On the Correctness of Coherence in the Determination of Preambles for Compressive Random Access

Jinho Choi and Nam Yul Yu
School of Electrical Engineering and Computer Science, GIST
e-mail: {jchoi0114, nyyu}@gist.ac.kr

***Abstract*—In this paper, we study a criterion for measurement matrices to understand the performance of sparse signal recovery. In particular, we consider the activity detection in compressive random access and derive an approximate upper-bound on the recovery error probability using pair-wise error probability (PEP) when the maximum likelihood (ML) detection is employed. From the approximate upper-bound, we obtain a performance criterion that can be used to choose a set of pilot or preamble sequences. Based on the derived performance criterion, we can show that the performance of Alltop sequences can be similar to that of Zadoff-Chu sequences.**

***Index Terms*—coherence, compressive sensing, performance criteria**

I. Introduction

For the Internet of Things (IoT), there has been a growing interest in machine-type communications (MTC) or machine-to-machine (M2M) communications in order to support connections for a number of devices to a network. In [1]–[3], MTC is studied within cellular systems. For MTC, random access is usually considered due to low control and signaling overhead [2], [4], [5]. In particular, a random access scheme, called random access procedure, has been proposed for MTC in the long term evolution-advanced (LTE-A) system [1].

In the random access procedure, devices that want to establish connections to an access point (AP) transmit randomly chosen preambles from a pool of preambles simultaneously. Thus, the AP needs to detect multiple preambles. Since only a fraction of devices are active, instead of employing an exhaustive search for multiple signal detection, it would be efficient to exploit the notion of compressive sensing (CS) [6]–[9] to derive low-complexity detection methods as in [10]–[13]. In *compressive random access*, since the sparsity of active devices is to be exploited for low-complexity multiple signal detection, the determination of a set of pilot sequences and a CS algorithm for sparse signal recovery are important to provide a good performance for given conditions including channels. In [14], [15], compressive random access is studied for joint activity detection and channel estimation over frequency-selective fading channels.

This work was supported by the GIST Research Institute (GRI) in 2016, and the Global University Project (GUP) in GIST.

Since the notion of CS is applied to compressive random access, CS theory helps to derive good compressive random access schemes. In general, most CS theory focuses on the conditions for recovery guarantee [9], [16], [17]. For example, with the coherence of a given measurement matrix, conditions for perfect recovery can be derived as in [18], [19]. On the other hand, when recovery errors are inevitable due to various reasons (including the presence of measurement errors or background noise) in some applications such as compressive random access, the *probability* of successful recovery would be important and conditions for a low probability of successful recovery might be desirable.

In this paper, we consider compressive random access where CS algorithms can be used for activity detection (or detecting active devices or users). Although it is generally expected to have a small number of active devices, the number of active devices can be greater than the maximum number of active devices that a CS algorithm can detect as the number of active devices is random. Thus, the performance criterion would be based on the probability of successful recovery. In this case, we argue that the coherence may not be a good criterion to determine a set of pilot or preamble sequences for compressive random access. To overcome this difficulty, we consider a criterion based on the average error probability. With this criterion, we expect that the performance of Alltop sequences can be similar to that of Zadoff-Chu (ZC) sequences, which are widely employed (e.g., for LTE-A in [1]).

Notation: Matrices and vectors are denoted by upper- and lower-case boldface letters, respectively. The superscripts T and H denote the transpose and complex conjugate, respectively. The p-norm of a vector $\mathbf{a}$ is denoted by $||\mathbf{a}||_p$ (If $p = 2$, the norm is denoted by $||\mathbf{a}||$ without the subscript). The superscript $\dagger$ denotes the pseudo-inverse. For a vector $\mathbf{a}$, $\mathrm{diag}(\mathbf{a})$ is the diagonal matrix with the diagonal elements from $\mathbf{a}$. For a matrix $\mathbf{X}$ (a vector $\mathbf{a}$), $[\mathbf{X}]_n$ ($[\mathbf{a}]_n$) represents the nth column (element, resp.). If n is a set of indices, $[\mathbf{X}]_n$ is a submatrix of $\mathbf{X}$ obtained by taking the corresponding columns. $\mathbb{E}[\cdot]$ and $\mathrm{Var}(\cdot)$ denote the statistical expectation and variance, respectively. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ ($\mathcal{N}(\mathbf{a}, \mathbf{R})$) represents the distribution of circularly symmetric complex Gaussian (CSCG) (resp., real-valued Gaussian) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

II. System Model

In this section, we consider a system model for a random access scheme that is similar to [1] and discuss pilot design.

Suppose that there is a set of L pre-determined pilots, denoted by $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_L\}$, where $\boldsymbol{\phi}_l$ is assumed to be an $N \times 1$ complex vector of unit norm (i.e., $||\boldsymbol{\phi}_l|| = 1$ for all l). In a cell, there is an access point (AP) and K devices for MTC. The AP can send a beacon signal so that active devices[1] can send pilot (or preamble) signals through a random access channel in order to establish connections. As in [1], we assume that an active device randomly chooses one of the L pilots and transmits it within a pilot slot of N-symbol length. Throughout the paper, we assume that $L > N$.

For convenience, let $\mathcal{A}$ denote the index set of active devices. Then, the received signal sequence or vector during the pilot slot becomes

$$\mathbf{y} = \sum_{k \in \mathcal{A}} \boldsymbol{\phi}_{l(k)} h_k d_k + \mathbf{n}, \tag{1}$$

where $l(k)$ is the pilot index that is chosen by device k, h_k is the (narrowband) channel coefficient from device k to the AP, d_k is the pilot amplitude from device k, and $\mathbf{n} \sim \mathcal{CN}(0, N_0\mathbf{I})$. For convenience, we assume that

$$h_k d_k = A > 0, \ \ k \in \mathcal{A}. \tag{2}$$

The determination of d_k as in (2) at device k is possible if time division duplexing (TDD) mode is assumed as the channel to the AP is the same as that from the AP, which can be estimated when the beacon signal from the AP is transmitted. Note that if $|h_k|$ is small, active device k may not transmit a pilot as the transmission power can be higher than a maximum transmission power. This active device may try to access in a next slot (i.e., this device becomes a back-logged device).

Let $M = |\mathcal{A}|$. From (1), we can also have

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{s} + \mathbf{n}, \tag{3}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1 \ \ldots \ \boldsymbol{\phi}_L] \in \mathbb{C}^{N \times L}$, which is called the measurement matrix, and $\mathbf{s} \in \Sigma_M$ is an M-sparse signal. Here,

$$\Sigma_M = \{\mathbf{x} \,\big|\, ||\mathbf{x}||_0 \leq M\}.$$

If M is sufficiently small, it is possible to recover the sparse signal $\mathbf{s}$ or the support of $\mathbf{s}$ using CS algorithms. The estimation of the support of $\mathbf{s}$ plays a key role in activity detection in compressive random access, because the AP is able to find the pilot sequences that are chosen by active devices from the support of $\mathbf{s}$.

It is noteworthy that the perfect estimation of $\mathbf{s}$ does not imply the perfect activity detection due to the cases that some active devices choose the same pilot in compressive random access, which is called pilot collision. To avoid pilot collision, a number of pilots or a large L is desirable. Unfortunately, as L increases, the performance of most low-complexity CS algorithms (e.g., the orthogonal matching pursuit (OMP) algorithm) is degraded as shown in [15] for compressive random access. Thus, it is important for compressive random access to find a set of a number of pilots (i.e., a large L) that can provide a good performance with a low-complexity CS algorithm. In compressive random access, the throughput is a key performance measure and the throughput depends on the probability of successful recovery (for a large L).

[1] Active devices are the devices that have data to send to the AP.

III. Pilot Design based on Coherence

Suppose that M is sufficiently small such that $M \ll L$ and $M < N$. Then, from [7], [9], [20], to find an approximate solution to (3), we can have the following problem:

$$\min_{\mathbf{s}} ||\mathbf{y} - \boldsymbol{\Phi}\mathbf{s}||^2 + \kappa||\mathbf{s}||_1, \tag{4}$$

where $\kappa > 0$ is the Lagrange multiplier. Although (4) is a convex problem, the complexity to solve this problem can be still high. Thus, various suboptimal but low-complexity approaches have been proposed to find approximate solutions such as greedy algorithms [9, Chp. 8].

The performances of various CS algorithms are generally decided by $\boldsymbol{\Phi}$. In compressive random access, we can find a good set of pilot sequences or $\boldsymbol{\Phi}$ under various criteria. For example, the coherence of $\boldsymbol{\Phi}$ can be used, which is the maximum of the absolute values of the inner products of two distinct columns of $\boldsymbol{\Phi}$, i.e.,

$$\mu(\boldsymbol{\Phi}) = \max_{l \neq m} |\boldsymbol{\phi}_l^{\mathrm{H}} \boldsymbol{\phi}_m|. \tag{5}$$

Compared to the restricted isometric constant (RIC) [6], [17], the coherence is easy to find. Thus, the coherence can be used a practical criterion to determine a set of pilots or $\boldsymbol{\Phi}$.

For a set of pilot sequences, ZC sequences can be considered as in LTE-A system [1]. For example, when N is a prime number, there are $N-1$ ZC root sequences [21]. For each ZC root sequence, we can have N sequences by circular shifting. These sequences are orthogonal to each other, while the correlation between root sequences is $\frac{1}{\sqrt{N}}$. Thus, the coherence of $\boldsymbol{\Phi}$ consisting of ZC sequences is $\frac{1}{\sqrt{N}}$. Note that since there are $N(N-1)$ ZC sequences, L can be up to $N(N-1)$.

We can also consider Alltop sequences [22], [23] with prime N. The coherence is also $\frac{1}{\sqrt{N}}$. Since there are N^2 possible Alltop sequences, L can be up to N^2. Therefore, a set of Alltop sequences can also be a good choice in terms of the coherence with a large number of pilot sequences.

As shown above, both sets of ZC sequences and Alltop sequences could be good candidates to build a pool of pilot sequences due to *i*) low coherence and *ii*) a number of available sequences (or a large L).

IV. PEP Analysis for a New Criterion

In this section, we consider the maximum likelihood (ML) approach to estimate $\mathbf{s}$ and derive an upper-bound on the error probability using the pair-wise error probability (PEP). Based on the PEP analysis, we propose a new criterion that can be used to determine a set of pilots for compressive random access.

The ML estimation of the M-sparse signal vector $\mathbf{s}$ can be given by

$$\hat{\mathbf{s}}_{\rm ml} = \underset{\mathbf{s}\in\Sigma_M}{\operatorname{argmax}} f(\mathbf{y}|\mathbf{s}) = \underset{\mathbf{s}\in\Sigma_M}{\operatorname{argmin}} ||\mathbf{y} - \mathbf{\Phi s}||^2. \quad (6)$$

Although the ML approach is computationally prohibitive, we can consider it[2] to decide the measurement matrix $\mathbf{\Phi}$ that can provide a best performance in terms of the error probability. Since a closed-form expression for the error probability may not be easy to find, we consider an upper-bound.

There are $D = \binom{L}{M}$ combinations of the index sets for the support of $\mathbf{s}$. Denote by $\mathcal{I}_i$ the ith index set and, for notational convenience, let $\mathbf{\Phi}_{(i)} = \mathbf{\Phi}_{\mathcal{I}_i}$. Suppose that $\mathcal{I}_p$ is the correct support of $\mathbf{s}$. Then, from (3) and (6), the PEP that the AP can erroneously decide $\mathcal{I}_q$, $q \neq p$, as the support of $\mathbf{s}$ from $\mathbf{y}$ using the ML approach is given by

$$\begin{aligned}\Pr(\mathbf{\Phi}_{(p)} \to \mathbf{\Phi}_{(q)}) &= \Pr(||\mathbf{y} - \mathbf{\Phi}_{(p)}\mathbf{x}||^2 > ||\mathbf{y} - \mathbf{\Phi}_{(q)}\mathbf{x}||^2) \\ &= \Pr(||\mathbf{n}||^2 > ||(\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)})\mathbf{x} + \mathbf{n}||^2) \\ &= \mathcal{Q}\left(\frac{A}{\sqrt{2N}}||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||\right),\end{aligned}$$

where $\mathbf{x} = \mathbf{1}A$.

Let $\mathcal{D}_{p,q} = (\mathcal{I}_p \setminus \mathcal{I}_q) \cup (\mathcal{I}_q \setminus \mathcal{I}_p)$. If $p \neq q$, the number of elements in $\mathcal{D}_{p,q}$ is an even number that is greater than or equal to 2, i.e., $|\mathcal{D}_{p,q}| = 2m$, where m is a positive integer. Since the sparsity of $\mathbf{s}$ is M, m is upper-bounded by M, i.e., $m \in \{1\ldots, M\}$.

Lemma 1. *Let* $\tilde{\mu}_{l',l''} = |\Re(\boldsymbol{\phi}_{l'}^{\rm H}\boldsymbol{\phi}_{l''})|$, $l', l'' \in \{1, \ldots, L\}$. *If*

$$\tilde{\mu}_{l_1,l_2} \leq \frac{m-1}{m(2m-1)}, \quad (7)$$

for all $l_1 \in \mathcal{I}_p \setminus \mathcal{I}_q$ *and* $l_2 \in \mathcal{I}_q \setminus \mathcal{I}_p$, *we have*

$$||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}|| \geq ||\boldsymbol{\phi}_{l_1} - \boldsymbol{\phi}_{l_2}||^2. \quad (8)$$

Proof: See Appendix A. ∎

According to Lemma 1, we can claim (8) when M is sufficiently small and $\tilde{\mu}_{l_1,l_2}$ is also small. Then, we can have the following approximation:

$$\begin{aligned}P_{\rm err} &\leq \frac{1}{D}\sum_{p=1}^{D}\sum_{q\neq p}\mathcal{Q}\left(\frac{A}{\sqrt{2N_0}}||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||\right) \\ &\approx C\sum_{l_1=1}^{L}\sum_{l_2>l_1}\mathcal{Q}\left(\frac{A}{\sqrt{N_0}}(1-\mu_{l_1,l_2})\right). \quad (9)\end{aligned}$$

Unfortunately, the condition in (7) is not valid for a large m and might be too restrictive. In order to show that (8) can be true for most measurement matrices, we can consider random sequences. In particular, suppose that the following random sequences are used for $\boldsymbol{\phi}_l$'s:

$$[\boldsymbol{\phi}_l]_n = \frac{1}{\sqrt{N}}e^{j\theta_{l,n}}, \quad (10)$$

where $\theta_{l,n} \in [0, 2\pi)$ is i.i.d.

[2]This is analogue to channel coding where the ML performance is considered to design good channel codes.

Lemma 2. *Suppose that* $\boldsymbol{\phi}_l$*'s are generated as* (10). *If* $|\mathcal{D}_{p,q}| = 2m$, *we have*

$$\Pr\left(\frac{||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2}{2m} \leq 1 - 2mt\right) \leq e^{-2Nt^2}, \ t \geq 0. \quad (11)$$

Proof: See Appendix B. ∎

The result in (11) shows that the probability that $||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 \leq 2$, $p \neq q$, is low. Thus, the approximation in (9) might be reasonable.

In (9), since $\mu(\mathbf{\Phi}) \geq \tilde{\mu}_{l_1,l_2}$, any PEP in (9) can be upper-bounded as

$$\Pr(\mathbf{\Phi}_{(p)} \to \mathbf{\Phi}_{(q)}) \leq \mathcal{Q}\left(\frac{A}{\sqrt{N_0}}(1 - \mu(\mathbf{\Phi}))\right). \quad (12)$$

This implies that the minimum coherence can be a reasonable criterion (as mentioned in Section III) to find a good measurement matrix for pilots in terms of the average error probability. Thus, according the coherence, two $\mathbf{\Phi}$'s of the same coherence may provide similar performances. Unfortunately, this is not true as will be shown in Section V.

We now consider a different criterion without relying on the coherence. From (9), using the Chernoff bound, we can have

$$P_{\rm err} \leq C\sum_{l_1=1}^{L}\sum_{l_2>l_1} e^{-\beta(1-\mu_{l_1,l_2})^2}, \quad (13)$$

where $\beta = \frac{A^2}{2N_0}$ and C is constant that does not depend on $\mathbf{\Phi}$ and β. For convenience, define Z as a discrete random variable that has the following probability mass function (pmf):

$$\Pr(Z = (1-\mu_{l_1,l_2})^2) = \frac{2}{L(L-1)}, \ l_1 \in \{1, \ldots, L\}, l_2 > l_1,$$

Then, $\frac{2}{L(L-1)}\sum_{l_1=1}^{L}\sum_{l_2>l_1} e^{-\beta(1-\mu_{l_1,l_2})^2}$ can be obtained by finding the following expectation:

$$\begin{aligned}\mathbb{E}[e^{-\beta Z}] &= e^{-\beta\bar{Z}}\mathbb{E}[e^{-\beta\tilde{Z}}] \\ &\leq e^{-\beta\bar{Z}}\mathbb{E}\left[1 - \beta\tilde{Z} + \frac{(\beta\tilde{Z})^2}{2}\right] \\ &= e^{-\beta\bar{Z}}\left(1 + \beta^2\mathbb{E}[\tilde{Z}^2]\right), \quad (14)\end{aligned}$$

where $\bar{Z} = \mathbb{E}[Z]$ and $\tilde{Z} = Z - \bar{Z}$. In (14), the inequality is due to $e^{-x} \leq 1 - x + \frac{x^2}{2!}$. Then, from (14), we can have the following criterion for $\mathbf{\Phi}$:

$$U_{\rm E}(\mathbf{\Phi}, \beta) = e^{-\beta\mathbb{E}[Z]}\left(1 + \beta^2{\rm Var}(Z)\right). \quad (15)$$

For example, for given $\mathbf{\Phi}'$ and $\mathbf{\Phi}''$, at β, we can say that $\mathbf{\Phi}'$ is better than $\mathbf{\Phi}''$ if

$$U_{\rm E}(\mathbf{\Phi}', \beta) < U_{\rm E}(\mathbf{\Phi}'', \beta).$$

For convenience, the criterion based on (15) is referred to as the PEP based criterion, while that based on the coherence is referred to as the coherence based criterion.

Note that the criterion in (15) is based on the ML approach to estimate $\mathbf{s}$ as in (6). Thus, if other suboptimal approaches are employed, the criterion in (15) may not be useful. However, for a CS algorithm of near ML performance under certain conditions, we may expect that the criterion in (15) can also be valid to choose a good measurement matrix or a good set of pilot sequences.

V. Simulation Results

For simulations, we assume that $N = 139$ and $L = 2N$. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \frac{A^2}{N_0}$. For performance comparisons, we consider $\mathbf{\Phi}$'s consisting of Alltop and ZC sequences. For $\mathbf{\Phi}$ consisting of ZC sequences, since $L = 2N$, we only use two root sequences. Thus, the resulting $\mathbf{\Phi}$, denoted by $\mathbf{\Phi}_{\text{ZC}}$, can be seen as a concatenation of two $N \times N$ unitary matrices. With Alltop sequences, we consider two different $\mathbf{\Phi}$'s. The first $\mathbf{\Phi}$ is obtained by randomly choosing Alltop sequences. This is called Alltop based $\mathbf{\Phi}$ with random columns and denoted by $\mathbf{\Phi}_{\text{A},1}$. The second $\mathbf{\Phi}$ is obtained as the ZC based $\mathbf{\Phi}$. Thus, this $\mathbf{\Phi}$, denoted by $\mathbf{\Phi}_{\text{A},2}$, can also be seen as a concatenation of two $N \times N$ unitary matrices. In addition, $\mathbf{\Phi}$ whose elements are complex random binary ($[\mathbf{\Phi}]_{n,l} \in \{\frac{1}{\sqrt{2N}}(\pm 1 \pm j)\}$) is considered, which is denoted by $\mathbf{\Phi}_{\text{R}}$.

In Table I, we show the values of the coherence and $U_{\text{E}}(\mathbf{\Phi}, \beta)$ of the four different $\mathbf{\Phi}$'s when the SNR is set to 14 dB. It is shown that $\mathbf{\Phi}_{\text{R}}$ has the highest coherence and the highest value of $U_{\text{E}}(\mathbf{\Phi}, \beta)$. From this, we can expect that the performance of $\mathbf{\Phi}_{\text{R}}$ might be the worst. According to the coherence, the performances of $\mathbf{\Phi}_{\text{A},1}$ and $\mathbf{\Phi}_{\text{A},2}$ may not be too different. However, according to the PEP criterion (or $U_{\text{E}}(\mathbf{\Phi}, \beta)$), we may expect that $\mathbf{\Phi}_{\text{A},2}$ performs much better than that of $\mathbf{\Phi}_{\text{A},1}$, while the performances of $\mathbf{\Phi}_{\text{A},2}$ and $\mathbf{\Phi}_{\text{ZC}}$ are similar.

TABLE I
The coherence and PEP based criteria for different measurement matrices with $(N, L) = (139, 278)$ when SNR = 14 dB.

Seq.	$\mathbf{\Phi}_{\text{R}}$	$\mathbf{\Phi}_{\text{A},1}$	$\mathbf{\Phi}_{\text{ZC}}$ and $\mathbf{\Phi}_{\text{A},2}$
PEP ($\times 10^{-4}$)	0.3658	0.2723	0.0351
Coherence	0.2802	0.0848	0.0848

As a suboptimal CS approach, we consider the OMP algorithm for activity detection by recovering the sparse signal, s, in (3). With 1000 runs for each[3] $\mathbf{\Phi}$, successful recovery rates are found. The results are shown in Fig. 1. As expected, $\mathbf{\Phi}_{\text{A},2}$ outperforms $\mathbf{\Phi}_{\text{A},1}$ although they have the same coherence. This demonstrates that the criterion in (15) could be used to choose the better measurement matrix among the measurement matrices that have the same coherence.

Fig. 2 shows the successful recovery rates for various values of SNR when $(N, L) = (139, 278)$ and $M = 24$. We can also confirm that $\mathbf{\Phi}_{\text{A},2}$ and $\mathbf{\Phi}_{\text{ZC}}$ can perform better than the others for a wide range of SNR.

[3]For complex random binary $\mathbf{\Phi}$, we consider 100 different realizations of $\mathbf{\Phi}$ and for each realization, 10 runs are used to obtain the successful recovery rate.

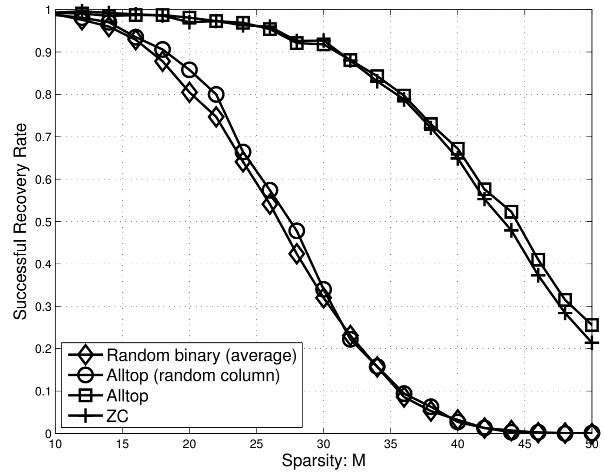

Fig. 1. The performances of successful recovery of 4 different measurement matrices for various values of M with $(N, L) = (139, 278)$ and SNR = 14 dB.

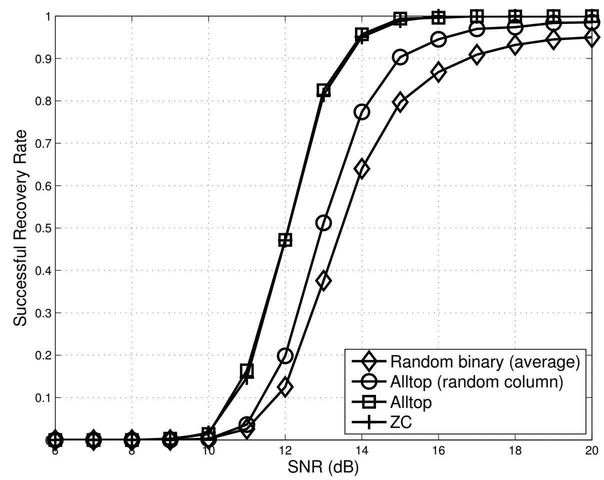

Fig. 2. The performances of successful recovery of 4 different measurement matrices for various values of SNR with $(N, L) = (139, 278)$ and $M = 24$.

VI. Conclusions

In this paper, we argued that the coherence may not be effective as a criterion in determining the set of pilots for compressive random access. To replace the coherence, we proposed a criterion based on the probability of successful recovery that can be used to choose a set of pilots. Based on this criterion, we found that the performance of Alltop sequences can be similar to that of ZC sequences, which are adopted in LTE-A system. Simulation results confirmed this as the successful recovery rate of compressive random access with Alltop sequences is similar to that with ZC sequences for a wide range of sparsity and SNR.

Appendix A
Proof of Lemma 1

Suppose that $m = 1$ or $\mathcal{D}_{p,q} = \{l_1, l_2\}$. Then, we have

$$
\begin{aligned}
||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 &= ||\boldsymbol{\phi}_{l_1} - \boldsymbol{\phi}_{l_2}||^2 \\
&= ||\boldsymbol{\phi}_{l_1}||^2 + ||\boldsymbol{\phi}_{l_2}||^2 - 2\Re\left(\boldsymbol{\phi}_{l_1}^{\mathrm{H}} \boldsymbol{\phi}_{l_2}\right),
\end{aligned}
$$

which results in

$$
2(1 - \tilde{\mu}_{l_1,l_2}) \leq ||\boldsymbol{\phi}_{l_1} - \boldsymbol{\phi}_{l_2}||^2 = ||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 \leq 2. \quad (16)
$$

Now suppose that $m = 2$ or $\mathcal{I}_p \setminus \mathcal{I}_q = \{l_{a(1)}, l_{a(2)}\}$ and $\mathcal{I}_q \setminus \mathcal{I}_p = \{l_{b(1)}, l_{b(2)}\}$, where $a(1)$ and $a(2)$ are two distinct indices and $b(1) \neq b(2)$ are also two distinct indices. In this case, we have

$$
\begin{aligned}
||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 &= ||\boldsymbol{\phi}_{l_{a(1)}} + \boldsymbol{\phi}_{l_{a(2)}} - \boldsymbol{\phi}_{l_{b(1)}} - \boldsymbol{\phi}_{l_{b(2)}}||^2 \\
&= 4 - 2\big(\Re(\boldsymbol{\phi}_{l_{a(1)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{a(2)}}) + \Re(\boldsymbol{\phi}_{l_{b(1)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{b(2)}}) \\
&\quad - \Re(\boldsymbol{\phi}_{l_{a(1)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{b(1)}}) - \Re(\boldsymbol{\phi}_{l_{a(1)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{b(2)}}) \\
&\quad - \Re(\boldsymbol{\phi}_{l_{a(2)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{b(1)}}) - \Re(\boldsymbol{\phi}_{l_{a(2)}}^{\mathrm{H}} \boldsymbol{\phi}_{l_{b(2)}})\big) \\
&\geq 4 - 2\big(\tilde{\mu}_{a(1),a(2)} + \tilde{\mu}_{b(1),b(2)} + \tilde{\mu}_{a(1),b(1)} \\
&\quad + \tilde{\mu}_{a(1),b(2)} + \tilde{\mu}_{a(2),b(1)} + \tilde{\mu}_{a(2),b(2)}\big). (17)
\end{aligned}
$$

It follows that

$$
||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 \geq 4 - 2\binom{4}{2} \max \tilde{\mu},
$$

where $\max \tilde{\mu}$ is the maximum of all possible $\tilde{\mu}_{l_1,l_2}$'s, $l_1 \neq l_2$. Then, in order to satisfy the following inequality:

$$
||\boldsymbol{\phi}_{l_{a(1)}} + \boldsymbol{\phi}_{l_{a(2)}}||^2 \leq ||\boldsymbol{\phi}_{l_{a(1)}} + \boldsymbol{\phi}_{l_{a(2)}} - \boldsymbol{\phi}_{l_{b(1)}} - \boldsymbol{\phi}_{l_{b(2)}}||^2,
$$

from (16) and (17), we can have the following condition:

$$
2 \leq 4 - 2\binom{4}{2} \max \tilde{\mu}.
$$

From the above results, for any $m \geq 1$, a sufficient condition for (8), we can have $2 \leq 2m - 2\binom{2m}{2} \max \tilde{\mu}$ or

$$
\max \tilde{\mu} \leq \frac{m-1}{\binom{2m}{2}} = \frac{m-1}{m(2m-1)}, \ m = 1, \ldots, M. \quad (18)
$$

From (18), we can find the condition in (7).

Appendix B
Proof of Lemma 2

We only provide a sketch of the proof. Let

$$
X_n = |[\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}]_n|^2. \quad (19)
$$

From (10), if $|\mathcal{D}_{p,q}| = 2m$, we have $\mathbb{E}[X_n] = \frac{2m}{N}$. In addition, we have $X_n \leq \frac{(2m)^2}{N}$ w.p. 1. Thus, X_n is a bounded random variable and $||\mathbf{\Phi}_{(p)} - \mathbf{\Phi}_{(q)}||^2 = \sum_{n=1}^{N} X_n$ is a sum of bounded independent random variables. From this, applying Hoeffding's inequality [24], we can readily obtain (11).

References

[1] 3GPP TR 37.868 V11.0, *Study on RAN improvments for machine-type communications*, October 2011.

[2] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Communications Magazine*, vol. 51, pp. 86–93, June 2013.

[3] M. Condoluci, M. Dohler, G. Araniti, A. Molinaro, and K. Zheng, "Toward 5G densenets: architectural advances for effective machine-type communications over femtocells," *IEEE Communications Magazine*, vol. 53, pp. 134–141, January 2015.

[4] K.-D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, pp. 373–377, Dec 2011.

[5] D. Niyato, P. Wang, and D. I. Kim, "Performance modeling and analysis of heterogeneous machine type communications," *IEEE Trans. Wireless Communications*, vol. 13, pp. 2836–2849, May 2014.

[6] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4203–4215, Dec 2005.

[7] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, April 2006.

[8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[9] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[10] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," *CoRR*, vol. abs/0903.1022, 2009.

[11] H. Zhu and G. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Communications*, vol. 59, pp. 454–465, February 2011.

[12] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Physical Communication*, vol. 5, no. 2, pp. 129–147, 2012.

[13] J.-P. Hong, W. Choi, and B. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Trans. Wireless Communications*, vol. 14, pp. 998–1010, Feb 2015.

[14] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *Proc. ISWCS 2013*, pp. 1–5, Aug 2013.

[15] J. Choi, "On the sparsity for random access in machine type communications under frequency-selective fading," in *Proc. IEEE ICC*, pp. 1–5, May 2016.

[16] E. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[17] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[18] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2231–2242, Oct 2004.

[19] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, pp. 6–18, Jan 2006.

[20] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, Feb 2006.

[21] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Trans. Inform. Theory*, vol. 18, pp. 531–532, Jul 1972.

[22] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Processing*, vol. 57, pp. 2275–2284, June 2009.

[23] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.

[24] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.