Downlink Scheduling using Compressed Sensing

Sibi Raj Bhaskaran*, Linda Davis†, Alex Grant†, Stephen Hanly*, Paul Tune*
* University of Melbourne, VIC-3010, Australia.
† University of South Australia, Adelaide, Australia.
{sibib,hanly,l.tune@unimelb},{alex.grant,linda.davis@unisa}.edu.au

***Abstract*— We propose a novel access technique for cellular downlink resource sharing. In particular, a distributed self-selection procedure is combined with the technique of *compressed sensing* to identify a set of users who are getting simultaneous access to the downlink broadcast channel. The performance of the proposed method is analyzed, and its suitability as an alternate access mechanism is argued.**

I. Introduction

Consider a downlink broadcast scenario encountered in cellular systems, where there is a base station which has data meant for many users. Voice traffic, which requires certain delay guarantees are not our concern here. Rather, we are looking at the data traffic, which can accrue some delay before being served. Let N be the total number of users (mobiles) in the system. The mobiles are distributed across the cell according to some distribution. Due to physical propagation laws, each user encounters an independent channel. Let us assume a block fading model, where the channel remains the same for a block and then it changes to an independent realization. For a timeslot consisting of a blocklength of L channel uses, and using single antennas at the source and destinations, we can write the received symbol at receiver n, for any symbol in the slot, as:

$$r_n = h_n x + z_n, \quad n = 1, 2, \ldots, N \tag{1}$$

where h_n is the complex fading coefficient of the path to receiver n (fixed for the entire slot), x is the complex data symbol, and z_n is complex Gaussian noise of variance σ_n^2. In this paper, we assume a flat-fading model with Rayleigh distributed path-amplitudes, a symmetric model with $|h_n|^2$ of mean γ, for all $n = 1, 2, \ldots, N$.

We will assume that N is quite large, say 100, or even 1000 users. In such a setting, not all N users need to be serviced in a particular block. Users that are not scheduled in one block can be scheduled in another. We will assume that only a relatively small number, K, will be serviced in any particular block, a fraction of the total user population, N. A special case that we consider is $K = 1$, which is the typical practical case in which the base station only sends to one user at a time, rather like TDM, except that here the selection will be based on channel quality [2], which is sometimes referred to as "opportunistic scheduling". But obtaining the channel quality in each block is an onerous task and the base station may decide to serve more than one user in a block, to reduce overhead, ensure fairness etc.

The problems that we address are channel feedback and user selection. If all mobiles are required to send back their channel state information to the base station, this entails a large signalling overhead on the system, proportional to N, the total user population. In the present paper, we propose a new method, based on ideas from compressed sensing [3], to provide a distributed approach to user selection. Our approach offers considerable savings when K is a small fraction of N, as it will be in many practical settings. For example, consider a wide-band OFDM setting in which there are a large number of time-frequency slots, and only a relatively small number of users are assigned to each time-frequency slot. In such a scenario, the feedback overhead of our scheme is of the order of $K \log N$.

Our scheme involves two phases, which we describe next.

II. Primary access phase: User self-selection and feedback via compressed sensing

Lets assume that there is some target for K, which we denote by K_{tar}. If the base station is to send to one user at a time (in each slot) then $K_{tar} = 1$, but for various operational reasons K_{tar} can be higher than 1.

The basic idea is that a large amount of overhead can be avoided if users are allowed to self-select based on their own channel measurement. If the set of selected users is small, then the feedback problem is much reduced, and we show that a particularly simple, distributed approach to feedback can then be followed, using principles from compressed sensing.

The users first independently self-select by comparing their channel quality with a threshold, with the threshold chosen to ensure that the number of selected users, K', satisfies $1 < K' << N$. It is possible to vary the threshold as a function of the long-term average gain of the user, or of its history of recent channel access, to ensure some sort of fairness across the users. In the present paper, we focus on a symmetric model and hence only consider fixed thresholds, the same for all users. Denote the set of self-selected users by $\mathcal{K}'$, of size K', and note that K' is random, since it is based on channel quality. The user's threshold is chosen so that $K' > K_{tar}$ with high probability, yet $K' << N$ remains true. We provide more details on the selection of the threshold in Section IV.

The next idea is to use a CDMA mechanism to allow feedback of the channel strengths of the selected users with minimal overhead. Here, we exploit the theory of compressed sensing, which shows that the amount of overhead required need only grow with $\log N$, due to the sparsity of the set of users that have been self-selected. That the self-selected users is sparse follows from the choice of the thresholds, designed to ensure that $K' << N$. In our method, each user is assigned a signature sequence vector for use on an uplink feedback channel. We assume a TDD setup so that the channel gain on the uplink is the same as that on the downlink.

For now, let M denote the dimension of the signature sequence vectors, with the choice of M discussed in more detail below. Thus, user n is assigned $\boldsymbol{a}_n$, a vector of length M. The random sequence can be a pseudo-noise sequence or entries of a Gaussian random vector. We use sequences which are generated iid according to Bernoulli($\frac{1}{2}$) with values ± 1. This is in fact a CDMA scheme with random signature sequences, where each user is identified by its signature sequence. However, there are a few differences from the standard models. At the receiver, we do not have knowledge of which users are taking part in the transmissions. Moreover, our signature sequences are of length M, where M is assumed to be far below N (see below). Finally, we do not intend for the sequences to provide any spectrum spreading: the symbols of the sequences will be sent at the baseband sampling rate. Let A denote the $M \times N$ matrix with nth column provided by the vector $\boldsymbol{a}_n$.

Let $h_n = |h_n| \exp(j\phi_n)$ denote the random channel gain for user n and let Γ_n denote the threshold used for user selection. Sparsity arises because Γ_n is set so that it is quite unlikely for user n to be self-selected: if $|h_n|^2 < \Gamma_n$ then the user transmits nothing, but if $|h_n|^2 > \Gamma_n$ (a low probability event) the user will transmit the symbol $\exp(-j\phi_n)$. An equivalent real-valued model for the uplink feedback channel is then

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{z} \tag{2}$$

where $\boldsymbol{z}$ has iid Gaussian distributed components, with mean 0 and variance σ^2, and $\boldsymbol{x}$ is given by:

$$x_n = \begin{cases} 0 & |h_n|^2 < \Gamma_n \\ |h_n| & |h_n|^2 > \Gamma_n \end{cases}$$

This model is restricted to the real numbers, which is helpful from a computational point of view. We assume appropriate scaling on the noise variance σ^2 to account for various signal to noise ratios.

The next step is for the base station to perform multi-user detection (MUD) on the above CDMA feedback channel. However, we propose a novel compressed sensing approach to the MUD problem. Traditional approaches such as the single user matched filter (SUMF), decorrelator detector, or even the linear minimum mean squared error (LMMSE) filter, will perform badly in this setting, unless M is of the same order as N. However, this is the sort of overhead that we want to avoid.

Suppose for now that K' is known, and $M(<< N)$ happens to satisfy the inequality $M \geq 2K' \log(N/K')$. Examples include $N = 100, K' = 5, M = 13$; $N = 100, K' = 15, M = 25$; or $N = 1000, K' = 15, M = 55$. In all of these examples, the SUMF will perform badly, the LMMSE will perform very close to that of the SUMF (and badly), and the decorrelator will be undefined. This is because these receivers do not use the fact that the input, $\boldsymbol{x}$, is sparse, and hence they "see" an overloaded system; see Section VI for more details.

The theory of compressed sensing says that we can recover $\boldsymbol{x}$ exactly in the noiseless case (at least for Gaussian measurement matrices) with an overwhelming probability, if

$$M \geq cK' \log_2(\frac{N}{K'}), \tag{3}$$

where c is some constant [3]. The recovery can be shown to be robust with respect to some amount of additive noise. In our experiments we use $c = \sqrt{2}$, which seems to give ample noise immunity in the signal levels that we consider. We can show that near *optimal* recovery is impossible for $c < c'$, where c' is a function of the measurement matrix and the noise variance. This constant is further examined in Section V.

When the number of measurements satisfy (3), and there is no noise, we can recover $\boldsymbol{x}$ exactly based on the following ℓ_1 minimization algorithm [3].

$$\arg\min_{x \in \mathbb{R}^N} ||x||_{\ell_1} \text{ subject to } Ax = y \tag{4}$$

In the case of exact recovery, we will know K' exactly at the end of the minimization. In practice, exact recovery is not possible, due to noise and uncertainty in the theoretical model itself. However, compressed sensing theory assures us that the base station can still track users with the higher channel coefficients, but without guaranteeing that it can deduce K' exactly. Here is where a heuristic choice can greatly help, as K' is not the quantity of interest; rather, the base station wishes to choose a smaller set of K_{tar} users. The base station can pick the users which have reconstructed channel magnitude greater than some specified threshold or pick the best K_{tar} users. The threshold can take into account the effect of noise as well as the possible guess on the number of users based on the reconstructed signal. The base station can also take into account the issue of fairness in its selection of users.

In the presence of noise, l_1 minimization is not always the best thing to do from a computational point of view, though it is a linear program. This is especially true if the number of users, N, is high, as this linear program has $O(N^3)$ complexity. We can resort to techniques which do l_2 minimization with a penalty term favoring solutions with low l_1 norms, see [4]. We use the gradient projection approach for sparse reconstruction (GPSR) described in [4] to reconstruct the channel magnitudes at the base station. This is a reasonably fast algorithm, and also avoids the need for an initial guess to start the algorithm. See Section VI on simulation study below for more details. The optimization is of the form,

$$\arg\min_{x \in \mathbb{R}^N} \frac{1}{2}||y - Ax||_2^2 + \tau ||x||_{\ell_1}, \tag{5}$$

where τ is a non-negative parameter.

The GPSR algorithm has a debiasing phase, where the non-zero values of the reconstructed vector are adjusted to get the best least square solution. In our channel estimation problem, using our *apriori* knowledge that the channel gains exceeded some threshold, we can use a hard threshold at the end of the algorithm to obtain a sparser vector. Then we can run the debiasing phase of GPSR on the vector obtained after thresholding.

III. Refinement Phase

Denote the solution to (5) by the vector $\hat{\boldsymbol{x}}$. If the noise is sufficiently low, then $\hat{\boldsymbol{x}}$ will have the same level of sparsity as $\boldsymbol{x}$ and will be very close to $\boldsymbol{x}$ in a mean-square sense. Thus, the base station can acquire not only the identities, but the channel strengths of the self-selected users.

The next step is for the base station to make a further selection of users. If $K_{tar} = 1$, and if channel quality is the required metric, then the base station simply picks the user n with the highest value of $\hat{x}_n$. The choice can also be modulated by long-term data rate, as in proportional fairness [5]. Once the users are selected, the base station can use the downlink to convey the selected set to the mobiles. Notice that not every user wishes to know the outcome of the selection procedure, only the self-selected ones need to be notified. So the information that each mobile seeks is not on who gained access, but whether it was selected by the base-station or not. This falls in the framework of *identification* via channels, pioneered in [6]. In particular, using an identification code instead of a transmission code, $\log\log N$ bits suffice to identify a selected user.

In case of multiple antennas and $K_{tar} > 1$, it is important for the base station to obtain phase or angle of arrival information from the users selected, to make a good joint selection of K users. We can still use our procedure to identify the users having a good channel norm. Obtaining the direction information is then no more onerous since the user identities have already been obtained from the first phase, and there are not many of them.

IV. Choosing the thresholds for self-selection

In the present paper, we assume that the channel coefficients $|h_n|$ are Rayleigh distributed, with parameter $\sqrt{\gamma/2}$, and thus $|h_n|^2$ is exponentially distributed with parameter $1/\gamma$. Due to the symmetry of this model, and for simplicity, we use a common threshold for the self-selection process, but this can easily be generalized. Thus, for a threshold of τ,

$$P(|h_n|^2 \geq \tau) = \exp(-\frac{\tau}{\gamma}). \tag{6}$$

By defining $\rho = \exp(-\frac{\tau}{\gamma})$, the number of users K' satisfying the threshold condition is given by a Binomial random variable.

$$P(K' = k) = \binom{N}{k} \rho^k (1-\rho)^{N-k} \tag{7}$$

The average number of chosen users in each block and its variance is given respectively by

$$E[K'] = N\rho \tag{8}$$

$$\mathrm{var}(K') = N\rho(1-\rho). \tag{9}$$

We can set the threshold so that $E[K']$ is greater than K_{tar}, say, for example, $E[K'] = 2K_{tar}$. Roughly speaking, a small fraction ρ of the users self-select, and ρ is chosen to ensure that $N\rho > K_{tar}$. It is clear that knowledge of N is useful in choosing the threshold.

Bounds on tail probabilities can be used to ensure that for almost all blocks K' is close to its expected value. Since a Binomial variable here can be seen as a sum of independent Bernoulli(ρ) random variables, we can apply Hoeffding's inequality to get,

$$P(|K' - N\rho| > Nt) \leq 2\exp(-2Nt^2) \tag{10}$$

which is diminishing exponentially in N. This concentration effect gives us enough flexibility to extend the scheme to variable and adaptive thresholds, although we will not make these extensions in the present paper.

V. Number of Measurements

The minimum number of measurements required for data recovery depends on the error metric that we use. For simplicity, assume that x is generated by picking the highest K entries of an iid Exponential(γ) random vector of length N. Assume that the non-zero magnitudes are generated as above and fixed. The performance of the decoder that we consider does not rely on any particular order of appearance of these magnitudes. Let us consider a fixed order, but this is not revealed to the decoder in the beginning.

The measurement matrix A is generated with iid entries, each chosen according to some distribution Q. Suppose we are interested in recovering x based on M measurements as given in (2). This problem is closely connected to the support recovery of x. See [7] for the details. Let $\hat{x}$ be the recovered vector using some method. The error event for support recovery can be explained by means of a random variable Φ, which is defined as,

$$\Phi = (\left(\prod_{i:x_i=0} 1_{\{\hat{x}_i=0\}}\right) \cdot \left(\prod_{i:x_i\neq 0} 1_{\{\hat{x}_i\neq 0\}}\right)) \tag{11}$$

where $1_{\{.\}}$ is the binary valued indicator function, which is unity when the argument is true. Since the decoder performance does not favour any particular choice of the support of x, $P_{error} = P(\Phi = 0)$.

Lemma 1: If P_{error} goes to zero with M, then

$$M \geq \frac{K\log(N/K)}{I_Q(W)} \tag{12}$$

where W is a MAC channel defined by,

$$y = c_1 \sum_{i=1}^{K} u_i + z \;; z \sim \mathcal{N}(0, \sigma^2) \tag{13}$$

and $I_Q(\cdot)$ is the sum-rate obtained over this MAC using a distribution Q at each input, i.e., $u_i \sim Q, \forall i$. Furthermore, for given N, the constant c_1 is independent of K and can be taken to be $\sqrt{\gamma \log N + \delta}$, for any fixed $\delta > 0$.
Our proof is a specialization of the converse results in [7], where a compound MAC model to find the minimum required measurements is discussed in detail. In our case, there is a finite probability that N iid Exponential(γ) random variables $X_i, i = \{1, \cdot\cdot, N\}$ have their maximum less than $\gamma \log(N) + \delta$, for some fixed $\delta > 0$. To see this,

$$P(\max X_i > \alpha) \leq NP(X_i > \alpha) \quad (14)$$
$$= N \exp(-\frac{\alpha}{\gamma}), \quad (15)$$

which is strictly less than one when $\alpha > \gamma \log(N)$.

Assume the event that all the K non-zero coefficients had squared magnitude less than c_1. Then, the rate at which one can successfully communicate the support of x over the channel in (2), is bounded by the sum-rate of a MAC channel given in (13), see [7] for details. For this MAC, using the converse to the coding theorem we can write,

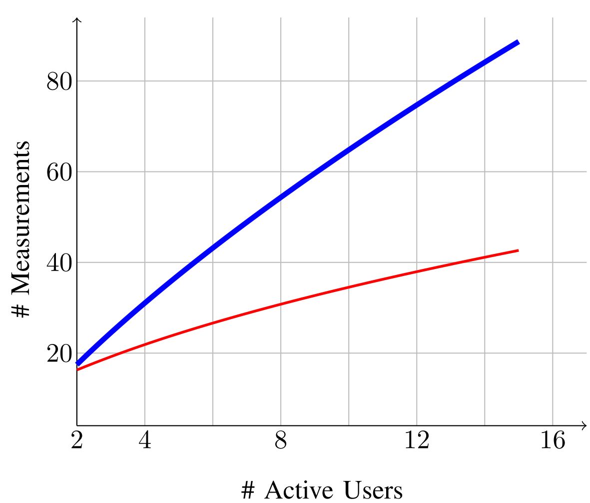
$$MI_Q(W) \geq K \log \frac{N}{K}. \quad (16)$$

In order to apply this result to our example, consider the event that K channel magnitudes are greater than a threshold τ, and each of them is less than c_1, which can be shown to have positive probability by choosing $c_1 = \tau + \gamma \log(N)$.

By choosing Q as Bernoulli($\frac{1}{2}$), the mutual information term in (12) can be bounded as [7]

$$I_{Bernoulli(\frac{1}{2})}(W) \leq \frac{1}{2} \log_2(\pi e \frac{K}{2}). \quad (17)$$

Thus we obtain a lower bound to the number of required measurements, which is plotted below and compared against (3) with $c = 1$.



We direct the interested reader to [7] where the number of required measurements as well as the speed of convergence is discussed.

VI. Simulation Study

We consider 1024 users, i.e., $N = 1024$. Let us assume an independent flat Rayleigh faded channel to each user. For illustration purposes, we generate an iid Gaussian random sequence of variance γ and zero mean. We set the threshold as a function of γ and the target number of users K. The figure shown below uses $\gamma = 0.25$ and the channel magnitude was compared against a threshold of 2.58. This is done so as to target the highest $5-15$ users. In all the plots of this section, the $x-$ axis shows the user index (active as well as inactive) $\{1, \cdot, \cdot, N\}$ and $y-$axis shows the channel magnitudes, actual (red color) value after thresholding, and reconstructions (blue color). The reconstruction used the GPSR routine [4] on 100 measurements taken at the base station through the uplink fading channel. The additive Gaussian noise was taken to have $\sigma^2 = 0.01$, corresponding to a situation where the pilot channel has power of a few dBs above the noise. The figure is obtained in one random run of the whole access phase. The

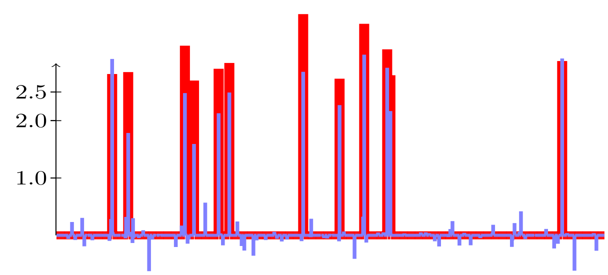


Fig. 1. Original Signal and Reconstruction from 100 measurements

bars in red represent the thresholded channel magnitudes, and those in blue correspond to the obtained reconstruction. The performance can be improved by putting more power in the pilot symbol. The following figure assumes $\sigma^2 = 10^{-4}$. Notice that the reconstruction is near perfect, and all the

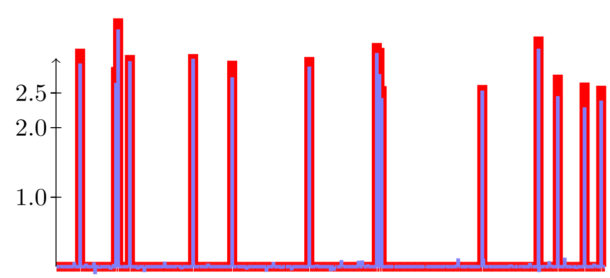


Fig. 2. Original Signal and Reconstruction from 100 measurements

active users are tracked with the channel coefficients almost same as the original. This in turn suggests that we can accommodate more active users by increasing the pilot signal power. Another way to boost the performance is to request for more measurements using a common downlink channel. So the scheme has lot of flexibility to be considered as a practical access scheme.

It is instructive to compare the performance of the compressed sensing detector with standard approaches in linear multi-user detection. As A has M rows and N columns with $M << N$, we cannot apply the standard decorrelator approach to decode the users. The other popular approach is the single user matched filter (SUMF). The matched filtering involves N matchings, one with each signature sequence. The performance is poor compared to the CS decoder, since there are a lot more possible users than the sequence length. Thus many sequences have high correlation among themselves, leading to lot of false detections. The matched filter output for the measurements corresponding to the data in Figure 2 is plotted below in Figure 3.

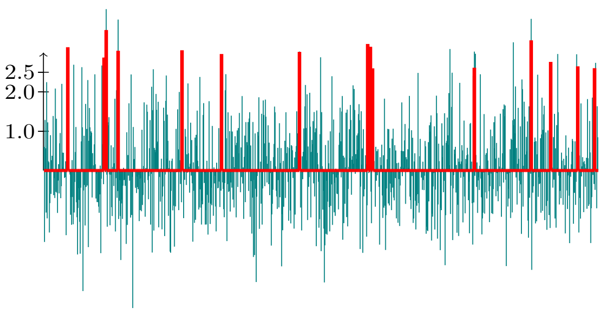


Fig. 3. Original Signal and MF output from 100 measurements

As expected, the SUMF will have reasonably high outputs for most of the active users, nevertheless, the output is significantly high for lot of inactive users too. This makes it difficult to decide which set of users are the active ones. The performance is not significantly improved by increasing the pilot power. This is due to the fact that interference from non-active users is the primary reason for getting false detections at the output of SUMF. To make it clear, notice that the received symbols can be written as,

$$y = \sum_{i \in \mathcal{K}'} a_i |h_i| + z \tag{18}$$

where a_k is the k^{th} column of A. For the matched filter corresponding to user k, by defining $a_k^{\dagger} = (a_k^T)^*$, we get,

$$a_k^{\dagger} y = |h_k| I[k \in \mathcal{K}'] + \sum_{\substack{i \in \mathcal{K}' \\ i \neq k}} a_k^{\dagger} a_i |h_i| + a_k^{\dagger} z. \tag{19}$$

Since the columns of A are not strictly orthogonal, the interference from other users is quite high with a large number of users. Even if k is not active, the second interference term can dominate and give a false positive. The only way to improve the SUMF performance is to make the signatures nearly orthogonal, but this involves increasing the measurements to a level comparable to N. Clearly, if the number of measurements are much less than N, the compressed sensing detector provides much better performance than the SUMF receiver.

VII. Final remarks and possible extensions

Our scheme can easily be generalized to systems with multiple antennas (MIMO systems). In this case, the channel state feedback problem becomes even more onerous, and our compressed sensing approach even more useful. Recently, [1] proposed methods based on quantized feedback of the channel-coefficients. The mobiles are assumed to have perfect knowledge of their respective channel coefficients. The feedback is in the form of channel direction indicator (CDI) and channel quality (magnitude) indicator(CQI). Using this feedback, K users are selected to maximize the sum-rate under zero-forcing beamforming (ZFBF) at the transmitter. Here the user-selection is centralized and inherently coupled to the utility maximization. In particular, each user needs to correctly send the CQI and CDI to the base station. The base station now performs a sub-optimal (still not far from optimal) low complexity search to select the set of K users. The remaining users and the information they sent are discarded. This operation is repeated over each block. Clearly, there is a lot of protocol overhead involved, and even more, most of the overhead information is discarded at the end.

Our approach is to partially decouple user-selection and utility maximization. While there is a price paid due to this decoupling, we argue that the cost is low. Even in the scheme of [1], it is clear that we can improve the performance by first choosing a subset of K users and then improving the quality of their channel estimates. This leads to more overhead, but the extra load is only for the selected K users, which is much less of a burden. This is the approach we take in the present paper. Our method can also be applied to the Multiple Access Channel (MAC), and of particular importance is the random MAC, where K out of N users choose to transmit in a time slot. Substantial overhead reduction can be achieved in this case by using our techniques for user-identification, but this is beyond the scope of the present paper.

Acknowledgements

We thank Arif Khan for explaining us the broadcast user-selection problem. The suggestions by Jamie Evans are gratefully acknowledged.

References

[1] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, September 2007.

[2] R. Knopp and P. A. Humbler, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, June 1995, pp. 331–335.

[3] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, no. 12, pp. 4203–4215, December 2005.

[4] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing*, vol. 1, no. 4, pp. 586–598, 2007.

[5] E.F. Chaponniere; P.J. Black; J.M. Holtzman; D.N. Tse, "Transmitter directed code division multiple access system using path diversity to equitably maximize throughput" *U.S. Patent 6449490*, September 10, 2002.

[6] R. Ahlswede and G. Dueck,"Identification via Channels ", *IEEE Trans. Info. Theory*,vol. 35, no. 1, pp. 15–29, Jan 1989.

[7] P. Tune , S. R. Bhaskaran and S. Hanly, "Number of Measurements in Sparse Signal Recovery", International Symposium on Information Theory 2009.