# Reliable Activity Detection for Massive Machine to Machine Communication via Multiple Measurement Vector Compressed Sensing

Fabian Monsees, Carsten Bockelmann and Armin Dekorsy
Department of Communications Engineering
University of Bremen, Germany
E-mail: {monsees,bockelmann,dekorsy}@ant.uni-bremen.de

***Abstract*—Compressed sensing based multiuser detection is a novel research field in massive machine to machine communication. Mainly focusing at decreasing signaling overhead, this approach implements sophisticated detection algorithms at the physical layer that jointly estimate activity and data. As a consequence, the reliability of the activity detection is crucial for the system performance as data is lost if users are erroneously classified as inactive. This paper introduces a novel approach to estimate node activity on a per frame basis by Multiple Measurement Vector Compressed Sensing approaches. This approach allows for reliable activity detection with complexity invariant of the length of the transmitted frame. Moreover, we are able to show that this approach works with only a few measurements available to the detector. In particular we demonstrate that reliable activity detection is possible if the number of observations is larger than the square root of the number of nodes in the system.**

## I. Introduction

Machine Type Communication (MTC) is one of the emerging fields for future communication systems. Besides human driven communications, MTC involves traffic between autonomous entities without human interaction in mind. MTC traffic differs in a variety of parameters from traffic caused by humans. Exemplary, many MTC applications such as sensor networks, smart meters or medical applications are of very low data rate and MTC devices are inactive for most of the time [1]. This renders MTC traffic to be sporadic in time which complicates the integration of MTC in high data rate cellular systems such as LTE. Sporadic and low data rate traffic easily leads to situations where signaling overhead is higher than the payload. A novel physical layer technique to reduce signaling for low data rate and sporadic MTC is to perform joint activity and data detection [2], [3], also summarized as Compressed Sensing based Multiuser Detection (CS-MUD). This method utilizes algorithms from sparse signal processing and Compressed Sensing to estimate the activity of MTC terminals and the corresponding data.

The core part of CS-MUD is to estimate the activity and the data of the nodes at the physical layer, which is a challenging task. The performance of the scheme with respect to correct activity detection is crucial for the system performance as data is lost if nodes are erroneously classified as inactive. A variety of works have analyzed the so-called false alarm and missed detection activity errors and some approaches to control these errors exist [4], [3]. In [4] the authors introduced a Bayes-Risk based detector as a generalized maximum-a-posteriori (MAP) detector that allows a weighting between false alarm and missed detection activity errors. In [5] the authors utilize coding of the data to improve the activity detection performed by a modified Orthogonal Matching Pursuit (OMP) algorithm.

Similar to [5] this paper focuses on improving the activity detection by exploiting knowledge about frame based transmissions. More specifically we utilize the knowledge that nodes are either active or inactive for the duration of a whole frame. The novelty of this approach is to cast the activity estimation as a Multiple Measurement Vector Compressed Sensing (MMV-CS) problem. Based on MMV-CS we show how activity estimation is done with very low complexity detection schemes. The major advantage of this approach is that the complexity does not depend on the length of the frame transmitted by the nodes. This reduces the complexity compared to previous schemes that estimated frame (in)activity on a symbol-by-symbol fashion. In [6] the authors used $L$ consecutive symbol-by-symbol MAP estimators and combined the result to obtain a per frame estimate for node (in)activity, which leads to high complexity especially for long frames. Besides its low complexity, MMV-CS allows us to derive bounds on the parameters for which activity estimation will succeed. We show in this paper that successful activity estimation is possible as long as the number of observations at the detector is larger than the square root of the number of nodes in the system. Most interestingly this relationship holds even if the activity in the system increases.

For reliable activity estimation, this paper introduces two different detection schemes. First, we set up a Matrix Matching Pursuit (MMP) algorithm as an extension of the well known Orthogonal Matching Pursuit (OMP) for solving underdetermined matrix problems. Compared to the OMP, the MMP does not require a matrix inversion. Beyond that and as the major contribution, we set up an approximate MAP detection scheme that refers to the solution of a regularized least-squares problem, which is fully or even overdetermined if the square root of the number of nodes is larger than the observations available at the detector. This relation allows for low-complexity solvers such as Successive Interference Cancellation (SIC).

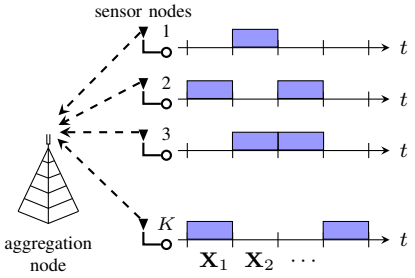


Fig. 1. Machine to machine uplink scenario with $K$ nodes transmitting a frames of length $L$ to a central aggregation node.

Finally, we demonstrate the performance of our schemes in different simulation setups. To demonstrate the performance with respect to varying node activity we use so called phase transition diagrams. Moreover, we investigate the stability with respect to frame asynchronicity and show that the performance is preserved even if nodes transmit asynchronously. We also compare the reliability with algorithms from literature.

## II. System Model

Throughout this work, the following detection model, also depicted in figure 1 is assumed. A set of $K$ nodes sporadically accesses the wireless channel to transmit modulation symbols to a central aggregation node. The sporadic nature of the nodes is captured using a simple traffic model parametrized by the node activity probability $p_a$, that we assume to be the same for all nodes. With this model a node starts transmitting at the beginning of a frame with probability $p_a$. With probability $1-p_a$ the node is silent for the duration of the whole frame. Active nodes modulate their data by using some type of modulation alphabet $\mathcal{A} \in \mathbb{C}$ which is of cardinality $M$ such as Phase Shift Keying (PSK) or Quadrature Amplitude Modulation (QAM). For joint detection of activity and data, we model inactive nodes as transmitting zeros instead of modulation symbols. More precise, we capture inactivity at the detector by using an augmented modulation alphabet, $\mathcal{A}_0 := \mathcal{A} \cup \{0\}$ that accounts for active and inactive nodes. Stacking the modulation symbols from active and inactive nodes together yields a multi-user vector $\mathbf{x} \in \mathcal{A}_0^K$, which is sparse if $p_a$ is sufficiently low. We further assume that nodes transmit frames of length $L$ to the central aggregation node. Collecting $L$ consecutive multiuser vectors yields the sparse matrix $\mathbf{X} \in \mathcal{A}^{K \times L}$. Here the $l$th column vector contains the modulation symbols for the $K$ nodes at time instance $l$. The $k$th row contains one transmitted frame from node $k$, respectively. As medium access and for multiuser detection we assume user specific random spreading via constant modulus sequences of the form $\mathbf{t}_k \sim \exp(j2\pi U) \in \mathbb{C}^K$, with $U \sim \mathcal{U}(0,1)$ being uniformly distributed. The spreading sequences are of constant modulus which simplifies transmission via low cost amplifiers. For the detection model, we collect the spreading sequences of length $m$ in the columns of the matrix $\mathbf{T} \in \mathbb{C}^{m \times K}$. This allows to state a per-frame detection model as follows

$$\mathbf{Y} = \mathbf{T}\mathbf{X} + \mathbf{W}. \tag{1}$$

Here we have omitted any time indexing for multiple frames to preserve clarity. The matrix $\mathbf{W} \in \mathbb{C}^{m \times L}$ contains i.i.d. samples from a symmetric circularly distributed white Gaussian noise process with variance $\sigma_n^2$. In cases where $m < K$, the system is underdetermined and relates to the theory of Compressed Sensing. More specifically, $\mathbf{Y}$ contains $L$ measurement vectors that carry information about the pattern of non-zero and zero elements in $\mathbf{X}$. In Compressed Sensing this relates to the the class of MMV-CS problems [7]. The formulation of (1) implicitly contains an assumption about synchronicity. Nodes are assumed to be synchronized in the network such that all frames start and end at the same time. In section IV we also demonstrate that the introduced algorithms are to some extend robust against frame asynchronicity.

### A. Activity Detection Model

In the following we sketch the detection model for the system described in (1). Our goal is to estimate node activity without estimating the underlying data. Data detection can be done after the set of active nodes has been estimated correctly. For activity estimation, we focus on the covariance matrix of the received signal $\mathrm{E}\left(\mathbf{Y}\mathbf{Y}^H\right) = \mathbf{T}\mathbf{D}\mathbf{T}^H + \sigma_n^2\mathbf{I}_m$, where $\mathbf{D} = \mathrm{E}\left(\mathbf{X}\mathbf{X}^H\right)$ is a diagonal matrix whose $d_{k,k}$th entry is one, if the $k$th node was active and zero otherwise. This assumption implicitly assumes uncorrelated data symbols with power $\sigma_k^2 = 1$ at the nodes. Our goal is to estimate the activity of the nodes, which is fully determined by the diagonal elements of the matrix $\mathbf{D}$. Thus, the detection task is to recover the position of the non-zero diagonal elements in $\mathbf{D}$. In this work we denote the sparsity $S \leq K$ for the number of active nodes within one frame. The covariance matrix is not available at the detector and we use the sample covariance matrix denoted as $\Phi_{YY}$ calculated over a frame of length $L$ to achieve a proper detection model

$$\Phi_{YY} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^H = \mathbf{T}\mathbf{D}\mathbf{T}^H + \Phi_{WW}. \tag{2}$$

Here, the matrix $\Phi_{WW} = \frac{1}{L}\mathbf{W}\mathbf{W}^H$ models the sample noise covariance matrix. This matrix obeys the class of Wishart distributions and can be separated as $\Phi_{WW} = \sigma_n^2\mathbf{I}_m + \mathbf{N}$, where $\sigma_n^2\mathbf{I}_m$ corresponds to the mean of the covariance matrix and $\mathbf{N}$ is a random variable with zero mean and variance $\sigma_n^4/L$. In the limit, $L \to \infty$, $\Phi_{WW}$ converges to its mean $\sigma_n^2\mathbf{I}_m$ while the remaining error, modeled by $\mathbf{N}$ converges to zero.

## III. Algorithms

In this section we introduce two algorithms for solving (2) in terms of finding the diagonal elements of $\mathbf{D}$.

### A. Matrix-Matching Pursuit

A very simple and straightforward approach is to incorporate a Matching Pursuit [8] (MP) algorithm that iteratively estimates the non-zero diagonal elements of $\mathbf{D}$. We subsequently denote this algorithms as Matrix Matching pursuit (MMP) as the underlying estimation task is over matrix dictionaries and not as usually over vector dictionaries. To illustrate this idea we rewrite the detection model (2) by neglecting the noise at this point as

$$\Phi_{YY} = \mathbf{T}\mathbf{D}\mathbf{T}^H = \sum_{k=1}^{K} d_{k,k}\mathbf{t}_k\mathbf{t}_k^H. \tag{3}$$

Here $d_{k,k}$ denotes the $k$th diagonal element of the matrix $\mathbf{D}$. The outer product $\mathbf{t}_k\mathbf{t}_k^H$ corresponding to an entry of an over complete dictionary also knows as atom. The MMP algorithm correlates the atoms with the sample covariance matrix $\Phi_{YY}$ and selects the atom with the highest correlation as active. As we a-priori know that the amplitude of $d_{k,k} = 1$ for active elements, we can avoid estimation. In the next step the MMP subtracts the contribution of the selected atom from the received signal. The algorithm repeats this procedure and correlates the atoms with the updated received signal and selects the next atom in the same manner as described before. The MP algorithm repeats this procedure until a pre-defined stopping criterion is met. The pseudo-code of this algorithm is demonstrated in listing 1. In this paper we assume that the algorithm has a-priori knowledge about the instantaneous sparsity $S$, which leads to the benchmark performance of the algorithm.

**Algorithm 1** Matrix Matching Pursuit (MMP)

$\mathbf{R}^0 = \Phi_{YY}$, $\Gamma^0 = \emptyset$, $v = 0$
**repeat**
  $v = v + 1$
  $i_{\max} = \arg\max_i \ \mathbf{t}_i^H\mathbf{R}^{v-1}\mathbf{t}_i$ {Matrix Correlation}
  $\hat{\mathbf{D}}_{i,i}^v = 1$
  $\mathbf{R}^v = \Phi_{YY} - \mathbf{T}\hat{\mathbf{D}}^v\mathbf{T}^H$
**until** $v = S$

The main advantage of the MMP is its low complexity and simplicity. The pseudo code of this algorithm shows that it mainly consists of a correlator bank composed of $K$ different correlators identifying the strongest user. In this setting we save the estimation of the value of the diagonal elements $d_{k,k}$ which we know to be either one or zero. If the amplitude of $d_{k,k}$ has to be estimated as well,e.g, in one tap fading channels, the MMP can easily be augmented by performing an estimation step after the matrix correlation.

### B. Approximate MAP Detection

In this section we formally state the MAP estimation problem (3) which we later solve via a suboptimal but efficient detector. To ease notation, we first note that $\mathbf{D}$ is solely determined by its main diagonal elements denoted as $\mathbf{d} := \mathrm{diag}(\mathbf{D})$. We formulate the MAP estimator for the problem

$$\begin{aligned}\Phi_{YY} &= \mathbf{T}\mathbf{D}\mathbf{T}^H + \Phi_{WW} \\ &= \mathbf{T}\mathbf{D}\mathbf{T}^H + \sigma_n^2\mathbf{I}_m + \mathbf{N}\end{aligned} \tag{4}$$

Here, the remaining noise $\mathbf{N}$ is of zero mean. The statistical properties of $\mathbf{N}$ are determined by $\Phi_{WW}$ which obeys a Wishart distributions as stated previously. Instead of using the Wishart variate, we employ a simpler Gaussian model as approximation for the remaining noise $\mathbf{N}$ to ease further calculations. Thus, we model the elements of $\mathbf{N}$ to be symmetric circular Gaussian distributed with variance $\sigma_n^4/L$. Note, that this approximation keeps mean and variance, but changes the type of the distribution.

The MAP estimate $\hat{\mathbf{D}}$ maximizes the a-posterori probability $p(\mathbf{D}|\Phi_{YY})$ which is usually reformulated via the Bayes-Rule to

$$\hat{\mathbf{D}} = \underset{\mathbf{d}\in\{0,1\}^K}{\arg\max}\ p(\Phi_{YY}|\mathbf{D})\Pr(\mathbf{D}). \tag{5}$$

The likelihood function $p(\Phi_{YY}|\mathbf{D})$ is solely determined by the statistics of the remaining noise $\mathbf{N}$ which we approximate by a Gaussian distribution as stated previously. With this assumption the likelihood function is proportional to a matrix variate Gaussian distribution

$$p(\Phi_{YY}|\mathbf{D}) \propto \exp\left[-\frac{L}{\sigma_n^4}\|\Phi_{YY} - \sigma_n^2\mathbf{I}_m - \mathbf{T}\mathbf{D}\mathbf{T}^H\|_F^2\right]. \tag{6}$$

From now on we summarize $\bar{\Phi}_{YY} = \Phi_{YY} - \sigma_n^2\mathbf{I}_m$ to preserve clarity. Additionally, we can write the prior in terms of the diagonal elements of $\mathbf{D}$ via

$$\Pr(\mathbf{D}) = (1-p_a)^{K-\|\mathbf{d}\|_0} \cdot p_a^{\|\mathbf{d}\|_0}, \tag{7}$$

where $\|\mathbf{d}\|_0$ is the so-called zero-"norm"[1] that simply counts the non-zero elements in $\mathbf{d}$. Taking $\log$ of the functional in (5) yields with likelihood function (6) and prior (7)

$$\hat{\mathbf{D}} = \underset{\mathbf{d}\in\{0,1\}^K}{\arg\min}\ \|\bar{\Phi}_{YY} - \mathbf{T}\mathbf{D}\mathbf{T}^H\|_F^2 + \lambda\|\mathbf{d}\|_0 \tag{8}$$

with $\lambda = \frac{\sigma_n^4}{L}\log\left(\frac{1-p_a}{p_a}\right)$.

The optimization problem (8) is still dependent on the matrix $\mathbf{D}$ even though $\mathbf{D}$ is solely determined by its main diagonal elements. The main focus now is to reformulate (8) into a vector optimization problem only depending on $\mathbf{d}$. To do so, we introduce the following augmented descriptions to come to an equivalent real system description

$$\hat{\mathbf{D}} = \underset{\mathbf{d}\in\{0,1\}^K}{\arg\min}\ \|\tilde{\bar{\Phi}}_{YY} - \tilde{\mathbf{T}}_1\tilde{\mathbf{D}}\tilde{\mathbf{T}}_2^T\|_F^2 + \lambda\|\mathbf{d}\|_0, \tag{9}$$

where $\tilde{\bar{\Phi}}_{YY} = \left[\mathrm{Re}\{\bar{\Phi}_{YY}^T\}, \mathrm{Im}\{\bar{\Phi}_{YY}^T\}\right]^T$ is the stacked matrix composed of real and imaginary part of $\bar{\Phi}_{YY}$. The matrices $\tilde{\mathbf{T}}_1$ and $\tilde{\mathbf{T}}_2$ are

$$\tilde{\mathbf{T}}_1 = \begin{bmatrix}\mathrm{Re}\{\mathbf{T}\} & -\mathrm{Im}\{\mathbf{T}\} \\ \mathrm{Im}\{\mathbf{T}\} & \mathrm{Re}\{\mathbf{T}\}\end{bmatrix},\ \tilde{\mathbf{T}}_2 = \begin{bmatrix}\mathrm{Re}\{\mathbf{T}\} & -\mathrm{Im}\{\mathbf{T}\}\end{bmatrix}. \tag{10}$$

[1] In fact the zero-"norm" is neither a norm nor a pseudo norm. However, the expression zero-"norm" is commonly used in Compressed Sensing contexts

The matrix $\tilde{\mathbf{D}}$ is composed of

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & \mathbf{0}_{K,K} \\ \mathbf{0}_{K,K} & \mathbf{D} \end{bmatrix}. \tag{11}$$

We now rewrite (9) using the $\mathrm{vec}\,(\cdot)$ operator which stacks the columns of a matrix. This turns the Frobenius matrix norm into an $\ell_2$ vector norm. Additionally, we make use of the identity $\mathrm{vec}\,(\mathbf{XYZ}) = \mathbf{Z}^T \otimes \mathbf{X}\,\mathrm{vec}\,(\mathbf{Y})$ [9] which allows us to reformulate (9) to

$$\hat{\mathbf{D}} = \underset{\mathbf{d}\in\{0,1\}^K}{\arg\min} \left\| \mathrm{vec}\left(\tilde{\Phi}_{YY}\right) - \tilde{\mathbf{T}}_2 \otimes \tilde{\mathbf{T}}_1\,\mathrm{vec}\left(\tilde{\mathbf{D}}\right)\right\|_2^2 + \lambda\|\mathbf{d}\|_0. \tag{12}$$

Writing $\boldsymbol{\varphi}_{YY} = \mathrm{vec}\,(\Phi_{YY})$ and $\boldsymbol{\Upsilon} = \tilde{\mathbf{T}}_2 \otimes \tilde{\mathbf{T}}_1$ expresses (12) finally as

$$\hat{\mathbf{D}} = \underset{\mathbf{d}\in\{0,1\}^K}{\arg\min} \|\boldsymbol{\varphi}_{YY} - \boldsymbol{\Upsilon}\mathbf{B}\mathbf{d}\|_2^2 + \lambda\|\mathbf{d}\|_0. \tag{13}$$

Here, the matrix $\mathbf{B}$ is used to obtain the transformation $\mathrm{vec}(\tilde{\mathbf{D}}) = \mathbf{B}\mathbf{d}$. This matrix is composed of two stacked matrices as

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0}_{2K,K} \\ \mathbf{B}_1 \end{bmatrix} \tag{14}$$

Where the row $[i(2K+1) - 2K], i = 1, \ldots K$ of $\mathbf{B}_1$ contains the $i$th row of the $K \times K$ identity matrix, while the remaining entries are zeros yielding the matrix $\mathbf{B} \in \{0,1\}^{4K^2 \times K}$.

The matrix $\boldsymbol{\Upsilon}\mathbf{B} \in \mathbb{R}^{2m^2 \times K}$ is of rank $\min\left[m^2, K\right]$ [10]. This fact plays a key role in our scheme as it shows that as long as $K \leq m^2$ holds, $\boldsymbol{\Upsilon}\mathbf{B}$ is of full rank which allows for fast implementation. To solve this problem efficiently, we employ SIC with sorted QR decomposition as pre-processing which has been shown to nearly achieve MAP performance [11] at low complexity.

We therefore assume $K \leq m^2$ which allows us to decompose $\boldsymbol{\Upsilon}\mathbf{B}$ using the *skinny* QR decomposition yielding the matrix $\mathbf{Q} \in \mathbb{R}^{2m^2 \times K}$ and the upper triangular matrix $\mathbf{R} \in \mathbb{R}^{K \times K}$ such that $\boldsymbol{\Upsilon}\mathbf{B} = \mathbf{QR}$ holds. Multiplying the argument of the $\ell_2$ norm in (12) with $\mathbf{Q}^T$ triangulizes the system to

$$\hat{\mathbf{d}} = \underset{\mathbf{d}\in\{0,1\}}{\arg\max} \|\tilde{\boldsymbol{\varphi}}_{YY} - \mathbf{R}\mathbf{d}\|_2^2 + \lambda\|\mathbf{d}\|_0, \tag{15}$$

with $\tilde{\boldsymbol{\varphi}}_{YY} = \mathbf{Q}^T\boldsymbol{\varphi}_{YY}$. The regularization term involving the calculation of a zero-"norm" can easily be scalarized and used as an additional penalty term in the calculation of the SIC. For further details on this approach the reader is referred to [4] Sec. II-C. The result of the optimization problem $\mathbf{d}$ is the estimate of the support set of active nodes and contains no information about the data transmitted. In the following, we summarize the pre-processing steps and SIC as MMV-SIC algorithm.

## IV. Performance Analysis

In the following, we exemplary show the performance of the algorithms with respect to activity errors by investigating an overloaded multiuser system analogous to the setup in section II. The activity error rate can easily be defined via the zero-"norm" as

$$\eta = \frac{1}{K}\,\mathrm{E}\left(\|\mathbf{d} - \hat{\mathbf{d}}\|_0\right), \tag{16}$$

where $\mathbf{d}$ is the true activity pattern and $\hat{\mathbf{d}}$ is the estimate at the output of the detector. Our results restrict to the performance with respect to activity detection. Once the set of active nodes is found by the activity detector, the underlying data can be estimated by known techniques from communications.

We perform our simulations using a 4-PSK modulation. However, the performance of the algorithm is independent of the cardinality of the modulation alphabet. Further, the channel is assumed to be AWGN at this point. The channel impulse responses can easily incorporated in the formulation of the system (1) via the matrix $\mathbf{T}$. Some approaches for channel estimation in the context of CS-MUD exist [12] but are out of the scope of this paper and subject to further research.

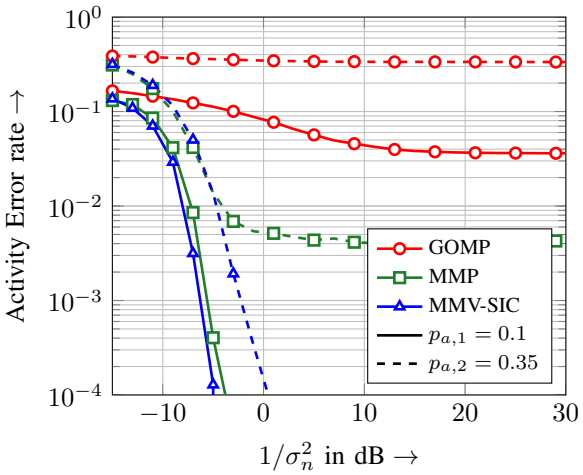


Fig. 2. Activity error rates in a setup with $m = 15$ observations, $K = 100$ nodes transmitting frames of $L = 1000$ symbols for different activity probabilities.

Figure 2 compares the activity error rates of the MMP and MMV-SIC with a Group Orthogonal Matching Pursuit (GOMP) [13] for a system with $K = 100$ nodes that are sporadically active and transmit data using spreading sequences of length $m = 15$. Conventionally, the system is overloaded by a factor of $\beta = 6.6$. Performance is investigated for two different node activity probabilities $p_{a1} = 0.1$ and $p_{a2} = 0.35$. In the first case nodes have a quite low activity probability and only $10\%$ of the nodes are simultaneously active on average. It can be seen that the performance of MMP and MMV-SIC is nearly the same for this setting. The GOMP algorithm does not achieve error rates below $\approx 4 \times 10^{-2}$. In the next setting, the activity probability has increased to $p_{a2} = 0.35$. In this setting, the MMP exhibits an error floor below $10^{-2}$. In contrast the MMV-SIC suffers only from a small SNR loss by $\approx 5$dB.

To get a better overview regarding $p_a$ and $m$, Fig. 3 shows a so-called phase transition diagram which is loosely based on the Donoho Tanner phase transition diagrams. The diagram is

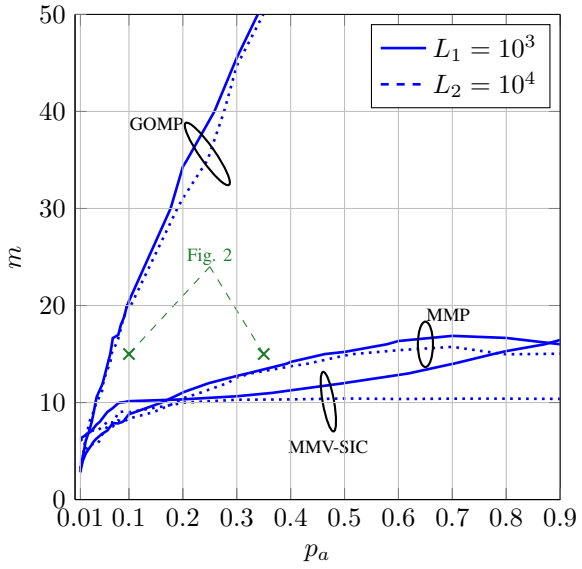


Fig. 3. Phase transition diagram for different frame length. The area above the curve denotes the tuples of $m$ and $p_a$ where the detector is able achieve an activity error rate of $\eta \leq 10^{-2}$

simulated at $1/\sigma_n^2 = 10$dB for two different frame lengths $L_1 = 10^3$ and $L_2 = 10^4$. The area above each curve denotes the tuples of $p_a$ and $m$, where the corresponding detector achieves at least an activity error rate of $10^{-2}$. The GOMP algorithm is very sensitive to variations of $p_a$, leading to a strong increase in the required spreading sequence length $m$. Increasing the length of the transmitted frame yields no significant performance gain. In contrast to that the MMP algorithm is only moderately affected by variations of $p_a$ and significantly outperforms GOMP. Finally, and most interestingly the MMV-SIC has even better performance than the MMP algorithm. For shorter frames $L_1 = 10^3$, the MMV-SIC requires more observations for higher node activity. The required number of observations is in this case approximately $m = 15$. This demand changes for longer frames. For $L = 10^4$ the MMV-SIC algorithm meets the activity error rate demand with only $m = 11$ observations for all activity probabilities shown. The system is overloaded by a factor of $\beta \approx 10$ in this case. This is a direct consequence of the fact that (13) is fully determined if $m = \sqrt{K} = 10$ holds. The fact that we need one further observation may be due to the fact that we incorporate noise during the measurement process. This simulation shows that the MMV-SIC very stable over over a wide range of activity probabilities.

All algorithms rely on the estimation of the sample covariance matrix and increasing $L$ yields a better estimate. To illustrate this effect, figure 4 shows the dependency of the detection performance on the frame length $L$. The plot shows $1/\sigma_n^2$ in dB that is required to meet an activity error rate constraint of $\eta \leq 10^{-2}$. The results in figure 4 incorporate a spreading sequence length of $m = 15$ and $K = 100$ nodes with an activity probability of $p_a = 0.35$. Curves for GOMP are not shown in this setting since GOMP does not achieve any reliable detection. Both algorithms first achieve the required error rate requirement at a frame length of $L = 300$. The performance increases as the frame length increased up to a nearly log-log-linear behavior which is due to the fact that the noise variance is averaged out with higher frame length. In particular, and as shown in Section III-B remaining noise variance is $\sigma_n^4/L$ which is a linear decrease in $L$.

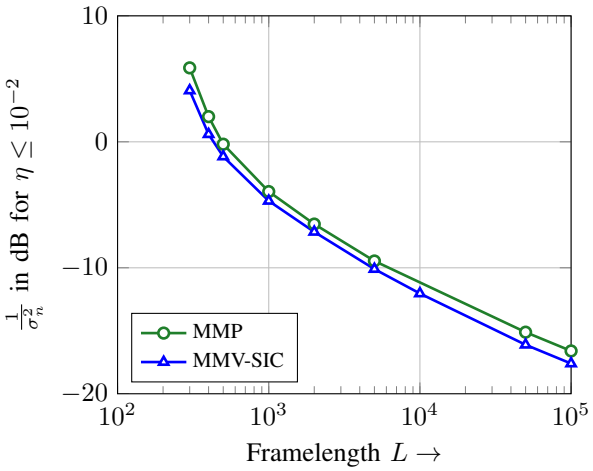


Fig. 4. $\frac{1}{\sigma_n^2}$ required to meet an activity error rate of $\eta \leq 10^{-2}$ for different frame-length $L$.

### A. Asynchronicity

Subsequently, we investigate the impact of frame asynchronicity on the activity detection performance for the introduced schemes. In this investigation nodes transmit their frames with a time shift to the clock assumed by the detector. This time shift is modeled in multiples of symbol durations and is restricted to the interval

$$\Delta_L \in \left[-\frac{\Delta_{L,\max}}{2}, \dots \frac{\Delta_{L,\max}}{2}\right]. \tag{17}$$

In this model $\Delta_{L,\max}$ is the highest possible time shift that can occur between two nodes in the network. $\Delta_L$ is modeled to be equally distributed between $\pm\frac{\Delta_{L,\max}}{2}$.

Fig. 5 illustrates asynchronous transmissions exemplary, where $\Delta_{L,3}$ exemplary shows the time shift between the clock at the detector and Node 3. As all nodes have different time shifts, the detector has to deal with crosstalk between frames which complicates the activity detection task. Previous and following transmissions may affect the activity estimation at the current frame.

Figure 6 shows the performance of the detection schemes under the impact of asynchronous transmissions. This investigation shows $1/\sigma_n^2$ in dB that is required to obtain an activity error rate $\eta \leq 10^{-2}$ as a function of the maximum relative asynchronicity $\Delta_{L,\max}/L$. The setup involves $K = 100$ nodes with an activity probability of $p_a = 0.35$ that spread their symbols with a spreading sequence length of $m = 15$, the length of the frames is set to $L = 1000$. It can be observed that MMP and MMV-SIC have very low performance loss even for severe time shifts of up to $\Delta_{L,\max}/L \approx 0.3$. Further

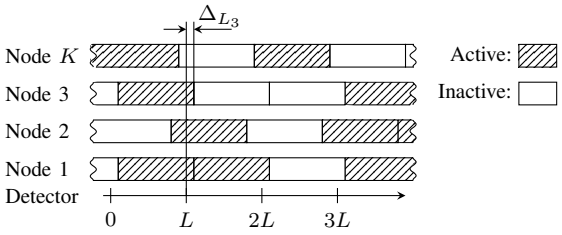


Fig. 5. Illustration of frame asynchronicity, where nodes are asynchronous to the clock assumed by the detector.

increasing the asynchronicity in the system decreases the performance where MMP fails to reach the required bound at $\Delta_{L,\max}/L = 0.5$ and MMV-SIC fails at $\Delta_{L,\max}/L = 0.7$, respectively.

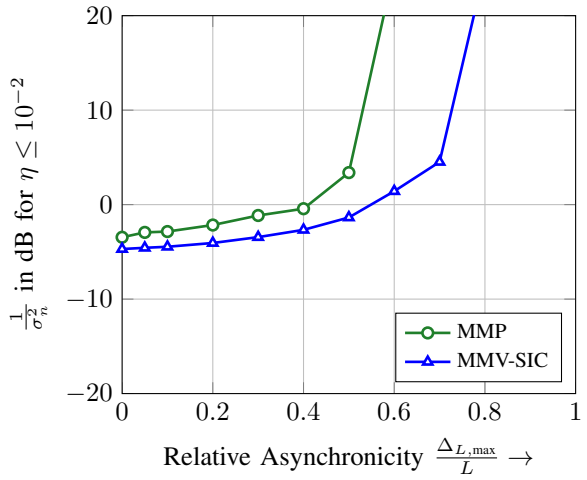


Fig. 6. Performance of the schemes under asynchronous transmissions in a setup with $K = 100$ nodes, $m = 15$ observations and a frame length of $L = 1000$. Per node activity probability is $p_a = 0.35$

## V. Conclusion

In this work we have shown how efficient low complexity activity estimation can be performed in machine to machine uplink communication where nodes sporadically transmit whole frames of data. The major contribution is to cast the activity estimation problem as an underdetermined multiple measurement vector Compressed Sensing problem. This allows for low complexity activity detection schemes with complexity that is invariant of the frame length. Moreover we have shown that reliable activity estimation is possible as long as the number of observations available to the detector is larger than the square root of the number of nodes. The introduced schemes are even robust to asynchronicity where reliable detection is possible even if time shifts up to 50% occur.

## Acknowledgment

This work was funded by the German Research Foundation (DFG) under grant DE 759/3-1.

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.

## References

[1] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 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 (ISWCS)*, Aachen, Germany, Nov., pp. 291 –295.

[3] C. Bockelmann, H. 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, Jun 2013.

[4] F. Monsees, C. Bockelmann, D. Wübben, and A. Dekorsy, "Compressed sensing bayes risk minimization for under-determined systems via sphere detection," in *77th IEEE Vehicular Technology Conference (VTC2013-Spring)*, Dresden, Germany, Jun 2013.

[5] H. Schepker, C. Bockelmann, and A. Dekorsy, "Improving greedy compressive sensing based multi-user detection with iterative feedback," in *2013 IEEE 78th Vehicular Technology Conference (VTC2013-Fall)*, Las Vegas, USA, Sep 2013.

[6] F. Monsees, C. Bockelmann, and A. Dekorsy, "Compressed sensing soft activity processing for sparse multi-user systems," in *9th IEEE Broadband Wireless Access Workshop co-located with IEEE Globecom 2013*, Atlanta, USA, Dec 2013.

[7] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

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

[9] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," October 2008, version 20081110.

[10] D. Cohen and Y. C. Eldar, "Sub-nyquist sampling for power spectrum sensing in cognitive radios: A unified approach," *arXiv:1308.5149*, 2013.

[11] D. Wübben, R. Böhnke, J. Rinas, K.-D. Kammeyer, and V. Kühn, "Efficient algorithm for decoding layered space-time codes," *IEE Electronic Letters*, vol. 37, no. 22, pp. 1348–1350, Nov 2001.

[12] H. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *10th International Symposium on Wireless Communication Systems (ISWCS 13)*, Ilmenau, Germany, Aug 2013.

[13] A. Majumdar and R. Ward, "Fast group sparse classification," *Canadian Journal of Electrical and Computer Engineering*, vol. 34, no. 4, pp. 136–144, Fall 2009.