

Exploiting Sparsity in Channel and Data Estimation for Sporadic Multi-User Communication

Henning F. Schepker, Carsten Bockelmann and Armin Dekorsy

Department of Communications Engineering

University of Bremen, Bremen, Germany

email: {schepker, bockelmann, dekorsy}@ant.uni-bremen.de

Abstract—Machine-to-Machine communication requires new physical layer concepts to meet future requirements. In previous works it has already been shown that Compressive Sensing (CS) detectors are capable of jointly detecting both activity and data in multi-user detection (MUD). To date, the investigations on CS MUD have omitted the channel estimation and assumed perfect knowledge. However, in a practical application the channel also has to be estimated, requiring that the joint estimation of activity and data is extended by channel estimation. Therefore, we will investigate how to adapt several approaches to channel estimation, such that they are suitable for this scenario. Additionally, we will provide simulation results and discuss their difference to the performance of CS MUD for perfect channel state information.

I. INTRODUCTION

The field of wireless Machine-to-Machine communication is expected to grow tremendously in the future. With low data rates and sporadic communication, these applications present new challenges for physical layer concepts, as system requirements differ from many current applications, such as high data rate access. Due to the sporadic communication, i.e., each transmitter being inactive most of the time, the system is more severely impacted by transmission overheads than continuous transmission, necessitating a reduction of the transmission overhead. One approach towards reducing the transmission overhead is to no longer signal user activity and to detect both the activity and the data at the receiver instead.

For the scenario of a sporadic uplink transmission in a sensor network, it has already been shown in [1]–[4] that reliable joint detection of both activity and data is possible using Compressive Sensing (CS) in Multi-User Detection (MUD) [5], [6]. However, these investigations are focused on activity and data detection, and for simplicity assume perfect channel state information (CSI). For practical implementation, as the channel coefficients also have to be estimated for time variant channels, this estimation has to be included in the joint data and activity detection. Therefore, the receiver actually has to simultaneously detect activity, data and channel coefficients.

In this paper, we will investigate different approaches to incorporate channel estimation in the joint detection of activity and data for sporadic communication. First, we modify the

This work was supported in part by the German Research Foundation (DFG) under grant DE 1858/1-1. Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

known approach of channel estimation based on pilots, such that a joint estimation of activity and channel coefficients is performed. For this joint estimation of activity and channel coefficients, we introduce a new greedy algorithm, which exploits the specific sparsity structure of this new approach. As an alternative, we will apply blind channel estimation to sporadic communication, following the ideas from [7]. Additionally, we introduce a semi-blind estimation, based on the blind estimation approach, in order to improve data detection accuracy.

II. SYSTEM MODEL

A. Random Coded Transmission

We consider a wireless uplink transmission, where K sensor nodes communicate with a central aggregation node. Here, we assume that the transmissions from the sensor nodes are *sporadic*, i.e., the sensor nodes are only active on occasion. As a model for sensor node activity, we assume that each sensor node is active for a short period of time with a given activity probability p_a . Further, we assume that this activity probability is identical for all sensor nodes and rather small, i.e., $p_a \ll 1$.

For the transmitter setup of the sensor nodes, we assume that an active node k_a transmits N_F modulated symbols $\mathbf{d}_{k_a} \in \mathcal{A}^{N_F}$, where \mathcal{A} is the modulation alphabet. Without loss of generality, we assume BPSK in the following and simplify the notation to a real-valued model. Other modulation schemes can easily be applied. An inactive node k_i does not transmit at all, thus we model the transmitted “symbols” as zeros, i.e., $\mathbf{d}_{k_i} \in \{0\}^{N_F}$. Therefore, each sensor transmits N_F symbols drawn from the so-called *augmented* alphabet $\mathcal{A}_0 = \{\mathcal{A} \cup 0\}$, which is the BPSK alphabet \mathcal{A} augmented and extended by the zero symbol to indicate inactivity.

We assume that before transmission each node k encodes the symbols \mathbf{d}_k using *random coding*. The symbol-specific code words $\mathbf{c}_{n,k} \in \mathbb{R}^F$ for node k and symbol n contain random Gaussian distributed values that are normalized, such that $\|\mathbf{c}_{n,k}\|_2 = 1 \forall n, k$. Therefore, the F code symbols \mathbf{s}_k contain the superposition of all the symbol-specific code words $\mathbf{c}_{n,k}$ multiplied with the corresponding data symbol $d_{n,k}$. The reason why random coding is applied here is that CS on average performs well for measurements over random Gaussian matrices [8], and random coding can represent different transmission techniques.

As a time discrete channel model, we assume that for each node the current channel is determined by L_h real-valued non-zero channel coefficients at unknown delays within a maximum delay of τ_c coded symbols. This means that within the delay of τ_c , the time discrete channel impulse response of node k is represented by sparse vector \mathbf{h}_k , where neither the values nor the position of the non-zero coefficients are known.

B. Detection and Estimation Models

Under the assumption of perfect CSI, or for estimated channels, the activity and data can be jointly detected based on

$$\mathbf{y} = \sum_{k=1}^K \mathbf{H}_k \mathbf{C}_k \mathbf{d}_k + \mathbf{n} = \mathbf{M} \mathbf{H} \mathbf{C} \mathbf{x} + \mathbf{n} = \mathbf{A}_h \mathbf{x} + \mathbf{n}, \quad (1)$$

where \mathbf{H}_k is the convolution matrix for the node-specific channel \mathbf{h}_k , \mathbf{x} is the stacked vector of all \mathbf{d}_k , $\mathbf{C}_k = [\mathbf{c}_{1,k}, \mathbf{c}_{2,k}, \dots, \mathbf{c}_{N_F,k}]$ and $\mathbf{M} = [\mathbf{I}, \mathbf{I}, \dots, \mathbf{I}]$. However, for unknown channels equation (1) is ill-defined. In order to estimate the channel in those cases, we have to rewrite (1) as a joint channel and activity estimation problem for known data \mathbf{d}_k

$$\mathbf{y} = \sum_{k=1}^K \mathbf{S}_k \mathbf{h}_k + \mathbf{n} = \mathbf{M} \mathbf{S} \mathbf{h} + \mathbf{n} = \mathbf{A}_x \mathbf{h} + \mathbf{n}, \quad (2)$$

where \mathbf{S}_k is the convolution matrix for \mathbf{s}_k and \mathbf{h} is the stacked vector of all \mathbf{h}_k . Using the two detection and estimation models, we define different approaches for joint estimation of activity, data and channel coefficients.

III. CHANNEL ESTIMATION

A. Channel Estimation Based on Pilot Sequence

The standard approach to channel estimation is based on pilot sequences. In sporadic communication, we have to estimate activity, data and channel coefficients from the received values. As estimation of all properties at once is not feasible, we separate these steps, as shown in Figure 1a). Firstly the channel estimation jointly estimates the activity and channel coefficients, and secondly the data detection based on this estimation is performed.

In this paper, we assume that N_F known BPSK pilot symbols are transmitted. For these pilot symbols there are two options, as depicted by the switch in Figure 1a): either transmitting them in a single frame as additional symbols on top of the data, or as an additional frame only consisting of pilot symbols. On the one hand, using a separate pilot frame of F_P code symbols causes additional overhead and effectively reduces the potential data rate by $\frac{F}{F+F_P}$. On the other hand, transmitting the pilots in the same frame does not increase the number of code symbols, but increases the interference between both data symbols and nodes, as the random code encodes each pilot symbol to all F code symbols within the frame. For this approach, before detecting the data in the second step the influence of the pilot symbols has to be removed from the received values using the estimated channel.

B. Blind Channel Estimation

An alternative to pilot based channel estimation is blind estimation following the ideas from [7]. This approach has minimal requirements on the transmission, as no pilot symbols are required. The blind estimation is solely based on the knowledge about the random codes and the statistics of the channel, e.g., the maximum channel delay of τ_c code symbols.

The main difference to channel estimation with pilot symbols is the lack of known symbols, which limits the reliability of the channel estimation. Thus, the first step is instead a joint activity and data estimation based on an initial channel assumption, as shown in Figure 1b). Here, we use an AWGN channel with an unknown delay within τ_c as an initial channel assumption to model the strongest channel coefficient. In order to incorporate the unknown delay in the CS MUD, the matrix \mathbf{A}_h in (1) has to be augmented with delay hypotheses for each possible delay τ within $0 \leq \tau \leq \tau_c$. This approach is explained in more detail in [4]. As the initial channel assumption is not an accurate model for the current channel, we can further improve the data detection accuracy of this first detection with an estimated channel. Thus, an estimation of the channel coefficients and then data detection is performed.

This process is similar to the iterative process described in [7]. However, in this scenario the number of non-zero channel coefficients is known for each node. Thus, we only need to perform one step of the iterative process estimating L_h channel coefficients for each active node.

C. Semi-Blind Channel Estimation

One of the problems with the initial channel assumption in blind estimation is that it does not accurately model the phase of the strongest channel coefficient. Therefore, if the strongest channel coefficient is actually negative, the data symbols will likely be detected with an incorrect sign. These errors are not detectable, as CS detection algorithms are based on either vector norms or absolute correlation, all of which are ambivalent to change in the sign of the detected vector. This motivates the use of a semi-blind approach.

The semi-blind estimation adds a simple channel estimation step to the blind channel estimation approach, as shown in Figure 1c). This step requires that at least a single BPSK pilot symbol is added to the data frame. For this the single node version of (2) is built for each node k , such that it only contains the sequences of node k , again augmented by delay hypotheses. Afterwards, a CS MUD detection for a channel cardinality of $L_{h, sb} = 1$ is applied, to determine the strongest channel coefficient. The sign, i.e., phase, of this strongest channel coefficient is then used as the sign of the initial AWGN channel model in the blind estimation, which eliminates the sign ambiguities.

IV. COMPRESSIVE SENSING MULTI-USER DETECTION

Both the systems defined in (1) and (2) are detection problems defined for an unknown sparse vector. Due to the low activity probability of the nodes, both the multi-user data vector \mathbf{x} and the multi-user channel vector \mathbf{h} are sparse.

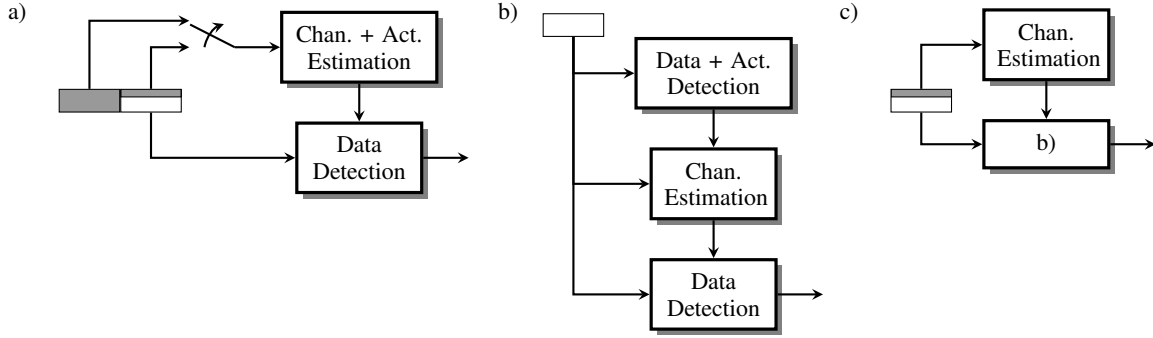


Fig. 1. Different approaches to joint activity, data and channel estimation. The received data frame is shown in white, and the received pilot frame or pilot layer is shown in gray.

This property allows for CS to be applied to (1) and (2) to determine the sparse vector, as the theory of CS is focused on the reconstruction of compressible signals by recovery of sparse signals even from under-determined equation systems [5], [6]. While there are many different approaches to CS detection, e.g., [9]–[11], in this paper we focus on CS MUD using greedy algorithms. These are in general more efficient, but less accurate, than solving convex optimization problems, such as [9].

In order to determine a well suited greedy algorithm for the different estimation and detection approaches, the available information about the sparse vectors has to be considered. The data vector \mathbf{d}_k of node k is either all zero or taken from the modulation alphabet \mathcal{A} . Therefore, \mathbf{x} in (1) is *block-sparse* or *group-sparse*, as it contains blocks of \mathbf{d}_k and only a few nodes are active. Therefore, the Group Orthogonal Matching Pursuit (GOMP) [12] is well suited, as this greedy algorithm exploits block-sparsity.

The stacked vector \mathbf{h} in (2) is also known to be sparse, as only the channel coefficients of an active node are modeled as non-zero. However, the channel coefficient vector \mathbf{h}_{k_a} of an active node k_a is also sparse, as each vector \mathbf{h}_{k_a} only has L_h non-zero entries. Therefore, we call the stacked vector \mathbf{h} *hierarchical sparse*, as at the first layer \mathbf{h} is block-sparse and at the second layer each active block \mathbf{h}_{k_a} is sparse. To exploit this special sparsity structure, we propose the Hierarchical Block Orthogonal Matching Pursuit (HBOMP), a new greedy algorithm.

A. Hierarchical Block Orthogonal Matching Pursuit

In order to explain the Hierarchical BOMP (HBOMP), we first introduce our notation: B is a set of block-indices and Γ is a set of vector-indices. \bar{B} and $\bar{\Gamma}$ are the corresponding complementary sets. Further, $f(k)$ is a set function that defines all vector-indices contained in block k , and $g(j)$ is a set function that defines the block-index of the block which contains vector-index j . \mathbf{A}_Γ specifies the sub-matrix which only contains those columns with vector-indices in Γ , and likewise \mathbf{h}_Γ contains only those elements of vector \mathbf{h} with vector-indices in Γ . \mathbf{h}^ℓ , \mathbf{A}^ℓ , B^ℓ and Γ^ℓ each 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

Algorithm 1 Hierarchical BOMP (HBOMP)

$$B^0 = \emptyset, \Gamma^0 = \emptyset, \ell = 1, \mathbf{r}^0 = \mathbf{y}$$

repeat

$$j_{\max} = \arg \max_j |\mathbf{A}_j^H \mathbf{r}^{\ell-1}| \text{ with } j \in f(\bar{B}^{\ell-1})$$

$$n = 0$$

repeat

$$n = n + 1$$

$$i_{\max} = \arg \max_i |\mathbf{A}_i^H \mathbf{r}^{\ell-1}| \text{ with } i \in f(g(j_{\max}))$$

$$\Gamma^\ell = \Gamma^{\ell-1} \cup i_{\max}$$

$$\hat{\mathbf{h}}_{\Gamma^\ell}^\ell = \mathbf{A}_{\Gamma^\ell}^\dagger \mathbf{y} \text{ and } \hat{\mathbf{h}}_{\bar{\Gamma}^\ell}^\ell = \mathbf{0}$$

$$\mathbf{r}^\ell = \mathbf{y} - \mathbf{A} \hat{\mathbf{h}}^\ell$$

$$\ell = \ell + 1$$

until $n = L_h$

$$B^\ell = B^{\ell-1} \cup g(j_{\max})$$

until $\ell > K_a L_h$

matrix of \mathbf{A} . For notational clarity, we simply denote the matrix \mathbf{A}_x as \mathbf{A} in the HBOMP.

The HBOMP shown in Algorithm 1 consists of two different loops, an outer loop based on the Block Orthogonal Matching Pursuit (BOMP) [12] and an inner loop based on the Orthogonal Matching Pursuit (OMP) [10]. During each iteration of the outer loop, the HBOMP selects the block j_{\max} that contains the highest correlation to the current residual \mathbf{r}^ℓ . After this block selection step, a regular BOMP would add all elements in block j_{\max} to $\Gamma^{\ell-1}$, thereby setting them as active. However, due to the hierarchical sparsity only few of the elements contained in the block are active. Thus, the HBOMP continues with the inner loop for the currently selected block j_{\max} . In each iteration of the inner loop, the HBOMP determines the strongest correlation within block j_{\max} to the current residual \mathbf{r}^ℓ and then adds the vector index i_{\max} to the set $\Gamma^{\ell-1}$. Afterwards, the HBOMP estimates the non-zero elements $\hat{\mathbf{h}}_{\Gamma^\ell}^\ell$ using LS estimation, and then updates the residual \mathbf{r}^ℓ .

In general, the correct number of iterations for greedy CS algorithms is not known prior to detection. However, in our scenario the sparsities at both layers are known for blind and semi-blind estimation approaches. Therefore, for the outer

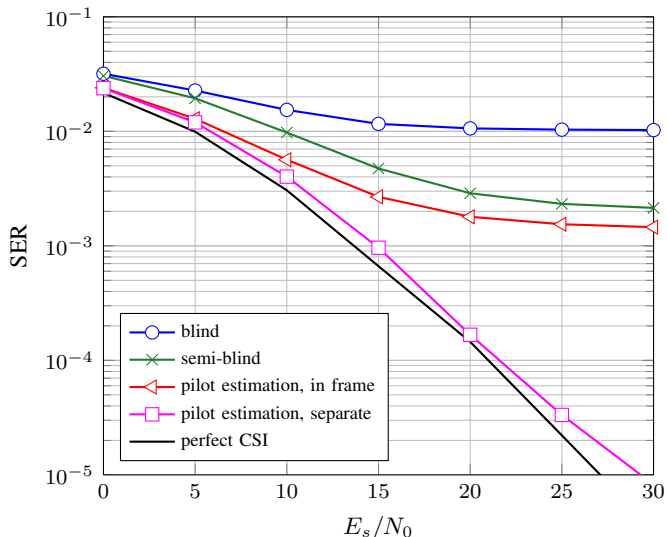


Fig. 2. Symbol Error Rate over the Augmented Alphabet for $K = 64$ and $F = 512$.

loop, the number of iterations is given by the number of active nodes K_a in the previous CS detection. Additionally, the inner loop has to perform a total of L_h iterations, as given by the channel statistics. For comparison purposes, we assume that this is also known for the detection in pilot based channel estimation.

V. SIMULATION RESULTS

In this section, we will discuss simulation results for the estimation approaches described in section III and compare them with a GOMP detection assuming perfect CSI. Unless otherwise noted, we set the code length of the additional pilot frame for pilot estimation to be the same as the data frame, i.e., $F_P = F$. For the semi-blind estimation one data symbol is replaced by a pilot symbol in the data frame. In these simulations the GOMP is used for activity and data detection and the HBOMP is used for activity and channel estimation. Prior to a detection or estimation, the system matrix is normalized to have an identical column norm for all columns, as described in [13].

As a simulation setup, we consider a transmission from $K = 64$ sensor nodes that each transmit a data frame of $N_F = 8$ symbols. The random code for each node k is given by code words of $F = 128$ or $F = 512$ i.i.d. real Gaussian distributed values, normalized such that each code word has unit norm. The channel is modeled by $L_h = 3$ i.i.d. real Gaussian distributed taps with the power profile $\sigma_h^2 = [0.873, 0.436, 0.218]$. These taps are located at equally distributed random but ordered delays within $\tau_c = 20$ code symbols. We assume that sensor nodes are only active with a probability of $p_a = 0.02$, so that the number of active nodes is on average much smaller than K .

Figure 2 shows simulation results for the symbol error rate (SER) over the augmented alphabet \mathcal{A}_0 for a frame length of $F = 512$ code symbols, i.e., equation (1) is fully determined. First, it should be noted that pilot estimation using a separate

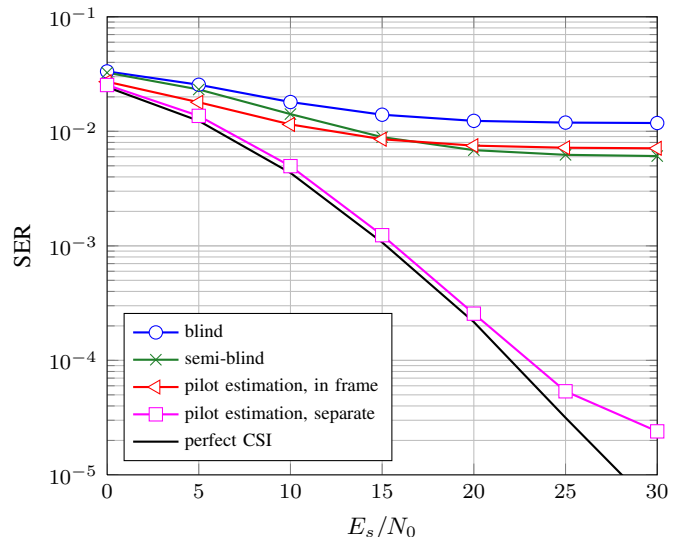


Fig. 3. Symbol Error Rate over the Augmented Alphabet for $K = 64$ and $F = 128$.

pilot frame almost achieves the performance for perfect CSI up to the high SNR region. This means that, with the transmission overhead of the pilot frame, the performance of previous investigations, e.g., [1]–[4], can almost be achieved. For all the approaches with no additional frame, the SERs are significantly higher and show a noticeable error floor in the high SNR region. Additionally, these results show that the semi-blind estimation improves the results of blind estimation and that pilot estimation has lower SER even for pilots within the data frame. The SER for pilot estimation can likely be improved with appropriate power-loading, but this is beyond the scope of this paper.

Figure 3 shows simulation results for the symbol error rate (SER) over the augmented alphabet \mathcal{A}_0 for a frame length of $F = 128$ code symbols, such that equation (1) is under-determined. In general, CS MUD is still able to reliably detect activity and data even in an under-determined system. This property is shown by the fact that the SER increases only slightly for perfect CSI. However, the same does not hold true for semi-blind estimation and pilot estimation. These two approaches have significantly increased SERs, due to the increased multi-user interference.

As the SER contains both errors due to incorrect data detection and errors due to incorrect activity detection, it is necessary to also analyze the performance in terms of activity errors. Two kinds of activity errors can be distinguished: on the one hand estimating an inactive element as active, called *false alarm*, and on the other hand estimating an active element as inactive, called *missed detection*. Figure 4 shows both the missed detection rate (MDR) and the false alarm rate (FAR) for a code length of $F = 512$ code symbols. From these results, we see that blind and semi-blind estimation both result in the exactly same number of activity errors. This is due to the ambivalence regarding the phase discussed in section III, as a change in the sign of the strongest channel coefficient

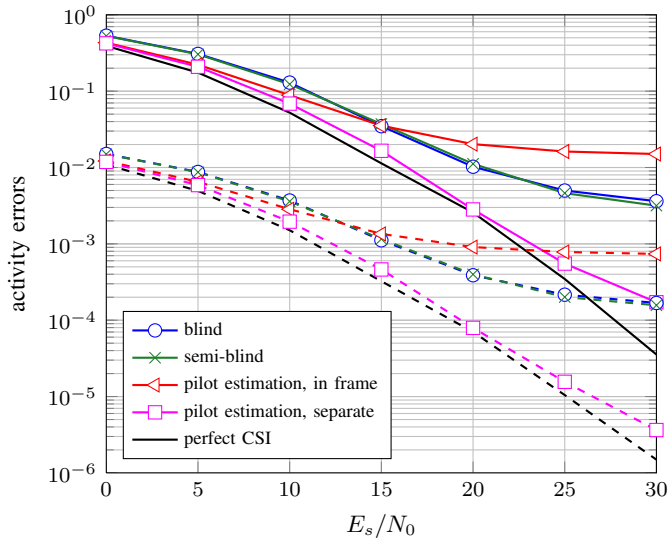


Fig. 4. Missed Detection Rate (solid) and False Alarm Rate (dashed) for $K = 64$ and $F = 512$.

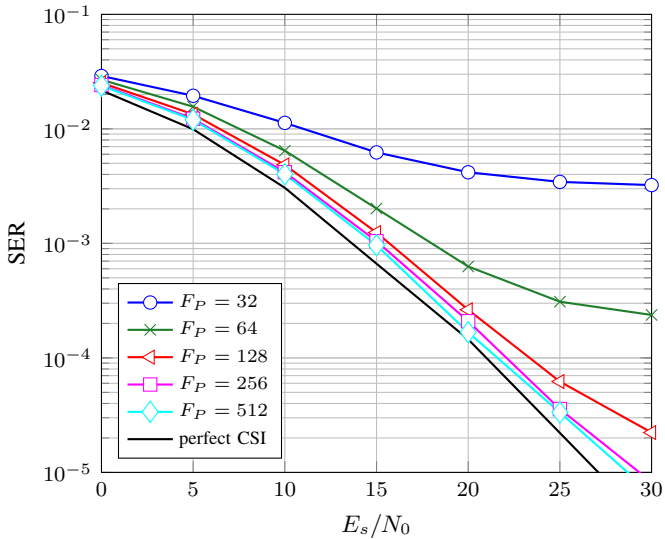


Fig. 5. Symbol Error Rates for pilot estimation using separate pilot frame, different sizes of F_P , $K = 64$, and $F = 512$.

in the semi-blind estimation does not influence the activity detection. Additionally, both blind and semi-blind estimation result in less activity errors than pilot estimation. Thus, pilot estimation results in more activity errors, but much fewer data errors.

Finally, Figure 5 shows simulation results for pilot estimation using a separate pilot frame of different code lengths F_P . These results show that the pilot frame can be shorter than the data frame without a large increase of the SER, as shown by the results for $F_P = 256$ and $F_P = 128$. However, once the separate pilot frame becomes too short, the activity and channel estimation become unreliable, significantly increasing the SER. Therefore, the length of the pilot frame F_P defines a tradeoff between detection accuracy and additional transmission overhead.

VI. CONCLUSION

In this paper, we investigated different approaches to estimate activity, data and channel coefficients in a sporadic multi-user transmission. For estimating all three properties, the best approach is to jointly estimate activity and channel coefficients, followed by a data detection. To reliably solve this estimation problem, we introduced a new greedy Compressive Sensing algorithm. Furthermore, this paper in principle shows that channel estimation for sporadic communication with very short packages is not only possible, but can achieve nearly the performance for perfect CSI. The only requirement for this performance is that an additional frame of pilot symbols are transmitted, slightly increasing the transmission overhead.

REFERENCES

- [1] 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.
- [2] H. F. Schepker and A. Dekorsy, "Sparse multi-user detection for CDMA transmission using greedy algorithms," in *8th International Symposium on Wireless Communication Systems*, Aachen, Germany, November 2011.
- [3] H. Schepker and A. Dekorsy, "Compressive sensing multi-user detection with block-wise orthogonal least squares," in *IEEE 75th Vehicular Technology Conference*, Yokohama, Japan, May 2012.
- [4] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Coping with CDMA asynchronicity in compressive sensing multi-user detection," in *IEEE 77th Vehicular Technology Conference*, Dresden, Germany, June 2013.
- [5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [6] 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.
- [7] M. S. Asif, W. Mantzel, and J. Romberg, "Channel protection: Random coding meets sparse channels," in *IEEE Information Theory Workshop*, Taormina, Sicily, October 2009.
- [8] 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.
- [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. Rezaifar, 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] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009.
- [12] A. Majumdar and R. K. Ward, "Fast group sparse classification," *Electrical and Computer Engineering, Canadian Journal of*, vol. 34, no. 4, 2009.
- [13] S. Rangan, A. Fletcher, and V. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1902 –1923, March 2012.