{FR}, respectively. In the table, the length of Fourier-based sequences is set to $M = M_{\mathrm{ZC}} - 1$. As stated in Remark 1, the maximum number of cells with no reuse of ZC sequences is $B_{\mathrm{ZC}} = \lfloor \frac{M_{\mathrm{ZC}}(M_{\mathrm{ZC}}-1)}{N} \rfloor$, which decreases as each cell requires more unique signatures with an increasing number of devices. In contrast, the Fourier-based sequences can supply unique signatures with

TABLE 2. Maximum number of cells supported by ZC and Fourier-based sequences with no reuse.

	$N = 500$		$N = 1000$		$N = 2000$	
M_{ZC}	B_{ZC}	B_{FR}	B_{ZC}	B_{FR}	B_{ZC}	B_{FR}
61	7	60	3	60	1	60
101	20	100	10	100	5	100
127	32	126	16	126	8	126
139	38	138	19	138	9	138
157	48	156	24	156	12	156
179	63	178	31	178	15	178

no reuse for up to M cells, regardless of N,[3] which demonstrates the practical advantage over the ZC sequences.

Fig. 5 depicts the distributions of intra- and inter-cell correlation magnitudes of tested non-orthogonal sequences over the sequence length, $M \in [60, 100, 140, 180]$ and $M_{\mathrm{ZC}} \in [61, 101, 139, 179]$, where $N = 500$ and $B = 19$. The figure shows that all the sequences have no significant difference in the average correlations, but they are quite different in the maxima. Random Gaussian and MUSA sequences have relatively high maxima in the intra- and inter-cell correlations, respectively. The intra-cell correlations of TF-based sequences are fairly low for the optimization effort, but the maximum inter-cell correlations are high due to no attempt for reduction. The ZC sequences of prime length $M_{\mathrm{ZC}} > 61$ exhibit the lowest maxima in the correlations, which are theoretically bounded by $\frac{1}{\sqrt{M_{\mathrm{ZC}}}}$ [47] if no sequences are reused. However, if $M_{\mathrm{ZC}} = 61$, only $\left\lfloor \frac{M_{\mathrm{ZC}}(M_{\mathrm{ZC}}-1)}{N} \right\rfloor = 7$ distinct sets are available, which requires some ZC sequences to be reused to support 19 cells, yielding the trivially high maximum inter-cell correlation of 1.

Meanwhile, Fig. 5 shows that the intra- and inter-cell correlations of the Fourier-based non-orthogonal sequences are sufficiently low with no shortage of sequence sets for each M, whether the subsampling index set Ω is found by (e) random subsampling, (f) incoherent harmonic design, or (g) genetic algorithm. In contrast to the random subsampling, using the algorithms for Ω reduces the maxima of intra-cell correlations significantly for the Fourier-based sequences, which justifies the efforts of the algorithms. In particular, the incoherent harmonic design algorithm presents the lowest maxima among them in the intra-cell correlation, since the algorithm attempts to minimize the correlation. However, the maxima of the corresponding inter-cell correlations are not so low, since the algorithm makes no attempt to reduce the correlation, as mentioned in Section IV.A. By contrast, the GA shows a good balance of low intra- and inter-cell correlations for the Fourier-based sequences, by attempting to reduce the correlations simultaneously.

Consequently, it is expected from Fig. 5 that the ZC and the Fourier-based non-orthogonal sequences will have good performance of sparse activity detection for multi-cell mMTC, thanks to their low intra- and inter-cell correlations.

[3] As N increases, the intra- and inter-cell correlations of the Fourier-based sequences will increase gracefully, but not reach the maximum 1, as long as $B \leq M$.

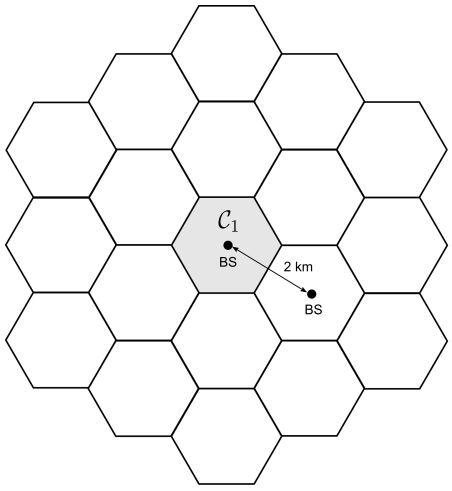

FIGURE 6. Multi-cell network structure of $B = 19$ hexagonal cells. Total $BN = 19 \times 500 = 9500$ devices are present in this multi-cell network, where $N = 500$ devices are uniformly distributed in each cell. In simulation, we consider sparse activity detection to be performed at the BS of $\mathcal{C}_1$.

B. PERFORMANCE OF CS-BASED ACTIVITY DETECTION

For multi-cell massive connectivity, we assume a cellular network of $B = 19$ hexagonal cells, as illustrated in Fig. 6, where each cell has a BS at its center and the BS-to-BS distance between neighbor cells is 2 km. Total $N = 500$ devices are uniformly distributed in an mMTC cell, where each device has its unique signature of length M for uplink grant-free access. As a result, total $BN = 9500$ distinct sequences are uniquely assigned to all devices in 19 cells. At each access trial, the probability $p_a^{(b)}$ of device activity for cell $\mathcal{C}_b$, $1 \leq b \leq B$, is uniformly at random between 0.05 and 0.15. Assuming that the cell $\mathcal{C}_1$ is at the center of the 19 cell cluster, we perform the sparse activity detection at the BS receiver of $\mathcal{C}_1$. The path loss of the wireless channel is modeled by $\beta_n^{(b)} = -128.1 - 36.7 \log_{10} d_n^{(b)}$ in dB, where $d_n^{(b)}$ is the distance in km from device n of $\mathcal{C}_b$ to the BS of $\mathcal{C}_1$. The transmit power of each active device is $\rho = 23$ dBm and the power spectral density of AWGN at the BS receiver of $\mathcal{C}_1$ is -169 dBm/Hz over 1 MHz. In simulations, the BS receiver of $\mathcal{C}_1$ performs CS-based activity detection by treating the inter-cell interference from $\mathcal{C}_2, \cdots, \mathcal{C}_{19}$ as noise. We would like to point out that the CS-based detection with no interference mitigation may not be practical, but it will make little difference in the order of performance among the tested non-orthogonal sequences.

For CS-based activity detection, the BS receiver of $\mathcal{C}_1$ deploys the simultaneous orthogonal matching pursuit (SOMP) algorithm [48], where the number of active devices in $\mathcal{C}_1$ is assumed to be known a priori at each access. Running the SOMP with the prior knowledge cannot be realistic in practice, but it will show the best achievable

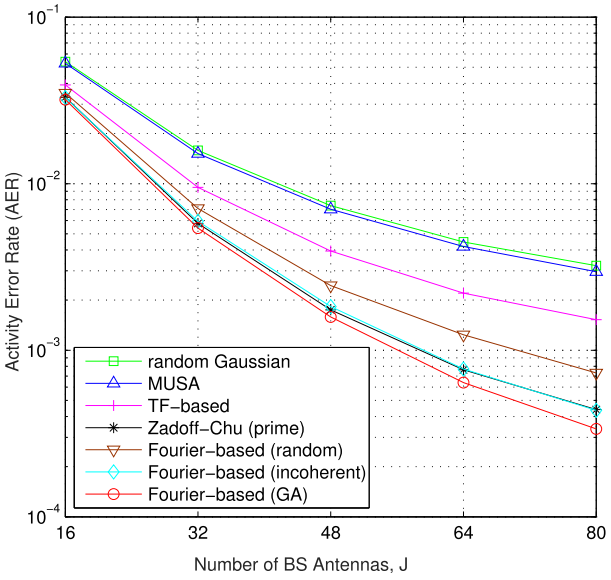

FIGURE 7. Performance of CS-based activity detection for tested non-orthogonal sequences over the number of BS antennas, where $M = 140$ and $M_{ZC} = 139$.

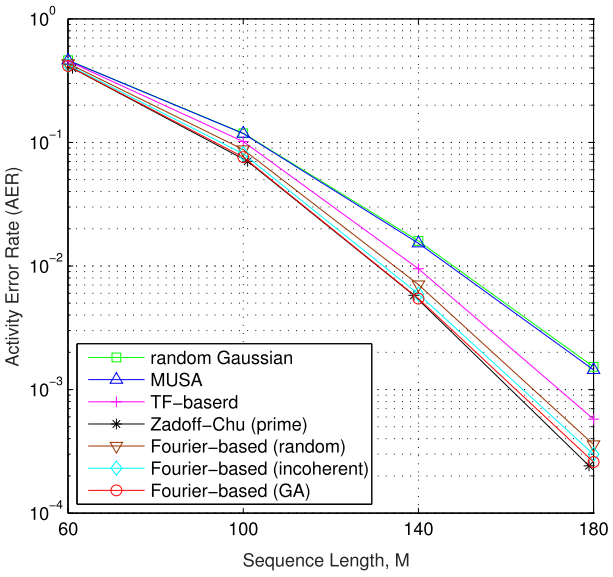

FIGURE 8. Performance of CS-based activity detection for tested non-orthogonal sequences over the sequence length, where $J = 32$. For the Zadoff-Chu sequences, $M_{ZC} \in [61, 101, 139, 179]$.

performance of tested sequences by the algorithm. Treating both undetected and false-alarmed devices as errors, we evaluate the activity error rate (AER), defined by the average of $\frac{\left|\mathcal{S}^{(1)} \backslash \widehat{\mathcal{S}}^{(1)}\right| + \left|\widehat{\mathcal{S}}^{(1)} \backslash \mathcal{S}^{(1)}\right|}{\left|\mathcal{S}^{(1)} \cup \widehat{\mathcal{S}}^{(1)}\right|}$, where $\mathcal{S}^{(1)}$ and $\widehat{\mathcal{S}}^{(1)}$ are true and detected sets of active devices at $\mathcal{C}_1$, respectively. In simulations, the AER is averaged over 10^5 access trials.

Fig. 7 shows the performance of CS-based activity detection at $\mathcal{C}_1$ for tested non-orthogonal sequences over the number of BS antennas, where $M = 140$ and $M_{ZC} = 139$. Also, Fig. 8 sketches the performance for $J = 32$ over the sequence length, where $M \in [60, 100, 140, 180]$ and $M_{ZC} \in [61, 101, 139, 179]$. In the figures, the Fourier-based sequences from our design show significantly lower AER than random Gaussian, MUSA, and TF-based sequences. If the subsampling index set Ω is found by the incoherent harmonic design or the genetic algorithm in Section IV, the Fourier-based sequences outperform those with random subsampling, which demonstrates the effectiveness of the algorithms. Moreover, the performance of the Fourier-based sequences using the GA is slightly better than those with the incoherent harmonic design algorithm, since the GA attempts to find Ω for reducing the intra- and inter-cell correlations simultaneously.

In Fig. 7, it appears that the Fourier-based sequences using the GA work better than the ZC sequences, but it comes from the difference of their sequence lengths, i.e., $M = 140$ and $M_{ZC} = 139$. As shown in Fig. 8, the detection performance of the Fourier-based sequences using the GA is similar to that of the ZC sequences for various sequence lengths. Note that we chose the sequence length M of the Fourier-based sequences as a *non-prime* integer from which no ZC-based non-orthogonal sequences can be obtained, which is to emphasize the benefit of taking arbitrary sequence length. In summary, the results of Figs. 7 and 8 show that the Fourier-based non-orthogonal sequences using the GA are comparable to the well known, algebraically designed ZC sequences for sparse activity detection in multi-cell massive connectivity, while providing the practical advantages of arbitrary sequence length and more sequence sets.

VI. CONCLUSION

In this paper, multiple sets of Fourier-based non-orthogonal sequences have been presented for uplink grant-free access in multi-cell mMTC. We constructed up to M distinct sets of the Fourier-based non-orthogonal sequences of length M, which are uniquely assigned to devices of up to M mMTC cells. We presented a design framework based on a partial Fourier matrix with masking operations and defined the correlation matrix to analyze the intra- and inter-cell correlations of multiple sequence sets simultaneously. Simulation results confirmed that the Fourier-based sequences show the excellent performance of sparse activity detection for uplink grant-free access in multi-cell mMTC, comparable to algebraically designed ZC sequences of prime length. In practice, this design framework provides additional advantages over the ZC sequences, taking arbitrary sequence lengths and offering more sequence sets for multi-cell mMTC. The main contribution of this paper is the design framework for multiple sets of Fourier-based non-orthogonal sequences of arbitrary length, which achieve the excellent performance of sparse activity detection and the practical advantages for multi-cell mMTC.

This paper employed the incoherent harmonic design and the genetic algorithms to find a subsampling index set for our design framework, in an effort to reduce the

intra- and inter-cell correlations of the Fourier-based sequences. In practice, each algorithm can be executed offline only once, with no need to be implemented with high complexity in real-time applications. Instead, the Fourier-based sequences can be efficiently generated by storing the index set and the masking sequence. We would like to point out that one can use any other algorithm or algebraic method to find the subsampling index set under our design framework. For example, a more advanced GA [49] or the reinforcement learning (RL) [50] can be considered to find the index set. Since using another method to find the index set makes little difference in the novelty of the contribution of this paper, we will leave it for future work. The design framework can also be applicable to tackle the pilot contamination problem [51] along with a novel pilot assignment method, e.g., [52], by providing a large number of non-orthogonal pilots for multi-cell massive MIMO, which will also be a future work.

REFERENCES

[1] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: Potential, challenges, and solutions," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 178–184, Mar. 2012.

[2] C. Bockelmann, N. Pratas, and H. Nikopour, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[3] *Service Requirements for Machine-Type Communications (MTC)*, Release 13, 3GPP document TS 22.368, V13.1.0, 2015.

[4] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, 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.

[5] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[6] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 332–336.

[7] S. Zhang, X. Xu, L. Lu, Y. Wu, and G. He, "Sparse code multiple access: An energy efficient uplink approach for 5G wireless systems," in *Proc. IEEE Global Commun. Conf. (IEEE GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 4782–4787.

[8] F. Wei and W. Chen, "A low complexity SCMA decoder based on list sphere decoding," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[9] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for Internet of Things," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, Nanjing, China, May 2016, pp. 1–5.

[10] S. Kang, X. Dai, and B. Ren, "Pattern division multiple access for 5G," *Telecommun. Netw. Technol.*, vol. 5, no. 5, pp. 43–47, May 2015.

[11] M. V. Jamali and H. Mahdavifar, "A low-complexity recursive approach toward code-domain NOMA for massive communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[12] I. Goodfellow, Y. Bengjo, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[13] N. Ye, X. Li, H. Yu, L. Zhao, W. Liu, and X. Hou, "DeepNOMA: A unified framework for NOMA using deep multi-task learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2208–2225, Apr. 2020.

[14] W. Kim, Y. Ahn, and B. Shim, "Deep neural network-based active user detection for grant-free NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2143–2155, Apr. 2020.

[15] T. Sivalingam, S. Ali, N. Huda Mahmood, N. Rajatheva, and M. Latva-Aho, "Deep neural network-based blind multiple user detection for grant-free multi-user shared access," 2021, *arXiv:2106.11204*.

[16] N. Ye, J. An, and J. Yu, "Deep-learning-enhanced NOMA transceiver design for massive MTC: Challenges, state of the art, and future directions," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 66–73, Aug. 2021.

[17] J. Pan, N. Ye, H. Yu, T. Hong, S. Al-Rubaye, S. Mumtaz, A. Al-Dulaimi, and I. Chih-Lin, "AI-driven blind signature classification for IoT connectivity: A deep learning approach," *IEEE Trans. Wireless Commun.*, early access, Jan. 31, 2022, doi: 10.1109/TWC.2022.3145399.

[18] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Standards Mag.*, vol. 3, no. 4, pp. 60–66, Dec. 2019.

[19] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[20] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[21] Y. Du, C. Cheng, B. Dong, Z. Chen, X. Wang, J. Fang, and S. Li, "Block-sparsity-based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7894–7909, Dec. 2018.

[22] 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.

[23] Z. Chen, F. Sohrabi, and W. Yu, "Multi-cell sparse activity detection for massive random access: Massive MIMO versus cooperative MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4060–4074, Aug. 2019.

[24] N. Y. Yu, "Multiuser activity and data detection via sparsity-blind greedy recovery for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 2082–2085, Nov. 2019.

[25] S. Jiang, X. Yuan, X. Wang, C. Xu, and W. Yu, "Joint user identification, channel estimation, and signal detection for grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6960–6976, Oct. 2020.

[26] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.

[27] K. Yang, Y.-K. Kim, and P. Vijay Kumar, "Quasi-orthogonal sequences for code-division multiple-access systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 982–993, May 2000.

[28] S. M. Hasan, K. Mahata, and M. M. Hyder, "Uplink grant-free NOMA with sinusoidal spreading sequences," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3757–3770, Jun. 2021.

[29] N. Y. Yu, "Binary Golay spreading sequences and Reed–Müller codes for uplink grant-free NOMA," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 276–290, Jan. 2021.

[30] N. Y. Yu, "Non-orthogonal Golay-based spreading sequences for uplink grant-free access," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2104–2108, Oct. 2020.

[31] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.

[32] *Physical Channel and Modulation*, 3GPP document TS 36.211 V13.1.0, Mar. 2016.

[33] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[34] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, Dec. 2010, Art. no. 560349.

[35] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing," *Signal Process.*, vol. 92, no. 4, pp. 999–1009, Apr. 2012.

[36] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "On the use of unit-norm tight frames to improve the average MSE performance in compressive sensing applications," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 8–11, Jan. 2012.

[37] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.

[38] C. Rusu, N. González-Prelcic, and R. W. Heath, "Algorithms for the construction of incoherent frames under various design constraints," *Signal Process.*, vol. 152, pp. 363–372, Nov. 2018.

[39] J. Kovačević and C. Amina, "An introduction to frames," *Found. Trends Signal Process.*, vol. 2, no. 1, pp. 1–94, 2008.

[40] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[41] Y.-J. Chen, Q. Zhang, Y. Luo, and Y.-A. Chen, "Measurement matrix optimization for ISAR sparse imaging based on genetic algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1875–1879, Dec. 2016.
[42] A. Elkelesh, M. Ebada, S. Cammerer, and S. ten Brink, "Decoder-tailored polar code design using the genetic algorithm," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4521–4534, Jul. 2019.
[43] M. A. Mazaideh and J. Levendovszky, "A multi-hop routing algorithm for WSNs based on compressive sensing and multiple objective genetic algorithm," *J. Commun. Netw.*, vol. 23, no. 2, pp. 138–147, Apr. 2021.
[44] N. Y. Yu, "Design of non-orthogonal sequences using a two-stage genetic algorithm for grant-free massive connectivity," 2021, *arXiv:2108.00361*.
[45] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The Next Generation Wireless Technology*. Amsterdam, The Netherlands: Elsevier, 2018.
[46] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx
[47] D. Sarwate, "Bounds on crosscorrelation and autocorrelation of sequences," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 720–724, Nov. 1979.
[48] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
[49] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Aug. 2002.
[50] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2015.
[51] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, Nov. 2017.
[52] T. H. Nguyen, T. V. Chien, H. Q. Ngo, X. N. Tran, and E. Bjornson, "Pilot assignment for joint uplink-downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8292–8297, Aug. 2021.

NAM YUL YU (Senior Member, IEEE) received the B.S. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1995, the M.S. degree in electronics and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2000, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2007.

From 2000 to 2003, he was with the Telecommunication Research and Development Center, Samsung Electronics, South Korea, where he worked on channel coding schemes for wireless communication systems. In 2007, he was a Senior Research Engineer at LG Electronics, South Korea, working on the standardization of the 3GPP-LTE. From 2008 to 2014, he was an Assistant/Associate Professor with the Department of Electrical Engineering, Lakehead University, Thunder Bay, ON, Canada. In 2014, he joined the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, where he is currently working as an Associate Professor with the School of Electrical Engineering and Computer Science. His research interests include communication and signal processing techniques for wireless communications. He has served as an Associate Editor for Sequences in IEEE TRANSACTIONS ON INFORMATION THEORY, from 2009 to 2011.