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 based MUD have omitted the channel estimation and assumed perfect channel knowledge. However, in practical applications the channel also has to be estimated, such that the joint detection of activity and data has to be extended by channel estimation. Therefore, in this paper we investigate how to adapt several approaches to channel estimation, such that they are suitable for this joint detection and estimation. Further, we provide simulation results, showing that low overhead channel estimation can be achieved, with only a small loss in detection accuracy compared to CS based 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 in continuous transmission, necessitating a strong reduction of the transmission overhead compared to current systems designed for high data rates. One approach towards reducing the transmission overhead is to avoid signaling of the activity, and to detect both activity and data in the physical layer processing at the receiver. This results in a random access for higher layers.

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) based Multi-User Detection (MUD) [5], [6]. However, these investigations for simplicity assume perfect channel state information (CSI), and so far the aspect of channel estimation has not been incorporated in the CS based MUD for sporadic communication. In order to include the channel estimation, the receiver simultaneously has to detect activity, data and channels.

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.

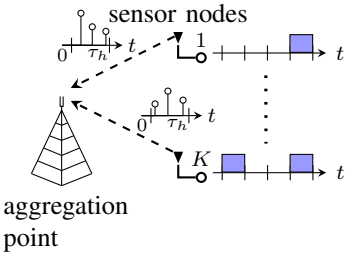

Fig. 1. Transmission scenario.

In this paper, we investigate different approaches to incorporate channel estimation in the joint detection of activity and data for sporadic communication. First, we adapt the approach of pilot based channel estimation, which is used in many commercial systems, e.g. LTE, to sporadic communication, by performing a joint estimation of activity and channels. For this joint estimation, we introduce a new greedy algorithm, which exploits hierarchical sparsity, i.e., block-sparse [7] per node and symbol-wise sparse for each active node. As an alternative method, we investigate blind channel estimation for sporadic communication, following the ideas from [8]. Further, we introduce a semi-blind channel estimation, based on the blind estimation approach, in order to improve data detection accuracy.

II. System Model

A. Transmission Setup

We consider a wireless uplink transmission, where K sensor nodes communicate with a central aggregation node, as shown in Figure 1. 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 frame with a given frame activity probability p_a. Without loss of generality, we assume that this frame activity probability is identical for all sensor nodes and rather small, i.e., $p_a \ll 1$.

For the data of the sensor nodes, we assume that an active node k_a transmits N_F modulated symbols $\mathbf{d}_{k_\mathrm{a}} \in \mathcal{A}^{N_\mathrm{F}}$ per frame, where $\mathcal{A}$ is the modulation alphabet. Without loss of generality, we assume BPSK modulation in the following and simplify the notation to a real-valued model. Other

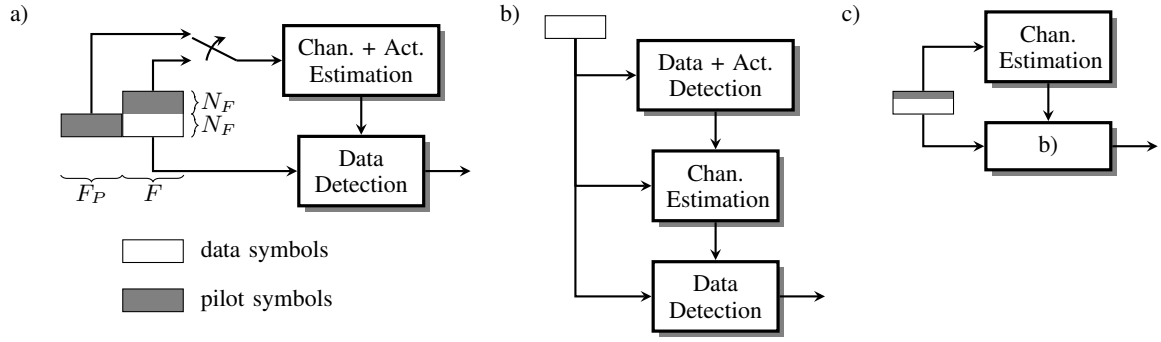

Fig. 2. Different approaches to joint activity, data and channel estimation. a) Pilot based channel estimation, b) blind channel estimation and c) semi-blind channel estimation.

modulations can easily be applied. An inactive node k_i does not transmit during this frame, and thus we model the data symbols as zeros, i.e., $\mathbf{d}_{k_\mathrm{i}} \in \{0\}^{N_\mathrm{F}}$. Therefore, for each node, a frame consists of N_F data 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 each node k encodes the symbols in $\mathbf{d}_k$ using *random coding* before transmission. This means that node k encodes the n^{th} data symbol $d_{n,k} \in \mathcal{A}_0$ with a node and symbol specific code word $\mathbf{c}_{n,k} \in \mathbb{R}^F$. The node k then transmits the superposition of the F transmit-symbols for each data symbol $d_{n,k}$, as given by

$$\mathbf{s}_k = \sum_{n=1}^{N_F} \mathbf{c}_{n,k} d_{n,k} = \mathbf{C}_k \mathbf{d}_k \,, \tag{1}$$

where $\mathbf{C}_k = [\mathbf{c}_{1,k},\, \mathbf{c}_{2,k},\, \ldots,\, \mathbf{c}_{N_\mathrm{F},k}] \in \mathbb{R}^{F \times N_F}$. Here, we assume that each code word $\mathbf{c}_{n,k}$ contains random Gaussian distributed values that are normalized, such that $\|\mathbf{c}_{n,k}\|_2 = 1 \; \forall n, k$. The reasoning for the application of random coding is that CS detection on average performs well for measurements over random Gaussian matrices [9], and other transmission techniques, such as CDMA and SC-FDMA, can be interpreted as specific, structured cases of random coding.

As a time discrete channel model, we assume that for node k the time discrete channel impulse response $\mathbf{h}_k$ is determined by L_h real-valued non-zero channel coefficients at unknown delays within a maximum delay of τ_h transmit-symbols, as shown in Figure 1. This means that for $L_\mathrm{h} < \tau_\mathrm{h}$ the channel impulse response $\mathbf{h}_k \in \mathbb{R}^{\tau_\mathrm{h}+1}$ is a sparse vector with L_h non-zero elements, where neither the values nor the position of the non-zero coefficients are known.

B. System Models for Data Detection and Channel Estimation

In sporadic communication, activity, data and channels have to be estimated based on the same received values $\mathbf{y} \in \mathbb{R}^{F+\tau_\mathrm{h}}$. As we cannot reliably estimate all these properties at once, we split the system model into two separate models. For known channels $\mathbf{h}_k$, e.g., perfect CSI, the system model is given as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{C}_k \mathbf{d}_k + \mathbf{n} = \mathbf{M}\mathbf{H}\mathbf{C}\mathbf{x} + \mathbf{n} = \mathbf{A}_\mathrm{h}\mathbf{x} + \mathbf{n} \,, \tag{2}$$

where $\mathbf{H}_k \in \mathbb{R}^{(F+\tau_\mathrm{h}) \times F}$ is the convolution matrix for channel $\mathbf{h}_k$ and $\mathbf{n} \in \mathbb{R}^{F+\tau_\mathrm{h}}$ is a vector of uncorrelated Gaussian noise $\mathcal{N}\big(0, \sigma_n^2\big)$. Additionally, the multi-user data vector $\mathbf{x} \in \mathbb{R}^{KN_F}$ is the stacked vector of all $\mathbf{d}_k$, $\mathbf{H} \in \mathbb{R}^{K(F+\tau_\mathrm{h}) \times KF}$ and $\mathbf{C} \in \mathbb{R}^{KF \times KN_F}$ are the block-diagonal matrices of all $\mathbf{H}_k$ and $\mathbf{C}_k$ respectively, and $\mathbf{M} = [\mathbf{I},\, \mathbf{I},\, \ldots,\, \mathbf{I}] \in \mathbb{R}^{F+\tau_\mathrm{h} \times K(F+\tau_\mathrm{h})}$, leading to $\mathbf{A_h} \in \mathbb{R}^{F+\tau_\mathrm{h} \times KN_F}$. As has been shown in the literature [1]–[4], CS based MUD applied to (2) is able to jointly detect activity and data.

For known data $\mathbf{d}_k$, e.g., pilot symbols, the system model is given as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{S}_k \mathbf{h}_k + \mathbf{n} = \mathbf{M}\mathbf{S}\mathbf{h} + \mathbf{n} = \mathbf{A}_\mathrm{x}\mathbf{h} + \mathbf{n} \,, \tag{3}$$

where $\mathbf{S}_k \in \mathbb{R}^{(F+\tau_\mathrm{h}) \times (\tau_\mathrm{h}+1)}$ is the convolution matrix for transmit sequence $\mathbf{s}_k$, $\mathbf{S} \in \mathbb{R}^{K(F+\tau_\mathrm{h}) \times K(\tau_\mathrm{h}+1)}$ is the block-diagonal matrix of all $\mathbf{S}_k$, the multi-user channel vector $\mathbf{h} \in \mathbb{R}^{K(\tau_\mathrm{h}+1)}$ is the stacked vector of all $\mathbf{h}_k$, and $\mathbf{A_x} \in \mathbb{R}^{F+\tau_\mathrm{h} \times K(\tau_\mathrm{h}+1)}$. In contrast to previous works, CS based MUD applied to (3) is able to jointly detect activity and channels.

III. Channel Estimation

A. Pilot Based Channel Estimation

In pilot based channel estimation, shown in Figure 2a), first we jointly estimate the activity and channels based on known pilot symbols, and afterwards a data detection based on the estimated activity and channels is performed. In this paper, we assume that N_F known BPSK pilot symbols are transmitted encoded with independent code words. For these pilot symbols there are two options, as depicted by the switch in Figure 2a):

- N_F pilot symbols are transmitted in a separate frame of length F_P, only containing the encoded pilot symbols. This increases the amount of transmit-symbols transmitted by each active node by F_P, and thus reduces the spectral efficiency by $F/(F + F_P)$.
- N_F encoded pilot symbols are transmitted simultaneously with the encoded data. This does not decrease the spectral efficiency, but increases the interference both between nodes and between the symbols of each node.

After estimating the channels, we remove the influence of the pilots from the received measurements before performing the data detection.

B. Blind Channel Estimation

For blind channel estimation, we follow the ideas from [8], where the blind estimation is solely based on the knowledge about the random codes and the statistics of the channel. The blind estimation is structured as shown in Figure 2b): First we perform a joint activity and data detection, based on an initial channel assumption. Afterwards, we perform a channel estimation utilizing the data detected in the previous step. Finally, we re-evaluate the data detection using the newly estimated channels.

For the initial detection, we assume a single tap channel with a positive phase at an unknown delay within τ_h to model the strongest channel coefficient. In order to incorporate the unknown delay in the detection, the matrix $\mathbf{A}_\mathrm{h}$ in (2) has to be augmented with delay hypotheses for each possible delay τ within $0 \leq \tau \leq \tau_\mathrm{h}$. This approach is explained in more detail in [4]. As the initial channel assumption is in general not an accurate model for the current channels, we can further improve the data detection accuracy of this first detection with estimated channels. Thus, based on the detected data an estimation of the channels and then a data detection based on these new channels is performed.

This process is similar to the iterative process described in [8]. However, based on our transmission assumptions, the number of non-zero channel coefficients is L_h 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

The main problem for blind channel estimation is that the initial channel assumption is not accurate in most cases. This is especially a problem if the phase of the strongest coefficient is not positive, as this will likely result in an incorrect sign for the detected data symbols. These errors are not detectable, as CS detection algorithms are based on either vector norms or absolute correlation, all of which are not influenced by a change in the sign of the detected vector. This motivates the approach of a semi-blind estimation.

The semi-blind estimation adds a simple channel estimation step prior to the blind channel estimation, as shown in Figure 2c). This step requires that a single BPSK pilot symbol $d_{\mathrm{p},k}$ is added to the data frame for all active nodes. Based on the pilot symbol of node k, we model the received values as being dependent only on $d_{\mathrm{p},k}$, with

$$\mathbf{y} = \mathbf{S}_{\mathrm{p},k}\mathbf{h}_k + \mathbf{n}\,, \tag{4}$$

where $\mathbf{S}_{\mathrm{p},k}$ is the convolution matrix for $\mathbf{s}_{\mathrm{p},k} = \mathbf{c}_{\mathrm{p},k}d_{\mathrm{p},k}$. Then we use (4) to estimate the phase of the strongest channel coefficient for node k. This estimated sign is then used as the sign of the initial single tap channel model for node k in the blind estimation.

IV. Compressive Sensing Multi-User Detection

The theory of CS is focused on the reconstruction of compressible signals by recovery of a sparse representation even from under-determined equation systems [5], [6]. 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. Thus, by applying CS detectors to (2) and (3), we can reliably recover these sparse vectors. While there are many different approaches to CS detection, e.g., [10]–[12], we focus on greedy algorithms. These are in general more efficient, but less accurate, than solving appropriate convex optimization problems, such as [10].

Algorithm 1 Hierarchical BOMP (HBOMP)

$B^0 = \emptyset$, $\Gamma^0 = \emptyset$, $\ell = 1$, $\mathbf{r}^0 = \mathbf{y}$
repeat
  $j_\mathrm{max} = \arg\max_j \left|\mathbf{A}_j^H \mathbf{r}^{\ell-1}\right|$ with $j \in f\left(\overline{B^{\ell-1}}\right)$
  $b = g\left(j_\mathrm{max}\right)$
  $n = 0$
  repeat
    $n = n + 1$
    $i_\mathrm{max} = \arg\max_i \left|\mathbf{A}_i^H \mathbf{r}^{\ell-1}\right|$ with $i \in f\left(b\right)$
    $\Gamma^\ell = \Gamma^{\ell-1} \cup i_\mathrm{max}$
    $\hat{\mathbf{h}}^\ell_{\Gamma^\ell} = \mathbf{A}^\dagger_{\Gamma^\ell}\mathbf{y}$ **and** $\hat{\mathbf{h}}^\ell_{\overline{\Gamma}^\ell} = \mathbf{0}$
    $\mathbf{r}^\ell = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}^\ell$
    $\ell = \ell + 1$
  until $n \geq L_\mathrm{h}$
  $B^\ell = B^{\ell-1} \cup b$
until $\ell \geq K_\mathrm{a} L_\mathrm{h}$

The greedy algorithm used depends on the sparsity structure of $\mathbf{x}$ or $\mathbf{h}$ respectively. The symbols in data vector $\mathbf{d}_k$ of node k are either all zero or all taken from the modulation alphabet $\mathcal{A}$. Therefore, $\mathbf{x}$ in (2) is *block-sparse* or *group-sparse* [7], as it contains blocks of $\mathbf{d}_k$ and only a few nodes are active. Thus, the Group Orthogonal Matching Pursuit (GOMP) [13] is well suited, as this greedy algorithm exploits block-sparsity.

The multi-user channel vector $\mathbf{h}$ in (3) is sparse, as only the channels of an active node are modeled as non-zero. However, $\mathbf{h}$ is not block-sparse, as the channel coefficient vector $\mathbf{h}_{k_a}$ of an active node k_a is also sparse, due to $L_\mathrm{h} < \tau_\mathrm{h}$. Therefore, we call the multi-user channel 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 a new greedy algorithm, the Hierarchical Block Orthogonal Matching Pursuit (HBOMP).

A. Hierarchical Block Orthogonal Matching Pursuit

In order to explain the HBOMP, we first introduce our notation: B is a set of block-indices and Γ is a set of vector-indices. $\overline{B}$ and $\overline{\Gamma}$ 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. $\mathbf{A}_\Gamma$ specifies the sub-matrix which only contains those columns with vector-indices in set Γ, i.e., $\mathbf{A}_j$ specifies the j^th column, and likewise $\mathbf{h}_\Gamma$ contains only those elements of vector $\mathbf{h}$ with vector-indices in Γ. $\mathbf{h}^\ell$, $\mathbf{r}^\ell$,$\mathbf{A}^\ell$, 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 matrix of $\mathbf{A}$. For notational clarity, we simply denote the matrix $\mathbf{A}_{\text{x}}$ 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) [13] and an inner loop based on the Orthogonal Matching Pursuit (OMP) [11]. During each iteration of the outer loop, the HBOMP selects the block b containing column $j_{\max}$, with the highest correlation to the current residual $\mathbf{r}^{\ell}$. After this block selection step, a regular BOMP would add all elements in block b to $\Gamma^{\ell-1}$, thereby setting them as active. However, due to the hierarchical sparsity only few of the elements contained in a block are active. Thus, the HBOMP continues with the inner loop for the currently selected block b. In each iteration of the inner loop, the HBOMP determines the strongest correlation within block b to the current residual $\mathbf{r}^{\ell}$ and then adds the vector index $i_{\max}$ to the set $\Gamma^{\ell-1}$. Afterwards, the HBOMP estimates the non-zero elements $\hat{\mathbf{h}}^{\ell}_{\Gamma^{\ell}}$ using LS estimation, and then updates the residual $\mathbf{r}^{\ell}$.

In general, the correct number of iterations for greedy CS algorithms is not known prior to detection. However, due to our assumptions, the number of non-zero channel coefficients is L_{h} is known. Additionally, for simplicity we use a known number of active nodes K_a in the simulations. For implementation an appropriate termination criterion has to be found.

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 length of the separate frame of pilot symbols 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 [14].

As a simulation setup, we consider a transmission from $K = 64$ sensor nodes that each transmit a data frame containing $N_{\text{F}} = 8$ data 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 $L_{\text{h}} = 3$ i.i.d. real Gaussian distributed taps with the power profile $\sigma^2_{\text{h}} = [0.873,\ 0.436,\ 0.218]$. These taps are located at equally distributed random but ordered delays within $\tau_{\text{h}} = 20$ transmit-symbols. We assume that sensor nodes are only active with a probability of $p_a = 0.02$, so that the number of active nodes K_a is on average much smaller than K, i.e., $K_a \ll K$.

For a fair comparison of the different approaches, we incorporate the loss of spectral efficiency due to the separate pilot frame as an SNR loss in the E_s/N_0. Therefore, for all approaches without an additional pilot frame the E_s/N_0 is given by $1/\sigma_n^2$. For the pilot estimation based on an additional pilot frame the E_s/N_0 is given by $F/(F+F_P)\cdot 1/\sigma_n^2$.

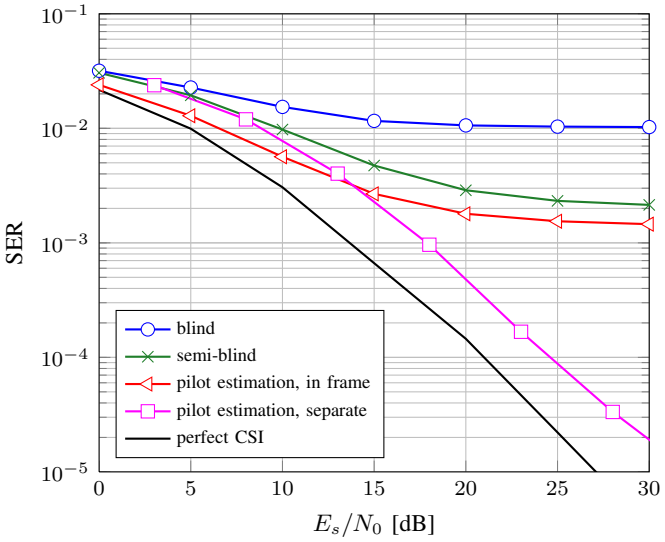

Fig. 3. Symbol error rate over the augmented alphabet for $K = 64$ and $F = 512$, i.e., fully determined case.

Figure 3 shows simulation results for the symbol error rate (SER) over the augmented alphabet $\mathcal{A}_0$ for a frame length of $F = 512$ transmit-symbols, i.e., model (2) is fully determined and model (3) is under-determined by a factor of 2.5. First, it should be noted that pilot estimation using a separate pilot frame has nearly a constant loss compared to the error rate for perfect CSI. This loss is mostly defined by the SNR loss due to the loss of spectral efficiency, as this is not considered in the perfect CSI case. This means that, with the transmission overhead of the pilot frame, the performance assuming perfect CSI, e.g., [1]–[4], can almost be achieved. Further, the results show that for pilot based channel estimation, in the low SNR region it is better to transmit the pilots within the data frame rather than in a separate frame, as the detection is noise limited. Secondly, all approaches with no additional pilot frame show a noticeable error floor in the high SNR region. For pilot estimation this is caused by the additional interference, which limits the achievable error rate. Additionally, these results show that the semi-blind estimation improves the results of blind estimation and that pilot estimation has lower SER for pilots within the data frame. The SER for pilot estimation with pilots in the data frame can likely be improved with appropriate power-loading, but this is beyond the scope of this paper.

In comparison to the previous simulations, we investigate a system setup with a frame length of $F = 512$ transmit-symbols, where the model (2) is under-determined by a factor of 4 and model (3) is under-determined by a factor of 10. For this setup, Figure 4 shows simulation results for the SER over the augmented alphabet $\mathcal{A}_0$. In general, CS MUD is still able to reliably detect activity and data with perfect CSI, even in an under-determined system. This property is shown by the fact that the SER increases only slightly for perfect

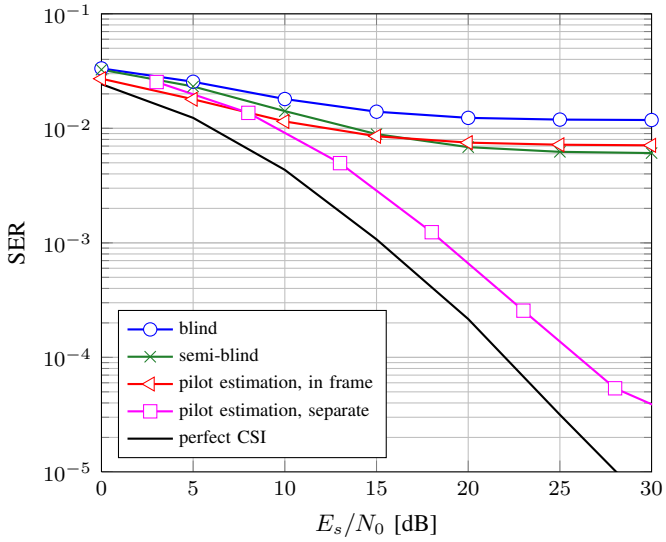

Fig. 4. Symbol error rate over the augmented alphabet for $K = 64$ and $F = 128$, i.e., under-determined case.

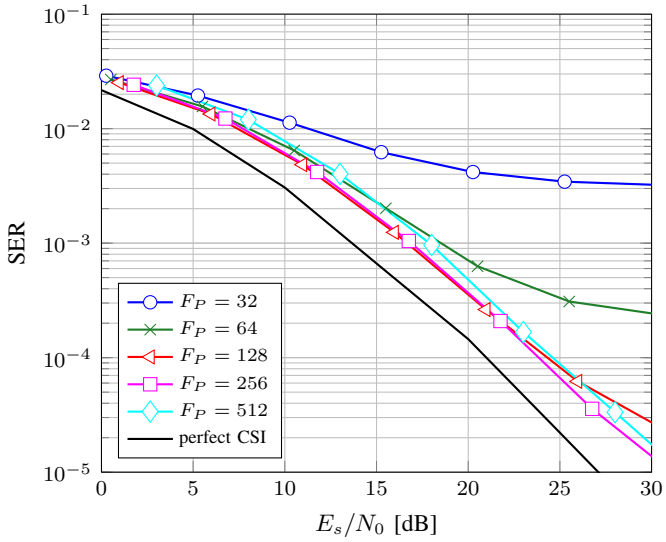

Fig. 5. Symbol error rates for pilot estimation using a separate pilot frame, different sizes of F_P, $K = 64$, and $F = 512$.

CSI compared to Figure 3. However, the same does not hold true for semi-blind and pilot estimation. These two approaches have significantly increased SERs due to increased multi-user interference. Comparing the pilot estimation for pilots within the data frame with the semi-blind estimation, we see that for the underdetermined case the semi-blind estimation is better for high SNR. As pilot estimation transmits N_F pilot symbols in the data frame instead of one, the interference is increased significantly.

Finally, we investigate the influence of the length of the separate pilot frame F_P for pilot estimation. Figure 5 shows the SER for different values of F_P. These results indicate that a moderate size of the additional pilot frame, i.e., $F_P = 128$, yields the best results, except for very high SNR. On the one hand, shorter pilot frames have higher spectral efficiency, but also a much higher error floor. On the other hand, longer pilot frames have much lower error rates with no noticeable error floor, but also a much lower spectral efficiency. Thus, the length of the pilot frame F_P defines a tradeoff between detection accuracy and spectral efficiency.

VI. Conclusion

This paper shows that joint activity and channel estimation for sporadic communication with very short packages is not only possible, but can almost achieve the error rate of perfect CSI. We investigated different approaches to estimating activity, data and channels in a sporadic multi-user transmission, which is typical for Machine-to-Machine communication. When avoiding signaling for activity, the best approach is to jointly estimate activity and channels based on pilot symbols, followed by a data detection. For high SNRs, these pilots should be transmitted in a separate frame, which causes a loss of spectral efficiency, while for low SNR the pilots should be transmitted within the data frame. To reliably perform this joint activity and channel estimation, we also introduced a new greedy Compressive Sensing algorithm (HBOMP).

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] 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] M. S. Asif, W. Mantzel, and J. Romberg, "Channel protection: Random coding meets sparse channels," in *IEEE Information Theory Workshop*, Taormina, Sicily, October 2009.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *Transactions on Emerging Telecommunications Technologies: Special Issue on Machine-to-Machine: An emerging communication paradigm*, vol. 24, no. 4, pp. 389 – 400, June 2013.