Dynamic and Sparsity Adaptive Compressed Sensing Based Active User Detection and Channel Estimation of Uplink Grant-Free SCMA

Mehmet Hakan DURAK^{1,2}, Ozgur ERTUG¹

¹ Dept. of Electrical-Electronics Engineering, Gazi University, Ankara, Turkey
² Dept. of Electrical-Electronics Engineering, Erzurum Technical University, Erzurum, Turkey

{mhdurak, ertug}@gazi.edu.tr

Submitted January 16, 2021 / Accepted May 11, 2021

Abstract. In uplink (UL) grant-free sparse code multiple access (SCMA) systems, unlike the conventional contentionbased transmission, users' activities should be known before data decoding due to sporadic transmission in massive machine-type communication (mMTC). Since compressed sensing (CS) is the theory of sparse signal reconstruction with fewer samples, this theory is a good solution to detect active users. In this paper, we propose the dynamic and sparsity adaptive compressed sensing (DSACS) based active user detection (AUD) and channel estimation (CE) of UL grant-free SCMA. Unlike most of the CS-based methods, sparsity knowledge or potential active user list is not needed in the proposed algorithm, which is already not known in the practical systems. The proposed algorithm adopts a stagewise approach to expand the set of accurate active users for adaptively achieve the sparsity level. It uses the temporal correlation of users' activity to improve performance and reduce complexity. Then, false detected users are eliminated with joint message passing algorithm (JMPA), and channel gains of the accurate active users are estimated again in CE with feedback. The simulation results show that the proposed method without sparsity knowledge is capable of achieving detection in various scenarios in case of sporadic transmission in mMTC.

Keywords

Sparse code multiple access (SCMA), compressed sensing, active user detection, activity detection, channel estimation, UL grant-free transmission, 5G, beyond 5G

1. Introduction

Massive connectivity, high spectral efficiency and low latency are crucial requirements to support the next generation communication system 5G and beyond 5G (B5G). With new concepts such as internet of things (IoT) and massive machine-type communication (mMTC), the importance of

these requirements has appeared. The current orthogonal multiple access (OMA) methods are based on orthogonal resource to each user exclusively to avoid interference between users [1]. Time division multiple access (TDMA), code division multiple access (CDMA) and orthogonal frequency division multiple access (OFDMA) are examples of these OMA methods. However, with massive connections, some of these multiple access methods may not suitable for the requirements in B5G because of orthogonality constraints. Therefore, non-orthogonal multiple access (NOMA) methods that multiple users transmit simultaneously by sharing the same resources have been proposed to improve system throughput and accommodate massive users. The current NOMA methods can be essentially classified as power and code domains [2]. Multi-user shared access (MUSA), pattern division multiple access (PDMA), interleave division multiple access (IDMA), and sparse code multiple access (SCMA) are methods investigated in the code domain [3].

SCMA [4] is one of the promising code domain NOMA techniques for 5G and B5G. SCMA is an upgraded form of the low-density spreading (LDS) used in the CDMA method characterized by sparse codebooks. These codebooks are regulated based on multi-dimensional constellations, which provide shaping gain [5]. Thanks to the shaping gain of the multidimensional constellation, SCMA has better performance compared to other NOMA methods [6]. In SCMA, users' data are directly mapped to sparse codewords in codebooks at different layers. The users' codewords share the same time-frequency resources of OFDMA, and due to this sparsity, low-complexity near-optimal detection is achieved with message passing algorithm (MPA) at the receiver. Such low-complexity receiver detection enables massive connectivity, a key feature of machine-type communication in 5G and B5G.

In a conventional wireless communication system, the request-grant procedure for the uplink transmission leads to high signaling overhead and latency, which is undesirable in 5G and B5G with massive connection [7]. Uplink (UL) grant-free multiple access systems are proposed to reduce

transmission latency and overhead related to control signals for scheduling. In UL grant-free multiple access scenario, users are permitted to transmit data at any time without request-grant procedure in pre-scheduled resources as shown in Fig. 1. Users can select their codebooks and pilot sequences and transmit their data in pre-scheduled resources without any request-grant procedure in UL grant-free SCMA. The pre-scheduled resource is referred to as contention transmission unit (CTU) that combines time, frequency, codebook and pilot for active users [8]. Therefore, it is required the receiver to be able to identify the active users, estimate their channels and decode their transmitted data without having the active user list.

The number of active users in the machine type communications is usually much less than the total potential users even during busy hours as illustrated in Fig. 1. Thus, the users can be considered as a sparse signal, which enables the activity detection. Since compressed sensing (CS) theory [9] is capable to reconstruct a sparse signal with less information, CS algorithms can be used to detect active users and estimate channel gains [10] to enable grant-free multiple access. Then MPA can be utilized to active users' data decode based on the active user list (AUL).

Much research has been done in recent years on joint user activity and data detection based on CS technology. One of the first examples of an active user detector for SCMA is presented in [8]. This detector identifies users who are active and decreases the number of potential users. Orthogonal pilot signals are allocated to different users, and the activity of the users is detected by active user detector (AUD) in the receiver. Then, with joint-MPA (JMPA), false detected inactive users are removed and active users' data is decoded. Hereafter, orthogonal matching pursuit (OMP) and compressive sampling matching pursuit (CoSaMP) have been applied to NOMA grant-free systems for sparse signal detections in [11], [12], and these algorithms have superior performance at low sparsity. However, since these methods require a potential active users list, it is unlikely to be applied in practice. Therefore, methods possible in practice that do not require the sparsity information have been developed in [13–15]. The channel state information is calculated after AUD and the data of active users are decoded with JMPA. In addition, a method that optimizes the sensing matrix according to the SNR status has been developed in [15]. In [16], [17], under the block-sparsity frame, the active user detection, and data decoding is performed for the case the activity rate does not change in adjacent time slots. This is the case for machine-type communication where statistically approximately 10% of users are active.

In the UL grant-free SCMA system for massive user connectivity, we propose an active user detection and channel estimation. We are motivated by the algorithm in [18] used for multi-user detection and data decoding for UL grantfree NOMA. Based on this algorithm, we developed a new method for active user detection and channel estimation for UL grant-free SCMA systems. In the study in [5], a method



Fig. 1. Illustration of a typical grant-free SCMA system.

was developed using CS to reduce the complexity of MPA. However, in this method, data of all users is decoded using MPA. The complexity of MPA increases as there are also inactive users in UL grant-free SCMA systems. In the proposed method, active user detection is performed at adjacent time slots. The proposed algorithm adopts a stagewise approach to expand the set of accurate active users for adaptively achieve the sparsity level of user activity. Then, the algorithm exploits the idea of backtracking to improve the estimate of the accurate active user set at each iteration for more accurate reconstruction. Thus, it is used the temporal correlation of active user sets to reduce the computational complexity and improve performance.

The rest of the paper is structured as follows. Section 2 introduces the system models including UL grant-free SCMA system, data and pilot transmission. The proposed algorithm is presented in Sec. 3. Section 4 illustrates the simulation results, and the conclusions are drawn in Sec. 5.

2. System Model

We consider an UL grant-free SCMA system with *J* potential users. The users can be active or inactive in compliance with their service specifications. In this system, active user detection is required before multiuser detection due to sporadic communication and lack of request-scheduling processes. In general, the pilot signal is transmitted together with the users' data signal for activity detection. In this section, the model of uplink grant-free SCMA transmitter and receiver as illustrated in Fig. 2 is described. And pilot and data transmission in the transmitter are explained in detail.

2.1 Uplink Grant-Free SCMA Transmitter

Consider an uplink grant-free SCMA system with *J* potential users, *A* active users and *K* orthogonal resource elements (REs), where K < A. Here, the overloading factor $\lambda = A/K$ is usually greater than 1 to allow massive connectivity. This is characteristic of NOMA methods.



Fig. 2. Uplink grant-free SCMA system model (J potential users and A active users).

In Fig. 2, the schematic diagram of the uplink grant-free SCMA system model can be seen. Each binary incoming bits from active users are directly mapped into complex codewords x_i from the predefined $N_{\rm CB}$ codebooks. The codewords of each codebook are sparse designed with N < Knon-zero elements. Each non-zero element is modulated to the orthogonal frequency division multiple access (OFDMA) subcarrier and overlapped with the subcarriers of other users. The same time-frequency resources are used to transmit different active users' codewords. Because of the sparsity of codewords, inter-user interference is reduced. Each codebook is linked with $N_{\rm P}$ pilot sequences. As shown in Fig. 2, active users map their data from their own codebook to codeword and take the pilot sequence belonging to the users' own codebook from the pilot pool and transmit it to the receiver together via the channel.

2.2 Pilot and Data Transmission

In the UL grant-free SCMA, the data of each user is encoded from the corresponding codebook, then the pilots are used to identify active users and estimate the channel at the receiver. The received pilot signal at the receiver is

$$y_{p} = \sum_{j=1}^{J} \alpha_{j} h_{j} \phi_{j} + \mathbf{w} = \mathbf{\Phi} \mathbf{h} + \mathbf{w}$$
(1)

where *J* is the registered user set, α_j for j = 1, ..., J, is a binary logical variable of user j, $\alpha_j = 1$ if j^{th} user is active, otherwise $\alpha_j = 0$ if j^{th} user is inactive. h_j and ϕ_j represent channel gain and pilot sequences of user *j*, respectively and **w** is complex white gaussian noise. Also, $\mathbf{\Phi} = [\phi_1, ..., \phi_J]$ and $\mathbf{h} = [\alpha_1 h_1, ..., \alpha_J h_J]$ are diagonal matrix and sparse matrix, respectively. The problem, therefore, is to find active users and to estimate channel gains among potential users by finding which α_J 's are 1. The received data signal at the receiver is

$$y_{\rm d} = \sum_{j=1}^{J} h_j x_j + \mathbf{w} \tag{2}$$

where h_j is channel vector of user j, x_j is transmitted vector of the user j chosen from assigned codebook and \mathbf{w} is additive white Gaussian noise with zero mean and σ^2 variance per element.

$$x_j = q_j(b_j) = [x_1, x_2, \dots, x_j]$$
 (3)

where q_j and b_j are referred to as the mapping matrix and the binary sequence, respectively.

2.3 Uplink Grant-Free SCMA Receiver

In [8], transmitted pilot and data signals from the transmitter are processed in a 3-step process in the UL grant-free SCMA receiver and the data of active users are decoded. These are active user detection (AUD), channel estimation (CE) and joint-MPA (JMPA).

In the AUD process, active users are detected in the received pilot signal. In (1) the received pilot signal can be expressed as $y_p = [y_1, y_2, \dots, y_L]$ where *L* is the length of the signal. $\mathbf{\Phi} = [\phi_1, \dots, \phi_J]$ is the pilot signal including *J* registered pilot sequence. Each element of the channel matrix **h** is associated with one pilot. $\mathbf{w} = [w_1, w_2, \dots, w_L]$ is the noise vector. The received pilot signal is rewritten as in (4).

$$y_{p} = \mathbf{\Phi}\mathbf{h} + \mathbf{w}$$

$$= \begin{bmatrix} \phi_{11} & \phi_{21} & \dots & \phi_{J1} \\ \phi_{12} & \phi_{22} & \dots & \phi_{J2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{1L} & \phi_{2L} & \dots & \phi_{JL} \end{bmatrix} \begin{bmatrix} \alpha_{1}h_{1} \\ \alpha_{2}h_{2} \\ \vdots \\ \alpha_{J}h_{J} \end{bmatrix} + \begin{bmatrix} w_{1} \\ w_{2} \\ \vdots \\ w_{L} \end{bmatrix}$$
(4)

In UL grant-free receiver structure, after a CS algorithm such as the focal underdetermined system solver (FO-CUSS) [19] process applied to the pilot signals in [8] for AUD, it is stated that the user is active if $|\mathbf{h}^2|$ is greater than a threshold value. The threshold value is set to 0.01. Channel gain of users also positively detected in AUD is estimated.

False detected inactive users can exist due to signal noise and fading channel in the AUD process. In the JMPA process, therefore, it is aimed to correct this false detection and then decode only the updated active users' data. In this process, it is assumed that the false detected inactive user virtually transmit zero codewords whose values are zero 0. We defined the zero codeword as $c_j^0 = 0$, then assumed the extended codebook as $\overline{C}_j = C_j \cup c_j^0$. The only difference of JMPA from MPA is that it implements the extended codebook and eliminates false detected inactive users. In JMPA the probability of constellation point zero is represented as

$$p(c_u^0) = \frac{\prod_m p_k(c_u^m)}{(\prod_m 1 - p_k(c_u^m)) + \prod_m p_k(c_u^m)}.$$
 (5)

If some users have a high probability of zero codeword, that user is considered a false detected inactive user and is removed from the active user list. After elimination of false detected users, the distance $P_k(c_u^m)$ between the received signal and the *k*-node is calculated by

$$P_k(c_u^m) = \exp\left\{-\frac{1}{\sigma_n^2} \left\| y_k - \sum_{m \in \zeta_k} h_{km} c_{k,m}^m \right\|^2\right\}.$$
 (6)

Then, according to the calculated distance, the message passed from functional node f_k to variable node v_a and the message passed from variable node v_a to functional node f_k represented as

$$I_{f_k \to v_a}^i(c_u^m) = \sum_{\sim c_u^m} P_k(c_u^m) \cdot \prod_{\tilde{l} \in \xi_k \setminus \{a\}} I_{v_l \to f_k}^{i-1}(c_l^m),$$

$$I_{v_a \to f_k}^i(c_u^m) = \prod_{n \in \xi_a \setminus \{k\}} I_{f_n \to v_a}^{i-1}(c_n^m)$$
(7)

where ξ_k and ζ_a are sets of the position of the nonzeros in column *k* and *a*, respectively, with index of iteration *i*. After *i* iterations the final output for each codeword is calculated as

$$Q(c_u^m) = \prod_{k \in \zeta_a} I^i_{f_k \to v_a}(c_u^m).$$
(8)

Since the number of users in MPA increases the complexity proportionally [20], eliminating the false detected inactive users in JMPA significantly reduces the complexity. In addition, the achivement of the AUD process also plays an important role in reducing the complexity of JMPA.

3. Proposed Algorithm

In this section, we developed a CS-based method for active user detection and channel estimation in UL grant-free SCMA systems. In this method, unlike the CS-based methods developed in the literature, it detects active users without potential active user list, i.e. sparsity prior knowledge is not required.

As can be seen in Fig. 3, active user list (1) is detected in DSACS-based AUD. After the AUD, the channel gains of active users (2) are estimated with AUL. In JMPA, if there are false detected users, they are eliminated and the updated AUL (3) comes back the channel estimation process again, and the updated AUL and channel gain information (4) is sent back to the JMPA for decode. Then, the data signal of active users is decoded (5) in JMPA by using the channel gain information of active users.

In MPA, the sub-optimal receiver of SCMA, its complexity increases considerably as the number of users increases [21]. In addition, since only a small part of the registered users in mMTC is active in a time slot, an active user detection is required for this receiver structure. And this problem can be considered as a compressed sensing reconstruction problem since there are few active users in a time slot.

In order to reconstruct the signal, the signal must also satisfy to the restricted isometry property (RIP). We assume the pilot matrix Φ satisfies the RIP of order *K* [22]. RIP is defined as

$$(1 - \delta_k) \|h^{[t]}\|_2^2 \le \|\phi_j h^{[t]}\|_2^2 \le (1 + \delta_k) \|h^{[t]}\|_2^2 \qquad (9)$$

for RIP constant $\delta_k \in (0, 1)$. RIP indicates that the pilot sequence is approximately orthogonal with sparsity of k.

In this paper, we used Zadoff-Chu (ZC) sequences as a pilot sequence due to their high autocorrelation and low cross-correlation properties [15]. In MTC, users are generally active for adjacent time slots rather than activity in the entire time block. Therefore, we performed the AUD process based on the number of active users in adjacent time slots and considered the activity as a high probability in previous time blocks. Received pilot signal in T adjacent time slots:

$$\mathbf{y}^{[t]} = \mathbf{\Phi}^{[t]} \mathbf{h}^{[t]} + \mathbf{w}^{[t]}$$
(10)

where $\mathbf{y}^{[t]}$, $\mathbf{\Phi}^{[t]}$ and $\mathbf{h}^{[t]}$ denote the received pilot signal, pilot sequence and channel matrix in the *t*-th time block, respectively, and $\mathbf{w}^{[t]}$ is the Gaussian noise vector in the *t*-th time block. The proposed algorithm step-by-step detects active users and estimates channel gains. Details of the proposed method are available in Algorithm 1.

At the beginning of the DSACS-based AUD, the number of active users initially estimated in the first time slot is considered l = 1 and initial support $I_c = \emptyset$. The support in the (t-1)th time slot can be used for the predicted support in the *t*th time block due to temporal correlation in time slots. Here, the absolute $|I_c|$ refers to the number of elements, and $\lceil a \rceil$ denotes the ceil function, so this function rounds off to the smallest integer greater than or equal to *a*.



Fig. 3. Proposed UL grant-free SCMA receiver structure including AUD, channel estimator, JMPA.

Algorithm 1. The Proposed DSACS-based Alg	gorithm.
---	----------

Inputs: Φ , $\mathbf{y}_{p}^{[t]}$ $I_c \leftarrow \emptyset$ for t = 1 to T do if t = 1 then l = 1else $l = \left\lceil |I_c|/3 \right\rceil$ end if $\begin{aligned} &i = 1, I^{[t](0)} = I_c \\ &h = \left(\Phi_{I^{[t](0)}}^{[t](0)}\right)^{\dagger} y^{[t](0)} \\ &r^{[t](0)} = y - \Phi_{I^{[t](0)}}^{[t]} \\ & \text{while } \|r^{[t](i-1)}\|_2^2 > N\sigma^2 \text{ do} \\ &P^{[t](i)} = \max\left(\left|\left(\Phi^{[t]}\right)^H r^{[t](i-1)}\right|, l\right) \\ & \cap_{I^{[t](i)}}^{[t](i)} = \sum_{I^{[t](i-1)}}^{[t](i-1)} p^{[t](i)} \right) \end{aligned}$ $C^{[t](i)} = I^{[t](i-1)} \cup P^{[t](i)}$ $I^{[t]} = \max \left(\left| \left(\Phi_{C^{[t](i)}}^{[t]} \right)^{\dagger} y^{[t]} \right|, l \right) \right.$ $r^{[t]} = y^{[t]} - \Phi_{I^{[t]}}^{[t]} \left(\Phi_{I^{[t]}}^{[t]} \right)^{\dagger} y^{[t]}$ if $||r^{[t]}||_2^2 \le N\sigma^2$ then Quit the iteration else if $||r^{[t]}||_2 > ||r^{[t](i-1)}||_2$ then l = l + 1else $I^{[t](i)} = I^{[t]}, r^{[t](i)} = r^{[t]}$ i = i + 1end if end while $I_c = I^{[t]}$ $h^{[t]} = (\Phi^{[t]}_{I^{[t]}})^{\dagger} y^{[t]}$ end for **Outputs:** $\mathbf{h}^{[t]}$, $\mathbf{I}^{[t]}$

After the initialization steps, in each iteration, the l indices of the users that best match the residual signal are selected as follows.

$$P^{[t](i)} = \max\left(\left| (\Phi^{[t]})^{\mathrm{H}} r^{[t](i-1)} \right|, l \right)$$
(11)

where $r^{[t](i-1)}$ is the residual vector in the last iteration. The max $(|(\Phi^{[t]})^{H}r^{[t](i-1)}|, l)$ function returns l indices, which have the largest absolute values in vector $(\Phi^{[t]})^{H}r^{[t](i-1)}$. Then, to correct the error estimation of the active user set in the last iteration, the preliminary support set and the estimated support in the last iteration are combined to form the candidate list *C* as shown in (12). This is called backtrack [18].

$$C^{[t](i)} = I^{[t](i-1)} \cup P^{[t](i)}.$$
(12)

Finally, the candidate vector is generated using the candidate list with the Least Squares (LS) method, and the l indices with the largest absolute value of this vector forms the final support in the *t*-th time slot as seen in (13).

$$I^{[t]} = \max\left(\left| (\Phi_{C^{[t](i)}}^{[t]})^{\dagger} y^{[t]} \right|, l \right).$$
(13)

Then, after obtaining the final support set in (13), the residual vector is calculated as follows:

$$r^{[t]} = y^{[t]} - \Phi_{I^{[t]}}^{[t]} (\Phi_{I^{[t]}}^{[t]})^{\dagger} y^{[t]}.$$
(14)

If the energy of the residual vector is greater than or equal to the energy of the residual vector in the (i - 1)th iteration, the next iteration starts and the estimated number of active users is gradually increased until the correct number of active users is reached. The transmitted channel gain vector belonging to active users is calculated by the LS method as in (15).

$$\hat{h}^{[t]} = \left(\Phi_{I^{[t]}}^{[t]}\right)^{\mathsf{T}} y^{[t]}. \tag{15}$$

The termination of iteration in the algorithm is determined as N times the energy of the noise.

$$\|r^{[t]}\|_2^2 \le N\sigma^2.$$
(16)

Calculating the number of active users as one third of the number of elements of the final support computed in the last iteration reduces complexity and makes the algorithm dynamic. This calculation can be changed according to the busy period of the system (such as active time, quiet time). As can be seen in Fig. 3, after the active user detection, the channel gain is calculated with AUL. Then the data signal of active users is decoded in JMPA using the channel gain information.

4. Simulation Results

In this section, simulation results are obtained by applying the proposed DSACS-based active user detection and channel estimation algorithm using MATLAB programming language. The basic set of simulation parameters is shown in Tab. 1. Rayleigh fading channel is used in the simulations and also Monte Carlo simulations (100000 iterations) have been used to better show the reality. All simulations were performed in 9 time blocks and the results in these time blocks were averaged in all graphs. In the simulations, a system with varying number of uniformly distributed active users from 4 to 16 and a total of 36 registered users is considered. Since a pilot sequence is required for each registered user, 36 ZC sequences were obtained with 6 cyclic shifts each of the 6 root ZC sequences. Users are divided into 6 groups and in these groups, 6 different codebooks and pilot sequences can be used for transmission. Moreover, each user use LTE UL Demodulated Reference Signal (DMRS) as their pilots' location [8].

Parameter	Setting
Registered users (J)	36
The number of pilot sequences $(N_{\rm P})$	36
The number of codebooks (N_{CB})	6
The length of pilot sequence (L)	23
Channel model	Rayleigh

Tab. 1. Set of basic simulation parameters.

There are several parameters that are important in the UL grant-free scenario. These are miss detection probability, false alarm probability and mean square error (MSE). Miss detection is the detection of an active user as inactive. False detection means that inactive users are detected as active users. MSE is a criterion for estimating channel gain information. High miss detection probability and high false alarm probability reduce receiver performance. The false alarm probability is reduced by JMPA, but miss detection is not compensated. Therefore, low miss detection probability and MSE values are the most important features for the receiver.

In Fig. 4, we compare the miss detection probability performance among OMP [23], FOCUSS [19], CoSaMP [24], DGOMP [13] and proposed DSACS-based algorithms at different activity levels. The proposed algorithm has the best performance compared to other methods when the number of active users is 6 in Fig. 4(a). While the number of active users is 10 as seen in Fig. 4(b), the performance of the proposed algorithm decreases relatively compared to other algorithms, but as the SNR increases, its performance increases. CoSaMP also performs well, but the sparsity is assumed to be known, which is not possible in practice.

10

Miss Detection Probability

0

FOCUS

Propos

5

10

SNR(dB)

Figure 5 shows the effect of pilot length on the probability of missed detection versus SNR. As can be seen, increasing the ZC pilot length, decreases the miss detection probability, thus increasing the performance of the method. However, it also increases complexity. Therefore, there is a trade-off between complexity and performance. In Fig. 5, it is seen that while the pilot length is L = 11 and L = 17, miss detection probability is very high. However, when the pilot length at low SNR is L = 23, it appears that there is not much difference compared to L = 29 and L = 37. For this reason, the pilot length value was used as L = 23 due to the complexity-performance trade-off in other simulations, as well.

Figure 6 shows the performance of the proposed method and other algorithms according to the number of active users of miss detection probability at $SNR = 10 \, dB$. The proposed method has a better performance in low sparsity, ie. a few active users. As the number of active users increases, the FOCUSS and DGOMP methods have a good performance. The performance of the proposed method is low but close to good. For this reason, it can be thought that the proposed method can be used in sporadic communication where the number of active users is low.



Fig. 4. Miss detection probability performance comparison among OMP, FOCUSS, DGOMP, CoSaMP and proposed DSACS-based algorithm according to different SNRs at two different activity levels.



Fig. 5. Miss detection probability performance of the proposed algorithm at different pilot lengths, number of active users = 6.



Fig. 6. Miss detection probability performance comparison among OMP, FOCUSS, DGOMP, CoSaMP and proposed DSACS-based algorithm according to the number of active users. SNR = 10 dB.



Fig. 7. False alarm probability performance of the proposed algorithm at different SNRs.

Figure 7 shows the performance of the false alarm probability value of the proposed algorithm according to the number of active users at different SNR values. When the SNR value increases, the false alarm probability of the proposed algorithm decreases, that is, its performance increases proportionally. As it increases the computational complexity, the high false alarm probability increases the complexity of MPA. This is prevented by JMPA and the updated AUL is sent back for channel estimation as feedback. As can be seen, as the number of active users increases up to 8, the false alarm probability decreases, and this probability increases after 8. As expected in compressed detection theory, the proposed method performs well when there is sparsity.

Figure 8 provides the MSE performance against SNR for the proposed and state-of-art algorithms. This graph gives the MSE of all algorithms in the channel gains of active users estimated from the received pilot signal. The proposed algorithm has about 3 dB and 6 dB and more gain compared to the DGOMP, the FOCUSS, and the OMP algorithms, respectively. However, it is seen that the CoSaMP algorithm also provides a gain of approximately 2 dB at low SNRs and approximately 5 dB at high SNRs compared to the proposed algorithm. Since the sparsity knowledge in the CoSaMP algorithm is quiet good in terms of usability and performance compared to other algorithms.

5. Conclusion

In this paper, DSACS-based active user detection and channel estimation without prior knowledge on user sparsity is proposed for UL grant-free SCMA systems, taking into account the temporal correlation of activity in adjacent time slots. The proposed algorithm adopts a stagewise approach to expand the set of accurate active users for adaptively achieve the sparsity level of user activity. Then, the algorithm exploits the idea of backtracking to improve the estimate of the accurate active user set at each iteration for more accurate reconstruction. Thus, it is used the temporal correlation of active user sets to reduce the computational complexity and



Fig. 8. MSE against SNR performance of the proposed and stateof-art algorithms, number of active users = 6.

improve performance. The simulation results show that the performance of the algorithm in different criteria is almost superior to the algorithms that have prior knowledge on user sparsity. The proposed method is highly practical with its high performance in systems where users are sparse, such as massive machine-type communication.

Acknowledgments

The authors would like to thank to Dr. Markku Juntti at Oulu University (Finland) for their comments and suggestions.

References

- ALAM, M., ZHANG, Q. Sequence block based compressed sensing multiuser detection for 5G. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. Abu Dhabi (United Arab Emirates), 2018, p. 1–6. DOI: 10.1109/GLOCOM.2018.8647750
- [2] HUSSAIN, Q., SOHAIB, S. Full duplex relaying in nonorthogonal multiple access system with advanced successive interference cancellation. *Radioengineering*, 2020, vol. 29, p. 654–663. DOI: 10.13164/re.2020.0654
- [3] DAI, L., WANG, B., YUAN, Y., et al. Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends. *IEEE Communications Magazine*, 2015, vol. 53, no. 9, p. 74–81. DOI: 10.1109/MCOM.2015.7263349
- [4] NIKOPOUR, H., BALIGH, H. Sparse code multiple access. In Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). London (United Kingdom), 2013, p. 332–336. DOI: 10.1109/PIMRC.2013.6666156
- [5] DURAK, M. H., ERTUG, O. CS-based multiuser detector for SCMA systems. In *Proceedings of the International Symposium on Networks*, *Computers and Communications (ISNCC)*. Istanbul (Turkey), 2019, p. 3–6. DOI: 10.1109/ISNCC.2019.8909149
- [6] WANG, B., WANG, K., LU, Z., et al. Comparison study of nonorthogonal multiple access schemes for 5G. In *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. Ghent (Belgium), 2015, p. 1–5. DOI: 10.1109/BMSB.2015.7177186

- [7] SHARMA, S., DEKA, K., HONG, Y. User activity detectionbased large SCMA system for uplink grant-free access. In *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops).* Shanghai (China), 2019, p. 1–6. DOI: 10.1109/ICCW.2019.8756846
- [8] BAYESTEH, A., YI, E., NIKOPOUR, H., et al. Blind detection of SCMA for uplink grant-free multiple-access. In *Proceedings* of the IEEE 11th International Symposium on Wireless Communications Systems (ISWCS). Barcelona (Spain), 2014, p. 853–857. DOI: 10.1109/ISWCS.2014.6933472
- [9] DONOHO, D. L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, vol. 52, no. 4, p. 1289–1306. DOI: 10.1109/TIT.2006.871582
- [10] AKRAM, F., RASHID, I., GHAFOOR, A., et al. Coherence optimized channel estimation for mm-wave massive MIMO. *Radioengineering*, 2020, vol. 29, no. 4, p. 625–635. DOI: 10.13164/re.2020.0625
- [11] WANG, B., DAI, L., ZHANG, Y. Dynamic compressive sensingbased multi-user detection for uplink grant-free NOMA. *IEEE Communications Letters*, 2016, vol. 20, no. 11, p. 2320–2323. DOI: 10.1109/LCOMM.2016.2602264
- [12] WANG, B., DAI, L., YUAN, Y., et al. Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access. In *Proceedings of the IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. Boston (USA), 2016, p. 1–5. DOI: 10.1109/VTCFall.2015.7390876
- [13] LIU, J., WU, G., LI, S., et al. Blind detection of uplink grant-free SCMA with unknown user sparsity. In *Proceedings of the IEEE International Conference on Communications (ICC)*. Paris (France), 2017, p. 1–6. DOI: 10.1109/ICC.2017.7997304
- [14] WANG, F., ZHANG, Y., ZHAO, H., et al. Active user detection of uplink grant-free SCMA in frequency selective channel. In *Proceedings* of the IEEE Vehicular Technology Conference (VTC Spring). Porto (Portugal), 2018, p. 1–6. DOI: 10.1109/VTCSpring.2018.8417833
- [15] GUO, S., WU, W., WU, X., et al. Dynamic pilot design and channel estimation based on structured compressive sensing for uplink SCMA system. In *Proceedings of the IEEE/CIC International Conference on Communications Workshops* in China (ICCC Workshops). Changchun (China), 2019, p. 87–92. DOI: 10.1109/ICCChinaW.2019.8849953
- [16] WANG, B., DAI, L., MIR, T., et al. Joint user activity and data detection based on structured compressive sensing for NOMA. *IEEE Communications Letters*, 2016, vol. 20, no. 7, p. 1473–1476. DOI: 10.1109/LCOMM.2016.2560180
- [17] DU, Y., CHENG, C., DONG, B., et al. Block-sparsity-based multiuser detection for uplink grant-free NOMA. *IEEE Transactions* on Wireless Communications, 2018, vol. 17, no. 12, p. 7894–7899. DOI: 10.1109/TWC.2018.2872594
- [18] XIAO, J., DENG, G., NIE, G., et al. Dynamic adaptive compressive sensing-based multi-user detection in uplink URLLC. In *Proceedings* of the IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Bologna (Italy), 2018, p. 1–6. DOI: 10.1109/PIMRC.2018.8581010

- [19] GORODNITSKY, I. F., RAO, B. D. Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 1997, vol. 45, no. 3, p. 600–616. DOI: 10.1109/78.558475
- [20] ZHANG, S., XU, X., LU, L., et al. Sparse code multiple access: An energy efficient uplink approach for 5G wireless systems. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. Austin (USA), 2014, p. 4782–4787. DOI: 10.1109/GLOCOM.2014.7037563
- [21] TAN, J., DING, W., YANG, F., et al. Compressive sensing based time-frequency joint non-orthogonal multiple access. In *Proceed*ings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Nara (Japan), 2016, p. 1–4. DOI: 10.1109/BMSB.2016.7521901
- [22] CANDES, E. J., TAO, T. Decoding by linear programming. *IEEE Transactions on Information Theory*, 2005, vol. 51, no. 12, p. 4203–4215. DOI: 10.1109/TIT.2005.858979
- [23] TROPP, J. A., GILBERT, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions* on *Information Theory*, 2007, vol. 53, no. 12, p. 4655–4666. DOI: 10.1109/TIT.2007.909108
- [24] NEEDELL, D., TROPP, J. A. CoSaMP iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 2009, vol. 26, no. 3, p. 301–321. DOI: 10.1016/j.acha.2008.07.002

About the Authors ...

Mehmet Hakan DURAK was born in Erzurum, Turkey in 1989. He received his B.Sc. degree in 2011 from Ataturk University, M.Sc. from Eskisehir Osmangazi University in 2015. His research interests include wireless communications, signal processing, sparse code multiple access (SCMA), code domain-NOMA and multiuser detection for wireless communications. He is currently a Ph.D. student in Electrical and Electronics Engineering at Gazi University, Ankara, Turkey, where he is working as a research assistant since 2015.

Özgür ERTUĞ was born in Ankara, Turkey in 1975. He received his B.Sc. degree in 1997 from University of Southern California, USA, M.Sc. degree from Rice University in 1999 and Ph.D. degree from Middle East Technical University in 2005. He is currently working as an associate professor in Electrical and Electronics Engineering Department of Gazi University. His main research interests lie in algorithm and architecture design as well as theoretical and simulation-based performance analysis of software-defined telecommunication and defense systems.