Joint User Activity and Data Detection Based on Structured Compressive Sensing for NOMA

Bichai Wang, Linglong Dai, Talha Mir, and Zhaocheng Wang

***Abstract*— Non-orthogonal multiple access (NOMA) has been regarded as one of the promising key technologies for future 5G systems. In the uplink grant-free NOMA schemes, dynamic scheduling is not required, which can significantly reduce the signaling overhead and transmission latency. However, user activity has to be detected in grant-free NOMA systems, which is challenging in practice. In this letter, by exploiting the inherent structured sparsity of user activity naturally existing in NOMA systems, we propose a low-complexity multi-user detector based on structured compressive sensing to realize joint user activity and data detection. In particular, we propose a structured iterative support detection algorithm by exploiting such structured sparsity, which is able to jointly detect user activity and transmitted data in several continuous time slots. Simulation results show that the proposed scheme can achieve better performance than conventional solutions.**

***Index Terms*— 5G, non-orthogonal multiple access (NOMA), multi-user detection (MUD), structured compressive sensing (SCS).**

I. Introduction

IN THE history of wireless communications, multiple access has become one of the key technologies to distinguish different generations of wireless systems [1]. For example, frequency division multiple access (FDMA) in 1G, time division multiple access (TDMA) in 2G, code division multiple access (CDMA) in 3G, and orthogonal frequency division multiple access (OFDMA) in 4G are adopted in the evolution history of wireless communications. From the perspective of design principle, these conventional multiple access schemes belong to the category of orthogonal multiple access (OMA), which can avoid or reduce inter-user interference by orthogonal resource allocation, and thus simple receivers can be used to realize signal detection. However, the number of supportable users is strictly limited by the number of available orthogonal resources in OMA, which makes it difficult to meet the key requirements of massive connectivity for 5G [1]. To address this issue, non-orthogonal multiple access (NOMA) has been actively investigated [1]–[3], which can realize overloading by introducing some controllable interferences at the cost of slightly increased receiver complexity. In this way, NOMA is able to satisfy the demand of massive connectivity for 5G.

Manuscript received March 22, 2016; accepted April 25, 2016. Date of publication April 28, 2016; date of current version July 8, 2016. This work was supported in part by the International Science & Technology Cooperation Program of China (Grant No. 2015DFG12760), the National Natural Science Foundation of China (Grant Nos. 61571270 and 61271266), the Beijing Natural Science Foundation (Grant No. 4142027), and the Foundation of Shenzhen government. The associate editor coordinating the review of this letter and approving it for publication was H. Mehrpouyan.

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wbc15@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn; mir_talha@hotmail.com; zcwang@tsinghua.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2016.2560180

In order to reduce transmission latency and signaling overhead, grant-free transmission is highly expected in uplink NOMA systems, where users can randomly transmit data without the complex scheduling procedure [1], [4]. However, in grant-free NOMA systems, the base station (BS) cannot obtain the user activity information before data transmission, which implies that user activity has to be detected in the grant-free NOMA systems. Some investigations have been done to realize user activity detection. Specifically, a blind detection of sparse code multiple access (SCMA) was proposed in [4], which could jointly realize user activity detection and channel estimation by sending pilots. A multi-user detection (MUD) method realized by switching between one of the typical compressive sensing (CS) reconstruction algorithms, i.e., orthogonal matching pursuit (OMP) algorithm [5], and one of the classical linear signal detection algorithms, i.e., linear minimum mean square error (LMMSE), was proposed in [6]. Such CS-based user activity detection could achieve substantial performance gain over conventional solutions by exploiting user activity sparsity. However, the detection is usually independently realized in different time slots. Generally, there exist some correlations among transmitted signals in different time slots, e.g., the user activity in several continuous time slots may be similar, which have not been fully exploited in existing methods.

In this letter, inspired by the observation of structured sparsity of user activity in massive connectivity, we propose a low-complexity MUD based on structured compressive sensing (SCS) for NOMA to further improve the signal detection performance. Specifically, according to the statistics of mobile traffics [7], the number of active users is usually much smaller than the number of all possible users even in the busy hours. Thus, the sparsity of user activity naturally exists in massive connectivity, which enables us to formulate the MUD problem under the CS framework [8]. Furthermore, as users generally transmit their information based on an uniformed frame structure, where each frame is usually composed of several continuous time slots (e.g., each resource block (RB) contains 7 OFDM symbols in LTE-Advanced standard [9], and a RB is the smallest granularity for resource scheduling), we consider a system where users are synchronized in a frame structure and are active or inactive in the entire frame [10]. Such consideration, i.e., the number of active users remains constant in several continuous time slots within a frame, has also been used in the recent work [10]. As a result, structured sparsity of user activity can be obtained in NOMA, which is usually neglected in existing works. Accordingly, we propose a structured iterative support detection (SISD) algorithm by exploiting such structured sparsity, which can jointly detect user activity and transmitted data in several continuous time slots. Simulation results show that the proposed scheme achieves better signal detection performance than conventional solutions without increasing the computational complexity.

The rest of this letter is organized as follows. The system model of the uplink NOMA scheme is briefly described in Section II. Section III presents the proposed SISD algorithm in detail, and simulation results are provided to investigate the performance of the proposed scheme in Section IV. Finally, conclusions are drawn in Section V.

II. System Model

We consider a typical uplink NOMA system with one BS and K users. For simplicity but without loss of generality, we assume that the BS and all users are equipped with a single antenna. After channel coding and modulation, the transmitted symbol x_k for active user k is taken from a complex-constellation set $\mathbb{X}$. Then, the transmitted symbol x_k is modulated onto a spreading sequence $\mathbf{s}_k$ of length N. Particularly, we consider the case of $N < K$, i.e., the overloaded system [1]. After that, signals from all active users are superimposed, and then are transmitted over N orthogonal OFDM subcarriers. The received signal on subcarrier n at the BS can be represented as

$$y_n = \sum_{k=1}^{K} g_{nk} s_{nk} x_k + v_n, \tag{1}$$

where s_{nk} is the nth component of spreading sequence $\mathbf{s}_k$, g_{nk} is the channel gain of user k on the nth subcarrier, and all of them are identically and independent distributed (i.i.d.) complex Gaussian variables with zero mean and unit variance, i.e, Rayleigh fading channels are considered in this letter, and such typical channel model has been widely used in wireless communications [6], [7]. v_n is the Gaussian noise on subcarrier n with zero mean and variance σ^2. We combine the received signals over all N subcarriers, and then the received signal vector $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{2}$$

where $\mathbf{x} = [x_1, x_2, \cdots, x_K]^T$ is the transmitted signal vector for all K users, $\mathbf{H}$ is an equivalent channel matrix of size $N \times K$, whose element h_{nk} in the nth row and the kth column equals to $g_{nk}s_{nk}$, and $\mathbf{v} = [v_1, v_2, \cdots, v_N]^T$ is the noise vector following the distribution $\mathcal{CN}\left(0, \sigma^2 \mathbf{I}_N\right)$.

III. Joint User Activity and Data Detection

As mentioned before, according to a report on mobile traffic [7], the number of active users does not exceed 10% of the total number of users even in the busy hours. This implies that even if there are a lot of potential users, which is a typical characteristic of massive connectivity in 5G, only a small part of users have data to send at the same time. Moreover, in the uplink grant-free systems, which usually lead to the so-called sporadic communication, where the active users only constitute a small subset of all potential users, the MUD problem is inherently a sparse signal recovery problem, which motivates us to use the powerful tool of CS to realize joint user activity and data detection [10]. Particularly, we

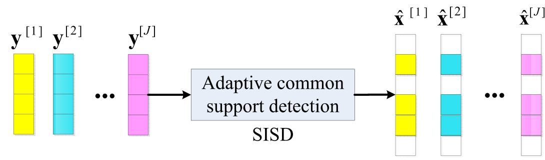

Fig. 1. Graphical representation of the proposed SISD algorithm.

consider that the transmitted symbols of active users are taken from the complex-constellation set $\mathbb{X}$, while the transmitted signals from inactive users are regarded as zero. Due to the naturally existing sparsity of user activity, the transmitted symbol vector $\mathbf{x}$ is sparse [6]. Furthermore, in this letter, we consider a system where users are synchronized in a frame structure and are active or inactive in the entire frame [10]. In this way, the set of active users remains unchanged during several continuous time slots within a frame. Particularly, we consider that the set of active users remains unchanged in J continuous time slots (e.g., $J = 7$ has been considered in LTE-Advanced standard [9]), and thus $\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \cdots, \mathbf{x}^{[J]}$ are sparse transmitted vectors. Therefore, the structured sparsity can be obtained, while such signal characteristic was usually not considered in the literature. Specifically, we have

$$\operatorname{supp}(\mathbf{x}^{[1]}) = \operatorname{supp}(\mathbf{x}^{[2]}) = \cdots = \operatorname{supp}(\mathbf{x}^{[J]}), \tag{3}$$

where $\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \cdots, \mathbf{x}^{[J]}$ are the transmitted symbol vectors in J continuous time slots, and $\operatorname{supp}(\mathbf{x}^{[j]})$ denotes the support of the vector $\mathbf{x}^{[j]}$, i.e., the index set of nonzero elements in $\mathbf{x}^{[j]}$.

At the receiver, the received signal vectors in J continuous time slots can be expressed as

$$\mathbf{y}^{[j]} = \mathbf{H}^{[j]}\mathbf{x}^{[j]} + \mathbf{v}^{[j]}, \quad 1 \le j \le J, \tag{4}$$

where $\mathbf{y}^{[j]}$ denotes the received signal vector in the jth time slot, $\mathbf{H}^{[j]}$ is the equivalent channel matrix in the jth time slot, which has the same form as the matrix $\mathbf{H}$ in (2). Particularly, we assume that the channel matrix remains unchanged during several continuous time slots within a frame, which is usually designed to be shorter than the channel coherence time. For example, by considering the typical system parameters in current LTE-Advanced standard [9] and the user velocity of 100 km/h, the channel matrix remains almost unchanged during 7 OFDM symbols. $\mathbf{v}^{[j]}$ is the Gaussian noise vector in the jth time slot. Note that (4) is a common model followed by CS-based channel access techniques in [6], [10], and references therein. Hence, if the channel matrix satisfies the restricted isometry property (RIP) with a high probability, the sparse signals can be reliably recovered based on CS theory [8].

In this letter, by exploiting the structured sparsity of $\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \cdots, \mathbf{x}^{[J]}$, we propose the SISD algorithm based on SCS to realize both user activity detection and data detection in J continuous time slots as shown in Fig. 1. The proposed SISD algorithm is developed from the classical ISD algorithm [8], [11]. Different from other greedy algorithms such as OMP [8], ISD can realize the adaptive support detection without prior knowledge of sparsity level, which is preferred in practice because it is usually difficult to exactly know how many users are active. However, the classical ISD algorithm

Algorithm 1 Proposed SISD Algorithm

Input:

Received signals: $\mathbf{y}^{[1]}, \mathbf{y}^{[2]}, \cdots, \mathbf{y}^{[J]}$;

Equivalent channel matrices: $\mathbf{H}^{[1]}, \mathbf{H}^{[2]}, \cdots, \mathbf{H}^{[J]}$.

Output:

Reconstructed sparse signals: $\mathbf{x}^{\hat{[1]}}, \mathbf{x}^{\hat{[2]}}, \cdots, \mathbf{x}^{\hat{[J]}}$.

1: $i = 0$ and $I^{(0)} = \emptyset$.

2: **while** $\text{Card}(I^{(i)}) < K - N$ **do**

3: $T^{(i)} \leftarrow (I^{(i)})^C := \{1, 2, \cdots, K\} \backslash I^{(i)}$;

4: **for** $j = 1$ to J **do**

5: $\mathbf{x}^{[j](i)} \leftarrow \min\limits_{\mathbf{x}^{[j](i)}} ||\mathbf{x}_{T^{(i)}}^{[j](i)}||_1 + \frac{1}{2\rho^{(i)}} ||\mathbf{H}^{[j]}\mathbf{x}^{[j]} - \mathbf{y}^{[j]}||_2^2$ s.t. $\mathbf{y}^{[j]} = \mathbf{H}^{[j]}\mathbf{x}^{[j]} + \mathbf{v}^{[j]}$;

6: **end for**

7: $\mathbf{x}_{\text{add}}^{(i)} = \sum\limits_{j=1}^{J} |\mathbf{x}^{[j](i)}|$, where $|\mathbf{x}^{[j](i)}| = [|x_1^{[j](i)}|, |x_2^{[j](i)}|, \cdots, |x_K^{[j](i)}|]^T$;

8: $\mathbf{w}^{(i)} = \text{sort}(\mathbf{x}_{\text{add}}^{(i)})$, $k = \min\limits_{k} |w_{k+1}^{(i)}| - |w_k^{(i)}| > \tau^{(i)}$, $\varepsilon^{(i)} = |w_k^{(i)}|$, $I^{(i+1)} \leftarrow \{k : |x_{\text{add},k}^{(i)}| > \varepsilon^{(i)}\}$;

9: $i = i + 1$.

10: **end while**

11: **return** $\mathbf{x}^{\hat{[j]}} = \mathbf{x}^{[j](i)}$, $j = 1, 2, \cdots, J$.

can only recovery one sparse signal vector from one received signal [11], which cannot make full use of the inherent structured sparsity among multiple signals. To this end, we propose the SISD algorithm as shown in **Algorithm 1** to recover multiple sparse signal vectors in a joint manner. Specifically, the main procedure of the proposed SISD algorithm at the ith iteration with the initial common support $I^{(0)} = \emptyset$ can be presented as follows.

1) (*Step 3*) Calculate the complementary set $T^{(i)}$ of the support $I^{(i)}$:

$$T^{(i)} \leftarrow (I^{(i)})^C := \{1, 2, \cdots, K\} \backslash I^{(i)}. \tag{5}$$

2) (*Step 5*) Update the estimated signal vectors in J continuous time slots separately:

$$\mathbf{x}^{[j](i)} \leftarrow \min_{\mathbf{x}^{[j](i)}} ||\mathbf{x}_{T^{(i)}}^{[j](i)}||_1 + \frac{1}{2\rho^{(i)}} ||\mathbf{H}^{[j]}\mathbf{x}^{[j]} - \mathbf{y}^{[j]}||_2^2, \quad \text{s.t. } \mathbf{y}^{[j]} = \mathbf{H}^{[j]}\mathbf{x}^{[j]} + \mathbf{v}^{[j]}, \tag{6}$$

where $\rho^{(i)} > 0$ is a proper parameter, which can be selected according to [11].

3) (*Step 7*) Add the absolute values of all J estimated signal vectors:

$$\mathbf{x}_{\text{add}}^{(i)} = \sum_{j=1}^{J} |\mathbf{x}^{[j](i)}|, \tag{7}$$

where $|\mathbf{x}^{[j](i)}| = [|x_1^{[j](i)}|, |x_2^{[j](i)}|, \cdots, |x_K^{[j](i)}|]^T$.

4) (*Step 8*) Update the support: $I^{(i+1)} \leftarrow$ support detection using $\mathbf{x}_{\text{add}}^{(i)}$ as the reference.

5) (*Step 9*) Update the number of iterations: $i \leftarrow i + 1$.

The iterative process stops when $\text{Card}(T^{(i)}) < K - N$. Particularly, in step 8, the new support $I^{(i+1)}$ can be calculated by

$$I^{(i+1)} \leftarrow \{k : |x_{\text{add},k}^{(i)}| > \varepsilon^{(i)}\}, \tag{8}$$

where $\varepsilon^{(i)}$ is the threshold in the ith iteration. In this algorithm, the support sets $I^{(i)}$ for different i's are not necessarily increasing and nested, i.e., we cannot guarantee $I^{(i)} \subset I^{(i+1)}$ for all i, which is beneficial for the reliable support recovery. In general, it is difficult to completely avoid wrong detections based on the previously obtained $x_{\text{add},k}^{(i)}$. For example, the absolute value of $x_{\text{add},k}^{(i)}$ may be very large, while the true value should be zero. In the proposed SISD algorithm, if the support sets are not monotonically increasing and nested, wrong detections may be removed in later iterations. As a result, $I^{(i)}$ is not very sensitive to the threshold $\varepsilon^{(i)}$, and thus it is easier to choose the proper parameter $\varepsilon^{(i)}$. Particularly, the selection of $\varepsilon^{(i)}$ can be based on the "first significant jump" [11]. More specifically, if the components of $\mathbf{x}_{\text{add}}^{(i)}$ are sorted in the increasing order, i.e., $|x_{\text{add},1}^{(i)}| \leq |x_{\text{add},2}^{(i)}| \leq \cdots \leq |x_{\text{add},K}^{(i)}|$, the first significant jump is the smallest k such that

$$|x_{\text{add},k+1}^{(i)}| - |x_{\text{add},k}^{(i)}| > \tau^{(i)}, \tag{9}$$

where $\tau^{(i)}$ can be selected according to [11]. Then, we have $\varepsilon^{(i)} = |x_{\text{add},k}^{(i)}|$.

Recall that in step 5, the estimated signal vector in the jth time slot can be obtained by (6), which can be regarded as the truncated basis pursuit (BP) problem [11]. In each iteration, the correct detections of the support in $I^{(i)}$ will help (6) to return a more reliable solution than the conventional BP algorithms. Based on this more reliable solution, a better recovery of the support will be available by calculating (8). In this way, the support detection procedure and the step of nonzero element recovery can promote each other to gradually improve the signal reconstruction accuracy. Even though the perfect signal recovery guarantee has not been theoretically found, [11] has provided a sufficient condition to improve the chance of accurate signal reconstruction.

The key difference between the conventional ISD algorithm and the proposed SISD algorithm is the way for support selection. Specifically, ISD independently calculates the support of each signal after updating the estimated signal vector, while SISD simultaneously updates the same support for J sparse signals in J continuous time slots by adding the absolute values of all J estimated signal vectors. By exploiting the key feature that J sparse signal vectors share the same support, the joint processing of $\mathbf{x}^{[1](i)}, \mathbf{x}^{[2](i)}, \cdots, \mathbf{x}^{[J](i)}$ to update the support can increase the robustness of the support detection and thus improve the signal recovery performance. Particularly, when $J = 1$, the proposed SISD algorithm reduces to the conventional ISD algorithm.

In addition, the proposed SISD algorithm enjoys the same order of computational complexity with the conventional ISD algorithm. Particularly, the main computational burden comes from step 5 to update the estimated signal vector via (6), which can be efficiently solved by some BP algorithms, e.g., YALL1 with a warm-start and a dynamic stopping rule [11]. Therefore, the computational complexity of the proposed SISD algorithm is similar with that of BP. Furthermore, if the

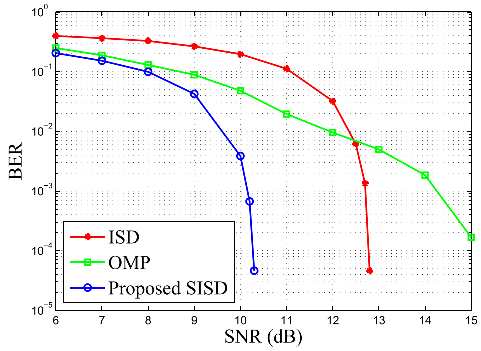

Fig. 2. BER performance comparison against SNR, where the overloading factor is 150% and the number of continuous time slots is $J = 7$.

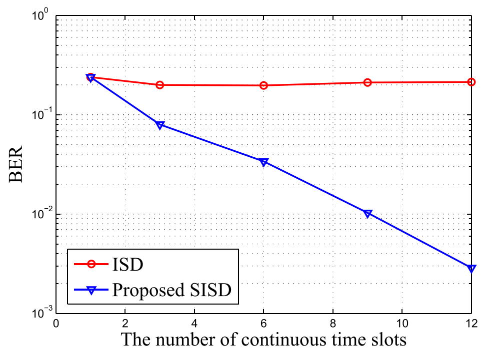

Fig. 3. BER performance comparison against the number of continuous time slots J, where the overloading factor is 150% and SNR = 9.5 dB.

tradeoff between the spectral efficiency and the detection complexity is considered, the subcarrier and power allocation should be optimized, where the impact of efficient user pairing can be significant in NOMA systems [3].

IV. Simulation Results

In this section, we investigates the signal recovery performance of the proposed SISD algorithm. The main simulation parameters are set as follows. The number of users is $K = 150$, the number of orthogonal subcarriers is $N = 100$, and thus the overloading factor is 150%. Spreading sequences are generated based on the pseudo random noise (PN) sequence. The rate of LDPC channel code is 2/3. We compare the BER performance of the proposed SISD algorithm, the classical ISD algorithm [11] and OMP algorithm [5] for MUD, where QPSK modulation is considered.

Fig. 2 shows the BER performance comparison against SNR. Particularly, the number of active users is 20, and the number of continuous time slots is $J = 7$. Simulation results show that compared with ISD, the proposed SISD algorithm has about 2.5 dB SNR gain, which is achieved by exploiting the structured sparsity of user activity. On the other hand, as shown in Fig. 2, the BER performance of the proposed SISD algorithm is better than that of OMP algorithm, e.g., about 5 dB SNR gain can be obtained when BER $= 10^{-4}$. What's more, the prior information of user activity sparsity is not necessary for the proposed SISD algorithm, while such information must be required by OMP algorithm [5].

Fig. 3 shows the BER performance comparison against the number of continuous time slots J, where the SNR is set as 9.5 dB. We can see from simulation results that the BER performance of ISD is not sensitive to J, which is obvious because the correlation among different sparse signals is not exploited by ISD, and thus the BER performance basically remains unchanged even though the number of continuous time slots is different. On the contrary, the BER performance of the proposed SISD algorithm becomes better with the increase of the number of continuous time slots J, which is caused by the exploiting of structured sparsity to generate the common support of different sparse signals. In addition, as we have discussed before, ISD can be regarded as the special case of SISD by setting $J = 1$, and such conclusion can be verified by the fact that the BER performance of ISD and the proposed SISD is the same when $J = 1$ as shown in Fig. 3.

V. Conclusions

In this letter, we have addressed the MUD problem of uplink grant-free NOMA systems for 5G, where user activity has to be detected at the BS. Particularly, by exploiting the structured sparsity of user activity in NOMA systems with massive connectivity, we have proposed the SISD algorithm to realize both user activity detection and data detection in several continuous time slots. Different from the conventional ISD algorithm that independently calculates the support for each signal, the proposed SISD algorithm simultaneously updates the same support for multiple sparse signals. Meanwhile, the computational complexity of SISD is similar with that of ISD. Simulation results show that the proposed SISD algorithm can achieve about 2.5 dB SNR gain compared with ISD. Therefore, the proposed SISD algorithm can improve the signal detection performance in NOMA systems with acceptable complexity.

References

[1] 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.

[2] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016. [Online]. Available: http://arxiv.org/abs/1503.05367

[3] Z. Ding, P. Fan, and V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access downlink transmissions," *IEEE Trans. Veh. Technol.*, Sep. 2015. [Online]. Available: http://arxiv.org/abs/1412.2799, doi: 10.1109/TVT.2015.2480766

[4] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 853–857.

[5] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[6] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, Jul. 2012.

[7] J.-P. Hong, W. Choi, and B. D. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 998–1010, Feb. 2015.

[8] Z. Han, H. Li, and W. Yin, *Compressive Sensing for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, Jul. 2013.

[9] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 12)*, document 3GPP TS-36.211, Jan. 2016.

[10] A. T. Abebe and C. G. Kang, "Iterative order recursive least square estimation for exploiting frame-wise sparsity in compressive sensing-based MTC," *IEEE Commun. Lett.*, Mar. 2016, doi: 10.1109/LCOMM.2016.2539255.

[11] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 462–491, Aug. 2010.