A Hybrid Approach for Multi-User Massive MIMO Sparse Channel Estimation based on Bayesian Recovery and Hard Thresholding

Javier García, Jawad Munir, Amine Mezghani and Josef A. Nossek
Institute for Circuit Theory and Signal Processing
Technische Universität München, 80290 Munich, Germany
E-Mail: {javier.garcia, jawad.munir, amine.mezghani, josef.a.nossek}@tum.de

***Abstract*—We propose an efficient sparse channel estimation algorithm based on the compressed sensing (CS) approach for large scale multi-user (MU) MIMO systems. The proposed scheme is a hybrid one comprising Bayesian and greedy methods. It can improve the estimation performance by incorporating the spatial channel knowledge that the neighboring antennas in an array share the same support. The pilot overhead can be reduced by utilizing the data symbols using a reliability measure for channel estimation. Moreover, the effect of interfering and non-interfering pilots on the estimation performance will be investigated. It will be shown that the proposed hybrid technique performs similar or better than the Bayesian method with substantially reduced complexity.**

***Index terms*— multi-user massive MIMO, OFDM, distributed channel estimation; compressed sensing, sparse recovery**

I. Introduction

The demand for higher throughput and capacity is increasing at a staggering rate for cellular networks [1]. The deployment of very large number of antennas at the base station (BS), also known as massive MIMO, is considered as key enabler to achieve these growing demands of throughput. Moreover, massive MIMO can also provide substantial improvement in both spectral and energy efficiency [2].

The spectral efficiency in massive MIMO relies on the good estimate of the channel impulse response (CIR) or channel state information (CSI) at the base station. The downlink CSI can be estimated from the uplink CSI by exploiting the channel reciprocity in time-division duplex (TDD) mode. The uplink CSI is typically acquired by having the users send pilots, based on which the base station estimates the CIR to each user. Channel estimation in massive MIMO system becomes quite challenging and computationally expensive due to the large number of channel coefficients to estimate. The related work on massive MIMO channel estimation is reported in [3]-[4].

Recently, there has been growing interest to apply compressed sensing (CS) based techniques for massive MIMO channel estimation in order to reduce the computational complexity [5], [6]. CS based techniques rely on the idea that the wireless channel between the user and the base station is expected to be sparse in nature due to the finite number of scatterers. In [7], Masood et al. applied a Bayesian framework, Support Agnostic Bayesian Matching Pursuit (SABMP) [7], to estimate the sparse channels in massive MIMO-OFDM system. Their work can improve the estimation performance by incorporating the a priori knowledge that the neighboring antennas in an array observe the channel with a similar support. It can also reduce the pilot overhead using the data symbols for channel estimation based on a reliability measure.

The main contribution of this paper is the development of the weighted version of iterative hard thresholding (WIHT) to estimate the sparse channel and extend the work [8] to the multi-user massive MIMO-OFDM systems. We will investigate the effect of pilots sequence on same sub-carriers (interfering pilots) and time division multiplexed pilots (non-interfering pilots) on channel estimation. It will be shown that the proposed scheme has similar or better performance than the SABMP algorithm with less computational complexity. This paper is organized as follows. Section II introduces the system model. Section III formulates the sparse recovery problem and presents SAMBP and WIHT algorithms. Section V describes the methodology to share the common support among the antennas. A data-aided version of the algorithm is presented in Section VI. Simulation results are presented in Section VIII and conclusions are given in Section IX.

II. System Model

Let us consider the uplink of a multi-user Massive MIMO wireless system using OFDM with N subcarriers. The base station has $R = M \times G$ receive antennas and there are U single-antenna users. The channel from user $u \in \{1, \ldots, U\}$ to receive antenna $r \in \{1 \ldots, R\}$ is denoted as $\mathbf{h}_u^r = [h_u^r[0], \ldots, h_u^r[L-1]]^T$, where L is the channel length.

At each time instant, each user generates an OFDM symbol $\boldsymbol{\mathcal{X}}_u \in \mathbb{C}^{N\times 1}$ containing N individual Q-QAM modulated samples. The transmitter applies IFFT to $\boldsymbol{\mathcal{X}}_u$, adds a cyclic prefix and converts it to the carrier frequency. Each antenna r in the base station receives a linear combination of the symbols from all users u. The cyclic prefix is removed and an IFFT is applied to obtain $\boldsymbol{\mathcal{Y}}^r$. This can be compactly written as:

$$\boldsymbol{\mathcal{Y}}^r = \sum_{u=1}^{U} \left[\mathbf{F}\mathbf{H}_u^r\mathbf{F}^H\boldsymbol{\mathcal{X}}_u\right] + \boldsymbol{\mathcal{Z}}^r, \tag{1}$$

where $\mathbf{H}_u^r \in \mathbb{C}^{N\times N}$ is a circulant matrix whose rows are cyclical shifts of the channel impulse response $\mathbf{h}_u^r$, and $\boldsymbol{\mathcal{Z}}^r \in$

$\mathbb{C}^{N\times 1}$ is AWGN noise. Due to the circulant property of $\mathbf{H}_u^r$, we have $\mathbf{F}\mathbf{H}_u^r\mathbf{F}^H = \mathrm{diag}\left[\underline{\mathbf{F}}\mathbf{h}_u^r\right]$, where $\underline{\mathbf{F}} \in \mathbb{C}^{N\times L}$ is a matrix built with the first L columns of an $N \times N$ DFT matrix. Plugging this into (1) yields:

$$\begin{aligned}\boldsymbol{\mathcal{Y}}^r &= \sum_{u=1}^{U} \left[\mathrm{diag}\left[\underline{\mathbf{F}}\mathbf{h}_u^r\right]\boldsymbol{\mathcal{X}}_u\right] + \boldsymbol{\mathcal{Z}}^r \\ &= \sum_{u=1}^{U} \left[\mathrm{diag}\left[\boldsymbol{\mathcal{X}}_u\right]\underline{\mathbf{F}}\mathbf{h}_u^r\right] + \boldsymbol{\mathcal{Z}}^r = \sum_{u=1}^{U}\left[\underline{\mathbf{A}}_u\mathbf{h}_u^r\right] + \boldsymbol{\mathcal{Z}}^r, \quad (2)\end{aligned}$$

where $\underline{\mathbf{A}}_u = \mathrm{diag}\left[\boldsymbol{\mathcal{X}}_u\right]\underline{\mathbf{F}} \in \mathbb{C}^{N\times L}$.

The channel estimation problem amounts to obtaining an estimate of the channels $\mathbf{h}_u^r$ for all users u and all receive antennas r. For this purpose, known pilot symbols need to be sent on some subcarriers. From (2) it is seen that each user suffers interference from all others, unless the pilot sequences are orthogonal to each other. However, only a few mutually orthogonal sequences are available (no more than the sequence length). Therefore, we focus on a scenario where all users send pilots simultaneously and the sequences are not necessarily orthogonal, but chosen pseudo-randomly and different for each user to minimize correlation.

If the pilot locations are different for each user, the interference will come from the other users' pilots; otherwise, it will come from their data symbols. The knowledge of the pilot sequences of the other users allows for an iterative interference cancellation technique which improves estimation performance and will be presented later in this paper. Therefore, all users have the same pilot locations, which are chosen pseudo-randomly. Let us denote the number of pilots by P, and the set of indices of the pilot locations by $\mathcal{P}$. Then, the received symbols in these locations are given by:

$$\overline{\boldsymbol{\mathcal{Y}}}^r = \sum_{u=1}^{U}\left[\mathbf{A}_u\mathbf{h}_u^r\right] + \overline{\boldsymbol{\mathcal{Z}}}^r, \quad (3)$$

where $\overline{\boldsymbol{\mathcal{Y}}}^r = \boldsymbol{\mathcal{Y}}^r(\mathcal{P}) \in \mathbb{C}^{P\times 1}$, $\mathbf{A}_u = \underline{\mathbf{A}}_u(\mathcal{P}) \in \mathbb{C}^{P\times L}$, and $\overline{\boldsymbol{\mathcal{Z}}}^r = \boldsymbol{\mathcal{Z}}^r(\mathcal{P}) \in \mathbb{C}^{P\times 1}$ are obtained by indexing the rows of the corresponding matrices with the pilot indices $\mathcal{P}$.

III. Sparse recovery of individual channels

As depicted in Fig. 1, the individual channels $\mathbf{h}_u^r$ can be assumed to be sparse, due to the distribution of the scatterers in a typical propagation scenario. This allows us to apply compressed sensing techniques to obtain a first estimate of these channels. To estimate the channel to each user, all the contributions from the other users are considered noise and the following sparse reconstruction problem is solved:

$$\overline{\boldsymbol{\mathcal{Y}}}^r = \mathbf{A}_u\mathbf{h}_u^r + \overline{\boldsymbol{\mathcal{Z}}}_u^r, \quad (4)$$

where $\overline{\boldsymbol{\mathcal{Z}}}_u^r = \overline{\boldsymbol{\mathcal{Z}}}^r + \sum_{\substack{u'\in\{1,\ldots,R\}\\ u'\neq u}} \mathbf{A}_{u'}\mathbf{h}_{u'}^r$ combines the noise and the interference. The result of this estimation will not be very accurate because of the high interference, but an iterative technique to improve the result will be introduced later.

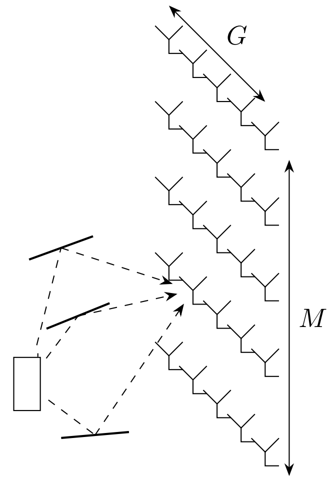

Fig. 1. Sparse distribution of scatterers in a Massive MIMO channel.

Sparse recovery is run independently for each user and antenna. Therefore, for the remainder of this section, the scripts u and r will be dropped for ease of notation. This work deals with greedy sparse recovery algorithms, which attempt to solve the optimization problem:

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}}\left\{\left\|\overline{\boldsymbol{\mathcal{Y}}} - \mathbf{A}\mathbf{h}\right\|_2^2\right\} \quad \text{s.t.} \quad \|\mathbf{h}\|_0 \leq T, \quad (5)$$

i.e., find the least-squares solution among all vectors with at most T active taps. There is another approach for sparse recovery, namely the ℓ_1 minimization algorithms [9], but their high computational complexity makes them less attractive for this Massive MIMO scenario, in which sparse recovery must be run a large amount of times.

Greedy sparse recovery algorithms need to know the sparsity T, but the performance of the algorithms is not degraded noticeably if this value is overestimated by a small amount. Thus, as proposed in [10], we choose T to be slightly higher than the number of elements of $\left|\mathbf{A}^H\overline{\boldsymbol{\mathcal{Y}}}\right|$ that are greater or equal than half its maximum value.

$$T = \left|\left\{j : \left|\mathbf{a}_j^H\overline{\boldsymbol{\mathcal{Y}}}\right| \geq \frac{1}{2}\left\|\mathbf{A}^H\overline{\boldsymbol{\mathcal{Y}}}\right\|_\infty\right\}\right|, \quad (6)$$

where $\mathbf{a}_j$ is the j-th column of $\mathbf{A}$.

We also note that the greedy sparse recovery algorithms admit the weighting $\mathbf{w} \in \mathbb{C}^{L\times 1}$ as input. This weighting is the a priori probability of each tap of the solution being active: $w[l] \triangleq p\left(h[l] \neq 0 \,\middle|\, \overline{\boldsymbol{\mathcal{Y}}}\right)$. It will play an important role in the collaborative estimation of the channels, but on the first run of sparse recovery there is still no information about these priors and they are all set to T/L.

A. Support Agnostic Bayesian Matching Pursuit (SABMP)

SABMP is a robust technique developed in [10], which provides a Bayesian estimate of $\mathbf{h}$ by performing a greedy search on the possible support sets $\mathcal{S}$ and applying the Theorem on Total Expectation. The support set is the set of indices of the active taps of the sparse vector: $\mathcal{S} = \{l \in \{0,\ldots,L-1\} : h[l] \neq 0\}$.

Algorithm 1 summarizes the SABMP method, where $\mathbf{J}_{\mathcal{S}} \in \mathbb{C}^{t \times L}$ is a binary matrix that selects the indices in the support set, $t = |\mathcal{S}|$, and $\mathbf{A}_{\mathcal{S}} = \mathbf{A}\mathbf{J}_{\mathcal{S}}^H \in \mathbb{C}^{K \times t}$. The scripts u and r have been dropped for ease of notation. More information and the derivation of this algorithm can be found in [10].

Algorithm 1 Greedy SABMP

Input: $\overline{\boldsymbol{\mathcal{Y}}}$, $\mathbf{A}$, $\mathbf{w}$, T
  Initialize: $\mathcal{S} = \{\}$, $\mathbf{p_T} = \mathbf{0}_{T\times 1}$
  for $t = 1 : T$ **do**
    $\mathbf{p}_{\text{temp}} = \mathbf{0}_{N\times 1}$
    for $n = 0 : (N-1)$, $n \notin \mathcal{S}$ **do**
      $\overline{\mathcal{S}} = \mathcal{S} \cup \{n\}$; $p\left(\overline{\mathcal{S}}\right) = \prod_{i\in\overline{\mathcal{S}}} w_l \prod_{j\notin\overline{\mathcal{S}}} (1-w_j)$
      $p_{\text{temp}}[n] = p\left(\overline{\mathcal{S}}\right) e^{-\frac{1}{\sigma_z^2}\left\|\left(\mathbf{I}-\mathbf{A}_{\overline{\mathcal{S}}}\left(\mathbf{A}_{\overline{\mathcal{S}}}^H\mathbf{A}_{\overline{\mathcal{S}}}\right)^{-1}\mathbf{A}_{\overline{\mathcal{S}}}^H\right)\overline{\boldsymbol{\mathcal{Y}}}\right\|_2^2}$
    end for
    find $n_{\max}$ **such that** $p_{\text{temp}}[n_{\max}] = \max[\mathbf{p}_{\text{temp}}]$
    $\mathcal{S} = \mathcal{S} \cup \{n_{\max}\}$; $p_T[t] = p_{\text{temp}}[n_{\max}]$
  end for
  Normalize: $\mathbf{p_T} = \mathbf{p_T}/\text{sum}(\mathbf{p_T})$
  $\hat{\mathbf{h}} = \sum_{t=1}^{T} p_K[t]\mathbf{J}_{\mathcal{S}_{(1:t)}}^H \left(\mathbf{A}_{\mathcal{S}_{(1:t)}}^H \mathbf{A}_{\mathcal{S}_{(1:t)}}\right)^{-1} \mathbf{A}_{\mathcal{S}_{(1:t)}}^H \overline{\boldsymbol{\mathcal{Y}}}$
Output: $\hat{\mathbf{h}}$

B. Weighted Iterative Hard Thresholding (WIHT)

The SABMP technique is very robust in a wide range of scenarios, but the greedy search over the support sets and the computation of a pseudo-inverse for each support set makes its complexity somewhat high, namely $\mathcal{O}\left(LPT^2\right)$ if the efficient implementation of pseudo-inverses of growing matrices proposed in [10] is used. Therefore, we propose an alternative much faster method that achieves good performance when the input priors $\mathbf{w}$ are accurate enough.

The proposed WIHT algorithm is based on Iterative Hard Thresholding (IHT) method [8]. IHT consists of a Steepest Descent approach in which a hard thresholding operator $H_T(\cdot)$ is applied after each step. This operator keeps the T maximum samples of its argument and sets all other samples to 0.

The proposed WIHT method runs a single iteration of standard IHT but applying the weighting $\mathbf{w}$ before thresholding:

$$\hat{\mathbf{h}}^{(1)} = H_T\left(\text{diag}[\mathbf{w}]\,\mathbf{A}^H\overline{\boldsymbol{\mathcal{Y}}}\right). \tag{7}$$

We found out that the values of the taps of the sparse vector $\mathbf{h}$ are generally not well approximated after only one iteration, but the support of $\hat{\mathbf{h}}^{(1)}$ usually coincides with that of $\mathbf{h}$. Therefore, Weighted IHT takes the support $\hat{\mathcal{S}}$ of $\hat{\mathbf{h}}^{(1)}$ and performs a Best Linear Unbiased Estimate (BLUE) over it, which reduces to a pseudo-inverse because the noise is assumed uncorrelated:

$$\hat{\mathbf{h}} = \mathbf{J}_{\hat{\mathcal{S}}}^H\left(\mathbf{A}_{\hat{\mathcal{S}}}^H\mathbf{A}_{\hat{\mathcal{S}}}\right)^{-1}\mathbf{A}_{\hat{\mathcal{S}}}^H\overline{\boldsymbol{\mathcal{Y}}}, \tag{8}$$

where $\mathbf{A}_{\hat{\mathcal{S}}}$ is obtained by taking the columns of $\mathbf{A}$ whose indices are in the support $\hat{\mathcal{S}} = \text{supp}\left(\hat{\mathbf{h}}^{(\text{init})}\right)$. This BLUE estimate approximates the conditional expectation over the observation and the computed support, $\hat{\mathbf{h}} \approx \text{E}\left[\mathbf{h}|\overline{\boldsymbol{\mathcal{Y}}}, \hat{\mathcal{S}}\right]$. The complexity of WIHT is given by this pseudo-inverse, and is $\mathcal{O}\left(PT^2\right)$, L times lower than that of SABMP.

Algorithm 2 Weighted Iterative Hard Thresholding (WIHT)

Input: $\overline{\boldsymbol{\mathcal{Y}}}$, $\mathbf{A}$, $\mathbf{w}$, T
  Weighting matrix: $\mathbf{W} = \text{diag}[\mathbf{w}]$
  Estimated support set: $\hat{\mathcal{S}} = \text{supp}\left[H_T\left(\mathbf{W}\mathbf{A}^H\overline{\boldsymbol{\mathcal{Y}}}\right)\right]$
  BLUE Estimate $\hat{\mathbf{h}} = \mathbf{J}_{\hat{\mathcal{S}}}^H\left(\mathbf{A}_{\hat{\mathcal{S}}}^H\mathbf{A}_{\hat{\mathcal{S}}}\right)^{-1}\mathbf{A}_{\hat{\mathcal{S}}}^H\overline{\boldsymbol{\mathcal{Y}}}$
Output: $\hat{\mathbf{h}}$

IV. CANCELLATION OF MULTI-USER INTERFERENCE

Each antenna runs sparse recovery separately for each user according to (5). This first estimation of the user-antenna channel $\hat{\mathbf{h}}_u^{r(\text{init})}$ is good enough for the single-user scenario considered in [10], but not very accurate in the multi-user case because of the interference from all other users. Therefore, a refinement step is proposed that consists of running sparse recovery again after subtracting from the observation vector the contribution of the current estimates of the channels to all other users. Mathematically, the following new observation vector is computed at each antenna for each user:

$$\overline{\boldsymbol{\mathcal{Y}}}_u^{r(\text{init})} = \overline{\boldsymbol{\mathcal{Y}}}^r - \sum_{\substack{u'\in\{1,\ldots,U\}\\ u'\neq u}} \mathbf{A}_{u'}\hat{\mathbf{h}}_{u'}^{r(\text{init})}, \tag{9}$$

and then sparse recovery is run again at each antenna for each user to compute the refined individual channel estimation:

$$\hat{\mathbf{h}}_u^{r(\text{ind})} = \arg\min_{\mathbf{h}_u^r}\left\{\left\|\overline{\boldsymbol{\mathcal{Y}}}_u^{r(\text{init})} - \mathbf{A}_u\mathbf{h}_u^r\right\|_2^2\right\} \text{ s.t. } \left\|\mathbf{h}_u^r\right\|_0 \leq T. \tag{10}$$

V. COLLABORATION BETWEEN ANTENNAS

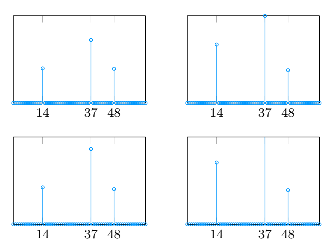

Fig. 2. Invariance of support set along neighboring antennas.

As introduced in [10] and depicted in Fig. 2, due to spatial correlation, the support sets of the channels from a given user u to the receiving antennas can be assumed to either have the same support set along the whole receiving array or a slowly varying support set which is constant along smaller regions of the array. The maximum distance along which the support set can be assumed constant is given by $d_{\max} = \frac{c}{10B}$, where c is the speed of light and B is the bandwidth [10].

This support set invariance can be exploited by adding a collaboration step to the channel estimation process, also proposed in [10]. After the individual channel estimation

$\hat{\mathbf{h}}_u^{r(\text{ind})}$ has been obtained from (10), a score $\tilde{w}_u^r[l]$ is computed for each tap of each channel $\hat{h}_u^{r(\text{ind})}[l]$. This score aims to roughly approximate the probability that the considered tap is active (nonzero), and is set to zero for all zero taps of $\hat{\mathbf{h}}_u^{r(\text{ind})}$. The nonzero taps get a score based on the rank of their magnitude in an ascending-order sorted list, normalized by the number of taps. This means that the highest tap in $\hat{\mathbf{h}}_u^{r(\text{ind})}$ gets a score of T/T, the second highest gets $(T-1)/T$, and so on:

$$\tilde{w}_u^r[l] = \begin{cases} 0 & \text{if } \hat{h}_u^{r(\text{ind})}[l] = 0 \\ t/T & \text{if } \left(\hat{h}_u^r r(\text{ind})\right)_{(T+1-t)} \equiv \hat{h}_u^{r(\text{ind})}[l], \end{cases} \quad (11)$$

where $(\cdot)_{(n)}$ denotes "n-th maximum sample". These scores are shared among antennas in an iterative fashion as explained in [10]: at each collaboration iteration, each antenna sends its scores to its neighboring antennas, and updates its own scores by averaging them with the ones it received from its neighbors.

After D collaboration iterations (chosen so that space invariance distance is not exceeded), sparse recovery algorithms are run again at each antenna, now using the updated scores $\tilde{\mathbf{w}}_u^{r(D)}$ as weighting input $\mathbf{w}$ for the algorithms. The only difference with respect to [10] is that, for the multi-user case, this new run of sparse recovery should be applied to a newly refined observation vector obtained by once again subtracting the estimated interference from all other users:

$$\overline{\boldsymbol{\mathcal{Y}}}_u^{r(\text{collab})} = \overline{\boldsymbol{\mathcal{Y}}}^r - \sum_{\substack{u' \in \{1,\ldots,U\} \\ u' \neq u}} \mathbf{A}_{u'} \hat{\mathbf{h}}_{u'}^{r(\text{ind})}. \quad (12)$$

The result of this run of sparse recovery is a new estimate of the channels $\hat{\mathbf{h}}_u^{r(\text{collab})}$.

VI. Refining the Estimate Using Reliable Carriers

A final improvement of the channel estimates can be achieved through the use of reliable carriers. Reliable carriers are data carriers in which we have high certainty that the transmitted symbol was not moved to a different constellation region by the noise, and therefore we can assume the equalized and demodulated value is correct. The geometric reliability measure proposed in [10] is used in this work:

$$R_{\text{geom}}\left(\hat{\boldsymbol{\mathcal{X}}}_u - Q\left[\hat{\boldsymbol{\mathcal{X}}}_u\right]\right) = \frac{\sqrt{2}d_{\text{min}} - \left|\hat{\boldsymbol{\mathcal{X}}}_u - Q\left[\hat{\boldsymbol{\mathcal{X}}}_u\right]\right|}{\sqrt{2}d_{\text{min}}} + \\ + \frac{\left|\hat{\boldsymbol{\mathcal{X}}}_u - Q\left[\hat{\boldsymbol{\mathcal{X}}}_u\right]\right|}{\sqrt{2}d_{\text{min}}} \cos\left(4\theta_{\hat{\boldsymbol{\mathcal{X}}}_u - Q[\hat{\boldsymbol{\mathcal{X}}}_u]} + \pi\right), \quad (13)$$

where $\hat{\boldsymbol{\mathcal{X}}}_u$ is the vector of equalized symbols, $Q[\cdot]$ obtains the closest constellation point to its argument, and $\theta_{\hat{\boldsymbol{\mathcal{X}}}_u - Q[\hat{\boldsymbol{\mathcal{X}}}_u]}$ is the angle to the closest constellation point. This reliability measure accounts for the fact that both a shorter distance to the closest constellation point and an angle to it closer to $\frac{2k+1}{4}\pi, k \in \{0,\ldots,3\}$ indicate a bigger gap between the distance to the closest constellation point and to the second closest, as depicted in Fig. 3.

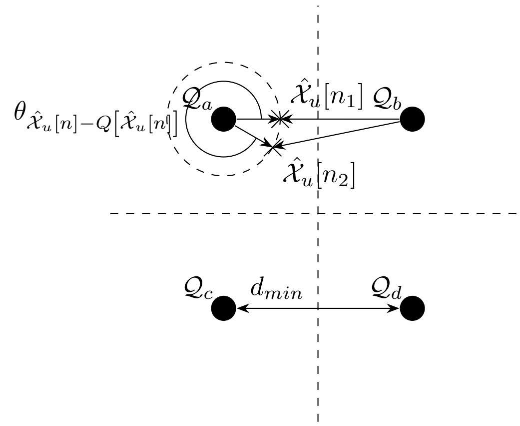

Fig. 3. Reliability measure of two received symbols

Here, another difference with respect to the single user scenario needs to be taken into account. When there is only one user, the channel can be equalized separately at each antenna and then a different set of reliable carriers chosen. However, in the multi-user case, this is no longer possible. At each subcarrier n, the received vector is:

$$\boldsymbol{\mathcal{Y}}[n] = \boldsymbol{\mathcal{H}}[n]\boldsymbol{\mathcal{X}}[n], \quad (14)$$

where $\boldsymbol{\mathcal{Y}}[n] \in \mathbb{C}^{R\times 1}$ stacks the received samples for subcarrier n at all antennas, $\boldsymbol{\mathcal{H}}[n] \in \mathbb{C}^{R\times U}$ is a matrix whose entry at row r and column u is the n-th sample of the DFT of the channel $\mathbf{h}_u^r$, and $\boldsymbol{\mathcal{X}}[n] \in \mathbb{C}^{U\times 1}$ stacks the symbols sent on subcarrier n by all users. Equalization can then only obtain a single estimate of $\hat{\boldsymbol{\mathcal{X}}}_u$ for all antennas. Therefore, the same set of reliable carriers must be chosen for all antennas.

Equalization does of course give different symbol estimations for each user, and therefore different sets of reliable carriers could be chosen for different users. However, this would degrade the effectiveness of the proposed iterative interference cancellation step (9), as the rows of the matrices $\mathbf{A}_u$ would correspond to different subcarriers for each user, and therefore the subtraction of interference would only be possible on the subcarriers that are simultaneously reliable for all users. It was found that better results are obtained if all users have the same set of reliable carriers. As the reliability measure is an approximation of the probability of no error, a joint reliability measure is proposed that is given by the product of the individual reliability measures:

$$R_{\text{geom}}[n] = \prod_{u=1}^{U} R_{\text{geom}}\left(\hat{\mathcal{X}}_u[n] - Q\left[\hat{\mathcal{X}}_u[n]\right]\right). \quad (15)$$

The set $\mathcal{R}$ of the K subcarriers with highest $R_{\text{geom}}[n]$ is chosen, and then a new matrix $\mathbf{A}_u^{(\text{rel})} = \underline{\mathbf{A}}_u(\mathcal{P} \cup \mathcal{R})$ is formed for each user, as well as a new observation vector:

$$\overline{\boldsymbol{\mathcal{Y}}}_u^{r(\text{rel})} = \boldsymbol{\mathcal{Y}}^r(\mathcal{P} \cup \mathcal{R}) - \sum_{\substack{u' \in \{1,\ldots,U\} \\ u' \neq u}} \mathbf{A}_{u'}^{(\text{rel})} \hat{\mathbf{h}}_{u'}^{r(\text{collab})} \quad (16)$$

for each antenna, which include both the pilots and the reliable carriers, and once again subtract the estimated interference from the other users. Finally, sparse recovery is run for a last time to obtain the definitive estimate of the channels $\hat{\mathbf{h}}_u^r$

VII. Computational Complexity

The SABMP technique needs to compute a pseudo-inverse of complexity $\mathcal{O}\left(P\,|\mathcal{S}|^2\right)$ for each calculated probability. With the efficient computation of pseudo-inverses of growing matrices proposed in [11], this can be reduced to $\mathcal{O}\left(P\,|\mathcal{S}|\right)$. This needs to be done L times for each possible support set size $|\mathcal{S}| \in \{1,\ldots,T\}$, resulting in an overall complexity of $\mathcal{O}\left(LPT^2\right)$ for SABMP.

The complexity of the WIHT algorithm is determined by the pseudo-inverse in (8), and given by $\mathcal{O}\left(PT^2\right)$. This is L times lower than SABMP.

In our simulations, we compare three different estimation schemes. The first one (SABMP-SABMP), already proposed in previous works for the single-user scenario, uses SABMP in all 4 stages of sparse recovery. The second scheme (SABMP-WIHT), which is our proposed hybrid method, uses SABMP for the individual estimate ($\hat{\mathbf{h}}_u^{r(\text{init})}$ and $\hat{\mathbf{h}}_u^{r(\text{ind})}$) but WIHT for the collaborative estimate ($\hat{\mathbf{h}}_u^{r(\text{collab})}$ and $\hat{\mathbf{h}}_u^{r(\text{rel})}$). As the complexity of WIHT is negligible compared to that of SABMP, this means that the proposed hybrid method executes twice as fast as the full SABMP approach (iterative collaboration and reliability measurements do not have a noticeable contribution to execution time).

The third compared approach (WIHT-WIHT) uses WIHT for all 4 sparse recovery runs, and is therefore L times faster than full SABMP.

VIII. Simulation Results

In this section, simulation results will be provided to evaluate the performance of the sparse recovery techniques in a multi-user scenario. Two multi-user ($U = 3$ users) scenarios have been tested: one where pilot sequences do not interfere with each other (for example, one in which TDMA, CDMA, FDMA or orthogonality are used to separate the pilots, as analyzed in [12] for MIMO-OFDM systems), and one in which they are chosen pseudo-randomly, to potentially allow for a higher number of users. A Massive MIMO-OFDM system with $R = M \times G = 20 \times 20$ receive antennas and $N = 512$ subcarriers was simulated. A convolutional code of rate 2/3 was used in all transmissions. The maximum channel length is $L = 64$, with $T = 3$ active taps. The channels were generated using the IlmProp program, developed by TU Ilmenau [13], which accurately models the correlation between antennas. A value of $K = 275$ reliable carriers was found to give best results. Apart from bit error rate, the Normalized Mean Square Error (NMSE) was used as a performance measure. This figure of merit is defined as:

$$\text{NMSE (dB)} = \frac{1}{IUR}\sum_{i=1}^{I}\sum_{u=1}^{U}\sum_{r=1}^{R}\frac{\left\|\hat{\mathbf{h}}_{u(i)}^r - \mathbf{h}_{u(i)}^r\right\|_2^2}{\left\|\mathbf{h}_{u(i)}^r\right\|_2^2}, \quad (17)$$

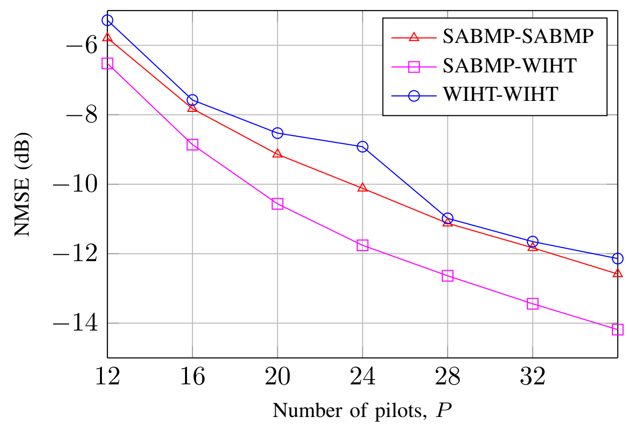

Fig. 4. NMSE vs number of pilots for the non-interfering pilots scenario.

where $I = 500$ is the number of Monte-Carlo runs.

A. Scenario with Non-interfering Pilot Sequences

The two first experiments were aimed at comparing the performance of SABMP and WIHT in the case where pilot sequences do not suffer interference from each other, due to orthogonality in the time domain.

1) Required Number of Pilots: First, the E_b/N_0 was fixed at 13 dB and the NMSE was plotted against the number of pilots P. The resulting curves are shown in Fig. 4. The three different settings described in Section VII were compared. The proposed hybrid SABMP-WIHT approach shows a significant gain with respect to the SABMP-SABMP method, which is due to both a low SNR regime and a low number of pilots needed to recover the channel. In this situation, even though SABMP is able to find the correct support, the computation of the probabilities of other support sets, $p\left(\mathcal{S}\,\middle|\,\overline{\boldsymbol{\mathcal{Y}}}\right)$ is not accurate, and too high weightings are applied to wrong support sets. Therefore, the actual values of the taps have a greater error than the output of WIHT, which only takes into account the most probable support set. The consequence is that this hybrid SABMP-WIHT approach needs about 6 fewer pilots to recover the channels.

2) BER performance: Next, an experiment was run to assess how this improvement in NMSE performance obtained by using the hybrid technique SABMP-WIHT translates into the BER domain. For this purpose, the number of pilots was fixed at $K = 24$, and the BER vs E_b/N_0 plot in Fig. 5 was obtained. It can again be verified that the hybrid approach that uses SABMP for the individual estimate and WIHT for the collaborative one can exploit the advantages of both techniques and needs about 0.5 dB less E_b/N_0 than SABMP-SABMP to achieve the same BER. In addition, SABMP-WIHT executes twice as fast because it only needs to run SABMP twice instead of four times, and the complexity of WIHT is negligible with respect to that of SABMP.

However, if WIHT is used for both the individual and the collaborative estimate, the graph (blue curve) shows that the performance degrades considerably, making the method

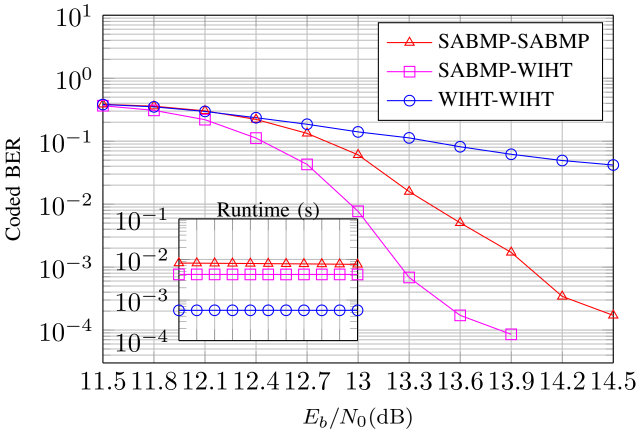

Fig. 5. BER vs SNR for the non-interfering pilots scenario.

unusable. This is because WIHT relies very heavily on good input priors $\mathbf{w}_u^r$, and these are not available during the first, individual estimate, when the receiver still has no information about the tap activation probabilities.

Therefore, for the non-interfering pilots case and the low SNR regime, SABMP-WIHT is the best option in all senses.

B. Scenario with Interfering Pilot Sequences

In the last experiment, the performance of the channel recovery techniques was tested in a scenario with $U = 3$ users and pseudo-random pilot sequences, using the iterative subtraction technique explained in Section IV. As all users send pseudo-random pilots on the same subcarriers, more pilots are needed to obtain accurate estimates of the channel. Fig. 6 plots the resulting BER vs SNR curve for a simulation was with $P = 80$ pilots.

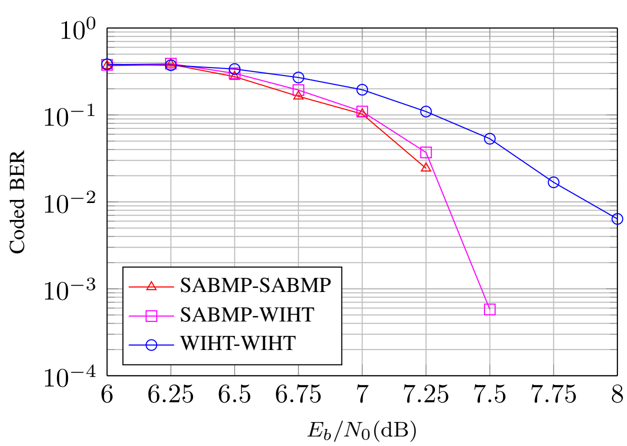

Fig. 6. BER vs SNR for the interfering pilots scenario

The plot shows that in this case, if SABMP is used in the first (individual) estimate, the results of using SABMP (red curve) or WIHT (pink curve) in the collaborative estimate are almost the same, unlike in the single-user scenario. The reason for this is that, due to pilot and data interference, a higher SNR is needed to recover the data correctly. The more complex SABMP technique becomes better at high SNR, and even slightly outperforms WIHT. Nevertheless, due to the good priors received by WIHT from collaboration, the hybrid approach SABMP-WIHT performs only slightly (less than 0.1 dB) worse than SABMP-SABMP. Therefore, the halved execution time may still make SABMP-WIHT a more attractive approach even in this case with interfering pilots.

IX. Conclusion

The high spectral efficiency in large scale MIMO systems is achieved by serving several users simultaneously through spatial multiplexing. However, it requires obtaining a good knowledge of the CSI of each user at the base station. In this paper, we have exploited the two characteristics of the propagation environment, i.e., sparsity and common support, to obtain a better estimate of CSI of each user to the base station. Two multi-user scenarios in a single cell were investigated: one with non-interfering pilots using time-division multiplexing, and one with interfering pilots. Simulation results show that the our scheme requires less number of pilots for non-interfering pilot sequences. Moreover, the computational complexity is substantially small compared with the fully Bayesian method.

References

[1] Nokia Solutions and Networks, "Technology vision 2020," June 2013.

[2] E. Larsson, F. Tufvesson, O. Edfor, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, Feb 2014.

[3] H. Q. Ngo, T. L. Marzetta, , and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *IEEE ICASSP*, May 2011.

[4] H. Jin, D. Gesbert, M. Filippou, , and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun*, vol. 31, February 2013.

[5] S. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, April 2013.

[6] X. Rao and V. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. on Signal Proc.*, June 2014.

[7] M. Masood and T. Y. Al-Naffouri, "Sparse reconstruction using distribution agnostic bayesian matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 61, Nov 2012.

[8] 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.

[9] E. J. Candes and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, 2006.

[10] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *Signal Processing, IEEE Transactions on*, January 2015.

[11] A. Ali, A. Al-Rabah, M. Masood, and T. Al-Naffouri, "Receiver-based recovery of clipped OFDM signals for PAPR reduction: A Bayesian approach," *IEEE Access*, vol. 2, Oct 2014.

[12] H. Minn and D. Munoz, "Pilot designs for channel estimation of mimo ofdm systems with frequency-dependent i/q imbalances," *Communications, IEEE transactions on*, vol. 58, August 2010.

[13] Technical University of Ilmenau. Ilmprop. [Online]. Available: http://www2.tu-ilmenau.de/nt/en/ilmprop//