Compressive Sensing Multi-User Detection with Block-Wise Orthogonal Least Squares

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

***Abstract*—One challenging future application in digital communications is the wireless uplink transmission in sensor networks. This application is characterized by sporadic transmissions by a large number of sensors over a random multiple access channel. To reduce control signaling overhead, we propose that sensors do not transmit their activity states; instead sensor activity is detected at the receiver. As sensors have low activity probabilities, the multi-user vector is in general sparse. This enables Compressive Sensing (CS) detectors to perform joint Multi-User Detection (MUD) of activity and data, by exploiting the sparsity. Since sensors are either active or inactive for several symbol durations, block-wise CS detection can be applied to improve the activity detection. In this paper, we introduce block-wise greedy CS MUD, compare it to symbol-wise greedy CS MUD, and show that statistically independent channels for each symbol further improve the activity detection for block-wise CS detection. Herein, we use Code Division Multiple Access (CDMA) as a multiple access scheme.**

I. Introduction

A future challenge in digital communications is designing efficient wireless sensor networks, where sensors transmit to a single aggregation node for data fusion. In general, sensors do not continuously transmit data, but rather infrequently transmit updates, whenever a measured value changes. Therefore, we can characterize the scenario as a sporadic transmission by multiple sensors over a random multiple access channel. Ideally, the sensors should only use a small amount of energy for transmissions. One possibility to achieve this is to avoid the exchange of control signals, whenever a sensor changes its activity state. This reduces both the signal processing complexity and the required transmit energy at the sensors. However, the activity state of each sensor is then no longer known at the receiver and has to be estimated. Thus, the tradeoff is additional complexity at the receiver, as it has to detect both the activity state of each sensor and the transmitted data of the active sensors. In the following, we will use the more general term "users" instead of "sensors".

To detect transmitted data over a random multiple access channel, Multi-User Detection (MUD) is commonly applied [1]. Existing MUD algorithms detect data of active users only. Therefore, the activity states of the users need to be known at the receiver. In order to achieve this, a user usually signals a transmission request to the aggregation node, which then needs to acknowledge this request and specify which resource the user may use, e.g., spreading sequence for Code Division Multiple Access (CDMA). If control signaling is avoided, each user uses a fixed transmission resource and the user activity has to be detected at the receiver. If we assume that the inactivity of a user is modeled as "transmitting" zero symbols, then the vector containing the transmitted symbols of all users is sparse, due to the low activity probability of each user. In general, sparse signals can be detected by applying Compressive Sensing (CS) theory [2], [3]. CS detectors enable joint detection of both user activity and transmitted data in CDMA systems, which are overloaded in regard to the total number of users [4].

The concept of sparse MUD for CDMA transmission [1, Chap. 1] using CS detectors has been introduced in [5]. The authors of [5] introduced CS MUD detectors that are based on convex optimization, using l_1-norm and l_2-norm minimization, and further extended the ideas to sparse sphere decoding. Aiming to reduce the complexity, [4] introduced CS MUD using greedy CS algorithms. In [4] we have shown that efficient greedy CS algorithms can reliably detect both activity and data in a CDMA transmission. The results in [4] indicate that the Orthogonal Least Squares (OLS) algorithm [6] is reliable for sparse MUD even in overloaded CDMA systems. However, it was shown in [4] that the symbol errors of the OLS are mainly caused by incorrect activity detection. This indicates that a feasible approach to improve the overall performance is to improve the activity detection.

In this paper, we introduce Block-wise Orthogonal Least Squares (BOLS) detection as sparse MUD for sporadic communication using CDMA. We verify the performance of joint activity and data detection and compare block-wise greedy detection with symbol-wise greedy detection (OLS). Additionally, we show that statistically independent channels for each transmitted symbol significantly improve the performance of block-wise greedy detection.

II. System Model

Consider a sporadic wireless uplink transmission from K users, e.g., sensors, to a single aggregation node, as shown in Figure 1. For this scenario, we assume that the transmission is organized in time frames, such that whenever a user transmits data, it transmits data for the duration of an entire frame. This means that the users can change their activity state on a per frame basis. Due to sporadic transmissions, each user does not transmit continuously, but rather each user is active only during a few frames. We assume that on average each user is active

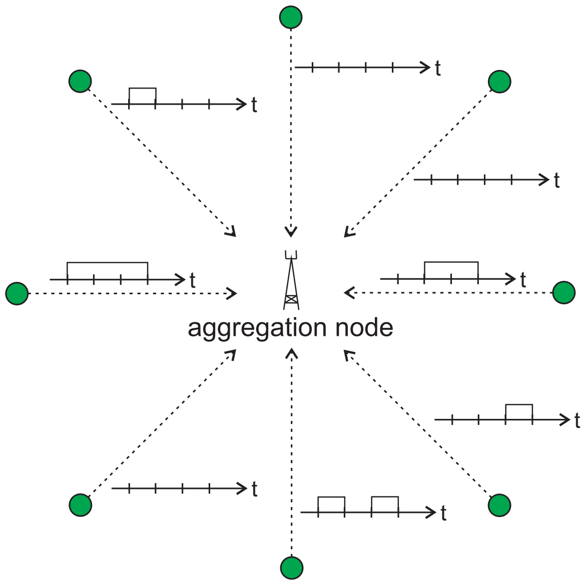

Fig. 1. Sporadic transmission from K users to a single aggregation node. Each time step corresponds to an entire frame duration. Each transmission is received with a relative delay of τ_k.

with a probability of $p_a \ll 1$, and that the activity of users is i.i.d. across both users and frames. Thus, we assume that the number of active users K_a for a frame follows a discrete binomial distribution $\mathcal{B}(K, p_a)$.

We define, that vector $\mathbf{x}$ contains the transmitted symbols of all K users during a frame. Additionally, we set the symbols for inactive users to zero and assume that all symbols of active users are taken from the discrete modulation alphabet $\mathcal{A}$. Thus, the elements of $\mathbf{x}$ are elements of the discrete augmented alphabet $\mathcal{A}_a = \{\mathcal{A} \cup 0\}$. Consequentially, vector $\mathbf{x}$ is sparse, as the symbols of each user are either all zero, or, with probability $p_a \ll 1$, all taken from the modulation alphabet $\mathcal{A}$. As the activity is defined per user, the symbols of the k^{th} user are either all zero or all non-zero. Therefore, all zero and non-zero elements of $\mathbf{x}$ appear in blocks of fixed length only. Thus, the vector $\mathbf{x}$ is not only sparse but also block sparse [7].

In this scenario, we apply CDMA [1, Chap. 1] as a channel access method for the uplink transmission. For this access method, we assume that all frames have a fixed length of F chips. We define L_k as the number of symbols, which the k^{th} user transmits during a frame. For the k^{th} user, L_k is determined by the spreading factor N_k, i.e., the length of the spreading sequence, with $L_k(N_k) = \lfloor F/N_k \rfloor$. Due to a channel impulse response $\mathbf{h}_k$, with a length of $\ell(\mathbf{h}_k)$ chips, and a (relative) delay of τ_k chips for the k^{th} user, the F transmitted chips are received as F' chips, where $F' = F + \max_k [\ell(\mathbf{h}_k) - 1 + \tau_k]$. The total number of symbols M is given by $M = \sum_{k=1}^{K} L_k$.

For the received chips $\mathbf{y} \in \mathbb{C}^F$ and the previously defined vector $\mathbf{x} \in \mathcal{A}_a^M$, the input-output relationship in the asynchronous chip-rate model is given by:

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

Here, $\mathbf{n} \in \mathbb{C}^F$ is additive white Gaussian noise, i.e., $\mathcal{N}_C(0, \sigma_{\mathbf{n}}^2)$. Each column of matrix $\mathbf{A} \in \mathbb{C}^{F' \times M}$ contains the spreading sequence $\mathbf{s}_k$ convolved with the current channel impulse response $\mathbf{h}_k$ of the k^{th} user [8].

For simplicity, we assume perfect channel state information at the receiver, and thus perfect knowledge of the matrix $\mathbf{A}$. Unless noted otherwise, we assume that the channels for all users are time invariant for the duration of a frame, i.e., we have block fading channels. Additionally, we assume i.i.d. channels across both users and frames.

For CDMA transmissions using a fixed bandwidth, the transmitted symbol-rate is determined by the spreading factor 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, state that the linear equation system (1) is under-determined.

III. Block Sparse Detection with Compressive Sensing

The main question for our system model is how the block sparsity of $\mathbf{x}$ can be efficiently exploited in joint detection of activity and data. To answer this question we will first look at how generic sparsity can be efficiently exploited in such a transmission.

In a nutshell, CS enables reliable detection of sparse signals, even in under-determined equation systems [2], [3]. The CS detection problem can be written in the form of (1), where the matrix $\mathbf{A}$ and the received chips $\mathbf{y}$ are given. Herein, the sparsity of $\mathbf{x}$ is determined by the number of non-zero elements

$$S_{\mathbf{x}} = \|\mathbf{x}\|_0 = |\{j : x_j \neq 0\}|\,. \tag{2}$$

In general, the detector knows neither the amount nor the position of the non-zero elements in the sparse vector $\mathbf{x}$. The CS theory defines different detectors that are able to recover the sparse solution of (1) even in underdetermined cases, i.e., $F' < M$. These detectors mainly fall into two categories: convex optimization and iterative greedy algorithms. Other approaches like the Iterative Hard Thresholding (IHT) algorithm [9] exist, but are not discussed here.

CS detectors using convex optimization are based on l_1-norm and l_2-norm minimization. The most commonly investigated detectors are Basis Pursuit De-Noising (BPDN) [10] and Least Absolute Shrinkage and Selection Operator (LASSO) [11]. Additionally, other variants exist [12], [13]. CS detectors using convex optimization have in common that the optimization is performed on the continuous alphabet $\mathbb{C}^M$. However, as previously stated in this transmission scenario vector $\mathbf{x}$ contains elements from the discrete augmented alphabet $\mathcal{A}_a = \{\mathcal{A} \cup 0\}$. For discrete alphabets $\mathcal{A}_a$, [5] introduces a LASSO based CS-MAP detector, which can be efficiently implemented with sphere decoding. However, the detectors in [5] are only sparsity aware and not sparsity exploiting, as they are not designed for the under-determined case of (1).

Greedy CS detectors are primarily variants of the Matching Pursuit (MP) algorithm [14]. The best known greedy CS

Algorithm 1 Orthogonal Least Squares (OLS)

$\Gamma^0 = \emptyset$, $l = 0$
repeat
  $l = l + 1$
  $i_{\min} = \arg\min_i \|\mathbf{y} - \mathbf{A}_{\Gamma_i^l}\mathbf{A}_{\Gamma_i^l}^\dagger \mathbf{y}\|_2$ with $\Gamma_i^l = \Gamma^{l-1} \cup i$
  $\Gamma^l = \Gamma^{l-1} \cup i_{\min}$
  $\hat{\mathbf{x}}_{\Gamma^l}^l = \mathbf{A}_{\Gamma^l}^\dagger \mathbf{y}$ **and** $\hat{\mathbf{x}}_{\overline{\Gamma}^l}^l = \mathbf{0}$
until $l = S_{\mathbf{x}}$

Algorithm 2 Block-wise Orthogonal Least Squares (BOLS)

$B^0 = \emptyset$, $l = 0$
repeat
  $l = l + 1$
  $i_{\min} = \arg\min_i \|\mathbf{y} - \mathbf{A}_{\Gamma(B_i^l)}\mathbf{A}_{\Gamma(B_i^l)}^\dagger \mathbf{y}\|_2$
  with $B_i^l = B^{l-1} \cup i$
  $B^l = B^{l-1} \cup i_{\min}$
  $\hat{\mathbf{x}}_{\Gamma(B^l)}^l = \mathbf{A}_{\Gamma(B^l)}^\dagger \mathbf{y}$ **and** $\hat{\mathbf{x}}_{\Gamma(\overline{B}^l)}^l = \mathbf{0}$
until $l = K_a$

algorithm is the Orthogonal Matching Pursuit (OMP) [15], with variants introduced in [6], [16]–[18]. The main difference to convex optimization is that greedy algorithms iteratively determine the position of the non-zero elements in $\mathbf{x}$ and perform data estimation for those non-zero elements. Moreover, greedy CS algorithms are in general less computationally expensive, compared to convex optimization [16], with the drawback of being less sampling efficient, i.e., greedy CS algorithms in general require a larger transmission overhead for reliable signal reconstruction [16].

To explain greedy CS algorithms, we use the following notation: 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 $\mathbf{x}_\Gamma$ contains only those 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. Herein, $\mathbf{A}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$, and $\mathbf{A}^H$ the Hermitian matrix of $\mathbf{A}$.

A. Orthogonal Least Squares

First, we will discuss symbol-wise greedy CS detection, as block-wise detection is based on symbol-wise algorithms. The Orthogonal Least Squares (OLS) algorithm [6], as given in Algorithm 1, iteratively determines the support of $\mathbf{x}$, i.e., the position of the non-zero elements in $\mathbf{x}$, by iteratively choosing columns of matrix $\mathbf{A}$. During each iteration and for each possible choice of a column i, OLS performs the orthogonal projection $\mathbf{A}_{\Gamma_i^l}\mathbf{A}_{\Gamma_i^l}^\dagger\mathbf{y}$ of $\mathbf{y}$ into the sub-space $\mathbf{A}_{\Gamma_i^l}$ given by all previous choices Γ^{l-1} and column i. Afterwards, OLS calculates the resulting residual $\mathbf{r}_i^l$ for each choice i, and then chooses the column $i_{\min}$, which yields the residual $\mathbf{r}_{i_{\min}}^l$ with the smallest Euclidian norm.

Greedy CS algorithms in general perform a joint detection of activity and transmitted data. The OLS, as a contrast, actually performs this detection in two stages. This is due to the fact that subsequent iterations only depend on the previous column choices Γ^{l-1}, but not the previous estimated vector $\hat{\mathbf{x}}^{l-1}$ or the previous residual $\mathbf{r}^{l-1}$. Thus, activity detection in later iterations is solely dependent on the activity detection of previous iterations, but not on the data estimations of previous iterations.

Ideally, the OLS executes exactly $S_{\mathbf{x}}$ iterations, as $S_{\mathbf{x}}$ non-zero elements are estimated after these iterations. However, the sparsity $S_{\mathbf{x}}$ of vector $\mathbf{x}$ is in general not known at the receiver, and therefore a different stopping criterion has to be used instead. Various stopping criteria exist, such as those discussed for the OMP in [19]. For the simulations, we assume perfect knowledge of $S_{\mathbf{x}}$ at the receiver. Therefore, we use $l = S_{\mathbf{x}}$ as a stopping criterion for the OLS. This assumption serves as a best case assumption for other stopping criteria, as they can never be better than the ideal stopping criterion.

B. Block-wise Orthogonal Least Squares

For block-wise detection we expand our notation: Let B be a set of indices of blocks and let $\overline{B}$ be the complementary set. Further, let $\Gamma(B)$ specify the indices of those columns contained in the blocks indexed by B. Using this notation, the BOLS is given in Algorithm 2.

While the selection of non-zero elements for the OLS is based on symbol-wise activity, the non-zero elements in $\mathbf{x}$ are given by block-wise activity in this transmission scenario. Thus, the OLS does not exploit the additional information that the user activity is constant for the duration of a frame, and thus the activity for all symbols of a user is the same. In order to exploit the block sparsity, the BOLS [20] algorithm, during each iteration, selects a block of columns, instead of a single column. Thus, the BOLS selects all columns $\Gamma(B_{i_{\min}})$ corresponding to the block $i_{\min}$ that yields the residual $\mathbf{r}_{i_{\min}}^l$ with the smallest Euclidean norm. As the BOLS only performs one iteration for each block of non-zero elements, i.e., each active user, the ideal stopping criterion is $l = K_a$. Thus, for our simulations we assume that the number of active users K_a is known, and use it for the ideal stopping criterion.

As the BOLS jointly detects the activity state for L_k symbols of the k^{th} user during each iteration, it is less computationally expensive than the OLS. First, for a given sparsity $S_{\mathbf{x}}$, the BOLS only computes K_a iterations to determine $S_{\mathbf{x}}$ non-zero elements, while the OLS computes $S_{\mathbf{x}}$ iterations. Secondly, during each iteration the OLS computes $M - l + 1$ different Least Squares (LS) estimations, while during each iteration the BOLS only computes $K - l + 1$ different LS estimations. Assume that for n columns of matrix $\mathbf{A}$ $LS(n)$ operations are required to compute the LS estimation. Further, assume that the total number of operations is mainly determined by the LS estimations. Then the number of operations of the OLS is given by $\sum_{l=1}^{S_{\mathbf{x}}}(M - l + 1)LS(l)$, whereas for the BOLS it is given by $\sum_{l=1}^{S_{\mathbf{x}}/L}(M/L - l + 1)LS(l \cdot L)$ for $L_k = L\ \forall k$.

The BOLS chooses all columns within block i based on the results of the orthogonal projections $\mathbf{A}_{\Gamma(B_i^l)}\mathbf{A}^{\dagger}_{\Gamma(B_i^l)}\mathbf{y}$. Therefore, the decision is more accurate for orthogonal sub-matrices $\mathbf{A}_{\Gamma(B)}$. For a CDMA transmission, orthogonal sub-matrices $\mathbf{A}_{\Gamma(B)}$ are more likely for independent channel realizations per symbol, i.e., the channels are time-variant. Therefore, we expect more accurate activity detection for statistically independent channels for each symbol.

IV. Simulation Results

In this section, simulation results for the symbol error rate (SER) are provided and evaluated. 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 for correctly detected activity. The following simulation setup was used: A set of $K = 128$ users is given, where each user is active with a probability of $p_a = 0.015$. Each user transmits a frame of $F = 256$ chips whenever active. For each user, Pseudo Noise (PN) sequences with $N_k = N\;\forall\,k$ chips are applied. Thus, $L_k(N) = \lfloor F/N \rfloor = L\;\forall\,k$ symbols can be transmitted per user. The transmitted symbols are Binary Phase Shift Keying (BPSK) modulated data bits. As a model for the instantaneous channel, six consecutive Rayleigh distributed coefficients with an exponentially decaying power profile are used. These channel realizations are constant for an entire frame, except when noted otherwise. It is assumed that the transmitted chips of all users arrive synchronously at the receiver, such that $\tau_k = 0\;\forall\,k$.

Whenever a user is active, this user transmits with an average energy per symbol of $E_S = 1$. Therefore, the signal-to-noise ratio at the transmitter is $E_S/N_0 = 1/\sigma_{\mathbf{n}}^2$, where $\sigma_{\mathbf{n}}^2$ is the variance of the additive white Gaussian noise.

For comparison purposes, we introduce a linear detection for known activity using LS estimation, i.e., the detector knows which users are active for every frame. As this detector has perfect knowledge of the position of the non-zero elements in $\mathbf{x}$ due to an oracle process, we call it "oracle LS". The oracle LS algorithm performs the LS estimation using $\mathbf{y}$ and the sub-matrix $\mathbf{A}_{\Gamma_{\text{oracle}}}$, where Γ_{oracle} are the indices of the non-zero elements of the transmitted vector $\mathbf{x}$, i.e., $\Gamma_{\text{oracle}} = \{i : x_i \neq 0\}$. Additionally, the oracle LS serves as a lower bound for both OLS and BOLS, since they can never be better than LS estimation for correctly detected active users.

Fig. 2 and 3 show simulation results for two different values of the spreading factor, $N = 128$ ($L = 2$) and $N = 32$ ($L = 8$). These figures show that the BOLS has a lower SER than the OLS in both cases. This is due to the fact that the BOLS exploits the block sparsity of $\mathbf{x}$. Note that for $N = 32$ the SER of the BOLS is lower than for $N = 128$, as opposed to both OLS and oracle LS, where the SER is slightly higher. Two different influences are responsible for this effect. On the one hand, the smaller the spreading factor N, the higher the condition number of $\mathbf{A}$, which makes both the activity and data detection less reliable. On the other hand, a smaller N results in a larger block size L, which provides more information for block-wise detection. This makes block-wise activity detection

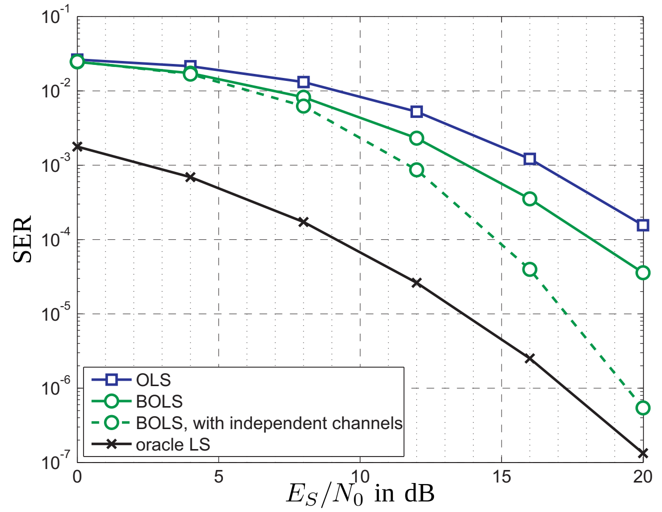

Fig. 2. SER simulation results for $K = 128$, $F = 256$, $N = 128$, $L = 2$ and $p_a = 0.015$. SER for statistically independent channels for each symbol is shown as a dashed line.

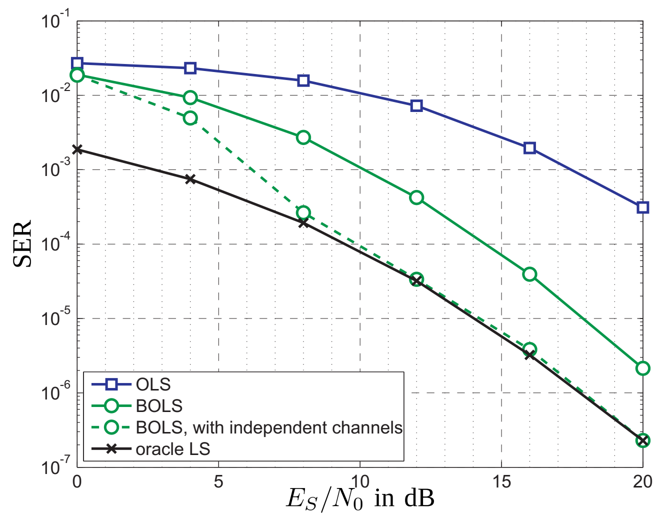

Fig. 3. SER simulation results for $K = 128$, $F = 256$, $N = 32$, $L = 8$ and $p_a = 0.015$. SER for statistically independent channels for each symbol is shown as a dashed line.

more reliable. For OLS and oracle LS, the condition number determines the SER, while the block length does not influence their performance, thus their SER increases slightly. For the BOLS, the influence of larger block lengths outweighs the influence of increased condition numbers, thus the SER is lower.

The SER of the BOLS in the case of statistically independent channels across symbols for all users is shown in Fig. 2 and 3 as a dashed line. In both cases, independent channels per symbol significantly improve the block-wise activity detection. For $N = 32$, the SER of the BOLS even reaches the oracle LS bound for high SNR. These results demonstrate that in the case of independent channels per symbol the BOLS rarely makes incorrect activity decisions for high SNR, so that the SER is mainly determined by bit errors. For $N = 128$, the

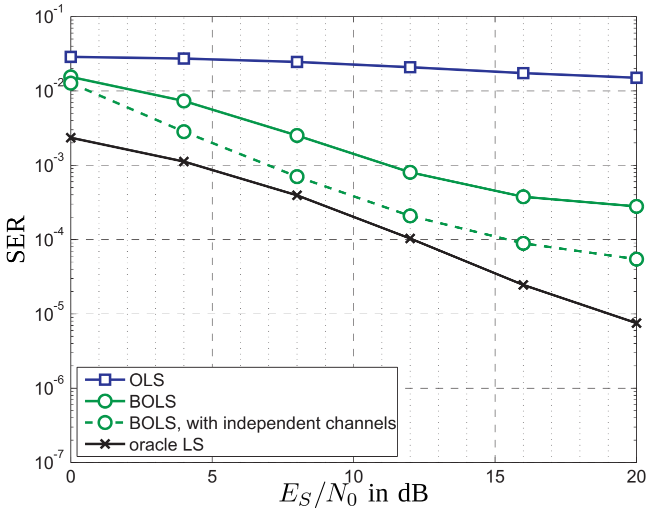

Fig. 4. SER simulation results for $K = 128$, $F = 256$, $N = 8$, $L = 32$ and $p_a = 0.015$. SER for statistically independent channels for each symbol are shown as a dashed line.

improvement of the SER of the BOLS is much smaller, as each user transmits fewer symbols per frame. However, even in this case the SER of the BOLS is much lower with independent channels for each symbol.

Similar results are derived in [21], where for Gaussian noise near oracle performance can be guaranteed for high SNR. However, the main difference is the assumption of normed column weights of $\mathbf{A}$ in [21]. Here, the simulations results indicate that independent channels suffice for near oracle performance in this transmission scenario and certain block sizes L.

Fig. 4 shows the SER for an even lower spreading factor $N = 8$ ($L = 32$). The SER of all algorithms for $N = 8$ is significantly higher than their respective SER for $N = 32$. The oracle LS and the OLS in particular are increased much more than for $N = 32$ and $N = 128$. This is due to an even larger increase of the condition number of matrix $\mathbf{A}$. For $N = 8$, the BOLS has an error floor behavior around an SNR of 20dB. This indicates that the influence of increased block length no longer outweighs the influence of a larger condition number of matrix $\mathbf{A}$. Thus, for the BOLS there is a certain value of the spreading factor N that yields the lowest SER for a given E_S/N_0.

V. Conclusion

In this paper, we investigated block-wise greedy CS detection for joint detection of data and activity in CDMA transmission. We have shown that using block-wise detection algorithms greatly improves the joint error rate of activity and data detection compared to symbol-wise detection. This improvement is even larger for an overloaded CDMA system, due to increased block lengths. However, this improvement is limited to certain problem dimensions. It was shown that for highly overloaded CDMA systems the error rate of block-wise detection is increased when decreasing the spreading factor even further. This indicates that for a given system setup there is a certain spreading factor that yields the lowest error rate for block-wise greedy detection. We also demonstrated that the performance of the BOLS improves in the case of statistically independent channels for each transmitted symbol. This property allows for near optimal activity detection, in certain under-determined cases.

References

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

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

[3] 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.

[4] H. F. Schepker and A. Dekorsy, "Sparse multi-user detection for CDMA transmission using greedy algorithms," in *Proceedings ISWCS*, November 2011.

[5] 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.

[6] 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.

[7] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.

[8] 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.

[9] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, p. 330, April 2010.

[10] 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.

[11] 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.

[12] E. J. Candès and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[13] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301–320, 2005.

[14] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[15] 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.

[16] 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.

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

[18] 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.

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

[20] A. Majumdar and R. K. Ward, "Fast group sparse classification," *Electrical and Computer Engineering, Canadian Journal of*, vol. 34, no. 4, 2009.

[21] Z. Ben-Haim and Y. C. Eldar, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 1032–1047, September 2011.