Sparse Multi-User Detection for CDMA Transmission Using Greedy Algorithms

Henning F. Schepker and Armin Dekorsy
Department of Communications Engineering
University of Bremen, Bremen, Germany
email: {schepker, dekorsy}@ant.uni-bremen.de

***Abstract*—A possible future application in communications is the wireless uplink transmission in sensor networks. This application is mainly characterized by sporadic transmission over a random access channel. Since each sensor has a low activity probability, the signal for Multi-User Detection (MUD) is sparse. Compressive Sensing (CS) theory introduces detectors that are able to recover sparse signals reliably. When applied to MUD, CS detectors perform a joint detection of both data and user activity. This allows for less control signaling, since information about the activity state of each user no longer need to be signaled. Additionally, CS detectors are able to reliably detect sparse signals even in under-determined systems. In the context of transmission with Code Division Multiple Access (CDMA) this property can be exploited by reducing the length of spreading sequences, which increases the symbol-rate for a given bandwidth, and making the CDMA system overloaded. In this paper we introduce the application of greedy CS detection algorithms to detect CDMA spread multi-user data.**

I. Introduction

Reconstruction of sparse signals has become the topic of much research recently, especially after Compressive Sensing (CS) theory, [1], [2], gained a lot of attention in the fields of applied mathematics and image processing. CS theory shows that reliable signal reconstruction far below the Nyquist sampling rate is possible provided that the signal is sparse. In communications, CS theory has been applied for instance to system parameter estimation [3] and channel estimation [4].

Another application of CS in communications is motivated by the future challenge of designing efficient wireless sensor networks, where the sensor nodes transmit to a single aggregation node for data collection. Such a transmission scenario is characterized by the sporadic communication of a large amount of sensors, i.e. each sensor has a low activity probability. In order to avoid control signaling each sensor transmits whenever it has new data without first sending control signals, i.e. the activity state of each sensor is not known at the receiver and has to be detected. From a detection perspective, the overall system has a large dimension, but only few sensors or users transmit at any given time.

Multi-User Detection (MUD) is a powerful technique to overcome the multi-access interference present in Code Division Multiple Access (CDMA) systems, e.g. [5], and has also been intensely investigated in the past for other commercial communication systems such as UMTS [6]. MUD algorithms simultaneously detect the transmitted symbols of all active users. However, these algorithms require knowledge about which user is active, leading to a large control signal overhead. The challenge is to design MUD algorithms that exploit user inactivity, while not requiring control information about each user's activity state.

In [7] the authors have introduced CS detectors for MUD in order to exploit the inactivity of users. These detectors are based on convex optimization and are well known in regression problems [8], [9]. Additionally the authors introduced a sparsity-promoting sphere decoder. However, the algorithms in [7] are designed for a fully loaded CDMA systems, i.e. the length of the spreading sequences is equal to the total number of users. Therefore they do not exploit the potential of CS theory to reliably solve under-determined linear equation systems by utilizing the knowledge that the signal is sparse. We can classify the MUD algorithms in [7] as sparsity aware, but not sparsity exploiting.

Besides convex optimization, [8], [9], CS also offers sparsity exploiting greedy algorithms, [10], [11]. The main advantage of these greedy algorithms is that they are less complex compared to convex optimization, [12]. However, it is still unknown how CS greedy algorithms perform for MUD for CDMA transmission. The contribution of this paper is to investigate greedy CS detection algorithms for joint detection of activity and transmitted data. In particular the algorithms Orthogonal Matching Pursuit (OMP) [10] and Orthogonal Least Squares (OLS) [11] will be investigated.

II. System Model and Problem Statement

Consider a wireless uplink transmission from K users to a single aggregation node using CDMA [5, Chap. 1], as shown in Figure 1. Herein, we assume that the transmission is structured in frames, each of a length of L chips, such that whenever a user transmits data it transmits an entire frame. Therefore, the activity of a user is defined for an entire frame. Furthermore, we assume that the number of active users K_a follows a discrete binomial distribution $\mathcal{B}(K, p_a)$, where $p_a \ll 1$ is the probability that the k^{th} user is currently active. This means that the activity of users is statistically independent across both users and frames. We further assume that the activity states of all users are *not* signaled to the receiver, rather they have to be detected by the receiver.

The spreading factor, i.e. the length of the spreading sequence, for the k^{th} user is denoted by N_k and the k^{th} user transmits $\lfloor L/N_k \rfloor$ symbols during a frame of L chips. In

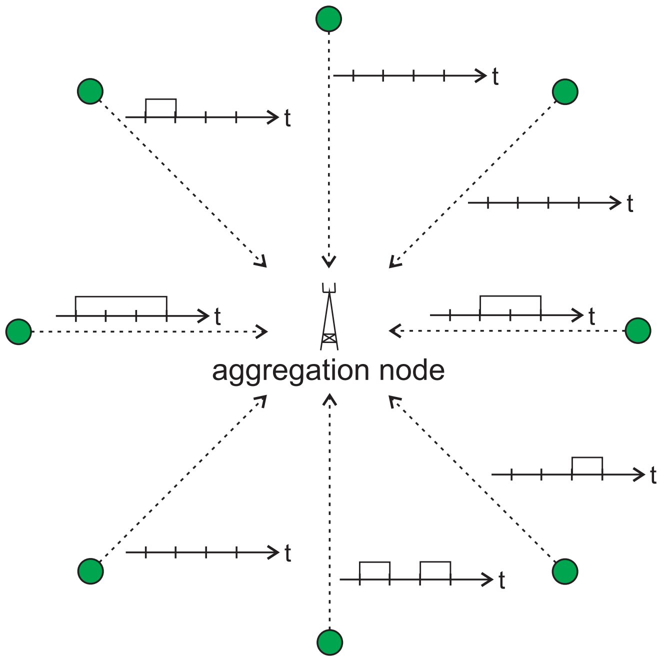

Fig. 1. Sporadic transmission from K users to a single aggregation node. Each time step corresponds to an entire frame duration. Each transmission is received with a relative delay of τ_k.

our signal model the symbol vector $\mathbf{x}$ contains the symbols of all users for a given frame, regardless of their activity states, and has a length of M, with $M = \sum_{k=1}^{K} \lfloor L/N_k \rfloor$. Since the symbols of each user are either all zero or with probability $p_a \ll 1$ all non-zero, the vector $\mathbf{x}$ is sparse. Due to a channel impulse response with a length of $L_{h,k}$ chips for the k^{th} user and a relative delay of τ_k chips for the k^{th} user, the L transmitted chips are received as L' chips, where $L' = L + \max_k (L_{h,k} - 1 + \tau_k)$.

For the symbol vector $\mathbf{x} \in \mathbb{C}^{M\times 1}$ of any given frame, the received vector $\mathbf{y}$ containing L' chips is given by the input-output relationship in the chip-rate model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}\,, \tag{1}$$

where $\mathbf{n} \in \mathbb{C}^{L'\times 1}$ is additive white Gaussian noise (AWGN), i.e. $\mathcal{N}(0, \sigma_{\mathbf{n}}^2)$, and $\mathbf{A} \in \mathbb{C}^{L'\times M}$ is a matrix representing the influence of the spreading sequences and the instantaneous channels for the current frame on the transmitted symbols.

We assume perfect channel state information, so that the matrix $\mathbf{A}$ is perfectly known at the receiver. The channel for each user has a multi-tap power delay profile and is time invariant within the duration of a frame, i.e. block fading channel. In addition we assume that the instantaneous channels are independent across users. In this transmission scenario each column of $\mathbf{A}$ contains the spreading sequence $\mathbf{s}_k$ of the k^{th} user convolved with the instantaneous channel impulse response $\mathbf{h}_k$ of the k^{th} user [6].

$$\mathbf{A} = \begin{pmatrix} r_{1,1,1} & 0 & 0 & r_{2,1,1} & 0 & 0 \\ r_{1,1,2} & 0 & 0 & r_{2,1,2} & 0 & 0 \\ r_{1,1,3} & r_{1,2,1} & 0 & r_{2,1,3} & r_{2,2,1} & 0 \\ 0 & r_{1,2,2} & 0 & 0 & r_{2,2,2} & 0 \\ 0 & r_{1,2,3} & r_{1,3,1} & 0 & r_{2,2,3} & r_{2,3,1} \\ 0 & 0 & r_{1,3,2} & 0 & 0 & r_{2,3,2} \\ 0 & 0 & r_{1,3,3} & 0 & 0 & r_{2,3,3} \end{pmatrix} \tag{2}$$

For illustration purposes, (2) shows matrix $\mathbf{A}$ in the case of two synchronous users and a frame length of $L = 6$ chips. The spreading factors for both users are $N_1 = N_2 = 2$ and both user transmit $\lfloor L/N_k \rfloor = 3$ symbols. The channels for both users have an impulse response of length $L_{h,1} = L_{h,2} = 2$ and the relative delays are $\tau_1 = \tau_2 = 0$, and thus $L' = 7$. Here $r_{k,l,i}$ denotes the i^{th} chip of the received sequence of the l^{th} symbol of the k^{th} user.

For CDMA transmission using a given bandwidth the transmitted symbol-rate is determined by the spreading factors N_k. If $N_k = N \;\forall\, k$ and $N < K$ we term the CDMA system to be overloaded with regard to the number of total users K, or from a CS perspective say the linear equation system (1) is under-determined.

III. Multi-User Detection with Compressive Sensing

CS enables the reliable detection of sparse signals even in under-determined systems [1], [2]. This section shall only provide a brief overview of those topics relevant to the content of this paper. For a thorough overview of CS see e.g. [13].

The CS detection problem can be written in the form of (1), where the matrix $\mathbf{A}$ and the noisy measurement $\mathbf{y}$ are given. The receiver in general knows neither the amount nor the position of the non-zero elements in the sparse vector $\mathbf{x}$. The sparsity of $\mathbf{x}$ is measured by the number of non-zero elements

$$S_{\mathbf{x}} = \|\mathbf{x}\|_0 = |\{j : x_j \neq 0\}|\,. \tag{3}$$

Different CS detectors have been proposed that mainly fall into two categories: either convex optimization, or greedy algorithms. In CS theory (1) is commonly solved by means of convex optimization, e.g. Basis Pursuit De-Noising (BPDN) [8] or Least Absolute Shrinkage and Selection Operator (LASSO) [9]. While the convex optimization is performed over $\mathbb{C}^M$, the values of $\mathbf{x}$ in our system model are from the augmented alphabet $\mathcal{A}_a = \{\mathcal{A} \cup 0\}$, where $\mathcal{A}$ is the modulation alphabet, e.g. $\mathcal{A} = \{-1, 1\}$ for Binary Phase Shift Keying (BPSK). Therefore, a discrete version of the LASSO optimization problem has been proposed in [7]. This optimization problem is given by

$$\hat{\mathbf{x}}_{\text{LASSO}} = \arg\min_{\mathbf{x}\in\mathcal{A}_a^L} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1\,, \tag{4}$$

where $\|\mathbf{x}\|_1 = \sum_{i=1}^{M} |x_i|$. The correct value of the regularization parameter λ is explained in detail in [7]. The optimization problem (4) can for example be solved with the method of sphere decoding.

Algorithm 1 Orthogonal Matching Pursuit (OMP)

$\mathbf{r}^0 = \mathbf{y}$, $\Gamma^0 = \emptyset$, $l = 0$
repeat
$l = l + 1$
$i_{\max} = \arg\max_i \left|\mathbf{A}_i^H \mathbf{r}^{l-1}\right|$ for $i \in \overline{\Gamma}^{l-1}$
$\Gamma^l = \Gamma^{l-1} \cup i_{\max}$
$\hat{\mathbf{x}}^l_{\Gamma^l} = \mathbf{A}^\dagger_{\Gamma^l}\mathbf{y}$ **and** $\hat{\mathbf{x}}^l_{\overline{\Gamma}^l} = \mathbf{0}$
$\mathbf{r}^l = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^l$
until $l = S_{\mathbf{x}}$ **or** $\left\|\mathbf{A}^H\mathbf{r}^l\right\|_\infty < \varepsilon$

Algorithm 2 Orthogonal Least Squares (OLS)

$\Gamma^0 = \emptyset$, $l = 0$
repeat
$l = l + 1$
$i_{\min} = \arg\min_i \left\|\mathbf{y} - \mathbf{A}_{\Gamma_i^l}\mathbf{A}^\dagger_{\Gamma_i^l}\mathbf{y}\right\|_2$ with $\Gamma_i^l = \Gamma^{l-1} \cup i$
$\Gamma^l = \Gamma^{l-1} \cup i_{\min}$
$\hat{\mathbf{x}}^l_{\Gamma^l} = \mathbf{A}^\dagger_{\Gamma^l}\mathbf{y}$ **and** $\hat{\mathbf{x}}^l_{\overline{\Gamma}^l} = \mathbf{0}$
$\mathbf{r}^l = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^l$
until $l = S_{\mathbf{x}}$ **or** $\left\|\mathbf{A}^H\mathbf{r}^l\right\|_\infty < \varepsilon$

Apart from convex optimization, several greedy algorithms exist in CS theory [10], [11] that can be used for out detection concept to jointly detect user activity and transmitted data. The advantage of these algorithms is that they are less complex compared to convex optimization, e.g. [12], but have the drawback of error propagation, since previous choices for user activity are not re-evaluated. As these algorithms perform data estimation on $\mathbb{C}^M$ we must perform a subsequent symbol decision on the augmented alphabet $\mathcal{A}_a$. In the following two greedy CS algorithms will be explained in detail.

A. Orthogonal Matching Pursuit

The basic concepts of greedy CS algorithms are introduced by the Orthogonal Matching Pursuit (OMP) algorithm [10], [14]. Several other greedy algorithms exist that are based on the OMP, such as Compressive Sampling MP (CoSaMP), Stagewise OMP (StOMP) and Probabilistic OMP (PrOMP) [12], [15], [16]. We herein use the OMP algorithm, since it is commonly used as a comparison.

For the algorithms the following notation is used: Γ is a set of indices for columns in matrix $\mathbf{A}$ and $\overline{\Gamma}$ is the complementary set containing all valid indices not in Γ. Furthermore $\mathbf{A}_\Gamma$ specifies the sub-matrix that only contains those columns with indices in Γ, and likewise $\mathbf{x}_\Gamma$ the vector containing only the elements of $\mathbf{x}$ with indices in Γ. Additionally $\mathbf{x}^l$, $\mathbf{A}^l$ and Γ^l specify the respective variable during the l^{th} iteration. Here, $\mathbf{A}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$ and $\mathbf{A}^H$ the Hermitian matrix of $\mathbf{A}$. With this notation the OMP algorithm is given by Algorithm 1.

During each iteration, the OMP algorithm chooses the column of $\mathbf{A}$ that has the highest correlation to the previous residual $\mathbf{r}^{l-1} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{l-1}$. Then the orthogonal projection into the sub-space not dependent on any of the columns in Γ^l is performed and the current residual $\mathbf{r}^l$ is calculated. The next iteration continues with residual $\mathbf{r}^l$. As the subsequent iteration depends on the previous residual $\mathbf{r}^{l-1}$, the performance of the OMP is highly influenced by both linear dependencies between columns of $\mathbf{A}$ and the additive noise $\mathbf{n}$.

The OMP performs these iterations until a stopping criterion is met. Ideally the algorithm is terminated after a number of iterations equal to a known sparsity $S_{\mathbf{x}}$, i.e. $l = S_{\mathbf{x}}$. However, since the sparsity is in general not known at the receiver, a stopping criterion based on the current residual $\mathbf{r}^l$ is used instead

$$\left\|\mathbf{A}^H\mathbf{r}^l\right\|_\infty < \varepsilon \tag{5}$$

for appropriately chosen ε [17].

B. Orthogonal Least Squares

Instead of using correlations to determine the support, i.e. the non-zero elements of $\mathbf{x}$ or the active users, the Orthogonal Least Squares (OLS) algorithm uses Least Square (LS) estimation of sub-spaces [11], [18], as given in Algorithm 2. During each iteration and for each possible choice of a column i the OLS performs the orthogonal projection $\mathbf{A}_{\Gamma_i^l}\mathbf{A}^\dagger_{\Gamma_i^l}\mathbf{y}$ of $\mathbf{y}$ into the sub-space $\mathbf{A}_{\Gamma_i^l}$ given by the previous choices Γ^{l-1} and the additional column i. Then OLS calculates the resulting residual $\mathbf{r}_i^l$ for each choice i and chooses the column $i_{\min}$ with the minimal Euclidian norm of the residual. Note that LS estimation, i.e. orthogonal projection, in the selection criterion leads to the minimal residual for each column i.

Note that both OMP and OLS are structurally similar to Successive Interference Cancellation (SIC) [5, pp. 344]. However, both OMP and OLS perform a joint detection of both user activity and transmitted data, while the SIC iteratively detects transmitted symbols only. Another difference to the SIC [5, pp. 344] is that both OMP and OLS return results over the continuous alphabet $\mathbb{C}^M$ for the transmitted data according to the LS estimation, while the SIC returns discrete symbols from $\mathcal{A}$. For the OLS the final result $\hat{\mathbf{x}}^l$ is estimated in the last iteration l and previous estimations do not directly affect the detection of the support of $\mathbf{x}$, i.e. the activity of the users. Thus, we can state that the OLS is a two stage detector, which first detects the support of $\mathbf{x}$ and then performs LS estimation of the transmitted data, as the choice of column $i_{\min}$ is not dependent on either $\hat{\mathbf{x}}^{l-1}$ or $\mathbf{r}^{l-1}$. For the OMP these two stages are executed in alternating order, since previous data estimations $\hat{\mathbf{x}}^{l-1}$ influence the detection of the support, due to dependency on the last residual $\mathbf{r}^{l-1}$.

IV. Simulation Results

In these simulations Pseudo Noise (PN) sequences with $N_k = N \;\forall\, k$ chips are used for each user, and the data bits are modulated with Binary Phase Shift Keying (BPSK). As a model for the multi-path channel six consecutive Rayleigh distributed coefficients with exponentially decaying power profile are used. The symbol-error-rate (SER) simulations are

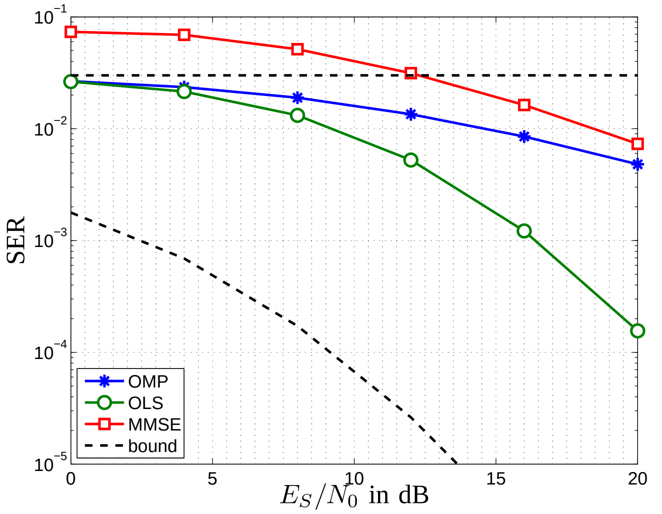

Fig. 2. SER simulation results for $K = 128$, $L = 256$, $N = 128$ and $p_a = 0.015$.

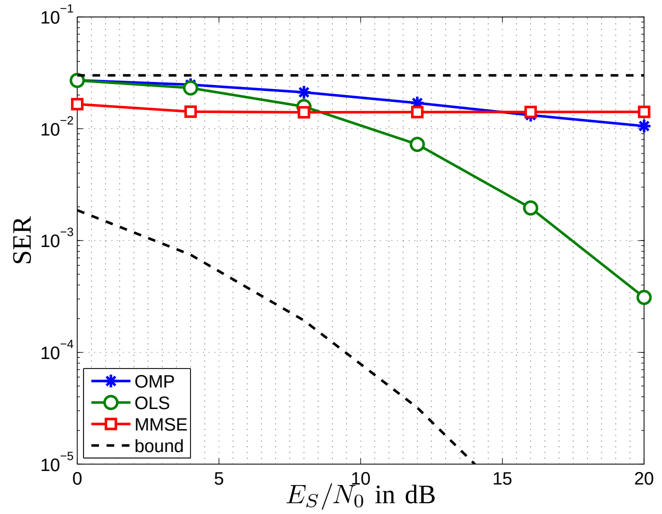

Fig. 3. SER simulation results for $K = 128$, $L = 256$, $N = 32$ and $p_a = 0.015$.

performed for a set of $K = 128$ users, where each user is active with a probability of $p_a = 0.015$, so that for example with a probability of $\Pr\{K_a \leq 6\} = 0.9967$ at most six users are active. The SER is measured over the entire vector $\mathbf{x}$ and contains both support errors, i.e. incorrect activity detection, and bit errors, i.e. incorrect data detection. In these simulations it is assumed that the frames of all users arrive synchronously at the receiver, such that the relative delay for all users is $\tau_k = 0\,\forall\, k$.

Here, the greedy algorithms terminate after $l = S_{\mathbf{x}}$ iterations. In general $S_{\mathbf{x}}$ is unknown at the receiver and needs to be signaled or estimated. For these simulations, we assume perfect knowledge of $S_{\mathbf{x}}$, as the simulation results for perfectly known $S_{\mathbf{x}}$ serve as a best case performance for other stopping criteria, such as (5).

Herein, whenever a user is active that user transmits with an average energy per symbol of $E_S = 1$. Therefore, the signal-to-noise ratio at the transmitter is $E_S/N_0 = 1/\sigma^2_{\mathbf{n}}$, where $\sigma^2_{\mathbf{n}}$ is the variance of the additive noise term.

For a known sparsity $S_{\mathbf{x}}$, both OMP and OLS return exactly $S_{\mathbf{x}}$ non-zero elements. Therefore, at most $2 \cdot S_{\mathbf{x}}$ support errors can occur and the SER for both OMP and OLS is upper-bounded by $\frac{2 \cdot S_{\mathbf{x}}}{M}$, assuming $2 \cdot S_{\mathbf{x}} \leq M$. As the sparsity is not constant in these simulations but determined by a discrete binomial distribution, an upper bound can be given by $\frac{2 \cdot \mathrm{E}\{S_{\mathbf{x}}\}}{M}$, where $E\{S_{\mathbf{x}}\} = E\{K_a\} \cdot \lfloor L/N \rfloor$ is the expected value of the sparsity $S_{\mathbf{x}}$. Therefore, with $E\{K_a\} = K \cdot p_a$, an upper bound of the SER is

$$\mathrm{SER} \leq \frac{2 \cdot \mathrm{E}\{K_a\} \cdot \lfloor L/N \rfloor}{K \cdot \lfloor L/N \rfloor} = 2 \cdot p_a\,. \tag{6}$$

Fig. 2 and 3 show simulation results for two values of the spreading factor, $N = 128$ and $N = 32$. In both figures the OLS has a much lower SER compared to the OMP. This is due to a better selection criterion, i.e. the minimal norm residual instead of the highest correlation, with the drawback of higher complexity. During iteration l, the OLS has to perform $M - l + 1$ different LS estimates using different sub-matrices as opposed to the same amount of vector products for the OMP.

Fig. 3 illustrates that especially the OLS has a low SER for a spreading factor much smaller than the number of users, i.e. $N < K$. In principle, this shows that the OLS is able to recover sparse data even in overloaded CDMA systems. When comparing the results for $\frac{N}{K} = 1$ (Fig. 2) and $\frac{N}{K} = \frac{1}{4}$ (Fig. 3), the SER only increases slightly for the OLS when reducing the spreading factor N.

In the figures we also included the results for Minimum Mean Square Error (MMSE) estimation [5, pp. 291] of (1) as a benchmark. At the output of the MMSE estimation a MAP symbol detector over the trinary alphabet $\mathcal{A}_a$ is applied. Here, for the MAP detector it is assumed that the a-priori probabilities of all symbols in alphabet $\mathcal{A}_a$ are the same. However, since for $p_a < 0.5$ on average more elements of $\mathbf{x}$ are zero than non-zero, the a-priori probabilities have to be changed according to p_a.

As shown in Fig. 2, the MMSE has a higher SER than both greedy algorithms in the fully loaded case for the range of E_S/N_0 shown here. For the overloaded CDMA system in Fig. 3 the MMSE has a better SER performance than both greedy algorithms for the low E_S/N_0 regime. This is due to the fact that the MMSE is aware of the noise and therefore is able to better adapt to it. However, as the MMSE is not aware of the sparsity in $\mathbf{x}$ it has a high error floor.

To determine a lower bound for the SER, we introduce an oracle Matching Pursuit (MP) algorithm that has perfect knowledge of the position of the non-zero elements in $\mathbf{x}$, due to an oracle process. Thus, the oracle MP algorithm performs the LS estimation using $\mathbf{y}$ and the submatrix $\mathbf{A}_{\Gamma_{\text{oracle}}}$, where Γ_{oracle} are the indices of the non-zero elements of the transmitted vector $\mathbf{x}$, i.e. $\Gamma_{\text{oracle}} = \{i : x_i \neq 0\}$.

Both Fig. 2 and 3 illustrate that the results for both OMP

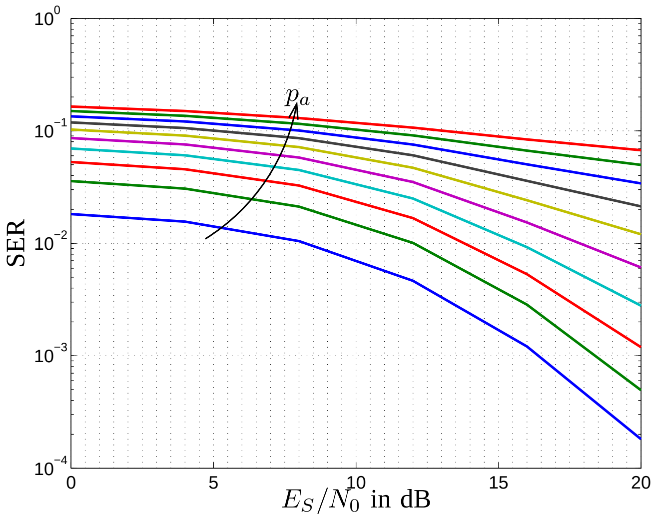

Fig. 4. SER simulation results using OLS for $K = 128$, $L = 256$, $N = 32$ and values of p_a from 0.01 to 0.1.

and OLS are significantly above the lower bound given by the oracle MP algorithm. As the oracle MP always uses the correct support for the LS estimation, the only errors occurring for the oracle MP are bit errors. This shows that the errors occurring for both OMP and OLS are primarily support errors and that bit errors are only a small portion of the total symbol errors for both OLS and OMP.

Finally fig. 4 shows SER results for the OLS in the overloaded case for different values of p_a. These results show that the concepts presented here are not only limited to small values of p_a, but rather that a low p_a improves the SER results. While a high p_a decreases the SER performance for the OLS, it will likely improve the performance for a MMSE, as the signal is less sparse on average.

V. Conclusion

In this paper we studied CS-MUD which jointly detects transmitted data and user activity. As CS detectors we applied the greedy algorithms OMP and OLS. The proposed solution can be applied for sporadic communication, e.g. sensor networks. The main results are that the OLS has a lower symbol-error-rate than the OMP in all simulations, and that the OLS is suitable for overloaded CDMA systems. However, the main weakness of these algorithms is in finding the correct support of the multi-user vector. This motivates further research. It should be noted that these results are not limited to CDMA transmission only. Sparse signal detection can also be applied to other technologies, such as multi-layer MIMO transmission.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[3] D. Angelosante, E. Grossi, G. B. Giannakis, and M. Lops, "Sparsity-aware estimation of CDMA system parameters," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–10, April 2010.

[4] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, November 2010.

[5] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, November 1998.

[6] A. Dekorsy and S. Brueck, "Is multiuser detection beneficial to mixed service UMTS networks?" *AEU - International Journal of Electronics and Communications*, vol. 59, no. 8, pp. 473–482, December 2005.

[7] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454–465, February 2011.

[8] S. S. Chen, D. L. Donoho, and M. A. Sanders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[10] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Signals, Systems and Computers*, vol. 1, pp. 40–44, November 1993.

[11] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, September 1989.

[12] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

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

[14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[15] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Stanford University, Stanford, CA, USA, Tech. Rep., 2006.

[16] A. Divekar and O. Ersoy, "Probabilistic matching pursuit for compressive sensing," Purdue University, West Lafayette, IN, USA, Tech. Rep., May 2010.

[17] J. A. Tropp, "Average-case analysis of greedy pursuit," in *Proc. SPIE Wavelets XI*, July 2005.

[18] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," University of Edinburgh, Edinburgh, Scotland, U.K., Tech. Rep., March 2007.