Compressive Random Access Using Multiple Resource Blocks for MTC

Jinho Choi, Kyungjun Lee, and Nam Yul Yu
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
e-mail: {jchoi0114, leekj9001, nyyu}@gist.ac.kr

***Abstract*—In this paper, we propose a compressive random access scheme using multiple resource blocks (RBs) to support massive connections for machine type communications (MTC) in 5G systems. The main advantage of the proposed scheme over conventional compressive random access schemes is that there can be more devices than preambles as each active device is to randomly choose a preamble for connection to an access point (AP) at the cost of preamble collisions. To mitigate the performance degradation due to preamble collisions, a block of subcarriers is divided into multiple RBs and an active device is to randomly choose an RB and a preamble within the chosen RB. For multiple signal detection, we consider a low-complexity compressive sensing (CS) algorithm. We also demonstrate that using multiple RBs in random access can offer the benefits of reducing the computation time with a multicore processor as well as improving the throughput.**

***Index Terms*—Internet of things, random access, machine type communications**

I. Introduction

For the Internet of Things (IoT), there has been a growing interest in machine-type communications (MTC) in 5th generation (5G) systems [1]. The applications of MTC are diverse from health care to smart grid, where a huge number of devices exist in the system, but only a few of them are active at a particular timing instance. Therefore, random access is suitable for MTC to accommodate a number of devices with a low probability of activity [1]–[3]. A random access scheme, called the random access (RACH) procedure, has also been proposed for a cellular system, i.e., the long term evolution-advanced (LTE-A) system [4]. The RACH procedure is a contention-based random access method, which is similar to the slotted ALOHA protocol [5]. In the RACH procedure, there are multiple preambles in a preamble pool, and a device can select a preamble randomly from the pool and transmit the selected one for access. Finally, it can be connected if there is no collision, i.e., the preamble is not transmitted by any other devices.

Recently, the notion of compressive sensing (CS) [6]–[8] is employed to exploit the sparsity of active devices for multiuser detection (MUD) in random access [9]–[11]. Among many devices present in MTC, only a few of them attempt to access the network by transmitting their preambles. The sparse activity of devices, which can be modeled as a sparse vector, allows the principle of CS to be effectively applied to MUD with low complexity.

In most conventional compressive random access[1] schemes [9]–[11], each device has a unique signature sequence that is used to distinguish its signal from other signals from different devices. Consequently, the number of devices is limited by the number of signature sequences. To support many devices in MTC, it can result in two difficulties: *i)* a long sequence is required to accommodate a number of devices regardless of sparse activity (a low throughput); *ii)* the complexity of CS algorithms for MUD can be high due to a large number of columns (or signature vectors) in a measurement matrix (a high computational complexity). To avoid the above difficulties, we can apply the approach used in the RACH procedure to compressive random access. That is, instead of assigning a unique signature to each device, a randomly selected preamble or pilot from a pool of predetermined preambles can be used when an active device is to transmit signals to a receiver or access point (AP). This approach is studied in [12]. Unfortunately, this approach suffers from preamble collision that happens when multiple active devices choose the same preamble.

In this paper, we generalize the compressive random access scheme in [12] using multiple resource blocks (RBs) in a multicarrier system. In the proposed scheme, we divide subcarriers into multiple groups, or RBs, not only to mitigate preamble collisions, but also to reduce the computational complexity of sparse signal recovery when there are a large number of devices from MTC. A device can randomly choose one of multiple RBs and then a preamble within the chosen RB for connection by random access. At the AP, parallel multiple CS-based detectors can be employed for MUD with low complexity. Both theoretical analysis and simulation results show that the throughput of compressive random access increases with the number of RBs within the capability of CS recovery for a fixed total number of subcarriers. In addition, through the parallel detection with a graphics processing unit (GPU) multicore processor, we demonstrate that the complexity of MUD decreases with the number of RBs.

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

[1]This term is used in [11] to refer to a random access scheme that allows a receiver to employ low-complexity CS algorithms for MUD.

Notation: 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). $\mathbb{E}[\cdot]$ and $\text{Var}(\cdot)$ denote the statistical expectation and variance, respectively. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

II. System Model

Suppose that there are K devices and each device becomes active with probability p_a, where p_a is the access probability. Each active device can choose one of M RBs for random access. In each RB, there are N preambles of length L. Throughout the paper, we assume a multicarrier system that has $J = LM$ subcarriers. Thus, an RB consists of L subcarriers and each preamble is a multi-carrier spread sequence of length L [13]. Furthermore, we consider the case that J is fixed in order to see the impact of the number of RBs on the performance and complexity (i.e., if the number of RBs increases, the number of subcarriers per RB decreases).

Let $\mathbf{C} = [\mathbf{c}_0 \ \ldots \mathbf{c}_{N-1}]$ denote a matrix of N preamble vectors where $\mathbf{c}_n$ denotes the nth preamble, which is a multi-carrier spread sequence of length L. Thus, the size of $\mathbf{C}$ becomes $L \times N$, and in general $L \leq N$, which implies that the preamble sequences can be correlated. We assume the same set of preamble sequences for all RBs. At an AP, the received signal vector over the mth RB is given by

$$\mathbf{y}_m = [y_{m,0} \ \ldots \ y_{m,L-1}]^{\mathrm{T}} = \mathbf{C}\mathbf{s}_m + \mathbf{n}_m, \qquad (1)$$

where $y_{m,l}$ represents the received signal through the lth subcarrier of the mth RB, $\mathbf{s}_m$ is a random access vector, and $\mathbf{n}_m = [n_{m,0} \ \ldots \ n_{m,L-1}]^{\mathrm{T}} \sim \mathcal{CN}(0, N_0\mathbf{I})$ is the background noise. Let $\mathcal{I}_{m,n}$ denote the index set of the active devices that choose the mth RB and the nth preamble. Then, the nth element of $\mathbf{s}_m$ is given by

$$[\mathbf{s}_m]_n = \sum_{k \in \mathcal{I}_{m,n}} h_{k,m} x_k, \qquad (2)$$

where $h_{k,m}$ denotes the (frequency-domain) channel coefficient from device k to the AP over the mth RB and x_k represents the signal from device k. In (2), we assume that the bandwidth of an RB is sufficiently narrow so that the frequency-domain channel gain remains unchanged within the bandwidth of an RB. The total number of active devices is given by $K_A = \sum_{m=1}^{M} \sum_{n=1}^{N} |\mathcal{I}_{m,n}|$ and the number of active devices that choose the mth RB is given by $K_A(m) = \sum_{n=1}^{N} |\mathcal{I}_{m,n}|$. Thus, we have $K_A = \sum_{m=1}^{M} K_A(m)$. In general, we expect to have $K_A(m) \ll N$ due to a low probability of activity.

The random access system with multiple RBs (i.e., $M > 1$) is referred to as the *multi-RB random access system* in this paper. The advantage of the multi-RB random access system over the random access system of a single big RB (of J subcarriers) is two-fold: *i)* the number of active devices per RB can be low, which helps not only to decrease the probability of preamble collision, but also to exploit the sparsity for low-complexity signal detection; *ii)* the AP can run M multi-signal detectors in parallel when it is equipped with a multicore processor (e.g., GPU) for a higher processing throughput.

III. Sparsity Exploiting Multi-Signal Detection

A. Impact of M on Probability of Preamble Collisions

The average number of the active devices that choose the mth RB is $\frac{Kp_a}{M}$. Thus, the device activity of each RB decreases with M. The probability of preamble collision in the mth RB becomes

$$P_{\mathrm{pc}}(m) = 1 - \prod_{k=1}^{K_A(m)} \left(1 - \frac{k}{N}\right) \approx 1 - e^{\frac{(K_A(m))^2}{2N}}.$$

Noting that $1 - e^{-\frac{x^2}{2N}}$ is concave in x and $\mathbb{E}[K_A(m)] = \frac{Kp_a}{M}$, we have

$$\mathbb{E}[P_{\mathrm{pc}}(m)] \approx \mathbb{E}[1 - e^{\frac{(K_A(m))^2}{2N}}] \leq 1 - e^{\frac{(Kp_a)^2}{2NM^2}}.$$

From this, we can see that the average probability of preamble collision (per RB) can decrease with M (assuming that NM is proportional to J). This demonstrates that in compressive random access with a finite set of preambles, the problem of preamble collisions can be mitigated using multiple RBs. We also study the impact of multiple RBs on the throughput in Section IV.

B. CS Algorithms for MUD

In (1), $\mathbf{C}$ is considered a measurement matrix of CS. Thus, for each RB, we can use a CS algorithm to detect the preambles transmitted by active devices that choose the RB. For MUD in each RB, we can consider the following problem:

$$\min ||\mathbf{s}_m||_p \text{ subject to } ||\mathbf{y}_m - \mathbf{C}\mathbf{s}_m||_2^2 \leq \epsilon,$$

where ϵ is given by the background noise. While l_0-norm minimization ($p = 0$) is desirable to estimate the sparse signal vector $\mathbf{s}_m$, it is known to be computationally intractable. Instead, l_1-minimization ($p = 1$, i.e., basis pursuit) is preferable due to the computational tractability [8], [14]. There are also various low-complexity greedy algorithms [8, Ch. 8] to estimate the sparse vector, $\mathbf{s}_m$, e.g., the orthogonal matching pursuit (OMP) [15], CoSaMP [16], and so on.

In CS theory, it is well known that if $\mathbf{C}$ satisfies the *restricted isometry property (RIP)*, stable and robust reconstruction of s-sparse signals can be guaranteed [14]. In particular, if the $L \times N$ matrix $\mathbf{C}$ is a random matrix with Gaussian or Bernoulli distributed entries, $\mathbf{C}$ obeys the RIP with high probability as long as $L \geq O(s \log(N/s))$ [14]. Moreover, the OMP algorithm presents the recovery guarantee for $\mathbf{C}$ having the RIP [17], under the complexity of $O(sLN)$ [18]. Throughout this paper, we use a random Gaussian matrix for $\mathbf{C}$ by assuming that the entries of each preamble are randomly generated by the Gaussian distribution. Also, the OMP algorithm is employed in CS recovery to allow multi-signal detection with low complexity for compressive random access.

There is another advantage of multiple RBs in terms of power control provided that each device is able to know its channel gains. In (2), we need to adjust the amplitude of x_k to overcome fading if the channel gain, $|h_{k,m}|$, is low. In general, each device can choose the amplitude of x_k to be inversely proportional to the channel gain to equalize the overall signal gain. In this case, if $|h_{k,m}|$ is low, the transmit power has to be high, which is not desirable. In the multi-RB random access system, this problem can be mitigated if an active device can choose the RB that has the highest channel gain. That is, active device k can choose the RB as follows: $m(k) = \mathrm{argmax}_m |h_{k,m}|$, where $m(k)$ denotes the index of the selected RB by active device k. Since the channel gains are random, the selection of RBs by active devices becomes also random and active devices could be uniformly distributed over multiple RBs in this case as well (provided that $h_{k,1}, \ldots, h_{k,M}$ are iid). From this, throughout the paper, we assume that the power control is adopted to equalize the overall signal gain as follows:

$$|h_{k,m(k)} x_k| = A > 0, \text{ for all active device } k.$$

C. Complexity

A GPU has a number of processor cores called streaming processors (SPs), and multiple SPs consist of a streaming multi-processor (SM) [19]. Since a GPU consists of a number of SMs, multiple tasks can be performed simultaneously.

For fast multi-signal detection, we consider to process multiple RBs in parallel using a GPU. By assigning an RB to a single core, multi-signal detection can be performed in each core individually and simultaneously. Thus, parallel processing using M cores can perform faster than the serial processing of central processing unit (CPU). Specifically, if we assume $L = \frac{J}{M}$ for fixed J and $N = cL$ for constant $c > 1$, the OMP algorithm for the mth RB has the complexity of $O(K_A(m)LN) = O\left(\frac{K_A(m)J^2}{M^2}\right)$. As M RBs are processed in parallel by a GPU with M cores, the complexity of OMP-based multi-signal detection remains as $O\left(\frac{K_A(m)J^2}{M^2}\right)$. Therefore, as M increases, the complexity of OMP-based multi-signal detection will be reduced *quadratically* by GPU parallel processing. In Section V, simulation results will be presented for multi-signal detection by a GPU parallel processing.

IV. Throughput Analysis

While compressive random access with a finite set of preambles is more suitable to support massive connections in MTC than conventional compressive random access schemes in [10], [11], it suffers from preamble collisions as mentioned earlier. To mitigate this problem, we considered multiple RBs and showed that the probability of preamble collision can decrease with M. In this section, we consider the throughput analysis to see the impact of M on the throughput in detail.

When a CS algorithm is used for MUD in the multi-RB random access system, there can be two different error events. The first error event, called *sparsity outage event*, occurs when the number of active devices is more than the maximum sparsity for each RB, which is denoted by D. Note that the maximum sparsity D means the maximum number of preambles that can be successfully recovered by CS detector. Clearly, D depends on a number of parameters including L and N, as well as the CS algorithm employed at the receiver. If $K_A(m) > D$, the CS algorithm cannot perform well and the AP may fail to detect the signals from the active devices that choose the mth RB. The second error event is a conventional error event that is due to collisions. Within an RB, if a preamble is chosen by multiple active devices, there might be preamble collisions. In this section, we consider the two error events to derive the average throughput.

Recall that $K_A(m)$ is the number of active devices that choose the mth RB. Then, $K_A(m)$ is a binomial random variable where the probability that $K_A(m) = d$ is given by

$$\Pr(K_A(m) = d) = \binom{K}{d} \left(\frac{p_a}{M}\right)^d \left(1 - \frac{p_a}{M}\right)^{K-d}.$$

Thus, the probability of the sparsity-outage event at the mth RB becomes

$$\begin{aligned} P_{1,m} &= \Pr(K_A(m) \geq D+1) \\ &= \sum_{d=D+1}^{K} \binom{K}{d} \left(\frac{p_a}{M}\right)^d \left(1 - \frac{p_a}{M}\right)^{K-d}. \end{aligned}$$

When $K_A(m)$ ($\leq D$) active devices attempt to access via the mth RB, the conditional throughput[2] per RB is given by

$$T(K_A(m)) = \frac{K_A(m)}{L} \left(1 - \frac{1}{N}\right)^{K_A(m)-1},$$

which is the expected number of the active devices that can successfully transmit preambles without collisions in the mth RB. Note that if $K_A(m) \geq D+1$, the sparsity-outage event happens. Thus, we can simplify as follows:

$$T(K_A(m)) = 0, \; K_A(m) \geq D+1.$$

The average throughput of the mth RB becomes

$$\begin{aligned} \bar{T}_m &= \mathbb{E}[T(K_A(m))] \\ &= \sum_{d=0}^{K} T(d) \Pr(K_A(m) = d) \\ &= \sum_{d=0}^{D} T(d) \Pr(K_A(m) = d) \\ &= \sum_{d=0}^{D} \frac{d}{L} \binom{K}{d} \left(1 - \frac{1}{N}\right)^{d-1} \left(\frac{p_a}{M}\right)^d \left(1 - \frac{p_a}{M}\right)^{K-d} \\ &= \frac{p_a K}{LM} \sum_{d=0}^{D-1} \binom{K-1}{d} \left(\frac{(N-1)p_a}{NM}\right)^d \left(1 - \frac{p_a}{M}\right)^{K-1-d}. \end{aligned}$$

In order to find an approximate throughput, we can consider the Poisson approximation of $\Pr(K_A(m) = d)$, which is given by

$$\Pr(K_A(m) = d) \approx \frac{\lambda^d e^{-\lambda}}{d!},$$

[2]The average number of successfully received preambles over L subcarriers.

where $\lambda = \frac{p_a K}{M}$. Then, it can be shown that

$$\bar{T}_m \approx \frac{\lambda e^{-\lambda}}{L} \sum_{d=0}^{D-1} \frac{\left(\lambda\left(1-\frac{1}{N}\right)\right)^d}{d!}$$

and an approximation of the total throughput is given by

$$\begin{aligned}\bar{T}(L,M,N) &= M\bar{T}_m \\ &\approx \frac{p_a K e^{-\frac{p_a K}{M}}}{L} \sum_{d=0}^{D-1} \frac{\left(\frac{p_a K}{M}\left(1-\frac{1}{N}\right)\right)^d}{d!}.\end{aligned}$$

Note that D is dependent on L, M, and N. Since $L = J/M$ and $\sum_{d=0}^{D-1} \frac{\left(\frac{p_a K}{M}\left(1-\frac{1}{N}\right)\right)^d}{d!} \le e^{\frac{p_a K (1-\frac{1}{N})}{M}}$, which might be tight if D is sufficiently large or $\frac{p_a K}{M}$ is sufficiently small, we can see that

$$\bar{T}(L,M,N) \le \frac{M p_a K e^{-\frac{p_a K}{M}}}{J} e^{\frac{p_a K (1-\frac{1}{N})}{M}} \approx \frac{p_a K}{J} M - \frac{p_a^2 K^2}{NJ}$$

if N is sufficiently large. In particular, if $N = cL$ for $c > 1$,

$$\begin{aligned}\bar{T}(L,M,N) &\lesssim \frac{p_a K}{J} M - \frac{p_a^2 K^2}{cLJ} \\ &= \left(1 - \frac{p_a K}{cJ}\right) \frac{p_a K}{J} M \qquad (3)\end{aligned}$$

which implies that the average throughput can increase with M under certain conditions (e.g., a low p_a and a large N). Since $L \ge 2D$ is required for unique D-sparse solution by CS recovery, it is valid to claim that the average throughput increases with M as long as $M \le \frac{J}{2D}$ for fixed J, where D is determined by a CS recovery algorithm for given L and N.

V. Simulation Results

For simulations, the total numbers of devices and subcarriers are set to $K = 2048$ and $J = 1024$, respectively. Also, we assume that the number of preambles per RB is twice the length of preambles, i.e., $N = 2L$. Then, in this case, the approximate total throughput of (3) becomes

$$\bar{T}(L,M,N) \approx 2(1-p_a)p_a M.$$

Each preamble is randomly generated by the Gaussian distribution, so the preamble matrix $\mathbf{C}$ is a random Gaussian matrix. At the receiver, we assume noiseless CS detection, where the OMP algorithm has been applied for MUD. By assuming that the OMP algorithm has no prior knowledge of sparsity, we stop the iteration if the l_2-norm of the residual vector is less than 10^{-5}. We assume that the preamble transmitted by an active device is successfully recovered if the OMP algorithm can recover its index.

Figures 1 and 2 show the total throughput of multi-signal detection for different values of the number of RBs with $p_a = 0.03$ and $p_a = 0.1$, respectively. The figures demonstrate that the total throughput of multi-signal detection increases with the number of RBs, as the theoretical analysis predicted in (3). The simulation results, along with the performance analysis, support our idea that the increase of RBs can result in a higher throughput when the total number of subcarriers is fixed.

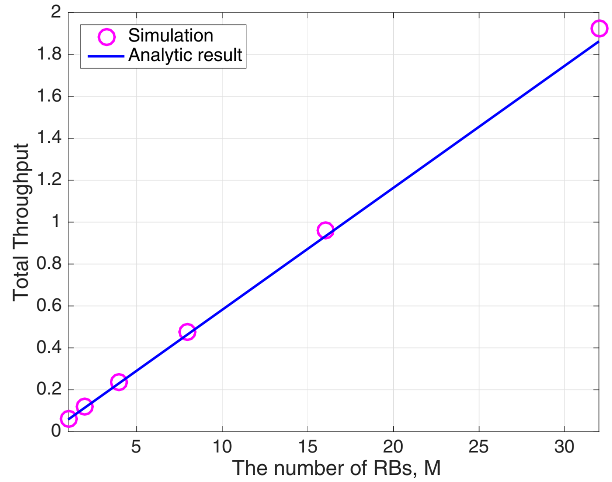

Fig. 1. Total throughput versus the number of RBs with $p_a = 0.03$

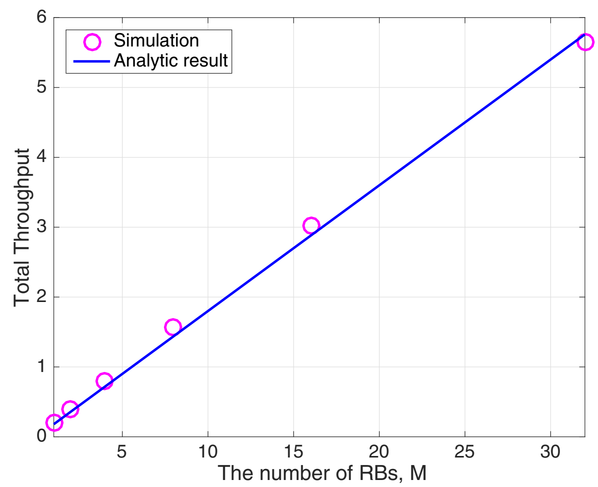

Fig. 2. Total throughput versus the number of RBs with $p_a = 0.1$

In Figure 3, we illustrate the total throughput over access probability p_a, where the number of RBs is set to $M = 32$. We can see that the throughput increases with p_a. According to (3), we can see that the throughput can increase with p_a when $p_a \in (0, \frac{cJ}{2K}]$. In this case, since $\frac{cJ}{2K} = \frac{1}{2}$, the increase of p_a results in the increase of the total throughput upto $p_a \le 0.1$.

Figure 4 sketches the computation time of multi-signal detection using a GPU (in particular, NVIDIA K20c GPU board is used). A GPU code launches M cores to process multi-signal detection. In addition, GPU code includes the parallel matrix-vector multiplications and matrix inversion in one RB. In this paper, we used the optimized *cuBLAS* library [20] by *NVIDIA* for parallel matrix and vector computations. As shown in Figure 4, when the number of active devices per RB is $K_A(m) = 3$, the computation time of multi-signal detection decreases quadratically with the number of RBs. Note that, as discussed in Section III.B, the actual computation time of GPU parallel processing is well approximated by the

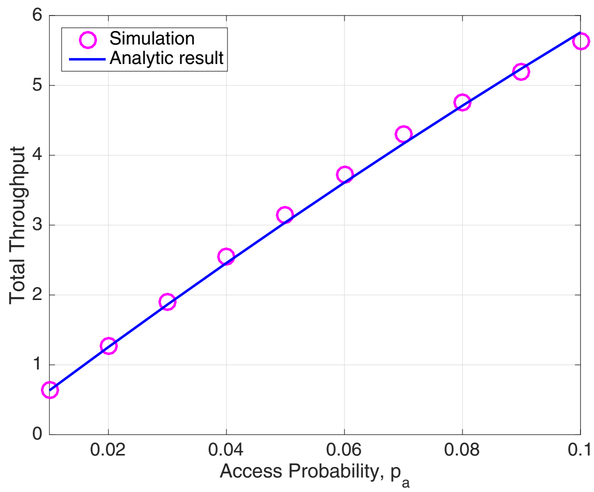

Fig. 3. Total throughput versus access probability with $M = 32$

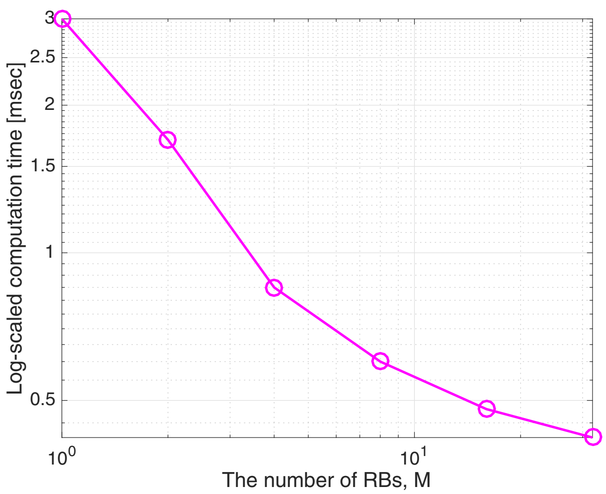

Fig. 4. Computation time over the number of RBs with $K_A(m) = 3$.

complexity of OMP, or $O\left(\frac{K_A(m)J^2}{M^2}\right)$. For $N = 2L$ and $L = \frac{J}{M}$ with a fixed J, it is obvious that the computation time decreases quadratically with M, since the size of the each RB, $L \times N$, decreases quadratically with M.

Consequently, we can observe from simulation results that using multiple RBs, compressive random access enjoys high throughput and low computation time with multicore GPU processing.

VI. Concluding Remarks

In this paper, we have proposed a compressive random access scheme using multiple RBs for MTC. The proposed random access divides subcarriers into multiple RBs to support massive connections using a low-complexity receiver at an AP. For random access to support a number of devices in MTC, a device randomly chooses one of multiple RBs and then a preamble within the chosen RB. In this proposed approach, since active devices can be uniformly distributed over multiple RBs, we can achieve not only a better performance (by mitigating preamble collisions), but also a lower complexity in MUD. Furthermore, at the AP, parallel multiple CS-based detectors can be employed for MUD with low complexity. Through theoretical analysis and simulations, we have shown that the throughput of the proposed compressive random access increases with the number of RBs. By conducting the parallel detection with a GPU multicore processor, we have also demonstrated that using multiple RBs in random access can offer the benefits of reducing the computation time as well as improving the throughput.

References

[1] 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.

[2] 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.

[3] 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.

[4] 3GPP TR 37.868 V11.0, *Study on RAN improvments for machine-type communications*, October 2011.

[5] B. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs: Prentice-Hall, 1987.

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

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

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

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

[10] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *Proc. International Symposium on Wireless Communication Systems (ISWCS 2013)*, pp. 1–5, Aug. 2013.

[11] G. Wunder, C. Stefanovic, P. Popovski, and L. Thiele, "Compressive coded random access for massive MTC traffic in 5G systems," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, pp. 13–17, Nov. 2015.

[12] J. Choi, "On the sparsity for random access in machine type communications under frequency-selective fading," in *2016 IEEE International Conference on Communications Workshops (ICC)*, pp. 517–521, May 2016.

[13] K. Fazel and S. Kaiser, *Multi-Carrier and Spread Spectrum Systems*. John Wiley & Sons, 2003.

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

[15] G. M. Davis, S. G. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Optical Engineering*, vol. 33, no. 7, pp. 2183–2191, 1994.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2009.

[17] L.-H. Chang and J.-Y. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, pp. 5702–5715, Sep. 2014.

[18] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4655–4666, Dec. 2007.

[19] NVIDIA Corporation, *CUDA C Programming Guide*. 2015.

[20] NVIDIA Corporation, *cuBLAS Library User Guide*. 2015.