

# Large-Scale Antenna-Assisted Grant-Free Non-Orthogonal Multiple Access via Compressed Sensing

Jun Fang, *Senior Member, IEEE*, and Yanlun Wu

**Abstract**—Support massive connectivity is an important requirement in 5G wireless communication system. For massive Machine Type Communication (MTC) scenario, since the network is expected to accommodate a massive number of MTC devices with sparse short message, the multiple access scheme like current LTE uplink would not be suitable. In order to reduce the signaling overhead, we consider an grant-free multiple access system, which requires the receiver facilitate activity detection, channel estimation, and data decoding in “one shot” and without the knowledge of active user’s pilots. However, most of the “one shot” communication research haven’t considered the massive MIMO scenario. In this work we propose a Multiple Measurement Vector (MMV) model based Massive MIMO and exploit the covariance matrix of the measurements to confirm a high activity detection rate.

**Index Terms**—5G, MTC, grant-free, Massive MIMO.

## I. INTRODUCTION

Machine-to-Machine (M2M) is an emerging paradigm for future communication systems. The machine type devices including smart grids, computers, FPGAs, etc, which transmit small data packets occasionally. Due to the sporadic communication and low data rate, these applications put forward new challenges for current LTE Random Access Channel (RACH), where the limited capacity and excessive control signals are not suitable for massive MTC. One approach for reducing signal overhead in physical layer is to facilitate activity and data jointly detection.

In the sporadic transmission scenario, the  $K$  sensor nodes are inactive most of the time, as shown in Fig. (1). Based 5G communication system, [1] and [2] introduce a general compressive multiple antenna random access and raised some crucial challenges to physical layer for sporadic communication. Inspired by the sparse user activity, many compressive sensing (CS)-based reliable joint detection both activity and data have been proposed in [3]–[7]. A random channel access facilitated by Code Division Multiple Access (CDMA) has been showed in [3], [4], which jointly estimating activity and data based on perfect channel information station (CIS). Pilot sequence based channel estimation has been showed in [5]–[7], however, the number of nonzero channel coefficients should be known as prior due to the use of Orthogonal Matching Pursuit (OMP) algorithm [5], [6]. Based OFDM systems, the

base station (BS) with a grant-free transmission procedure has been showed in [8], [9], which also exploit CS for activity detection. Notice that almost all of these methods formulated the detection model as a Single Measurement Vector (SMV) problem which has a poor performance in low SNR. However, in this paper, we making use of the covariance matrix of the measurements, instead of the measurements themselves to achieve a high activity detection rate.

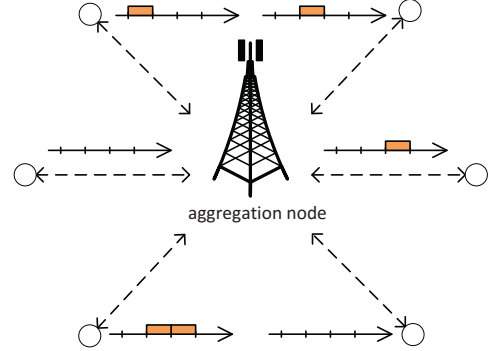


Fig. 1. sporadic communication scenario.

On the other hand, as M2M providing ubiquitous communication environment towards Internet of Things (IoT), the number of M2M users increasing rapidly. But how to increase system capacity? One of the most efficient equipment is massive MIMO system. Massive MIMO is one of the most important technology in 5G system design [2], such system can significantly improve the channel capacity and energy efficiency. The work in [10] represented that upgrading the BS with Massive MIMO can serve 40 percent more M2M devices in a single cell. Unfortunately, many of the sporadic communication have not considered large-scale antennas at aggregation node. In this paper, we consider a massive MIMO wireless uplink transmission for grant-free non-orthogonal multiple access sporadic communication. In the sporadic traffic scenario, we consider the  $K$  sensor nodes communicate with a central aggregation node which we assumed that it equipped with a very large number of antennas, all sensor nodes transmit a small data package occasionally. Our goal is to achieve “one shot” communication [2], i.e., user detection, channel estimation and data detection completed in one time slot. Without any handshaking processing, we will use the sparsity of the active user and non-orthogonal training sequence to avoid excessive signaling overhead. Furthermore, the activity detection is the

Jun Fang and Yanlun Wu are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China, Email: JunFang@uestc.edu.cn

This work was supported in part by the National Science Foundation of China under Grant 61522104.

first and foremost step of “one shot” communication, a reliable detection is crucial for system performance. According making use of the statistically uncorrelated of Massive MIMO channel information, we show in this paper that the accurate activity detection rate will go to 1 as antennas  $M \rightarrow \infty$ .

## II. SYSTEM MODEL

### A. Transmission Setup

We consider a sporadic uplink transmission in TDD massive MIMO system, as shown in figure 1. In order to model the activity of the sensor node, we assume that all sensor nodes have an identical active probability  $p_a \ll 1$  which is independent of users and data, and the number of active users obeys discrete Binomial distribution  $\mathcal{B}(K, p_a)$  which remain the same in the coherence time. Of course, when the nodes number  $K$  increase large, the number of active users is quite small. For an active node  $k_a$ , we assume that it transmits a symbol vector  $\mathbf{d}_{k_a} \in \mathcal{A}^N$  per frame, where  $\mathcal{A}$  denotes the modulation alphabet. For an inactive node  $k_i$ , the symbol vector modeled as zero. Consequently, the augmented alphabet  $\mathcal{A}_0 := \{\mathcal{A} \cup 0\}$  denotes the modulated symbol of both active and inactive nodes.

Different from conventional model in [4], [5], we assume that the aggregation node equipped with a large number of antennas  $M$  which communicates with  $K$  sensor nodes each employs with single-antenna (assuming  $K \ll M$ ). Notice that sensor nodes with single-antenna is simple and inexpensive and can be easily extended to nodes with multiple antennas. The received signal at the aggregation node is can be written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W} \quad (1)$$

where  $\mathbf{H} \in \mathbb{C}^{M \times K}$  denotes the flat-fading channel matrix,  $\mathbf{W} \in \mathbb{C}^{M \times L}$  is the noise matrix with i.i.d entries  $\mathcal{N}(0, \sigma_w^2)_{\mathbb{C}}$ .  $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_K^T]^T$  and  $\mathbf{x}_k \in \mathbb{R}^{L \times 1}$ ,  $k = 1, \dots, K$  is the transmitted symbol vector of the  $k$ th node and  $L$  is the length of symbols. Here, we assume that each node encodes the transmit symbol with a random Gaussian code, owing to the good performance over Gaussian matrices in CS and the other transmission techniques can be seen as a specific case of random coding [5], such as CDMA and SC-FDMA. Further more, we assume that the transmission is synchronized.

In order to implement activity detection, channel estimation and data decoding in “one shot” [2], we split the formulation of (1) into two separate section. First, for known pilot dictionary, we detect the activity and estimate CIS. Then, after channel estimation, complete data decoding.

### B. Massive MIMO Channel Model

We assume the aggregation node employs the uniform linear array (ULA) with the antenna spacing  $d$  at the aggregation node is half of signal wavelength  $\lambda$  (i.e.,  $d = \frac{1}{2}\lambda$ ), which has been studied for MIMO channel estimation [11], [12], where the angle spread is limited in the virtual angular domain. Then

the channel vector from the  $k$ th node to the aggregation node can be written as

$$\mathbf{h}_k = \frac{1}{\sqrt{P}} \sum_{p=1}^P g_{kp} \mathbf{a}(\theta_p) \quad (2)$$

where  $P$  is the i.i.d paths from node  $k$  to the aggregation node, which related to geometric propagation environment only.  $g_{kp}$  is the  $k$ th channel’s attenuation, which is independent over path direction  $p$  and assumed to satisfy  $g_{kp} \sim \mathcal{CN}(0, 1)$ .  $\mathbf{a}(\theta_p)$  is the  $M \times 1$  steering vector which is given by

$$\mathbf{a}(\theta_p) = [1 \ e^{-j2\pi \frac{D}{\lambda} \cos(\theta_p)} \ \dots \ e^{-j2\pi \frac{(M-1)D}{\lambda} \cos(\theta_p)}]^T \quad (3)$$

where  $\theta_p \in [-\pi/2, \pi/2]$  is a random angle of arrival (AOA). For massive MIMO system with large arrays, recent research in channel measurement has shown that the channel characteristics approximate the favorable propagation condition fairly well [11], [12], which show that as long as there are enough paths in (2), all the channel vectors can be approximated by Gaussian model:

$$\mathbf{h}_k = \mathbf{R}_k^{1/2} \mathbf{h}_{Wk}, k = 1, 2, \dots, K, \quad (4)$$

where  $\mathbf{h}_{Wk} \sim \mathbb{C}(\mathbf{0}, \mathbf{I}_M)$  and  $\mathbf{R}_k^{1/2} \triangleq \mathbb{E}\{\mathbf{H}_k \mathbf{H}_k^H\}$ . Under the most favorable propagation, the elements of channel vector that different users to the same antenna  $\mathbf{h}_m$  can be approximated by independent random variables with zero means. In the rest of this paper, we will hold the assumption of favorable propagation that

$$\begin{aligned} E[\mathbf{h}_m(i)] &= 0 \\ E[\mathbf{h}_m(i) \mathbf{h}_n(j)] &= \sigma_i^2 \delta(i-j)(m-n), \quad 1 < m, n \leq M. \end{aligned} \quad (5)$$

## III. ACTIVITY DETECTION

In this section, we consider the activity and channel information are jointly estimated. In conventional, channel estimation employs orthogonal training sequence, i.e.  $L \geq K$  where  $L$  represents the length of pilot symbols. As  $K$  increases, the training overhead is unaffordable. Furthermore, we consider the grant-free communication where the cumbersome hand-shaking procedure is no longer applicable for 5G. As the study in [5], the communications from sensor nodes to aggregation node are always sporadic, i.e., sensor nodes transmit signals occasionally. By the theorem of compressed sensing, the sparsity of sporadic communications can be leveraged to reduce the overhead of nodes accessing (i.e.  $L < K$ ) and a grant-free communication can be achieved. Note that the training sequences used for sensor nodes in a slot are not allocated by the aggregation node. However, the aggregation node knows the codebook of training symbols.

In the stage of activity detection, the received pilot signal at aggregation node is

$$\mathbf{Y}_p = \mathbf{H}\mathbf{S}^H + \mathbf{W} \quad (6)$$

taking a simply transposition of (6), we can get

$$\mathbf{Y}_p^H = \mathbf{S}\mathbf{H}^H + \mathbf{W}^H \quad (7)$$

where  $\mathbf{W}^H \in \mathbb{C}^{L \times M}$  is the noise matrix with i.i.d entries  $\mathcal{N}(0, \sigma_w^2)_{\mathbb{C}}$  and  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K]$  is the  $L \times K$  training dictionary with  $L$  is the length of the pilot sequences. Note that  $L < K$ , the dictionary is underdetermined.  $\mathbf{H}^H = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_M]$  is a  $K \times M$  matrix, where  $\mathbf{h}_m$  is a sparse vector

$$\mathbf{h}_m = \begin{cases} h_m^{(k)} & k \in S_0 \\ 0 & \text{elsewhere} \end{cases}$$

and they share a common sparsity  $\|\mathbf{h}_m\|_0 = D$ , where  $D$  denotes the sparse level of the active nodes,  $S_0 = \{i_1, i_2, \dots, i_D\}$  denotes the index of nonzeros. Consequently,  $\mathbf{Y}^H$  contains  $M$  measurement vectors which share a common sparse support, this is a MMV problem in sparse representation, it has been well discussed and showed a good performance in [13]–[15].

#### A. Activity Detection Model

In the section of activity detection, our goal is to recovery the sparse support of  $\mathbf{H}$  without estimate the underlying data, after detecting the active nodes correctly, channel estimation can be done with the least squares (LS) approach.

The activity detection is the first step of "one shot" communication and it is crucial for system performance. To avoid data losing caused by fallacious classification, we consider the detection based on covariance matrix of the measurement to reduce the interference of noise, and the statistic characteristics of noise variance usually can be seen as a prior. Based on MMV model in (7), let  $\mathbf{y}_m^H \in \mathbb{C}^{L \times 1}$ ,  $m = 1, \dots, M$  denote a set of  $M$  antennas' measurement vectors, and  $\mathbf{h}_m^H \in \mathbb{C}^{K \times 1}$ ,  $m = 1, \dots, M$  are  $M$  unknown channel vectors with the uniform sparsity  $D$ , we have

$$\mathbf{y}_m^H = \mathbf{S} \mathbf{h}_m^H + \mathbf{w}_m^H, \quad 1 \leq m \leq M. \quad (8)$$

We are interested in the correlation matrix of received signal, which is given by

$$E(\mathbf{y}_m^H \mathbf{y}_m) = E(\mathbf{S} \mathbf{h}_m^H \mathbf{h}_m \mathbf{S}^H) + E(\mathbf{w}_m^H \mathbf{w}_m) \quad (9)$$

Denote the sample covariance matrix as  $\Phi_{yy}$ , we can get a simplified formulation as

$$\Phi_{yy} \triangleq \frac{1}{M} \sum_{m=1}^M \mathbf{y}_m^H \mathbf{y}_m = \mathbf{S} \mathbf{R}_{hh} \mathbf{S}^H + \Phi_{ww} \quad (10)$$

where  $\mathbf{R}_{hh} = \frac{1}{M} \sum_{m=1}^M \mathbf{h}_m^H \mathbf{h}_m$ . With  $M \rightarrow \infty$ , exploiting (5), we can get that  $\mathbf{R}_{hh}$  goes to a diagonal matrix with  $r_{kk} \neq 0$  corresponding an active node and  $r_{kk} = 0$  if the  $k$ th node is inactive, so the activity detection fully determined by the nonzero position of the matrix  $\mathbf{R}_{hh}$ .  $\Phi_{ww} = \frac{1}{M} \sum_{m=1}^M \mathbf{w}_m^H \mathbf{w}_m$  is a matrix obeys the Wishart distribution and can be written as  $\Phi_{ww} = \sigma_w^2 \mathbf{I}_L + \mathbf{N}$ , where  $\sigma_w^2 \mathbf{I}_L$  denotes the mean of the noise covariance matrix and  $\mathbf{N}$  is a random variable with zeros mean and variance  $\sigma_w^4/M$ , so  $\mathbf{N}$  converge to zero with  $M \rightarrow \infty$ .

Using the property of Khatri-Rao product, upon vectorization of (10), we have

$$\mathbf{z} \triangleq \text{vec}(\Phi_{yy}) = (\mathbf{S} \odot \mathbf{S}) \mathbf{r}_{hh} + \mathbf{n} + \mathbf{e} \quad (11)$$

$$\mathbf{x} \triangleq \mathbf{z} - \mathbf{n} = (\mathbf{S} \odot \mathbf{S}) \mathbf{r}_{hh} + \mathbf{e} \quad (12)$$

where  $\mathbf{n} = \text{vec}(\Phi_{ww})$  is a  $L^2 \times 1$  vector. With the assumption that the noise variance is known, then the activity detection can be reformulated as (12) by finding the support of  $\mathbf{r}_{hh} \in \mathbb{C}^{K \times 1}$ .  $\mathbf{r}_{hh}$  is a  $D$  sparse vector with non-zero elements comprised in the vector  $\text{diag}(\mathbf{R}_{hh})$ , whose non-zero elements are given by

$$\sigma_i^{2[M]} \triangleq \frac{1}{M} \sum_{m=1}^M [\mathbf{h}_m^H]_i^2, \quad i \in S_0 \quad (13)$$

Owing to the finite number of antennas  $M$ , an additive noise-like term  $\mathbf{e} \in \mathbb{C}^{L^2 \times 1}$  arising which can be divided into two parts  $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2$ , the elements of  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are given respectively

$$\frac{1}{M} \sum_{i \neq j} \sum_{m=1}^M \mathbf{s}_{p,i} \mathbf{s}_{q,j} [\mathbf{h}_m^H]_i [\mathbf{h}_m^H]_j, \quad 1 \leq p, q \leq L, \quad 1 \leq i, j \leq K. \quad (14)$$

$$\frac{1}{M} \sum_{p \neq q} \sum_{m=1}^M [\mathbf{w}_m^H]_p [\mathbf{w}_m^H]_q, \quad 1 \leq p, q \leq L. \quad (15)$$

They reflect the cross-correlation of  $\mathbf{y}_m$  for finite number of antennas. This noise-like term will goes to arbitrarily small when the number of antennas  $M \rightarrow \infty$ .

To approach the activity detection, we have convert the MMV model in (7) to SMV model in (12) by recovering the sparse vector  $\mathbf{r}_{hh}$ , where  $\mathbf{S} \odot \mathbf{S}$  is the efficient measurement matrix with dimension  $L^2 \times K$ , as we show shortly, the dimension gain in measurement matrix could efficiently improve the system capacity of active nodes.

We focus on recover the support of sparse vector  $\mathbf{r}_{hh}$ , note that there is a constraint for vector  $\mathbf{r}_{hh}$  to be non-negative, we consider the constraint  $l_1$  norm regularized quadratic programming:

$$\begin{aligned} \min_{\mathbf{r}} \quad & \left( \frac{1}{2} \|(\mathbf{S} \odot \mathbf{S}) \mathbf{r} - \mathbf{x}^{[M]}\|_2^2 + \lambda \|\mathbf{r}\|_1 \right) \\ \text{s.t.} \quad & \mathbf{r} \geq 0 \end{aligned} \quad (16)$$

where  $\lambda$  a regularization parameter. This problem, known as LASSO [16], whose performance has been studied in [17], [18] including the sufficient conditions for exact sparsity recovery and examined a variety of approaches proposed for parameter estimation in LASSO.

#### B. Theoretical Results

In this section, we will give some theoretical results of training code design and activity detection. For ease of notation, we assume the columns of training dictionary  $\mathbf{S}$  to have  $\|\mathbf{s}_i\|_2 = 1$ . Notice that this assumption has no impact on the recovering performance. The mutual coherence of matrix  $\mathbf{S}$  is  $\mu_{\mathbf{S}}$ , which is defined as:

$$\mu_{\mathbf{S}} \triangleq \max_{i \neq j} |\mathbf{s}_i^H \mathbf{s}_j| \quad (17)$$

Let  $\mu_{\mathbf{S}_{KR}}$  indicates the mutual coherence of  $\mathbf{S} \odot \mathbf{S}$ , combine the property of Kronecker products that  $(\mathbf{a}_i \otimes \mathbf{b}_i)^H (\mathbf{a}_j \otimes \mathbf{b}_j) = (\mathbf{a}_i^H \mathbf{a}_j) (\mathbf{b}_i^H \mathbf{b}_j)$ , we can obtain

$$|[\mathbf{S}_{KR}]_i^H [\mathbf{S}_{KR}]_j| = |(\mathbf{s}_i \otimes \mathbf{s}_i)^H (\mathbf{s}_j \otimes \mathbf{s}_j)| = (|\mathbf{s}_i^H \mathbf{s}_j|)^2 \quad (18)$$

Hence the connection between  $\mu_{\mathbf{S}_{KR}}$  and  $\mu_{\mathbf{S}}$  can be written as:

$$\mu_{\mathbf{S}_{KR}} = \mu_{\mathbf{S}}^2 \quad (19)$$

Based on mutual coherence of a matrix, many recovery performance in different situations have been studied [19]–[21]. Follows from the study of [22], solving the sparse support problem via  $\min_{\mathbf{x}, \mathbf{Sx}=\mathbf{y}} \|\mathbf{x}\|_1$  yields a sufficient condition based on mutual coherence that

$$|S_0| < \frac{1}{2} \left(1 + \frac{1}{\mu_{\mathbf{S}}}\right) \quad (20)$$

Consequently, through the mutual covariance matrix and Khatri-Rao product, the coherence of measurement matrix from  $\mu_{\mathbf{S}}$  becomes  $\mu_{\mathbf{S}}^2$ , owing to the fact that  $\mu_{\mathbf{S}} < 1$ , the capacity of system active nodes  $|S_0|$  has been raised.

Further more, the support recovery condition in (20) suffers from the Welch bound [23], which is defined as:

$$\mu_{\mathbf{S}} \geq \sqrt{\frac{K-L}{(K-1)L}}. \quad (21)$$

If the maximum number of active node  $|S_0|$  has determinant, we can easily calculate the lower bound for the needed length of pilot sequence from (19), (20), and (21), which is given by

$$L > \frac{2K|S_0| - K}{K + 2|S_0| - 2} \quad (22)$$

Note that there are many construction matrix nearly meeting the Welch bound (21), such as the Gold/Gold-like codes, FZC codes, Kasami codes, etc. In sporadic communication system, via choosing a proper codec we can obtain the shortest length of pilot to ensure active nodes communication and meanwhile improve the efficiency of data transmission in a coherence time.

In the following analysis, we will show that the probability of active node detection will great than  $1 - \alpha\beta^{-M}$ , which increase overwhelming with the number of antennas  $M$ .

According to the study of [24] (*Theorem 7*), if  $|S_0| < \frac{1}{2} \left(1 + \frac{1}{\mu_{\mathbf{S}}^2}\right)$ , the optimal solution  $\mathbf{r}^*$  to the constrained LASSO is unique and it satisfies  $\text{Supp}(\mathbf{r}^*) = S_0$  when the following two conditions hold:

$$\|\mathbf{e}\|_2 < \lambda \frac{1 + \mu_{\mathbf{S}}^2 - 2\mu_{\mathbf{S}}^2 D}{1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D}, \quad (\text{Event } E1) \quad (23)$$

$$\sigma_{\min}^{2[M]} > \frac{\|\mathbf{e}\|_2 + \lambda}{1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D}, \quad (\text{Event } E2) \quad (24)$$

where  $\sigma_{\min}^{2[M]} = \min\{\sigma_i^{2[M]}, i \in S_0\}$ .

We first state the probability of successful recovery of sparse by solving the LASSO [24], which can be directly used in our derivation:

*Theorem 1:* The probability of successful recovery of sparse support by solving the LASSO, denotes as  $P_s$ , satisfies:

$$\begin{aligned} P_s &\geq \mathcal{P}(E1 \cap E2) \\ &= \mathcal{P}\left(\|\mathbf{e}\|_2 < \lambda \frac{1 + \mu_{\mathbf{S}}^2 - 2\mu_{\mathbf{S}}^2 D}{1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D}, \sigma_{\min}^{2[M]} > \frac{\|\mathbf{e}\|_2 + \lambda}{1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D}\right) \\ &\geq \mathcal{P}\left(\|\mathbf{e}\|_2 < c_1, \sigma_{\min}^{2[M]} > c_2\right) \\ &\geq \prod_{i=1}^D \mathcal{P}(\sigma_i^{2[M]} > c_2) - \sum_{i=1}^{L^2} \mathcal{P}\left(|[\mathbf{e}]_i| \geq \frac{c_1}{L}\right) \end{aligned} \quad (25)$$

where

$$c_1 \triangleq \lambda \frac{1 + \mu_{\mathbf{S}}^2 - 2\mu_{\mathbf{S}}^2 D}{1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D}, \quad c_2 \triangleq \lambda \frac{2(1 + \mu_{\mathbf{S}}^2) - 3\mu_{\mathbf{S}}^2 D}{(1 + \mu_{\mathbf{S}}^2 - \mu_{\mathbf{S}}^2 D)^2} \quad (26)$$

There is an important inequality in our derivation which is shown in [25]:

*Lemma 1:* Let each of  $x_i$  and  $y_i, i = 1, \dots, k$  be uncorrelated zero mean Gaussian random variables with variance  $\sigma_x^2$  and  $\sigma_y^2$  respectively. Then

$$\mathcal{P}\left(\left|\sum_{i=1}^k x_i y_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma_x \sigma_y (2\sigma_x \sigma_y k + t)}\right)$$

Notice from (14), we know that  $|[\mathbf{e}_1]_i| \leq \frac{\|\mathbf{S}\|_{\infty, \infty}^2}{M} \sum_{i \neq j} \sum_{m=1}^M |[\mathbf{h}_m^H]_i [\mathbf{h}_m^H]_j|$ , if  $|[\mathbf{e}_1]_i| \geq C_1$ , then

$$\begin{aligned} \sum_{i \neq j} \sum_{m=1}^M |[\mathbf{h}_m^H]_i [\mathbf{h}_m^H]_j| &\geq \frac{C_1 M}{\|\mathbf{S}\|_{\infty, \infty}^2} \Rightarrow \\ \sum_{i_0 \neq j_0} |[\mathbf{h}_m^H]_{i_0} [\mathbf{h}_m^H]_{j_0}| &> \frac{C_1 M}{\|\mathbf{S}\|_{\infty, \infty}^2 D(D-1)} \end{aligned} \quad (27)$$

where  $i_0, j_0 \in 1, \dots, D, i_0 \neq j_0, D = |S_0|$ . According to (27), we can get that

$$\begin{aligned} \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1) \\ \leq \mathcal{P}\left(\sum_{m=1}^M |[\mathbf{h}_m^H]_{i_0} [\mathbf{h}_m^H]_{j_0}| > \frac{C_1 M}{\|\mathbf{S}\|_{\infty, \infty}^2 D(D-1)}\right) \end{aligned} \quad (28)$$

Under the assumption that  $[\mathbf{h}_m^H]_{i_0}$  and  $[\mathbf{h}_m^H]_{j_0}$  are uncorrelated zero mean Gaussian random variables, we can use Lemma 1 to obtain

$$\begin{aligned} \mathcal{P}\left(\sum_{m=1}^M |[\mathbf{h}_m^H]_{i_0} [\mathbf{h}_m^H]_{j_0}| > \frac{C_1 M}{\|\mathbf{S}\|_{\infty, \infty}^2 D(D-1)}\right) \\ \leq 2\exp\left(-\frac{Mt_1^2}{2\sigma_{i_0} \sigma_{j_0} (2\sigma_{i_0} \sigma_{j_0} + t)}\right) \\ \leq 2\exp\left(-\frac{Mt_1^2}{2\sigma_{\max}^{(1)} \sigma_{\max}^{(2)} (2\sigma_{\max}^{(1)} \sigma_{\max}^{(2)} + t_1)}\right) \end{aligned} \quad (29)$$

where  $t_1 = \frac{C_1 M}{\|\mathbf{S}\|_{\infty, \infty}^2 D(D-1)}$ ,  $\sigma_{\max}^{(k)}$  denotes the  $k$ th largest element in the set of  $\{\sigma_i\}_{i=1}^{|S_0|}$ .

Similarly, from (15) and Lemma 1, we can say that

$$\begin{aligned}
& \mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2) \\
& \leq \mathcal{P}\left(\sum_{m=1}^M |[\mathbf{w}_m^H]_p [\mathbf{w}_m^H]_q| > \frac{C_2 M}{L(L-1)}\right) \\
& \leq 2\exp\left(-\frac{Mt_2^2}{2[\sigma_w]_p[\sigma_w]_q(2[\sigma_w]_p[\sigma_w]_q + t_2)}\right) \\
& \leq 2\exp\left(-\frac{Mt_2^2}{2[\sigma_w]_{\max}^{(1)}[\sigma_w]_{\max}^{(2)}(2[\sigma_w]_{\max}^{(1)}[\sigma_w]_{\max}^{(2)} + t_2)}\right) \quad (30)
\end{aligned}$$

where  $t_2 = \frac{C_2 M}{L(L-1)}$ ,  $[\sigma_w]_{\max}^{(k)}$  denotes the  $k$ th largest element in the set of  $\{[\sigma_w]_i\}_{i=1}^L$ .

Now, we assume that  $C_1 + C_2 = \frac{c_1}{L}$ , notice that  $|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \leq |[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i|$  and we can get  $\mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \geq \frac{c_1}{L}) \leq \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \geq \frac{c_1}{L})$ . Besides,  $\mathcal{P}(|[\mathbf{e}_1]_i| \leq C_1)\mathcal{P}(|[\mathbf{e}_2]_i| \leq C_2) \leq \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \leq C_1 + C_2)$ , then we can obtain

$$\begin{aligned}
\mathcal{P}(|[\mathbf{e}_i]| \geq \frac{c_1}{L}) &= \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \geq \frac{c_1}{L}) \\
&\leq \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \geq \frac{c_1}{L}) \\
&= \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| \geq C_1 + C_2) \\
&= 1 - \mathcal{P}(|[\mathbf{e}_1]_i| + |[\mathbf{e}_2]_i| < C_1 + C_2) \\
&\leq 1 - \mathcal{P}(|[\mathbf{e}_1]_i| < C_1)\mathcal{P}(|[\mathbf{e}_2]_i| < C_2) \\
&= 1 - (1 - \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1))(1 - \mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2)) \\
&= \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1) + \mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2) - \\
&\quad \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1)\mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2) \quad (31)
\end{aligned}$$

Define that  $\delta_1 \triangleq \left(\frac{t_1^2}{2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)}(2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)} + t_1)}\right)$  and  $\delta_2 \triangleq \left(\frac{t_2^2}{2[\sigma_w]_{\max}^{(1)}[\sigma_w]_{\max}^{(2)}(2[\sigma_w]_{\max}^{(1)}[\sigma_w]_{\max}^{(2)} + t_2)}\right)$ , combining with (29) and (30), we get

$$\begin{aligned}
& \sum_{i=1}^{L^2} \mathcal{P}(|[\mathbf{e}_i]| \geq \frac{c_1}{L}) \\
& \leq L^2 \left[ \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1) + \mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2) - \right. \\
& \quad \left. \mathcal{P}(|[\mathbf{e}_1]_i| \geq C_1)\mathcal{P}(|[\mathbf{e}_2]_i| \geq C_2) \right] \\
& \leq 2L^2 (e^{-\delta_1 M} + e^{-\delta_2 M} - e^{-(\delta_1 + \delta_2)M}) \quad (32)
\end{aligned}$$

**Lemma 2:** Let  $x_i, i = 1, \dots, M$  denote independent zero mean Gaussian random variables with variance  $\sigma_i^2$ . Assume  $0 < C < \sigma_{\min}^2$ , then, there exist  $\beta > 1$  such that  $\mathcal{P}(\frac{1}{M} \sum_{i=1}^M x_i^2 > C) \geq 1 - \beta^{-M}$ .

*Proof 2:* Denote  $p_x \triangleq \frac{1}{M} \sum_{i=1}^M x_i^2$ . Then  $\mathcal{P}(p_x > C) = \mathcal{P}(\sum_{i=1}^M x_i^2 > CM) = \mathcal{P}(\sum_{i=1}^M z_i^2 > \frac{CM}{\sigma_i^2})$  where  $z_i$  denote i.i.d zero mean standard normal variables. Therefore  $\sum_{i=1}^M z_i^2$  is a Chi-Squared random variable with  $M$  degrees

of freedom. Observe

$$\begin{aligned}
\mathcal{P}(p_x > C) &= 1 - \mathcal{P}\left(\sum_{i=1}^M z_i^2 \leq \frac{CM}{\sigma_i^2}\right) \\
&\geq 1 - \exp\left(\frac{tCM}{\sigma_i^2}(1+2t)^{-M/2}\right) \\
&\geq 1 - \exp\left(\frac{tCM}{\sigma_{\min}^2}(1+2t)^{-M/2}\right) \quad (33)
\end{aligned}$$

where  $t > 0$ . This equation is based on Chernoff Bound and Moment-generating function of a Chi-Squared that random variable with  $M$  degrees of freedom is given by  $(1 - 2t)^{-M/2}, t < 1/2$ . Now define the function  $\beta(t) \triangleq \exp(-\frac{2tC}{\sigma_{\min}^2})(1+2t)$  we can write from (33) that

$$\mathcal{P}(p_x > C) \geq 1 - (\beta(t))^{-M/2} \quad (34)$$

We want to ensure that  $\exists t > 0$ , such that  $\beta(t) > 1$ . Notice that if  $\beta \triangleq \exp(-\frac{2tC}{\sigma_{\min}^2})(1+2t) > 1$ , we will get  $C < \gamma(t)$ , where  $\gamma(t) \triangleq \frac{\sigma_{\min}^2}{2t} \log(1+2t)$ . We can easily find that  $\gamma(t)$  is a decreasing function in  $t$  for  $t > 0$ . Using the limit theorem of equivalent infinitesimal we can get that  $\gamma(0) = \sigma_{\min}^2$ . Given that  $C < \sigma_{\min}^2$ , then indeed  $\exists t_0 > 0$  such that  $C < \gamma(t_0)$  which in turn implies  $\beta(t_0) > 1$ . Hence, we conclude that  $\mathcal{P}(\frac{1}{M} \sum_{i=1}^M x_i^2 > C) \geq 1 - \beta^{-M}$  for  $\beta = \sqrt{\beta(t_0)} > 1$ .

Then from (34), we can say that

$$\prod_{i=1}^{|S_0|} \mathcal{P}(\sigma_i^{2[M]} > c_2) \geq \prod_{i=1}^{|S_0|} (1 - \beta_i^{-M}) \quad (35)$$

where  $\sigma_i^{2[M]}$  is defined in (13) and  $\beta_i > 1$ .

**Theorem 2:** The probability of successful recovery of sparse support by solving the LASSO (16) is greater than  $1 - \alpha\gamma^{-M}$  with  $\gamma > 1$ .

*Proof:* From (32) and (35), we can rewrite the recovery probability  $\mathcal{P}_s$  as

$$\begin{aligned}
\mathcal{P}_s &\geq \prod_{i=1}^D \mathcal{P}(\sigma_i^{2[M]} > c_2) - \sum_{i=1}^{L^2} \mathcal{P}(|[\mathbf{e}_i]| \geq \frac{c_1}{M}) \\
&\geq \prod_{i=1}^{|S_0|} (1 - \beta_i^{-M}) - 2L^2 (e^{-\delta_1 M} + e^{-\delta_2 M} - e^{-(\delta_1 + \delta_2)M}) \\
&\geq 1 - \sum_{i=1}^{|S_0|} \beta_i^{-M} - 2L^2 (e^{-\delta_1 M} + e^{-\delta_2 M} - e^{-(\delta_1 + \delta_2)M}) \\
&\geq 1 - D\beta_{\min}^{-M} - 2L^2 (e^{-\delta_1 M} + e^{-\delta_2 M} - e^{-(\delta_1 + \delta_2)M}) \quad (36)
\end{aligned}$$

notice that  $\delta_1 > 0, \delta_2 > 0$ . Define  $1 < \gamma < \min(\beta_{\min}, e^{\delta_1}, e^{\delta_2})$ , then

$$\begin{aligned}
\mathcal{P}_s &\geq 1 - D\gamma^{-M} - 4L^2\gamma^{-M} + 2L^2e^{-(\delta_1 + \delta_2)M} \\
&\geq 1 - (D + 4L^2)\gamma^{-M} \quad (37)
\end{aligned}$$

Note that when the aggregation node with large antenna arrays (i.e.,  $M \rightarrow \infty$ ), the recover probability goes to 1 with sparsity level restricted upto  $\frac{1}{2}(1 + \frac{1}{\mu_{\frac{1}{2}}^2})$ , this will greatly improves the performance of present sporadic communication which have not been considered in massive MIMO system.

#### IV. CHANNEL ESTIMATION AND DATA DECODING VIA LS

After estimating the activity correctly, aggregation node will get a new  $L \times K_a$  matrix  $\hat{\mathbf{S}}$ , which contains  $K_a$  active nodes training sequences occupation in current slot. Denote  $\hat{\mathbf{H}}$  as a new  $M \times K_a$  matrix which contains the support vectors only. Then we can rewrite the  $M \times L$  received signal as

$$\mathbf{Y}_p = \hat{\mathbf{H}}\hat{\mathbf{S}}^H + \mathbf{W} \quad (38)$$

To simplify system design, we implement channel estimation  $\hat{\mathbf{H}}$  combining with received pilot signal  $\mathbf{Y}_p$  and known training matrix  $\hat{\mathbf{S}}$  according to Least Squares (LS) approach

$$\hat{\mathbf{H}} = \mathbf{Y}_p \hat{\mathbf{S}}^\dagger = \mathbf{Y}_p \hat{\mathbf{S}} (\hat{\mathbf{S}}^H \hat{\mathbf{S}})^{-1} \quad (39)$$

After channel estimation, the transmit data can be decoded through received signal at aggregation node. The received signal matrix can be written as

$$\mathbf{Y}_d = \hat{\mathbf{H}}\mathbf{D} + \mathbf{W} \quad (40)$$

where  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N]$  is a  $K_a \times N$  matrix comprises  $N$  modulation symbols of  $K_a$  active sensor nodes. Also, the symbol vector for an inactive node  $k_i$  is modeled as zeros rather than modulation symbols over one slot.

For known channel  $\hat{\mathbf{H}}$ , a LS approach for data decoding can be obtained from (40)

$$\hat{\mathbf{D}} = \hat{\mathbf{H}}^\dagger \mathbf{Y}_d = \hat{\mathbf{H}}^H (\hat{\mathbf{H}} \hat{\mathbf{H}}^H)^{-1} \mathbf{Y}_d. \quad (41)$$

#### V. SIMULATION RESULTS

In the simulations we consider  $K = 64$  users, all the users are synchronized to the aggregation synchronization signal. Here we suppose the coefficients of each row in matrix  $\mathbf{H}$  are independent random variables with zero means as described in Section II-B. We consider a random Gaussian code which is normalized for pilot symbols.

##### A. Performance of Activity Detection

As described in Section II, firstly, we should complete activity detection according to (16) using training dictionary. We compare the activity detection performance of our covariance matrix method with three traditional MMV algorithms:

- 1) MSBL [14]: An extension of Sparse Bayesian Learning (SBL) for the SMV model to the MMV model. For a fairly comparison, we set the true noise variance as the noise variance parameter value and freeze it.
- 2) BOMP [26]: As mentioned at [15], the MMV model can be transformed to a block SMV model. By letting  $\mathbf{y} = \text{vec}(\mathbf{Y}_p) \in \mathbb{C}^{LM \times 1}$ ,  $\mathbf{T} = \mathbf{S} \otimes \mathbf{I}_M$ ,  $\mathbf{h} = \text{vec}(\mathbf{H}) \in \mathbb{C}^{KM \times 1}$ ,  $\mathbf{w} = \text{vec}(\mathbf{W})$ , we can transform the MMV model to block SMV model:  $\mathbf{y} = \mathbf{T}\mathbf{h} + \mathbf{w}$ , which can be solved by block-OMP (BOMP).
- 3) MFOCUSS [27]: The regularized M-FOCUSS is developed for noisy data. We set the p-norm to 0.8 as suggested by the authors.

In comparison to the activity detection error rates of these different approaches, we investigate a system with the number of antennas  $M = 128$  at aggregation node. For channel matrix

$\mathbf{H}$ , the  $|S_0|$  non-zero columns were randomly selected in each trial. For training matrix  $\mathbf{S}$ , we use a length-20 random code for each pilot symbol. Since that the length of pilot sequence is less to the number of users  $K$ , the training matrix is non-orthogonal. The probability estimates are computed by 1000 Monte Carlo Runs and a successful detection is defined as recovering the entire true support.

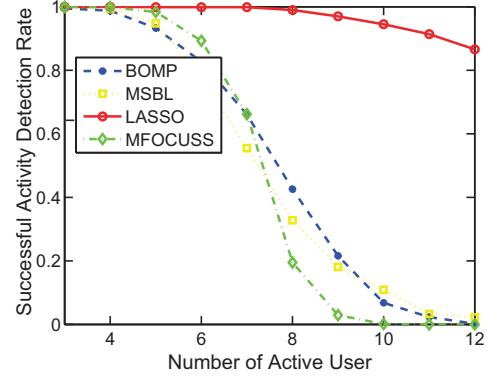


Fig. 2. Percentage of successful active user detection, as a function of  $|S_0|$ . Here  $L=20$ ,  $M=128$ ,  $\text{SNR}=0$ .

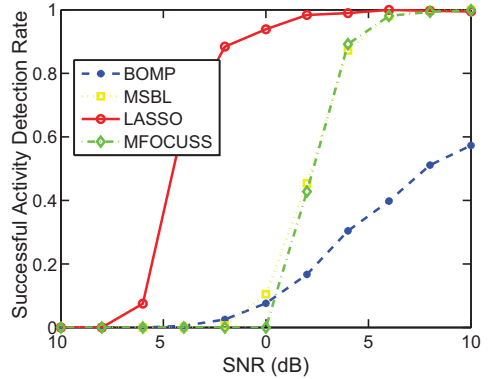


Fig. 3. Percentage of successful active user detection, as a function of SNR. Here  $L=20$ ,  $M=128$ ,  $|S_0|=10$ .

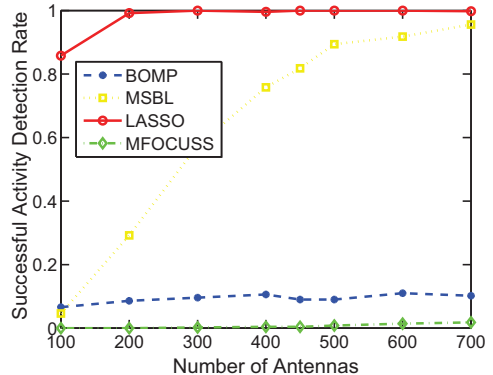


Fig. 4. Percentage of successful active user detection, as a function of  $M$ . Here  $L=20$ ,  $|S_0|=10$ ,  $\text{SNR}=0$ .

In Fig. 2, We plot the activity detection performance as a function of active users  $|S_0|$ . The results show that for less than 4 active users, all the algorithms mentioned above have an almost perfect detection rate, but as the number of active users increase over 6, only the LASSO based on covariance matrix method remain a high detection rate. Consider the same experimental setup as before and fixed  $|S_0| = 10$ , assume an active user transmits with an average energy per symbol of  $E_S = 1$ , then the signal to noise ratio (SNR) is  $1/\sigma_w^2$ , Fig. 3 shows the activity detection rate as a function of SNR from -10 dB to 10 dB. It can be seen that there is a distinct advantage of LASSO at low SNR, at almost 0dB, LASSO based covariance matrix has reached a very high detection rate. In Fig. 4, we plot the detection rate as a function of the antennas number  $M$  with  $|S_0| = 10$ . The observation is that given this level of sparsity and the particular noise, the detection rate of MSBL and LASSO increase as  $M$  becomes large, but MFOCUSS and BOMP are almost incapable of recovering the entire true active users at all values of  $M$ . Owing to this LASSO method based on covariance matrix, there are a perfect detection rate as the number of antennas equipped at BS becomes large. It demonstrates that this method is very satisfied with the massive MIMO system.

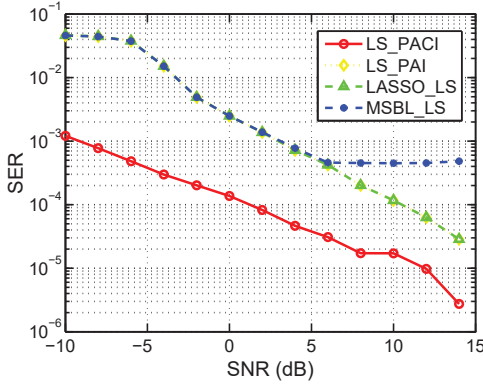


Fig. 5. Symbol error as a function of SNR. Here  $L=20$ ,  $T=40$ ,  $M=500$ ,  $|S_0|=6$ .

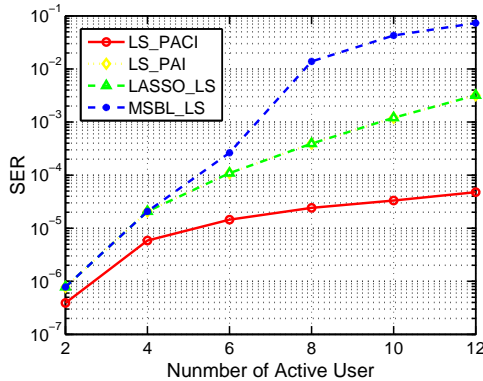


Fig. 6. Symbol error as a function of  $|S_0|$ . Here  $L=20$ ,  $T=40$ ,  $M=500$ ,  $\text{SNR}=10$ .

In Fig. 5 and Fig. 6, we simulated the average Symbol Error Rate (SER) over the augmented alphabet  $\mathcal{A}$  for a length

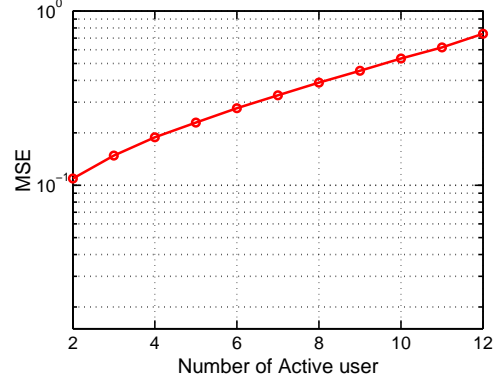


Fig. 7. Channel estimation error, as a function of  $|S_0|$ . Here  $L=20$ ,  $M=500$ ,  $\text{SNR}=10$ .

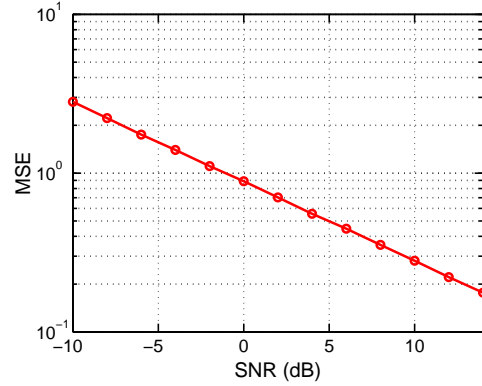


Fig. 8. Channel estimation error, as a function of SNR. Here  $L=20$ ,  $M=500$ ,  $|S_0|=6$ .

of  $N = 40$  transmit-symbols, the symbols are also encoded by random coding. Due to the data decoding is the second stage of “one shot” communication, the performance of SER is the result of missed detection and channel estimation error. In order to study the contribution of the channel and activity estimation error, we measured the SER for perfect activity and channel information (PACI) through LS and perfect activity information (PAI) through LS for channel estimation and data decoding, which are compared with the activity detection based LASSO and MSBL. The simulation results indicate that the performance of SER with and without PAI doesn't has significant differences in low activity and high SNR, with the increase of noise and active users, LASSO based activity detection shows certain advantages. We can also observe that there is a large gap between PACI and other cases in SER. This is mainly due to the LS based channel estimation with non-orthogonal training matrix suffers significantly, so the performance of SER is mainly limited by channel estimation. In Figs. 7 and Figs. 8, we plot the Mean Squared Error to explore the performance of channel estimation based LS:

$$\mathcal{MSE} = \sum_{ka \in |S_0|} \frac{\|\mathbf{h}_{ka} - \hat{\mathbf{h}}_{ka}\|_2^2}{\|\mathbf{h}_{ka}\|_2^2} \quad (42)$$



## VI. CONCLUSION

In this paper, we have proposed a massive MIMO wireless uplink transmission based grant-free non-orthogonal multiple access for 5G. Numerical results show that our proposed model achieves a significant performance improvement through exploiting statistical information about the unknown massive MIMO channel information.

## REFERENCES

- [1] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Brink, I. Gaspar, N. Michailow, A. Festag *et al.*, “5gnow: non-orthogonal, asynchronous waveforms for future mobile applications,” *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 97–105, 2014.
- [2] G. Wunder, H. Boche, T. Strohmer, and P. Jung, “Sparse signal processing concepts for efficient 5g system design,” *Access, IEEE*, vol. 3, pp. 195–208, 2015.
- [3] F. Monsees, C. Bockelmann, D. Wubben, and A. Dekorsy, “Sparsity aware multiuser detection for machine to machine communication,” in *GLOBECOM Workshops*, 2012, pp. 1706 – 1711.
- [4] C. Bockelmann, H. F. Schepker, and A. Dekorsy, “Compressive sensing based multi-user detection for machine-to-machine communication,” *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 389–400, 2013.
- [5] H. F. Schepker, C. Bockelmann, and A. Dekorsy, “Exploiting sparsity in channel and data estimation for sporadic multi-user communication,” in *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on*. VDE, 2013, pp. 1–5.
- [6] Y. Beyene, C. Boyd, K. Ruttik, C. Bockelmann, O. Tirkkonen, and R. Jantti, “Compressive sensing for mtc in new lte uplink multi-user random access channel,” in *AFRICON, 2015*. IEEE, 2015, pp. 1–5.
- [7] G. Wunder, P. Jung, and C. Wang, “Compressive random access for post-lte systems,” in *IEEE International Conference on Communications Workshops*, 2014, pp. 539–544.
- [8] B. Wang, L. Dai, Y. Yuan, and Z. Wang, “Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access,” in *Vehicular Technology Conference*, 2015.
- [9] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, “Blind detection of scma for uplink grant-free multiple-access,” in *International Symposium on Wireless Communications Systems*, 2014, pp. 853–857.
- [10] Y. Mehmood, N. Haider, W. Afzal, U. Younas, I. Rashid, and M. Imran, “Impact of massive mimo systems on future m2m communication,” in *Communications (MICC), 2013 IEEE Malaysia International Conference on*. IEEE, 2013, pp. 534–537.
- [11] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up mimo: Opportunities and challenges with very large arrays,” *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, 2013.
- [12] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 264–273, 2013.
- [13] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [14] D. P. Wipf and B. D. Rao, “An empirical bayesian strategy for solving the simultaneous sparse approximation problem,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3704–3716, 2007.
- [15] Z. Zhang and B. D. Rao, “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 912–926, 2011.
- [16] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [17] M. J. Wainwright, “Sharp thresholds for high-dimensional and noisy sparsity recovery using constrained quadratic programming (lasso),” *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2183–2202, 2009.
- [18] M. Schmidt, “Least squares optimization with l1-norm regularization,” *CS542B Project Report*, pp. 14–18, 2005.
- [19] Z. Ben-Haim, Y. C. Eldar, and M. Elad, “Coherence-based performance guarantees for estimating a sparse vector under random noise,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.
- [20] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [21] J. A. Tropp, “Corrigendum in ”just relax: Convex programming methods for identifying sparse signals in noise”,” *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [22] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ?” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–202, 2003.
- [23] L. R. Welch, “Lower bounds on the maximum cross correlation of signals (corresp.),” *Information Theory, IEEE Transactions on*, vol. 20, no. 3, pp. 397–399, 1974.
- [24] P. Pal and P. Vaidyanathan, “Pushing the limits of sparse support recovery using correlation information,” *Signal Processing, IEEE Transactions on*, vol. 63, no. 3, pp. 711–726, 2015.
- [25] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, “Toeplitz compressed sensing matrices with applications to sparse channel estimation,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.
- [26] J. Fang and H. Li, “Block-sparsity pattern recovery from noisy observations,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3321–3324.
- [27] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, 2005.