IEEE International Conference on Communications 2014
Workshop on Massive Uncoordinated Access Protocols

Compressive Random Access for Post-LTE Systems

GERHARD WUNDER[1,2], PETER JUNG[1,2], CHEN WANG[1,2]

[1]Fraunhofer Heinrich Hertz Institute (HHI), Berlin, Germany
[2]Technische Universität Berlin, Germany

July 10, 2014

Abstract

We introduce a random access procedure where control and data information is transmitted in the same "access" slot. The key idea is data-overlayed control signalling together with a dedicated frequency area for compressive measurements exploiting sparse channel profiles and, potentially, sparse user activity. This architecture is resource-efficent since otherwise pilots have to be suitably placed in the time-frequency grid for every potential user. We analyze the achievable rates depending on the key design parameters and show by simulations that sparse signal processing algorithms are indeed "strong" enough to retrieve the information symbols out of the induced noise. Moreover, for the very high dimensional receive space applied in this paper, the number of detected users is only limited by the sheer complexity rather than performance.

5GNOW is a European research project supported by the European commission within FP7 ICT Call 8.
www.5gnow.eu

I. INTRODUCTION

5G networks will have to accomodate traffic classes with very diverse (virtually contradicting) requirements in the uplink [1], [2] ranging from super-high rate video streaming, low latency real-time control applications, and, particularly, sporadic traffic with efficient (in other words "fast") and scalable random access. Sporadic traffic generating devices (e.g. machine–type communication (MTC) in the *Internet of Things* (IoT)) are most of the time inactive but regularly access the Internet for minor/incremental updates with no human interaction. Interestingly, smartphones apps show a similar behaviour (e.g. weather forecasts, stock prices, navigation position, location-dependent context information etc.) resulting in significant control signaling growth and network congestion threat.

Sporadic traffic will dramatically increase in the 5G market and, obviously, cannot be handled with the bulky 4G random access procedures. Two major challenges must be addressed: 1) unprecedent number of devices asynchroneously access the network over a limited resource and 2) the same resource carries control signalling *and* payload. Dimensioning the channel according to classical theory results in a severe waste of resources which, even worse, does not scale towards the requirements of the IoT. On the other hand, since typically user activity, channel profiles and message sizes are compressible within a very large receive space, sparse signal processing methodology is a natural framework to tackle the sporadic traffic.

In this paper, we will outline a new fast and scalable access architecture using a suitable sparse signal processing concept to efficiently deal with the sporadic traffic and control signalling problem. By doing so, e.g. MTC traffic would be removed from standard uplink data pipes with drastically reduced signalling overhead improving operational capabilities and network performance as well as user experience and life time of autonomous MTC nodes. In fact the ratio of control and data can be actually reversed by such concepts to approach a value below 5% within 1ms sub-frame. Notably, while we focus on fast handling of sporadic traffic, the derived architecture is suitable for other traffic classes with ultra-low latency requirements as well.

Notations: $\|x\|_{\ell_q} = (\sum_i |x_i|^q)^{1/q}$ is the usual notion of ℓ_q–norms, denote with $\text{supp}(x) := \{i : x_i := \langle e_i, x\rangle \neq 0\}$ the support of x in a given fixed (here canonical) basis $\{e_i\}_{i=1}^n$. The size of its support is denoted as $\|x\|_{\ell_0} := |\text{supp}(x)|$. W is the (unitary) Fourier matrix with elements $(W)_{kl} = n^{-\frac{1}{2}} e^{-i2\pi kl/n}$ for $k, l = 0 \ldots n-1$, hence, $W^{-1} = W^*$ where W^* is the adjoint of W. We use here also $\hat{x} = Wx$ to denote Fourier transforms and $\odot$ means point–wise product. I_n is the identity matrix in $\mathbb{C}^n$, $\text{diag}(x)$ is some arbitrary diagonal matrix with $x \in \mathbb{C}^n$ on its diagonal.

Preliminary work: In [3], [4] the key findings have been explored that in an under–determined system the signal components can be indeed identified if 1) the measurements are suitably constructed and 2) the signal space is sparse, i.e. only a limited number of elements in some given basis are non-zero. It has soon been recognized that this can be exploited for random access performing user identification and data detection in one step [5]. Sparsity can then be incorporated into ML estimation by a sparsity promoting term. However, a multipath channel is not involved. This problem has been recently extended to asynchronous fading channels [6] and to asynchronous multipath block fading channels [6], [7] with known multipath channel. Here, the problem is modelled by virtual nodes [7] which can be exploited in the detection. Hence, recent concepts deal either with data or channel estimation and an overall architecture which includes identification, channel estimation, asynchronicity and data detection in "one shot" is missing. Moreover, our formulation in terms of random partial circulant matrices seems new to the best of our knowledge. Finally, we mention that this communication model *without sparsity constraints* has been thoroughly investigated in [8] where it is shown that mean squared error (MSE) is minimal if

$$PP^* \odot [W\text{diag}(\underbrace{1...1}_{\tau \text{ times}}, \underbrace{0...0}_{n-\tau \text{ times}})W^*] \quad (1)$$

is a scaled identity. Here, τ is the delay spread, and $P = [p_1, \ldots, p_K]$ is the matrix of pilot sequences sent by the K users. Based on this condition several optimal designs have been presented in [8].

II. THE DESIGN PROBLEM

A. Sparsity-aware Random Access

The design problem is as follows: terminals seek access to the system over the physical layer random access channel (PRACH). We will assume that in the presence of a common broadcast channel some rough synchronization is available. Each user has a data and control signal part which overlap in time with each other to allow for "one shot" detection. Obviously, loosely speaking, to allow for channel equalization, the control must somehow "close" to the data to probe the channel conditions "nearby". Conversely, detectibility becomes erronous the more control is interfering with the data, i.e. when separation between pilots and data can not be achieved in a stable sense. This seems an contradicting, irresolvable task at first sight. However, with sparse signal processing we can cope with that task as follows:

Let us introduce a control part (C-PRACH) and data part (D-PRACH) of the PRACH. We assume that the data signalling leaves out some (observation) window $\mathcal{B}$ in the frequency domain for the identification of users and corresponding channel estimation. Otherwise, user identification performance would be severly degraded and not compliant with current standards. Actually, where this observation window lies is immaterial, but we assume here that it complies with the standard LTE PRACH in the middle of the bandwidth for simplicity. This is beneficial for designing sequences with certain correlation properties. However, as said, we also want to estimate the channel for the user in the D-PRACH. To do so we will spread the signalling over the whole signal space (underlay signalling) and collect it back within the observation window. The design task we are facing is to keep as much structure to ensure proper identification but incorporate as much randomness in the signal to enable proper channel estimation. The approach has some intriguing advantages:

1) Instead of placing distinguishable pilot patterns in the data section for a large number of users based on the classical sampling theorem the control signals can overlap in the large space but are then also distinguishable in a small observation window tailored to the actual number of degrees of freedom. This allows easy overloading of the channel as well.
2) The control is necessarily interfering with the data (and vice versa) but due to the spreading of the control signal the noise from the data is scaled down by a factor defined by the size of the observation window.

On the other hand, it is desired to keep this window at moderate size since the sparse signal reconstruction methods come with increased complexity as compared to standard demodulation and estimation techniques. Even more, exploiting channel sparsity in this way shows that increasing the windows size beyond a certain level no further significant reconstruction gain can be achieved. This phase transition is well–known phenomenon in many compressed sensing algorithms.

In the following sections the details are outlined.

III. COMPRESSIVE RANDOM ACCESS

A. Transmitter operations

The transmitter in our model sends out a binary data signal and an underlay signal taken to perform user identification and channel measurement.

1) Underlay control signalling: We start by considering first the following generic single–user system model. Let $p \in \mathbb{C}^n$ be an pilot (preamble) sequence from a given set $\mathcal{P} \subset \mathbb{C}^n$ (C-PRACH) and $x \in \mathbb{C}^n$ be an unknown (uncoded) data vector $x \in \mathcal{X}^n \subset \mathbb{C}^n$ (D-PRACH). Note that in our system n is large. Both are transmitted simultaneously and use the same resources. We set:

$$\alpha := \frac{1}{n}E\|p\|_{\ell_2}^2 \quad \text{and} \quad \alpha' := 1-\alpha = \frac{1}{n}E\|x\|_{\ell_2}^2 \tag{2}$$

Hence, the control signalling fraction of the power is α and, due to the random zero–mean nature of x we have $\frac{1}{n}E\|p+x\|_{\ell_2}^2 = 1$, i.e. the total transmit power is unity. We will use a cyclic model, which is achieved with OFDM–like signaling and the use of an appropriate cyclic prefix and restrict our model here to time–invariant channels. The vector $h \in \mathbb{C}^{T_{cp}+1}$ denotes the sampled channel impulse response where T_{cp} is the length of the cyclic prefix. Even in the unsynchronized setting T_{cp} is chosen to match the expected maximum delay spreads. Therefore, we assume to have a–priori support knowledge on h: (i) bounded support, i.e. $\text{supp}(h) \subseteq [0, \ldots, T_{cp}]$ due to the cyclic prefix and, most importantly, (ii) sparsity, i.e. $\|h\|_{l_0} \leq k$. For ease of exposition we assume $\|h\|_{\ell_2}^2 = 1$. In an OFDM system the FFT–size n is then chosen as $n \gg T_{cp}$ and therefore, $[h, 0] \in \mathbb{C}^n$ denotes the corresponding zero-padded channel impulse response. The AWGN is denoted as $e \in \mathbb{C}^n$ with $E(ee^*) = \text{SNR}^{-1} \cdot I_n$. The received signal is then:

$$y = \text{circ}(h)(p + x) + e \tag{3a}$$

$$y_{\mathcal{B}} = \Phi y \tag{3b}$$

Here, $\text{circ}(h) \in \mathbb{C}^n$ denotes the circulant matrix with h on its first column; Φ denotes the overall measurement matrix (to be specified later on) referring to the dedicated frequency band $\mathcal{B}$. With $|\mathcal{B}|$ we will denote the size of this band, i.e. the number of subcarriers in $\mathcal{B}$. For circular convolutions we have $\text{circ}(h)p = \sqrt{n} \cdot W^*(\hat{h} \odot \hat{p})$ so that we finally get:

$$y = \sqrt{n} \cdot W^* \left[(\hat{h} \odot (\hat{p} + \hat{x}) \right]) + \hat{e} \tag{4}$$

Here, e and $\hat{e}$ are statistically equivalent. We assume that the number of users is limited by size $|\mathcal{P}|$ of the preamble pool $\mathcal{P}$ and the data of these multiple users is spread in an FDMA fashion over the bandwith outside $\mathcal{B}$.

2) Pilot sequence set $\mathcal{P}$: As in LTE we can use Frank–Zadoff–Chu (FZC) sequences with length $n_{\text{ZC}} = |\mathcal{B}|$ defined by

$$f^u(l) = e^{-i\pi u l(l+1)/n_{\text{ZC}}},\ 0 \leq l < n_{\text{ZC}}, \tag{5}$$

where u is the root index. We consider contention–based RACH, where every user who wants to send a preamble chooses first a signature randomly from the set of available signatures $\mathcal{P} = \{1, ..., 64 - n_{\text{CF}}\}$, with n_{CF} is a given number of reserved signatures for contention–free RACH. Every element of $\mathcal{P}$ is assigned to index (u, v), such that the preamble for each user is obtained by cyclic shifting the u-th FZC $f^{(u,v)}(l) = f^u((l + v\, n_{\text{CS}}) \bmod n_{\text{ZC}})$ where $v \in \{1, ..., \lfloor n_{\text{ZC}}/n_{\text{CS}} \rfloor\}$ is the cyclic shift index, n_{CS} is the cyclic shift size and n_{ZC} is the preamble length which is set to be fixed for all user. Because there can only exists $\lfloor n_{\text{ZC}}/n_{\text{CS}} \rfloor$ number of preambles that can be generated at maximum from the root u, the assignment from the set $\mathcal{P}$ to (u, v) depends on the parameter n_{CS} and the size of set $\mathcal{P}$. As long as the number of preambles from cyclic shifting of the u-th sequence by $v\, n_{\text{CS}}$ is less than the size of $\mathcal{P}$, the next root index $u + 1$ is taken iteratively.

For a given FZC sequence f let p be the corresponding control signalling of length n, i.e. the l-th component $\hat{p}_l$ of its Fourier transform is defined as:

$$\hat{p}_l = \hat{f}((l - m_0) \bmod n_{\text{ZC}}) \tag{6}$$

where m_0 is some offset. Hence $\hat{p}$ consists of the periodic extension of the elements of $\hat{f}$. On the other hand within $\mathcal{B}$ the sequences look exactly the same as in LTE. Outside they might be modified by phase rotations to fulfill further properties such as better peak-to-average power ratio (PAPR) characteristics.

B. Receiver operations

In this paper, we resort to the standard identification process by correlating with the pilot sequences. Hence, we do not exploit sparse user activity at this stage. The primary goal is to estimate the data vector x from the observations y whereby also the vector h of channel coefficients is unknown. We follow here the simple strategy and estimate separately first the channel coefficients $\tilde{h}$ under sparsity assumptions whereby treating the data as additional noise. In a second phase the data $\tilde{x}$ conditioned onto $\tilde{h}$ will be estimated. Obviously, this procedure can then be iterated with or without data decoding. Notably, similar to user activity, sparsity of the messages is also not considered at this stage. However, the data message could be, for example, sensor readings which are often intrinsically compressible, i.e. represented by correlated measurements.

1) Identification: The goal is the limitation to a relatively small observation frequency window, i.e. frequency indices in the set $\mathcal{B}$ of cardinality $m = |\mathcal{B}|$. Practically, the frequency window should be known globally. However, for the channel estimation step, explained later on, $\mathcal{B}$ needs only to be known to the channel estimator. Let us abbreviate with $P_{\mathcal{B}} : \mathbb{C}^n \to \mathbb{C}^m$ the corresponding projection matrix, i.e. the submatrix of I_n with rows in $\mathcal{B}$. For identifying which preamble is in the system we consider the FFT of y and use the frequencies in $\mathcal{B}$, i.e. the vector:

$$\hat{y}_I := \sqrt{n} \cdot P_{\mathcal{B}} \left[\hat{h} \odot (\hat{p} + \hat{x}) \right] + P_{\mathcal{B}} \hat{e} \tag{7}$$

The standard decision variable is calculated as $d = W^* \left(\overline{(\hat{p}_u)} \odot \hat{y}_I \right)$ indicating that $|d\,(l)|^2 > d_{th}$ then l-th path path is present. A more sophisticated method is given and analyzed in [9]. However, since we spread the power over the entire bandwidth the detection performance (y-axis) over SNR (x-axis) curve will shifted in the x-axis by a factor of $|\mathcal{B}|/n$. Since, this detection process has been thoroughly investigated in the literature we will not further elaborate on it.

2) Illumination and Partial Fourier Sampling: It is known fixed Fourier windowing has several drawbacks (which will explained later on) and to introduce certain randomization we consider pointwise multipliers $\xi \in \mathbb{C}^n$ in time domain and we denote the corresponding $n \times n$ diagonal matrix as $M_\xi := \mathrm{diag}(\xi)$. Summarizing, the $m \times n$ sampling matrix $\Phi = P_{\mathcal{B}} W M_\xi$ will be considered. With the bilinear model in (3) we get for the observation vector y_B:

$$\hat{y}_B = \sqrt{n} \cdot P_{\mathcal{B}} W M_\xi W^* \left[\hat{h} \odot (\hat{p} + \hat{x}) \right] + P_{\mathcal{B}} \hat{e} \tag{8}$$

$$= \sqrt{n} \cdot P_{\mathcal{B}} \left[\hat{\xi} * (\hat{h} \odot (\hat{p} + \hat{x})) \right] + P_{\mathcal{B}} \hat{e} \tag{9}$$

$$= \sqrt{n} \cdot P_{\mathcal{B}} \cdot \mathrm{circ}(\hat{\xi}) \left[\hat{h} \odot (\hat{p} + \hat{x}) \right] + P_{\mathcal{B}} \hat{e} \tag{10}$$

The $m \times n$ matrix $P_{\mathcal{B}} \cdot \mathrm{circ}(\hat{\xi})$ is usually called *partial circulant matrix*. Note that $\hat{h}$ is sparse in the Fourier domain.

3) Channel Estimation: For sparse channel vectors h with $\mathrm{supp}(h) \subseteq [0, \dots, T_{cp}]$ and $\|h\|_0 \leq k$, the ℓ_1–penalized least–squares estimation:

$$\tilde{h} = \arg\min_h \|\Phi\, \mathrm{circ}(p) h - y\|_{\ell_2}^2 + \lambda \|h\|_{\ell_1} \tag{11}$$

is preferable which occurs as Lagrangian for the basis pursuit denoising (BPDN). Although this method is used often in practise its depends on a some choice of the regularization parameter λ. Later on, we will directly consider the BPDN approach in Sec.IV, see (17) below. Also several greedy methods exists for sparse reconstruction. In particular, for CoSAMP [10] explicit guarantees in reconstruction performance are known and could be used in the next sections instead of BPDN.

4) Estimation of Sparse Data Symbols: Once the channel is estimated the correponding pilot signal is subtracted for the recieved signal. Denote the error of this operation as:

$$d := \tilde{h} - h \tag{12}$$

Hence, the received signal is given by

$$\hat{y} = \sqrt{n} \cdot (\hat{h} + \hat{d}) \odot \hat{x} + \sqrt{n} \cdot \hat{d} \odot \hat{p} + \hat{e} \tag{13}$$

which a set of parallel channels each with power $E(|\hat{x}_k|^2) = 1 - \alpha$ and $|\hat{p}|^2 = \alpha$ and $E(|e_k|^2) =$SNR^{-1}.

IV. Performance Analysis

A. Compressed Channel estimation

Let us collect known results related to the measurement (sampling) matrix Φ with the focus on sparse or compressible signals. Let $\Sigma_k := \{x \in \mathbb{C}^n \,:\, \|x\|_{\ell_0} \leq k\}$ denotes the k–sparse vectors. Given a vector x, the ℓ_1–error of its best k–term approximation is:

$$\sigma_k(x)_{\ell_1} := \min \{ \|x - \tilde{x}\|_{\ell_1} \,:\, \tilde{x} \in \Sigma_k \} \tag{14}$$

and obviously this is an adaptive strategy (i.e. x has to be known to compute the minimizer). According to [11] a non–adaptive estimator Q (or an approximation/decoding strategy), which solely operates on *noiseless* measurements $y = \Phi x$, is called (instance–) optimal if it achieves the same scaling law in k as the best k–term approximation, i.e. $\|Q(y) - x\|_{\ell_2} \leq ck^{-1/2} \cdot \sigma_k(x)_{\ell_1}$ for some constant C and all x. For sparse signals $x \in \Sigma_k$ this implies perfect reconstruction, since $\sigma_k(x)_{\ell_1} = 0$.

1) RIP and Performance Guarantees:: The most well–known property here is the *restricted isometry property* (RIP): the matrix Φ is called RIP of order k (k-RIP) if exists $0 \leq \delta_k < 1$ such that

$$||\|\Phi x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2| \leq \delta_k \|x\|_{\ell_2}^2 \tag{15}$$

for all $x \in \Sigma_k$. Obviously, the RIP constant δ_k is hard to check (it is usually NP–complete [12]) but there is the upper bound $\delta_k \leq k \cdot \mu(\Phi)$ with respect to the coherence $\mu(\Phi)$ which can be straightforward computed from the columns $\{\phi_l\}_{l=1}^n$ of Φ:

$$\mu(\Phi) = \max_{k \neq l} \frac{|\langle \phi_k, \phi_l \rangle|}{\|\phi_l\|_{\ell_2} \cdot \|\phi_k\|_{\ell_2}} \tag{16}$$

The Welsh–bound states $\mu(\Phi) \in [\sqrt{\frac{n-m}{m(n-1)}}, 1]$. Now, there is a well–known result [4] on the quadratically constrained ℓ_1–minimization problem:

$$\tilde{x} = \arg\min \|x\|_{\ell_1} \text{ s.t. } \|y - \Phi x\|_{\ell_2} \leq \epsilon \tag{17}$$

and we call this as the Q_1–estimator for x given y, i.e. $\tilde{x} = Q_1(y)$. If Φ is $2k$–RIP with $\delta_{2k} < \sqrt{2} - 1$ and $\|e\|_{\ell_2} \leq \epsilon$ then:

$$\|Q_1(\Phi x + e) - x\|_{\ell_2} \leq c_1 \frac{\sigma_k(x)_{\ell_1}}{k^{1/2}} + c_2 \epsilon \tag{18}$$

with $c_1 = 2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}$ and $c_2 = 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}$ (in particular, for $\delta_{2k} = .2$ this gives $c_1 = 4.2$ and $c_2 = 8.5$). It is known that $\delta_{2k} \leq 1/\sqrt{2}$ is a necessary condition. However, according sufficiency, meanwhile several improvements in the δ_{2k}–bound for Q_1 have been found. For example, in [13] the bound has been improved to $\delta_{2k} \leq 3/(4 + \sqrt{6}) \approx 0.4652$. Similar bound exists for *COSAMP* [10], i.e. $\delta_{4k} \leq \sqrt{\frac{2}{5+\sqrt{73}}} \approx 0.3843$ [14]. Various RIP results are also known for *Iterative Hard Thresholding* and matching pursuit algorithms [15].

Up this point we have mentioned uniform reconstruction guarantees (for any x) for a given matrix Φ and these are related to its RIP–constant δ_{2k}. However, it is still difficult to constructively design measurements matrices with sufficiently small RIP constants. The familiar approach to cope with this is to generate random matrices according to certain random model and to ensure that with overwhelming (exponential) probability a randomly selected matrix is "good". For example, several results for row–independent or column–independent models of isotropic subgaussian random vectors are shown in [16]. In both cases there is the rule: if $m \geq c\delta^{-2} k \log(n/k)$ then $\delta_k \leq \delta$ with probability $\geq 1 - 2\exp(-c'^2 m)$ with c and c' depending only on the maximum subgaussian norm of the rows (or the columns) of Φ.

The RIP–properties of random *partial circulant matrices*, matrices of the form $P_{\mathcal{B}} \cdot \text{circ}(\hat{\xi})$ for fixed $\mathcal{B}$ which occur in 10, have been investigated in [17]. In particular, the situation is considered where $\hat{\xi}$ (or ξ itself, see remarks in this citation) is a $\{+1, -1\}$–valued i.i.d. Rademacher sequence or Steinhaus–sequence (uniform distributed on the torus) and $m = O(k^{3/2} \log^{3/2} n)$ has been achieved. The desired linear scaling in k has been achieved recently in [18] with

$$m \geq c\delta^{-2} k \log^2 k \log^2 n \tag{19}$$

where ξ (or $\hat{\xi}$) can be any i.i.d. zero–mean and unit–variance subgaussian random vector. Thus, the important point here is that linear scaling in k is achieved independently of the positions in $\mathcal{B}$, i.e. only in terms of $m = |\mathcal{B}|$. On the other hand, all these previous RIP–results for $P_{\mathcal{B}} \cdot \text{circ}(\hat{\xi})$ with fixed $\mathcal{B}$ are *not universal*, i.e. results are only valid when the vectors are sparse in the canonical basis. Universality with $m = \mathcal{O}(\mu^2 k \cdot \log(n))$ is achieved again if $\mathcal{B}$ is not fixed anymore but itself *random*. This result is based on the incoherence $\mu = \mu(U, \Psi)$ between two unitary matrices U and Ψ where measurement is performed via $\Phi = P_{\mathcal{B}} \cdot U$ [19]. In [20] non–uniform recovery results are presented for the case $U = \mathrm{circ}(\hat{\xi})$ and they turned out to be universal (not depend on Ψ) but stability is not considered (robustness against noise). A refined analysis has been achieved in [21] and settles the RIP–context giving stability. It is based on the results in [22]: For any unitary matrix U, the measurement matrix $P_{\mathcal{B}} \cdot U$ with $\mathcal{B}$ chosen uniformly at random with cardinality $m = |\mathcal{B}|$ such that:

$$m \geq C\delta^{-2}\mu^2 k \log^5(n) \tag{20}$$

has RIP with probability $\geq 1 - \mathcal{O}(n^{-1})$ and $\delta_{2k} \leq \delta$, where $\mu = \mu(U, \mathrm{Id})$ is the incoherence between U and the identity (standard basis). Based on this and some more details on μ for random (unitary/orthogonal) convolutions (ξ is pure i.i.d. phase, like Rademacher or Steinhaus sequences, then $U = \mathrm{circ}(\hat{\xi})$ is unitary and incoherent to any Ψ) the result in [21] is: For any unitary Ψ the measurement matrix $P_{\mathcal{B}} \cdot \mathrm{circ}(\hat{\xi})\Psi$ with $\mathcal{B}$ chosen uniformly at random with cardinality $m = |\mathcal{B}|$ such that:

$$m \geq C\delta^{-2} \min(k \log^6 n, k^2 \log n) \tag{21}$$

has RIP with probability $\geq 1 - \mathcal{O}(n^{-1})$ and $\delta_{2k} \leq \delta$ ($\log^5$ for a fixed probability). Thus, in contrast to the results given above for fixed $\mathcal{B}$ we would have here universality. Summarizing, it seems that universality and fixed $\mathcal{B}$ can not achieved simultaneously without further randomization.

B. Rate Scaling

Define $B(h,p) = \mathrm{circ}(p)h = \mathrm{circ}([h,0])p$. The system equation (3) is then equivalent to:

$$y = \sqrt{\alpha}\left[\Phi B(h,p) + \underbrace{\sqrt{\frac{1-\alpha}{\alpha}}\Phi B(h,x) + \sqrt{\frac{1}{\alpha}}\Phi e}_{=:z}\right] \tag{22}$$

With usual assumptions (zero mean etc.) the energy of z can be estimates like:

$$E(\|z\|_{\ell_2}^2) = \frac{1-\alpha}{\alpha} E(\|\Phi B(h,x)\|_{\ell_2}^2) + \frac{m \cdot \mathrm{SNR}^{-1}}{\alpha} \tag{23}$$

$$= \frac{m\,(1-\alpha)}{\alpha} + \frac{m \cdot \mathrm{SNR}^{-1}}{\alpha} \tag{24}$$

Let us define for a given pilot power α the channel estimation error as $d(\alpha) = \tilde{h} - h$ and take ℓ_1–decoding $\tilde{h} = Q_1(y/\sqrt{\alpha})$. From (18) follows with $c_2(\delta_{2k}) = \frac{4\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}$:

$$\begin{aligned}\frac{1}{n}E(\|d(\alpha)\|_{\ell_2}^2) &= \frac{1}{n}E(\|Q_1(y/\sqrt{\alpha}) - h\|_{\ell_2}^2) \\ &\overset{(18)}{\leq} \frac{c_2(\delta_{2k})^2}{n} E(\|z\|_{\ell_2}^2) \\ &\leq \frac{m \cdot c_2(\delta_{2k})^2}{n}\left(\frac{1-\alpha}{\alpha} + \frac{\mathrm{SNR}^{-1}}{\alpha}\right)\end{aligned}$$

With these definitions we can show the following:

Theorem 1. *Let the channel impulse response be k–sparse and use* (17) *(BPDN) as the channel estimate. The achievable rate per subcarrier is lower bounded by:*

$$R(\alpha) \geq \log(1 + \mathit{SNR} \cdot (1-\alpha)) - \log\left(1 + \frac{m}{n} \cdot c_2(\delta_{2k})^2 \left(\mathit{SNR} \cdot \frac{1-\alpha}{\alpha} + \frac{1}{\alpha}\right)\right)$$

for a fixed sampling Φ obeying a RIP–constant $\delta_{2k} < \sqrt{2} - 1$

Proof: Using eqn. (13) we can proceed exploiting an idea similar to [23, Theorem 1] (details omitted) linking the achievable mutual information to MSE estimation. By our illumination procedure it is then easy to see that the capacity loss of each subchannel in eqn. (13) can be upperbounded by:

$$\Delta R\,(\alpha) \leq \log\left(1 + \text{SNR}\cdot\frac{1}{n}E(\|d(\alpha)\|_{\ell_2}^2)\right) \tag{25}$$

Hence, we have:

$$\Delta R\,(\alpha) \leq \log\left(1 + \frac{m}{n}c_2(\delta_{2k})^2\cdot\left(\text{SNR}\cdot\frac{1-\alpha}{\alpha} + \frac{1}{\alpha}\right)\right)$$

■

Obviously, for a random sampling approach the rate gap can only hold with a certain probability. Briefly explained, let c_3 some constant and choose for a given $\delta < \sqrt{2}-1$ the number of measurements m sufficiently above the phase transition, i.e. $m = c_3\delta^{-2}k\log^a n$ for some exponent $a \geq 1$ (see here (19), (20) or (21)). Then follows with overwhelming probability $\delta_{2k} \leq \delta$ and we have also $c_2(\delta_{2k}) \leq c_2(\delta)$. The rate gap essentially depends then on k and n since we have:

$$\frac{m}{n}c_2(\delta_{2k})^2 \leq c_3\frac{\log^a(n/k)}{n/k}\cdot\frac{c_2(\delta)^2}{\delta^2} \tag{26}$$

C. Multiuser case

Once the principal sparse estimation principles are established, we can extend the framework for the estimation of multiuser signals and channels. Let $h := [[h_1, 0], \ldots, [h_K, 0]]^T \in \mathbb{C}^{Kn}$ and $x := [x_1, \ldots, x_K]^T \in \mathbb{C}^{Kn}$ be the stacked channel coefficients and data symbols of the identified users. The received signal is again:

$$\begin{aligned} y &= \Phi\sum_{u=1}^{K}\text{circ}(h_u)(p_u + x_u) + e && (27)\\ &= \Phi\left[D(p)h + C(h)x\right] + e && (28)\end{aligned}$$

where $D(p) := [\text{circ}(p_1), \ldots, \text{circ}(p_K)] \in \mathbb{C}^{n\times Kn}$ and $C(h) := [\text{circ}([h_1, 0]), \ldots, \text{circ}([0, h_K])] \in \mathbb{C}^{n\times Kn}$ are the corresponding compound matrices. If assume each user–channel vector h_u to be k–sparse then h is $K\cdot k$ sparse. Hence, the overall performance is then

$$\begin{aligned} R\,(\alpha) \geq\; &\log\left(1 + \text{SNR}\cdot(1-\alpha)\right)\\ &-\log\left(1 + \frac{m}{n}\cdot c_2(\delta_{2Kk})^2\left(\text{SNR}\cdot\frac{1-\alpha}{\alpha} + \frac{1}{\alpha}\right)\right)\end{aligned}$$

provided that all the multiuser sequences are shift-orthogonal on $\mathbb{C}^n$.

V. SIMULATIONS

We assume for comparability a sampling frequency of 30.72MHz equal to LTE-A. The FFT size is $n = 24576$ and the cyclic prefix is then 8192 samples which corresponds roughly to $100\mu s$. Assuming a maximum cell size of 1.5km the delay is at most $5\mu s$. With these considerations we set $\tau = 300$ with at most $k = 6$ pathes. The observation window size of C-PRACH is 839 corresponding to standard PRACH. D-PRACH consists of the rest of the subcarriers outside the window. It is assumed that the data of the users is equally distributed over the subcarriers outside the window. The data is assumed to be BPSK-modulated. We assume that for preamble detection the standard correlation detector is applied. After preamble detection the channel is estimated by the outlined procedure and subtracted form the received signal. Then the data is detected subcarrier by subcarrier.

In the first simulation in Fig. 1 we investigate the MSE performance over α: $\alpha = 0$ means no control power (full data power) so that reliable channel estimation is impossible, while $\alpha = 1$ means full control power (no data power) enabling the best possible channel estimation for our control signaling scheme. Obviously, the MSE performance becomes better the more power is allocated to the pilots. Clearly there is no tradeoff since with increasing control

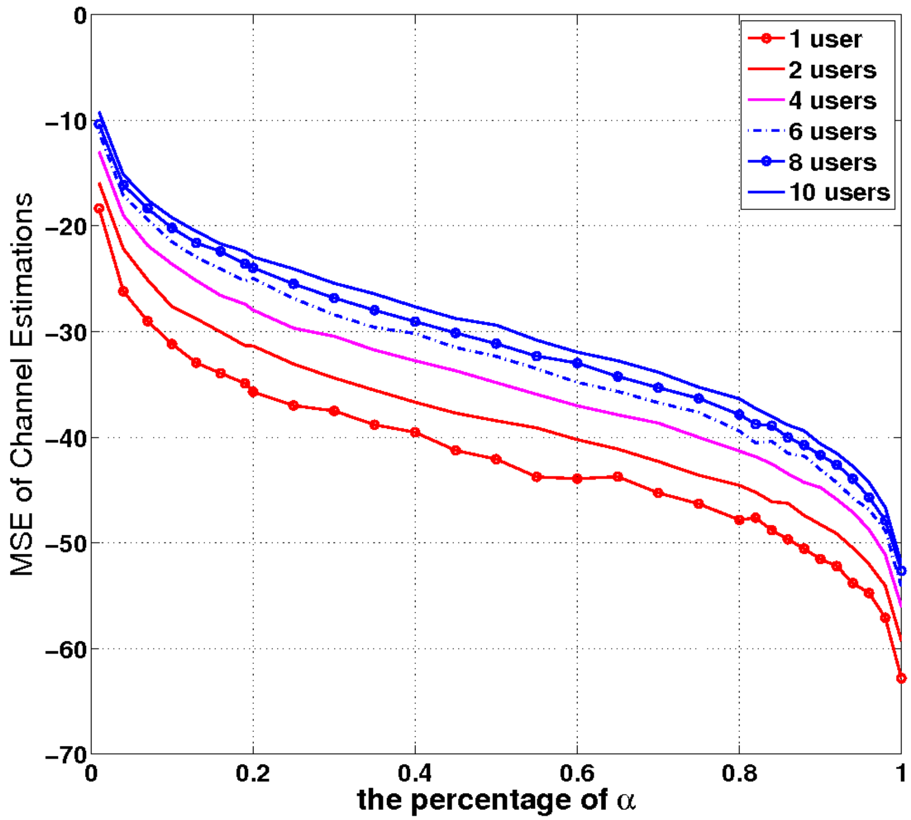

Fig. 1. MSE performance over $0 \leq \alpha \leq 1$ (i.e. zero control power to zero data power) for up to 10 users

power, the data power (representing noise for the channel estimation scheme) is reduced at the same time so that close to $\alpha = 1$ MSE is reducing faster and faster which is intuitively the reason why the curve changes its characteristics from convex to concave. The main simulation is in Fig. 2 showing (data) symbol error rates (SER) over α. Here, for α approaching unity data symbol detection is not possible since SNR is reduced to zero. On the other hand, we see, that irrespectively of the numbers of user there is some α allowing for SERs below 10% for all user numbers. The curve follows the predicted shape in Theorem 1 although, as outlined, the optimum α as well as the actual SER values cannot be suitably upperbounded or estimated due to the weak bounds on the RIP properties. However, from the simulations we can actually infer that the algorithms work well and some proper, i.e. non-trivial, detection performance can be achieved.

VI. Conclusions

In this paper, we provided some initial ideas how to enable random access for many devices in a massive machine-type scenario. In the conceptional approach as well as the actual algorithms sparsity of wireless channels plays an essential role. We plan to generalize this concept in future work to incorporate sparsity of user activity and/or messages as well.

VII. Acknowledgements

The work of G. Wunder was supported by the 5GNOW project supported by the European Commission under grant 318555 (FP7 Call 8). Peter Jung was supported by DFG–grant JU 2795/2-1. We also thank the reviewers for their comments.

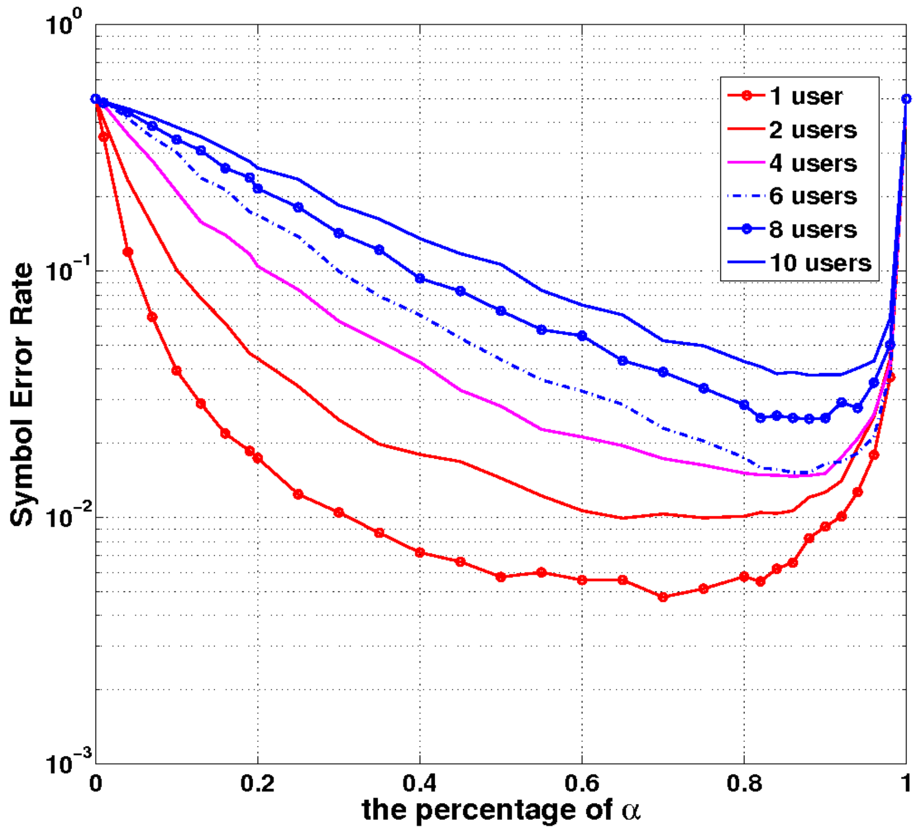

Fig. 2. SER performance over $0 \leq \alpha \leq 1$ (i.e. zero control power to zero data power) for up to 10 users

REFERENCES

[1] G. Wunder, M. Kasparick, S. ten Brink, F. Schaich, T. Wild, I. Gaspar, E. Ohlmer, S. Krone, N. Michailow, A. Navarro, G. Fettweis, D. Ktenas, V. Berg, M. Dryjanski, S. Pietrzyk, and B. Eged, "5GNOW: Non-Orthogonal, Asynchronous Waveforms for Future Mobile Applications," *IEEE Communications Magazine, 5G Special Issue*, February 2014, to appear.

[2] ——, "5GNOW: Challenging the LTE Design Paradigms of Orthogonality and Synchronicity," in *Proc. IEEE Vehicular Technology Conference (VTC'13 Spring) - Workshop Mobile and Wireless Communication Systems for 2020 and beyond*, Dresden, Germany, June 2013. [Online]. Available: http://arxiv.org/abs/1212.4034

[3] D. L. Donoho, "Compressed sensing," *IEEE T. Inform. Theory.*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences*, vol. 346, no. Paris, Serie I, pp. 589–592, May 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1631073X08000964

[5] H. Zhu and G. Giannakis, "Exploiting Sparse User Activity in Multiuser Detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454–465, 2011.

[6] L. Applebaum, W. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous Code-Division Random Access Using Convex Optimization," *Physical Communication*, vol. 5, no. 2, p. 129147, February 2012.

[7] C. B. H. F. Schepker and A. Dekorsy, "Coping with CDMA Asynchronicity in Compressive Sensing Multi-User Detection," in *Proc. 77th IEEE VTC Spring*, Stuttgart, Germany, June 2013.

[8] S. Stanczak, G. Wunder, and H. Boche, "On Pilot-Based Multipath Channel Estimation for Uplink CDMA Systems: An Overloaded Case," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 512–519, 2006.

[9] K. Lee, J. Kim, J. Jung, and I. Lee, "Zadoff-Chu Sequence based Signature Identification for OFDM systems," *IEEE Trans. on Wireless Communications*, 2014, to appear.

[10] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, 2008. [Online]. Available: http://front.math.ucdavis.edu/0803.2392

[11] A. Cohen, W. Dahmen, and R. A. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, Jul. 2008. [Online]. Available: http://www.ams.org/journal-getitem?pii=S0894-0347-08-00610-3

[12] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, 1995. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/S0097539792240406

[13] S. Foucart, "A note on guaranteed sparse recovery via l1-minimization," *Applied and Computational Harmonic Analysis*, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520309001158

[14] ——, "Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants," *Approximation Theory XIII: San Antonio 2010*, pp. 1–14, 2012. [Online]. Available: http://link.springer.com/chapter/10.1007/978-1-4614-0772-0_5

[15] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1063520309000384

[16] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing, Theory and Applications*, Y. Eldar and G.Kutyniok, Eds. Cambridge University Press, 2012, ch. 5. [Online]. Available: http://arxiv.org/abs/1011.3027

[17] H. Rauhut, J. Romberg, and J. a. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, Mar. 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1063520311000649

[18] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *arXiv preprint arXiv:1207.0235*, pp. 1–23, 2012. [Online]. Available: http://arxiv.org/abs/1207.0235

[19] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007. [Online]. Available: http://stacks.iop.org/0266-5611/23/i=3/a=008?key=crossref.bdb1d0e3390f5616a4395b50273198a9

[20] J. Romberg, "Compressive sampling via random convolution," *PAMM*, vol. 2010012, pp. 2 010 011–2 010 012, 2007. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/pamm.200700036/full

[21] ——, "Compressive Sensing by Random Convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, p. 1098, 2009. [Online]. Available: http://link.aip.org/link/SJISBI/v2/i4/p1098/s1&Agg=doi

[22] M. Rudelson, R. Vershynin, and R. V. On, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, Nov. 2007.

[23] G. Caire, N. Jinal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.