

Efficient Coordinated Recovery of Sparse Channels in Massive MIMO

Mudassir Masood, Laila H. Afify, and Tareq Y. Al-Naffouri*, *Member, IEEE*,

Abstract—This paper addresses the problem of estimating sparse channels in massive MIMO-OFDM systems. Most wireless channels are sparse in nature with large delay spread. In addition, these channels as observed by multiple antennas in a neighborhood have approximately common support. The sparsity and common support properties are attractive when it comes to the efficient estimation of large number of channels in massive MIMO systems. Moreover, to avoid pilot contamination and to achieve better spectral efficiency, it is important to use a small number of pilots. We present a novel channel estimation approach which utilizes the sparsity and common support properties to estimate sparse channels and require a small number of pilots. Two algorithms based on this approach have been developed which perform Bayesian estimates of sparse channels even when the prior is non-Gaussian or unknown. Neighboring antennas share among each other their beliefs about the locations of active channel taps to perform estimation. The coordinated approach improves channel estimates and also reduces the required number of pilots. Further improvement is achieved by the data-aided version of the algorithm. Extensive simulation results are provided to demonstrate the performance of the proposed algorithms.

Keywords—*massive MIMO, large-scale antenna array, sparse channel estimation, distributed channel estimation, distribution agnostic.*

I. INTRODUCTION

The deployment of multiple antennas in a wireless communication system offers key advantages to its performance in terms of power gain, channel robustness, diversity, and spatial multiplexing. The use of multiple antennas in rich scattering environments provides effective utilization of the scarcely available spectrum resources. As a result, multiple-input-multiple-output (MIMO) technology has gained much interest in the research community.

Installing extra antennas to a MIMO system can introduce substantial enhancements in both link reliability and data throughput of the system [1]. Specifically, the use of very large antenna arrays (in the order of multiples of a hundred antennas) has been found to be beneficial to overcome problems encountered in traditional MIMO settings. Such systems, known as *massive MIMO* or *large-scale MIMO* [2],[3], also have the potential to scale down the transmission power because of the use of small active antennas with very low power.

Authors are with the Department of Electrical Engineering, King Abdullah University of Science & Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia, e-mail: mudassir.masood@kaust.edu.sa, laila.afify@kaust.edu.sa, tareq.alnaffouri@kaust.edu.sa

Tareq Y. Al-Naffouri is also associated with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Kingdom of Saudi Arabia.

Moreover, in massive MIMO systems fast fading is averaged out and intracell interference almost vanishes. Thus, using large antenna arrays can play a key role in exploiting the true potential of traditional MIMO systems while at the same time overcoming several challenges. Massive MIMO is therefore considered as an emerging key technology that can meet the growing demands of current wireless systems. For interested readers, some other advantages of adding more antennas to the base station (BS) have been discussed in [4]–[6].

In order to benefit from the advantages of massive MIMO systems, we need to determine the channel impulse response (CIR) for each transmit-receive link. In a typical massive MIMO system, a BS is equipped with a large antenna array and communicates with several users resulting in a large number of channels that need to be estimated. This results in a substantial increase in complexity which causes performance limitations at the BS. Obviously, obtaining CIR requires training data (pilots) to be sent by the users. It is known that the number of pilot symbols required is proportional to the total number of users [4]. Therefore, as the number of users increases, there is a higher chance that the pilot sequences in the neighboring cells interfere with each other. This pilot contamination problem is a major limiting factor for the massive MIMO systems [7], [8]. However, pilot contamination could be reduced if the reserved number of pilot tones is reduced. Therefore, in a multi-user scenario there is a need to reduce the number of pilots without affecting the CIR quality. Hence the development of efficient channel estimation techniques for massive MIMO that are computationally less complex and require less number of pilots is a challenge that needs to be thoroughly addressed.

Massive MIMO channel estimation is similar to the MIMO channel estimation. Existing literature includes several methods proposed for channel estimation in MIMO systems [9]–[13]. However, it is difficult to directly adopt these approaches for a number of reasons. For example,

- 1) There is a need to reduce the number of pilots.
- 2) All received (thousands of) signals in a massive MIMO system can not be processed efficiently at one central processor. Therefore, there is a need for methods/algorithms which are *a)* distributed; *b)* computationally efficient; and *c)* require little communication overhead.
- 3) The antenna arrays could spread over a large space making it quite different in its model than a regular compact MIMO receiver.

Recent works have indicated increased interest in the problem of massive MIMO channel estimation (see for example [7], [8], [14]–[16]). Most of these algorithms make use of the channel statistics. However, these statistics are usually

not known and therefore some kind of assumption is made about the distribution of channel taps. Moreover, some of the techniques involve computationally expensive operations like inversion of channel covariance matrices which is not reasonable for the massive MIMO scenario.

It is well known that many wireless channels have impulse response that is sparse in the sense that they have very few significant paths. For example, see [11], [13], [17]–[24] and the references therein. We would also like to add that in massive MIMO since a large number of antennas has to be placed usually it is difficult due to several space, structural, aesthetic constraints that antennas are positioned far from each other. Antenna separation is bound to decrease as we increase the number of antennas. This means that for antennas that are close to each other the times of arrival will be similar however the amplitudes and phases of the paths will be different, implying common support. The ideas that we will put forward in this paper take advantage of the above-mentioned two properties.

In this paper, we propose a set of algorithms for channel estimation in massive MIMO. Specifically, we consider a base-station equipped with a large number of antennas serving several single-antenna user-equipments (UE). Our approach makes use of the fact that the wireless channels between a UE and base-station antennas are expected to be sparse and that neighboring antennas observe channels with similar support (i.e., sparsity pattern) but not necessarily the same fading along the active taps. The antennas share information with their neighbors to reach a decision on the most probable support. Decisions are made in a distributed manner with low complexity and communication overhead. In summary, the set of algorithms we propose in this paper has the following distinctive features:

- 1) It utilizes the sparsity of the CIR and the fact that channel supports for neighboring antennas are approximately the same.
- 2) It is Bayesian in nature. It utilizes the sparsity of CIR and acknowledges the Gaussianity of the additive noise but is agnostic to the distribution of the active taps of the CIR.
- 3) It has a distributed nature requiring limited communication between neighboring antennas. One version of the algorithm requires only integer communication between antennas.
- 4) It has a data-aided extension that identifies reliable carriers and uses them to further reduce the number of pilots and enhance the CIR estimate.

The distributed Bayesian algorithm we develop in this paper is based on the Support Agnostic Bayesian Matching Pursuit algorithm (SABMP) developed by the authors in [25].

The remainder of this paper is organized as follows. In Section II, we present the system model and formulate the channel estimation problem. In Section III, we present a simple channel recovery method and propose enhancements to it that are required for the development of our coordinated recovery algorithms proposed in Section IV. A data-aided version of this algorithm is presented in Section V. Simulation results are discussed in Section VI and Section VII concludes the paper.

A. Notation

We denote vectors with small-case bold-face letters (e.g., \mathbf{x}), matrices with upper-case, bold-face letters (e.g., \mathbf{X}), and reserve calligraphic notation for symbols in frequency domain (e.g., \mathcal{X}). We use \mathbf{x}_i to denote the i^{th} column of matrix \mathbf{X} and $x(j)$ to denote the j^{th} entry of vector \mathbf{x} . We also use $\mathbf{X}_{\mathcal{S}}$ to denote the sub-matrix formed by the columns $\{\mathbf{x}_i : i \in \mathcal{S}\}$, indexed by the set \mathcal{S} . We use $\hat{\mathbf{x}}$, \mathbf{x}^* , \mathbf{x}^{T} , and \mathbf{x}^{H} to respectively denote the estimate, conjugate, transpose, and conjugate transpose of the vector \mathbf{x} . Finally we use $\text{diag}(\mathbf{x})$ to transform the vector \mathbf{x} into a diagonal matrix with the entries of \mathbf{x} spread along the diagonal.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Transmission Model

We consider a MIMO-OFDM system in which the BS is equipped with a large two-dimensional antenna array consisting of $R = M \times G$ antennas distributed across M rows and G columns.¹ The base station serves a number of single-antenna terminals. Orthogonal frequency division multiplexing (OFDM) is adopted as the signaling mechanism. In an OFDM system, serially incoming bits are divided into N parallel streams and mapped into a Q -ary QAM alphabet $\{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_Q\}$. This results in an N -dimensional data vector $\mathcal{X} = [\mathcal{X}(1), \mathcal{X}(2), \dots, \mathcal{X}(N)]^{\text{T}}$. The equivalent time-domain signal \mathbf{x} is obtained by taking the inverse Fourier transform of the data vector, i.e.,

$$\mathbf{x} = \mathbf{F}^{\text{H}} \mathcal{X}, \quad (1)$$

where \mathbf{F} is an $N \times N$ unitary discrete Fourier transform (DFT) matrix whose (k, l) entry is given by

$$f_{k,l} = \frac{1}{\sqrt{N}} \exp(-j \frac{2\pi}{N} kl). \quad (2)$$

A cyclic prefix is inserted at the beginning of each symbol and the resulting signal is transmitted.

B. Channel Model

It is known that most wireless channels can be modeled as discrete multipath channels with large delay spread and very few significant paths as scatterers are sparsely distributed in space (see Fig. 1). This makes the CIR sparse [23], [26], [27]. Thus, for each transmit-receive link, we need only estimate a few significant multipath channel gains, which has the potential to reduce the pilot overhead substantially. We explicitly mention the sparsity property as property 1.

Property 1: *The channel impulse response is sparse.*

¹Depending on the value of M and G , the antennas could have a linear or a rectangular configuration. Further, we would like to stress that while we confine our attention to rectangular configurations for convenience, our approach applies to any one-, two- or three-dimensional configuration of antennas as explained at the end of Sec. V.

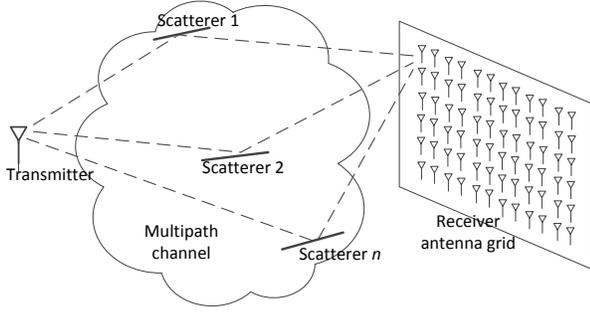


Fig. 1: In most wireless channels scatterers are sparsely distributed and the resulting channel impulse response is sparse.

Let $\mathbf{h}^r \in \mathbb{C}^L$ denote the CIR which models the channel between a typical single antenna user and the receive antenna $r = (m, g)$ where $m \in \{1, 2, \dots, M\}$ and $g \in \{1, 2, \dots, G\}$ as shown in Fig. 2. We assume that \mathbf{h}^r is sparse and is modeled

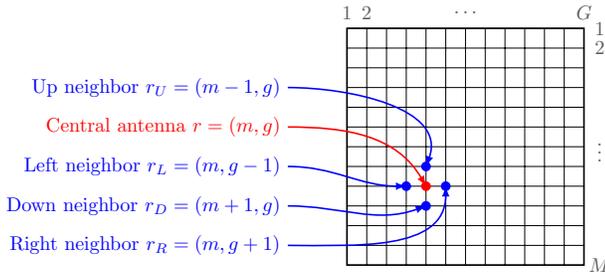


Fig. 2: 2-D antenna grid of size $M \times G$. An arbitrarily selected antenna is highlighted in red along with its neighboring antennas in blue. In this context, the red antenna is the central antenna r and r_U, r_R, r_D , and r_L are its 4-neighbors.

as [28]

$$\mathbf{h}^r = \mathbf{h}_A^r \odot \mathbf{h}_B^r, \quad (3)$$

where \odot indicates element-by-element multiplication. The vector \mathbf{h}_A^r consists of elements that are drawn from some distribution² and \mathbf{h}_B^r is a Bernoulli random vector with independent entries that are distributed as [28]

$$P(h_B^r(i) = j) = \begin{cases} \lambda_i, & \text{for } j = 1. \\ 1 - \lambda_i, & \text{for } j = 0. \end{cases} \quad (4)$$

In other words, the entries of \mathbf{h}_B^r form a collection of independent (and possibly non-identically distributed) Bernoulli random variables. Thus, \mathbf{h}^r is an L -tap discrete-time sparse

²We put no restriction on the distribution of \mathbf{h}_A^r which could be Gaussian or not. The distribution might even be unknown and the coefficients of \mathbf{h}_A^r need not be iid. The implementation in this paper is agnostic to the distribution of channel coefficients.

channel, where no assumption whatsoever is made about the distribution of its non-zero complex-valued coefficients.

The received signal at the r th antenna is best described in the frequency domain and is given by

$$\mathcal{Y}^r = \text{diag}(\mathcal{X})\mathcal{H}^r + \mathcal{W}^r, \quad (5)$$

where \mathcal{Y}^r is obtained from the time-domain received signal by removing the cyclic prefix and pre-multiplying by the Fourier matrix \mathbf{F} . The noise $\mathcal{W}^r \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$ is the frequency-domain noise vector of dimension $N \times 1$ and \mathcal{H}^r is the $N \times 1$ channel frequency response vector i.e.,

$$\mathcal{H}^r = \mathbf{F} \begin{bmatrix} \mathbf{h}^r \\ \mathbf{0}_{N-L \times 1} \end{bmatrix} = \underline{\mathbf{F}}\mathbf{h}^r \quad (6)$$

where $\underline{\mathbf{F}}$ is the truncated Fourier matrix of size $N \times L$ formed by selecting the first L columns of \mathbf{F} . Using (6), we can rewrite (5) as

$$\mathcal{Y}^r = \text{diag}(\mathcal{X})\underline{\mathbf{F}}\mathbf{h}^r + \mathcal{W}^r = \mathbf{A}\mathbf{h}^r + \mathcal{W}^r, \quad (7)$$

where $\mathbf{A} \triangleq \text{diag}(\mathcal{X})\underline{\mathbf{F}}$ is an $N \times L$ matrix.

C. Spatial Channel Model

The large number of antennas in massive MIMO can be arranged in different configurations. For example, *a*) linear, *b*) planar (rectangular), and *c*) cylindrical (circular). Our algorithm is capable of working on any configuration as will be explained later in the paper. However, for convenience, we adopt the uniform rectangular array.

In massive MIMO it is reasonable to assume that antenna elements in the same vicinity will observe almost same echoes from different scatterers and therefore the corresponding channels will have common support. For the wireless system under study, the signal bandwidth, operating frequencies and antenna separation controls the supports commonality across the large arrays. Specifically, the time difference of arrival $\Delta\tau$ of a wavefront to two antennas separated by a distance d satisfies $\Delta\tau \leq \frac{d}{C}$, where C is the speed of light. The authors in [18] suggest that two channel taps are resolvable if the time difference of arrival is larger than $\frac{1}{10BW}$ where BW is the signal bandwidth. Thus, let d_{\max} be the distance between the farthest antennas of an array (d_{\max} is a function of the antenna spacing d and the number of antennas in the array), then it follows easily from the above that the antenna array might exhibit one of the two possible scenarios:

1) *The array is spatially invariant with respect to the CIR support if $\frac{d_{\max}}{C} \leq \frac{1}{10BW}$* : Here, all \mathbf{h}^r 's will have same sparsity pattern, which means the amplitudes of the channel taps might be different but the positions of the most significant taps (MST) will not change. This assumption is motivated by the fact that for closely spaced antenna elements, the times of arrival are quite close, though the paths amplitudes and phases could be different. Therefore, different antenna elements will experience almost the same echoes from the different scatterers. In other words, the support of the channels will not change as we move from one antenna to another throughout the array but the tap strengths might be different as

evident from Fig. 3. We call such arrays space-invariant arrays (SIA).

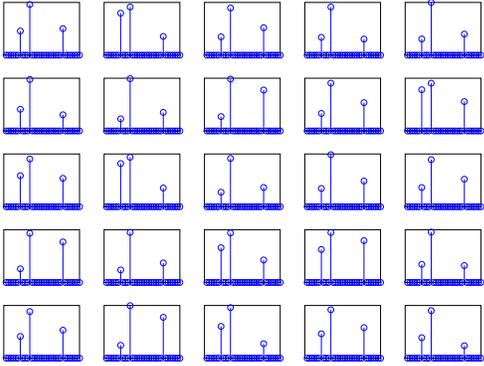


Fig. 3: CIRs of a 5×5 section of a space-invariant antenna array. Plots show the tap strengths on y-axis with respect to the tap locations on x-axis for antennas in this space-invariant antenna array section. Note that the support is invariant across the array but the taps strengths fade differently across the array.

2) The array is spatially variant with respect to the CIR support if $\frac{d_{\max}}{C} > \frac{1}{10BW}$: In this case, the channel support varies across the array. Note that such variation takes place slowly and, therefore, it is safe to assume that the following property is always valid.

Property 2: Any central antenna and its 4-neighbors have approximately common support.

See Fig. 4 for an example where the neighboring antennas have approximately the same support. We call such arrays space-variant arrays (SVA).

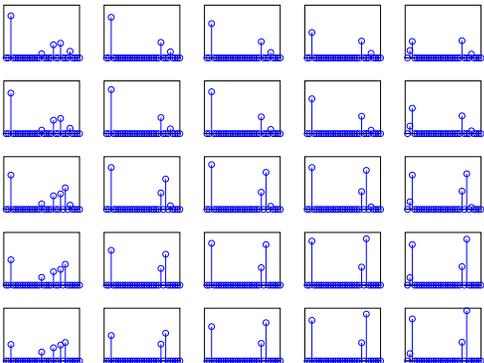


Fig. 4: CIRs of a 5×5 section of a space-variant antenna array. Plots show the tap strengths on y-axis with respect to the tap locations on x-axis for antennas in this space-variant antenna array section. Note that neighboring antennas have *approximately* the same support.

In Table I we classify antenna arrays of three different dimensions as either SIA or SVA. Specifically, the table illustrates the relationship between the maximum resolvable distance (d_{\max}) and the dimensions of the arrays for three different communications standards. For instance, in the 3GPP LTE standard, the distance between two antenna elements on the far ends of a 10×10 array is $9d < d_{\max}$, thus the array is SIA. Whereas, for a 50×50 array, the distance is $49d > d_{\max}$ causing the array to be SVA. Note that in this table the distance between two adjacent antennas is assumed to be $d = \lambda/2$ where λ is the signal wavelength.

While most of the available research in MIMO channel estimation deals with the space-invariant case (for example, [11], [18], [29], [30]), very limited research has been conducted for the space-variant scenario. Similarly, the literature related to the estimation of space-variant sparse channels in massive MIMO is limited (e.g., see [31] and the references therein). The approach we pursue in this paper is capable of dealing with both the space-variant and space-invariant cases.

D. Pilots

Pilots are needed for channel estimation where the transmitter reserves K subcarriers for pilots and uses the remaining $N - K$ carriers for data transmission. Let \mathcal{P} denote the set of indices of pilot carriers. Using (7), the received pilots at receive antenna r are then given by

$$\mathbf{y}^r(\mathcal{P}) = \mathbf{A}(\mathcal{P})\mathbf{h}^r + \mathbf{W}^r(\mathcal{P}) \quad (8)$$

where $\mathbf{y}^r(\mathcal{P})$ and $\mathbf{W}^r(\mathcal{P})$ are $K \times 1$ vectors formed, respectively, by selecting the K entries of \mathbf{y}^r and \mathbf{W}^r indexed by \mathcal{P} . Similarly, $\mathbf{A}(\mathcal{P})$ is a $K \times L$ matrix formed by selecting the rows of \mathbf{A} indexed by \mathcal{P} . Solving (8) for \mathbf{h}^r obviously requires that we at least have more pilots than the channel delay spread (i.e., $K \geq L$), which impacts the spectral efficiency of the system. Here, however, we use the sparse nature of the channel and the fact that adjacent antennas have almost the same support (i.e., properties 1 and 2) to substantially reduce the number of pilots needed as promised by the compressed sensing theory [32], [33].

Several pilot placement schemes have been suggested for OFDM channel estimation. It is best to allocate the pilots uniformly in conventional OFDM channel estimation (which does not make use of sparsity) [34]–[37]. However, when the channel is sparse a random assignment of pilots has been observed to be optimal [38], [39].

With this model, we are now ready to tackle the problem of channel estimation. We do that in three steps spread over three sections

- 1) Bayesian channel estimation at each antenna,
- 2) Distributed channel estimation, and
- 3) Data-aided channel estimation.

III. SPARSITY-AWARE DISTRIBUTION AGNOSTIC BAYESIAN CHANNEL ESTIMATION

Consider the linear regression model presented in (8). For notational convenience, we will drop the superscript r and

Standard	Bandwidth (BW)	Center frequency (f_c)	$d_{\max} = \frac{c}{10BW}$	$d = \frac{\lambda}{2}$	10×10 array	50×50 array	100×100 array
CDMA2000	1.25 MHz	1 GHz	24 m	0.150 m	SIA	SIA	SIA
3GPP LTE	20 MHz	2.6 GHz	1.5 m	0.058 m	SIA	SVA	SVA
UWB	500 MHz	3 GHz	0.06 m	0.050 m	SVA	SVA	SVA

TABLE I: Wireless Systems Parameters

the symbol \mathcal{P} unless these are required for clarity. Hence (8) becomes

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w}, \quad (9)$$

where \mathbf{y} and \mathbf{w} are vectors of dimension $K \times 1$, \mathbf{h} is a vector of dimension $L \times 1$ and \mathbf{A} is a matrix of dimension $K \times L$. Here we are interested in performing Bayesian estimation of the wireless CIR \mathbf{h} . Bayesian approaches assume a prior distribution, however, given the dynamic nature of wireless channels, it is usually impossible to characterize the distribution. Moreover, such an assumption is usually not suitable as it does not reflect the reality and might result in performance degradation. Additionally, even if the distribution is known, it is very difficult to estimate the distribution parameters (e.g., mean and variance for Gaussian), especially when the channel statistics are not i.i.d. In that respect, the use of distribution agnostic Bayesian sparse signal recovery (SABMP) developed by the authors in [25], [40] is quite attractive, as it provides Bayesian estimates even when the prior is non-Gaussian or unknown.

A. Simple Channel Estimation using SABMP

The set of channel estimation algorithms that we propose in this paper (Sec. IV and V) use a modified version of the SABMP algorithm proposed by the authors in [25], [40]. The modifications to SABMP required for the development of our distributed and data-aided channel estimation methods are proposed in Sec. III-B and III-C. However, before presenting the modifications we consider it essential to quickly go through the steps followed by the SABMP algorithm. In that respect, we briefly describe a straightforward approach for sparse channel estimation using SABMP. In this approach, all channels \mathbf{h}^r are estimated independently using the SABMP algorithm. Since no collaboration takes place among antennas in this approach, it is not possible to take advantage of property 2 mentioned earlier.

To estimate the $L \times 1$ sparse channel \mathbf{h} , from the $K \times 1$ observations vector \mathbf{y} related by the linear regression model given in (9), SABMP pursues an MMSE estimate of \mathbf{h} given \mathcal{Y} which is formally defined by

$$\hat{\mathbf{h}}_{\text{MMSE}} \triangleq \mathbb{E}[\mathbf{h}|\mathcal{Y}] = \sum_{\mathcal{S}} p(\mathcal{S}|\mathcal{Y}) \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}]. \quad (10)$$

Here the sum is executed over all possible 2^L support sets of \mathbf{h} . However, computing this sum is a challenging task when the channel delay spread (L) is large because the number of possible support sets can be extremely large and the computational

complexity will become unrealistic. To have a computationally feasible solution, this sum can be approximated by considering only those support sets which include the most significant taps with high probability. These few support sets correspond to the sets with significant posteriors $p(\mathcal{S}|\mathcal{Y})$. Let \mathcal{S}_d be the set of supports for which the posteriors are significant. Hence, (10) can be approximated by³

$$\hat{\mathbf{h}}_{\text{AMMSE}} = \sum_{\mathcal{S} \in \mathcal{S}_d} p(\mathcal{S}|\mathcal{Y}) \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}]. \quad (11)$$

We could determine \mathcal{S}_d and $\hat{\mathbf{h}}_{\text{AMMSE}}$ in a greedy manner using the dominant support selection metric defined as the log posterior

$$\nu(\mathcal{S}) \triangleq \ln p(\mathcal{S}|\mathcal{Y}) = \ln p(\mathcal{Y}|\mathcal{S})p(\mathcal{S}). \quad (12)$$

The greedy algorithm of SABMP starts by first finding the best support of size 1. This requires evaluating $\nu(\mathcal{S})$ for $\mathcal{S} = \{1\}, \dots, \{L\}$, i.e., a total of $\binom{L}{1}$ search points. Let $\mathcal{S}_1 = \{\alpha_1\}$ be the optimal support. Now, the optimal support of size 2 is found. Ideally, this involves a search over a space of size $\binom{L}{2}$. To reduce the search space, however, the greedy approach looks for the tap location $\alpha_2 \neq \alpha_1$ such that $\mathcal{S}_2 = \{\alpha_1, \alpha_2\}$ maximizes $\nu(\mathcal{S}_2)$. This involves $\binom{L-1}{1}$ search points (as opposed to the optimal search over $\binom{L}{2}$ points). The process continues in this manner by forming $\mathcal{S}_3 = \{\alpha_1, \alpha_2, \alpha_3\}$ and so on. Therefore, \mathcal{S}_d , the set of dominant support sets is composed of support sets that are incremental in nature and is given by⁴

$$\begin{aligned} \mathcal{S}_d &= \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{T_{\max}}\}, \\ \mathcal{S}_d &= \{\{\alpha_1\}, \{\alpha_1, \alpha_2\}, \{\alpha_1, \alpha_2, \alpha_3\}, \dots, \{\alpha_1, \alpha_2, \dots, \alpha_{T_{\max}}\}\}. \end{aligned} \quad (13)$$

The development of the SABMP algorithm in [25] assumes that the taps of \mathbf{h} are activated with equal probability λ (i.e., i.i.d. Bernoulli with probability λ). However, here we consider the case where some taps are more probable than others (based on the available information), and hence it is desirable to assign those taps a higher probability. This requires us to assume an independent and non-identically distributed Bernoulli behavior

³Note that $\sum_{\mathcal{S} \in \mathcal{S}_d} p(\mathcal{S}|\mathcal{Y}) < 1$ since $\mathcal{S}_d \subset \mathcal{S}$. This would render the estimate in (11) biased. To ensure an unbiased estimate, we normalize $p(\mathcal{S}|\mathcal{Y})$ so that $\sum_{\mathcal{S} \in \mathcal{S}_d} p(\mathcal{S}|\mathcal{Y}) = 1$.

⁴In (13), T_{\max} refers to the maximum number of non-zero elements in the sparse \mathbf{h} . T_{\max} is selected to be slightly larger than the expected number of active taps in the estimated CIR using the de Moivre-Laplace theorem. For details, readers are referred to [25].

for the unknown sparse vector and therefore the prior is given by

$$p(\mathcal{S}) = \prod_{i \in \mathcal{S}} \lambda_i \prod_{j \in \{1, \dots, L\} \setminus \mathcal{S}} (1 - \lambda_j), \quad (14)$$

where, λ_i is the probability that the i th tap of \mathbf{h} is active. Moreover, the likelihood is approximated as

$$p(\mathcal{Y}|\mathcal{S}) = \exp\left(-\frac{1}{2\sigma_w^2} \|\mathbf{P}_{\mathcal{S}}^{\perp} \mathcal{Y}\|_2^2\right), \quad (15)$$

where, $\mathbf{P}_{\mathcal{S}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathcal{S}} = \mathbf{I} - \mathbf{A}_{\mathcal{S}} (\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}})^{-1} \mathbf{A}_{\mathcal{S}}^H$ is the projection matrix and $\mathbf{A}_{\mathcal{S}}$ is a matrix formed by selecting columns of \mathbf{A} indexed by support \mathcal{S} . Substituting (14) and (15) in (12) yields

$$\begin{aligned} \nu(\mathcal{S}) \triangleq \ln p(\mathcal{S}|\mathcal{Y}) &= \left(-\frac{1}{2\sigma_w^2}\right) \|\mathbf{P}_{\mathcal{S}}^{\perp} \mathcal{Y}\|_2^2 + \sum_{i \in \mathcal{S}} \ln \lambda_i \\ &+ \sum_{j \in \{1, \dots, L\} \setminus \mathcal{S}} \ln(1 - \lambda_j) \end{aligned} \quad (16)$$

Now the only term that is left to be evaluated in (11) is $\mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}]$. Note that it is difficult or even impossible to evaluate this quantity because the distribution of the active taps of \mathbf{h} is unknown. Therefore, we replace it by the best linear unbiased (BLUE) estimate given by

$$\mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}] \leftarrow (\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}})^{-1} \mathbf{A}_{\mathcal{S}}^H \mathcal{Y}. \quad (17)$$

This provides us all the required quantities to evaluate $\hat{\mathbf{h}}_{\text{AMMSE}}$. Note that all parameters including σ_w^2 , $\boldsymbol{\lambda} = \{\lambda_i\}_{i=1}^L$ and the possible size of support T_{\max} need not be known and are estimated by the algorithm. Specifically, λ_i 's are initialized as

$$\lambda_i = \frac{1}{L} \left| \left\{ j : |\mathbf{a}_j^H \mathcal{Y}| \geq \frac{1}{2} \|\mathbf{a}^H \mathcal{Y}\|_{\infty} \right\} \right|,$$

where \mathbf{a}_j is the j th column of the matrix \mathbf{A} . Moreover, σ_w^2 is initialized simply as a scaled version of the variance of the received signal i.e., $\sigma_{\mathcal{Y}}^2$. Finally, T_{\max} is selected to be slightly larger than the expected number of active taps in the estimated CIR using the de Moivre-Laplace theorem. Note that our algorithm is robust to these initial estimates and can find right support even if these parameters are initialized away from their true values. For more details the interested readers are referred to [25].

By following this greedy approach, each antenna node estimates the corresponding approximate sparse CIR (11) in a distribution agnostic manner. A detailed statement of the greedy algorithm is presented in Table II.

In addition to the non-iid generalization above, we develop in the following two necessary modifications to the SABMP algorithm. Specifically, we modify SABMP to output *a*) the channel estimation error covariance matrix and *b*) the marginal probabilities of the detected MSTs that are needed for the distributed and data-aided versions of the channel estimation algorithms proposed in Sec. IV and V respectively.

```

1: procedure GREEDY( $\mathbf{A}, \mathcal{Y}, \boldsymbol{\lambda}, \sigma_w^2, T_{\max}$ )
2:   initialize  $J \leftarrow \{1, 2, \dots, L\}, i \leftarrow 1$ 
3:   initialize empty sets  $\mathcal{S}_{max}, \mathcal{S}_d, p(\mathcal{S}_d|\mathcal{Y}), \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_d]$ 
4:    $J_i \leftarrow J$ 
5:   while  $i \leq T_{\max}$  do
6:      $\Omega \leftarrow \{\mathcal{S}_{max} \cup \{\alpha_1\}, \mathcal{S}_{max} \cup \{\alpha_2\}, \dots, \mathcal{S}_{max} \cup \{\alpha_{|J_i|}\} \mid \alpha_k \in J_i\}$ 
7:     compute  $\{\nu(\mathcal{S}_k) \mid \mathcal{S}_k \in \Omega\}$ 
8:     find  $\mathcal{S}_* \in \Omega$  such that  $\nu(\mathcal{S}_*) \geq \max_j \nu(\mathcal{S}_j)$ 
9:      $\mathcal{S}_d \leftarrow \{\mathcal{S}_d, \mathcal{S}_*\}$ 
10:    compute  $p(\mathcal{S}_*|\mathcal{Y}), \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_*]$ 
11:     $p(\mathcal{S}_d|\mathcal{Y}) \leftarrow \{p(\mathcal{S}_d|\mathcal{Y}), p(\mathcal{S}_*|\mathcal{Y})\}$ 
12:     $\mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_d] \leftarrow \{\mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_d], \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_*]\}$ 
13:     $\mathcal{S}_{max} \leftarrow \mathcal{S}_*$ 
14:     $J_{i+1} \leftarrow J \setminus \mathcal{S}_*$ 
15:     $i \leftarrow i + 1$ 
16:  end while
17:  return  $\mathcal{S}_d, p(\mathcal{S}_d|\mathcal{Y}), \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}_d]$ 
18: end procedure

```

TABLE II: Support Agnostic Bayesian Matching Pursuit Algorithm (SABMP)

B. Error Covariance and Estimation Error

The channel estimation error and the covariance could be computed as follows.

Let,

$$\tilde{\mathbf{h}} = \hat{\mathbf{h}}_{\text{AMMSE}} - \mathbf{h} \quad (18)$$

be the error vector and $\mathbf{R}_{\tilde{\mathbf{h}}} \triangleq \text{cov}[\tilde{\mathbf{h}}|\mathcal{Y}]$ represent the error covariance matrix. The trace of $\mathbf{R}_{\tilde{\mathbf{h}}}$ i.e., $\text{Tr}[\mathbf{R}_{\tilde{\mathbf{h}}}]$ gives the MMSE estimation error. In order to evaluate $\mathbf{R}_{\tilde{\mathbf{h}}}$, let us define the error vector $\tilde{\mathbf{h}}_{\mathcal{S}} = \hat{\mathbf{h}}_{\mathcal{S}} - \mathbf{h}$ for a given support \mathcal{S} , where $\hat{\mathbf{h}}_{\mathcal{S}} = \mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}]$. Let the corresponding error covariance matrix be $\mathbf{R}_{\tilde{\mathbf{h}}|\mathcal{S}} \triangleq \text{cov}[\tilde{\mathbf{h}}|\mathcal{Y}, \mathcal{S}]$. Then $\mathbf{R}_{\tilde{\mathbf{h}}}$ could be expressed in terms of $\mathbf{R}_{\tilde{\mathbf{h}}|\mathcal{S}}$ by summing it over the dominant support set \mathcal{S}_d and is given by

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \sum_{\mathcal{S} \in \mathcal{S}_d} p(\mathcal{S}|\mathcal{Y}) \mathbf{R}_{\tilde{\mathbf{h}}|\mathcal{S}}. \quad (19)$$

Since we replace $\mathbb{E}[\mathbf{h}|\mathcal{Y}, \mathcal{S}]$ with a BLUE estimate, the conditional error covariance matrix will be $\mathbf{R}_{\tilde{\mathbf{h}}|\mathcal{S}} = (\mathbf{A}_{\mathcal{S}}^H \mathbf{C}^{-1} \mathbf{A}_{\mathcal{S}})^{-1}$ [41] (where $\mathbf{C} = \sigma_w^2 \mathbf{I}$ is the noise covariance matrix). Combining this fact with (19) yields

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \sigma_w^2 \sum_{\mathcal{S} \in \mathcal{S}_d} p(\mathcal{S}|\mathcal{Y}) (\mathbf{A}_{\mathcal{S}}^H \mathbf{A}_{\mathcal{S}})^{-1}. \quad (20)$$

Note that the calculation of covariance matrix involves a matrix inversion term which is a computationally expensive task. However, we would like to highlight that these inverses are available as part of intermediate calculations in the SABMP algorithm and hence do not pose any additional computational burden. The error covariance matrix and the estimation error play a vital role in the development of the data-aided approach

	\mathcal{S}_d^r	$\mathcal{S}_{T_{\max}}^r$
1	$\{\alpha_1\}$	$\{\alpha_1\}$
2	$\{\alpha_1, \alpha_2\}$	$\{\alpha_2\}$
3	$\{\alpha_1, \alpha_2, \alpha_3\}$	$\{\alpha_3\}$
4		$\{\alpha_1, \alpha_2\}$
5		$\{\alpha_1, \alpha_3\}$
6		$\{\alpha_2, \alpha_3\}$
7		$\{\alpha_1, \alpha_2, \alpha_3\}$

TABLE III: Sets \mathcal{S}_d^r and $\mathcal{S}_{T_{\max}}^r$ for $T_{\max} = 3$.

presented in Sec. V.

C. Finding Marginals

The marginal probabilities are not directly available at the output of SABMP and could be computed from the posteriors $p(\mathcal{S}|\mathcal{Y})$, $\forall \mathcal{S} \in \mathcal{S}_d$ in a simple manner described below.

Let $T^r = \{\alpha_1^r, \alpha_2^r, \dots, \alpha_{T_{\max}}^r\}$ be the set of MST locations of channel \mathbf{h}^r as detected by SABMP algorithm. Then the marginal probabilities of α_i^r , $\forall i \in \{1, 2, \dots, T_{\max}\}$ could be computed as

$$p(\alpha_i^r|\mathcal{Y}) = \sum_{\alpha_i^r \cap \mathcal{S} \neq \emptyset} p(\mathcal{S}|\mathcal{Y}), \quad (21)$$

where the sum is evaluated over all $\mathcal{S} \in \mathcal{S}_L^r$ satisfying the condition $\alpha_i^r \cap \mathcal{S} \neq \emptyset$ where \mathcal{S}_L^r contains all $2^L - 1$ support sets that could be created for \mathbf{h}^r (recall that \mathbf{h}^r is a vector of length L). Among these $2^L - 1$ support sets there are only $2^{T_{\max}} - 1$ support sets which involve purely the T_{\max} detected non-zero locations. Let us denote the set of these $2^{T_{\max}} - 1$ support sets by $\mathcal{S}_{T_{\max}}^r$. We assert that only the support sets present in $\mathcal{S}_{T_{\max}}^r$ have significant posteriors $p(\mathcal{S}|\mathcal{Y})$ as compared to the others which have very small values. This follows from the findings of [25] (see Fig. 7 therein). Thus, we can evaluate the sum in (21) over $\mathcal{S} \in \mathcal{S}_{T_{\max}}^r$ to find the marginal probability of each detected non-zero location. However, the SABMP algorithm returns \mathcal{S}_d^r , which unlike $\mathcal{S}_{T_{\max}}^r$ contains only T_{\max} support sets (see (13)). Therefore, we modify the SABMP algorithm so that it outputs $p(\mathcal{S}|\mathcal{Y})$ for all supports in $\mathcal{S}_{T_{\max}}^r$. For illustration purpose Table III provides an example of the support sets \mathcal{S}_d^r and $\mathcal{S}_{T_{\max}}^r$ when $T_{\max} = 3$. Please note that for convenience of notation, we shall from now on use $\lambda(\alpha_i^r)$ as a shorthand notation for $p(\alpha_i^r|\mathcal{Y})$.

Since $\mathcal{S}_{T_{\max}}^r$ has more support sets, this modification obviously results in increased computational complexity. However, utilizing the available intermediate information in SABMP helps to compute the marginalized posterior probabilities $p(\mathcal{S}|\mathcal{Y})$ in an efficient manner. Specifically, note that for the example of Table III, we require posteriors of $\{\alpha_2\}$, $\{\alpha_3\}$, $\{\alpha_1, \alpha_3\}$, and $\{\alpha_2, \alpha_3\}$ in addition to those returned by SABMP. However, it follows from the explanation given in Sec. III-A, that the posteriors for $\{\alpha_2\}$, $\{\alpha_3\}$, and $\{\alpha_1, \alpha_3\}$ are already available to the algorithm by virtue of the intermediate computations. Therefore, the only missing computation which has to be performed additionally is that

of $\{\alpha_2, \alpha_3\}$. The same reasoning applies for larger support sizes. Therefore, we state that the increase in computational complexity is not significant.

For ease of reference, we name the modified version of SABMP as RS1. It has the following additional features:

- it considers the non-zero taps to follow independent and non-identically distributed Bernoulli behavior (as highlighted in Sec. III-A)
- it returns the error covariance of our estimate which is needed for the data-aided part (Sec. III-B)
- it outputs the belief/probability that a given tap is active (Sec. III-C).

This algorithm along with SABMP will be used in the discussion that follows to develop the coordinated channel recovery algorithms.

IV. COORDINATED CHANNEL ESTIMATION

In the coordinated channel recovery method, the receive antennas collaborate with each other to take advantage of property 2 and estimate the MST locations jointly. In order to realize this coordinated method, we assume baseband processing at each receive antenna with an additional processor on each baseband card to implement the collaboration strategy described in this section. At the heart of the collaboration strategy followed by the proposed method is the following simple information-sharing step.

Sharing Step: Each antenna acting as a central antenna r_C receives information from its direct 4-neighbors $\mathcal{N} = \{r_U, r_D, r_R, r_L\}$.⁵

It is obvious that repetitive application of this sharing step would result in information diffusion throughout the antenna grid. For example, in the first iteration r_C receives information from just the first tier of antennas (i.e., the neighboring 4 antennas). In just two iterations r_C receives information from 12 antennas (i.e., the first and second antenna tiers). Therefore, with the help of this simple step each antenna is able to incorporate information from its neighbors to enhance its decision about the MSTs of its channel. This ultimately helps in estimating the channels accurately. There are two advantages of this stepwise collaboration mechanism:

- 1) It gives us the flexibility to control the number of collaborators for each receiver which is essential for the space-variant case (Sec. II-C2).
- 2) The collaboration mechanism is computationally efficient as the antennas do not all collaborate with each other at the same time. This also reduces the communication overhead.

We now present two algorithms for channel estimation based on this simple stepwise information sharing strategy.

⁵For the elements lying at the edges of the array the number of neighbors are different. We use \mathcal{N} to denote the set of neighbors irrespective of the position of r and therefore $2 \leq |\mathcal{N}| \leq 4$.

A. Algorithm 1: Marginal-based Channel Estimation using Pilots

We seek to solve the problem mentioned in (8). The proposed algorithm starts by estimating the sparse channels \mathbf{h}^r at each receive antenna r using the RS1 algorithm. We initialize the algorithm by assuming that all taps of \mathbf{h}^r have equal active probability λ_{init} , i.e.,

$$P(h_B^r(l) = 1) = \lambda_{\text{init}}, \quad l \in \{1, 2, \dots, L\}, \forall r. \quad (22)$$

Let $T^r = \{\alpha_1^r, \alpha_2^r, \dots, \alpha_{T_{\text{max}}}^r\}$ be the set of MSTs of channel \mathbf{h}^r as detected by RS1. Here α_i^r is the location of the i th detected tap of receiver r . Note that since λ_{init} is same throughout the array, the number of detected MSTs, i.e., T_{max} , will also be same for all receivers in the array.⁶ In other words, the cardinality $|T^r| = T_{\text{max}}, \forall r$. However, the actual tap locations might differ from one antenna to another. Therefore, it is not necessary that $|T^{r_1} \cap T^{r_2}| = T_{\text{max}}$, for $r_1 \neq r_2$. Along with the MSTs, RS1 also returns the marginals $\lambda(\alpha_i^r) \triangleq P(h_B^r(\alpha_i^r) = 1), t \in \{1, 2, \dots, T_{\text{max}}\}$. At this point, we invoke the sharing step mentioned previously and share the marginals. Hence, each antenna, acting as central antenna r_C , collects these marginals from its 4-neighbors and computes the average marginal for each tap given by

$$\lambda(\alpha_i^{r_C}) = \begin{cases} \sum_{j \in \mathcal{N}^+} \lambda(\alpha_i^j) / |\mathcal{N}^+|, & \text{if } \alpha_i^{r_C} \in \bigcup_{j \in \mathcal{N}^+} T^j \\ \lambda_{\text{small}}, & \text{otherwise} \end{cases}, \quad (23)$$

where $\mathcal{N}^+ = \mathcal{N} \cup r_C, i \in \{1, 2, \dots, L\}$ and $\lambda(\alpha_i^{r_C})$ can be seen as the updated marginal of the i th tap detected at r_C . Here, λ_{small} is an arbitrarily small value assigned to those taps which have not been detected by any of the receivers in \mathcal{N}^+ ; it is highly probable that these taps have almost zero gains. Note that (23) is performed at each antenna as each antenna is the center of some neighbors. Moreover, each antenna repeats these sharing and averaging steps D times where D is selected based on whether the array under consideration is classified as SIA or SVA. This repetition allows each antenna to utilize the observations of distant antenna tiers to bolster its support estimates. Note that when $D = 1$, information from only the immediate neighbors is taken into consideration and for $D = 2$, information belonging to the neighbors of neighbors is also incorporated in the computations. Therefore, in this fashion, higher values of D make it possible to extend the scope of information sharing to distant antennas.

In the space-invariant array case, since the MST locations do not vary across the array, contribution from as many antennas as possible will always strengthen our belief in these locations. Therefore, we may select $D = \max(M, G)$ which equals to the largest dimension of the antenna array. This particular choice of D ensures that each antenna receives information from every other antenna in the array. However, one might not need to select such high value of D and a smaller number of iterations might be sufficient based on the problem parameters such as

the observation size (K) and the sparsity (n) of the channels. In fact we could establish a loose lower bound on D in the noise free case as a function of these quantities using lemma 1 in [42]. According to this lemma, if observations from q antennas are used to recover n -sparse channel vectors using K pilots then for a unique solution the following relationship holds

$$n \leq \lceil (K + q)/2 \rceil - 1, \quad (24)$$

where $\lceil \cdot \rceil$ denotes the ceiling operation. This yields the lower bound on q which is given by

$$q > 2n - K. \quad (25)$$

Furthermore, it could be easily deduced that the total number of antennas that take part in the D th sharing and averaging step is $2D(D+1)+1$. Thus relating this number with q above we conclude that, to guarantee a unique solution in the noise free case, D must satisfy

$$2D(D+1)+1 > 2n - K, \quad (26)$$

which simplifies to

$$D > \sqrt{n - \frac{K}{2} - \frac{1}{4}} - \frac{1}{2}. \quad (27)$$

For a detailed account of the lemma and its requirements please refer to [42].

In the space-variant case, depending upon how fast the MST locations (support) change across the array, we might or might not gain from sharing. Specifically, if the change in support is fast, using higher values of D would degrade the estimates. On the other hand, if the support changes very slowly, we expect that the neighborhood around a given antenna would behave *approximately* as SIA. Therefore, we select a value for D such that collaboration among antennas in that neighborhood would improve the estimates. In fact, from the discussion in Sec. II-C, we can determine the value of D which will ensure that all sharing and averaging takes place among antennas having same support. Specifically, the number of tiers (also D) selected for sharing could be represented in terms of the distance between two adjacent antennas d and the signal bandwidth BW . Note that the distance between the farthest antennas in tier 1 (i.e., when $D = 1$) is $2d$. Similarly, for tier 2 this distance is $4d$. In general, the distance is directly related to D and is given by $2Dd$. Now to ensure space-invariance for antennas up to tier D , it follows that we should require $2Dd \leq \frac{C}{10BW}$ (see Sec. II-C). Therefore,

$$D \leq \frac{C}{20 \cdot d \cdot BW},$$

or

$$D = \left\lfloor \frac{C}{20 \cdot d \cdot BW} \right\rfloor, \quad (28)$$

where $\lfloor \cdot \rfloor$ denotes the floor operation. The value of D in (28) will ensure that sharing happens among antennas whose support is approximately the same. That said the number of pilots should also be large enough such that (27) is also

⁶The value of T_{max} is selected to be slightly larger than the expected number of active MSTs in the estimated CIR using the de Moivre-Laplace theorem which relies on λ_{init} .

Algorithm 1 Marginal-based Channel Estimation using Pilots

- 1) Initialize $P(h_B^r(l) = 1) = \lambda_{\text{init}}, l \in \{1, 2, \dots, L\}, \forall r$.
- 2) Run RS1 at each antenna to estimate its λ . (Fig. 5a)
- 3) Each antenna, acting as central antenna, receives marginals from its neighbors. (Fig. 5b)
- 4) Each antenna computes average marginals (λ'). (see (23) and Fig. 5c)
- 5) Repeat steps 3-4 above, D times. (Fig. 5d)
- 6) All antennas re-estimate channels using these marginals as new priors with SABMP algorithm.

satisfied.

At the end of D iterations each antenna has a new set of marginals which are used as new priors with SABMP to get the final sparse CIR estimate. This final estimate is more accurate as the antennas have shared their information to strengthen their beliefs about the locations of the active taps. We call this algorithm the *Marginal-based Algorithm*. A graphical description of the algorithm is given in Fig. 5 and a summary of the steps followed by the algorithm is presented in Algorithm 1.

B. Algorithm 2: Reduced Communication and Computational Cost – Integer-based Channel Estimation

We would like to point out that sharing the marginals vectors among the receiver antennas puts a high communication overhead on the massive-MIMO system. This is because the marginals are floating point numbers and communicating these numbers requires complex signalling. This increases the communication overhead between antennas. However, if just integers are shared, the communication cost could be reduced significantly. We therefore, propose a variant of Algorithm 1 which uses integers for communication among receiver antennas. Since we are not interested in sharing marginals, we do not calculate these and rely on the original SABMP algorithm. Therefore, this algorithm has an additional advantage of low computational complexity as marginals are not calculated.

The algorithm starts by estimating channels at each receiver using SABMP and depends completely on the amplitudes of the estimated MSTs. Following the same reasoning given in the previous section, let $T^r = \{\alpha_1^r, \alpha_2^r, \dots, \alpha_{T_{\max}}^r\}$ be the set of MSTs of channel \mathbf{h}^r as detected by SABMP and $\mathbf{h}^r(T^r)$ be the corresponding amplitudes. Based on these amplitudes, we define an integer metric for each tap which we call *score* and denote it by ψ . For a given T^r , the highest score is assigned to the channel tap with maximum amplitude in absolute sense. Similarly, the tap (in T^r) with minimum amplitude gets the least score. Specifically, since there are T_{\max} MSTs, we assign a score of T_{\max} to the channel tap with maximum amplitude, a score of $T_{\max} - 1$ to the second highest tap and so on until a score of 1 is assigned to the tap with the smallest amplitude among these T_{\max} taps. Therefore, if $|h^r(\alpha_1^r)| > |h^r(\alpha_2^r)| > \dots > |h^r(\alpha_{T_{\max}}^r)|$ then $\psi(\alpha_1^r) = T_{\max}, \psi(\alpha_2^r) = T_{\max} - 1, \dots, \psi(\alpha_{T_{\max}}^r) = 1$ where $\psi(\alpha_i^r)$ is the score of the i th detected tap of receiver r . All other $L - T_{\max}$ tap locations are assigned a score of zero.

Algorithm 2 Integer-based Channel Estimation using Pilots

- 1) Run SABMP at each antenna
- 2) Each antenna receives scores from its neighbors
- 3) Each antenna computes average scores (29)
- 4) Repeat steps 2-3 above, D times
- 5) Each antenna computes a belief vector (30)
- 6) All antennas re-estimate channels using these belief vectors in place of Bernoulli priors with SABMP algorithm

Once each receiver has assigned scores to its detected MSTs, we are ready to invoke the sharing step. Thus, each antenna acting as a central antenna r_C collects these scores from its 4-neighbors and computes the average score for each tap given by

$$\psi(\alpha_i^{r_C}) = \begin{cases} \left\lceil \frac{\sum_{j \in \mathcal{N}^+} \psi(\alpha_i^j)}{|\mathcal{N}^+|} \right\rceil, & \text{if } \alpha_i \in \bigcup_{j \in \mathcal{N}^+} T^j \\ 0, & \text{otherwise} \end{cases} \quad (29)$$

where $i \in \{1, 2, \dots, L\}$. The sharing and averaging process is repeated D times and all the related discussion in marginal-based algorithm applies to this algorithm as well. The averaging step of (29) is similar to the averaging step of Algorithm 1 given in (23) except that we round up the averaging result to the nearest largest integer. This ensures that the resulting score is always an integer. However, note that the rounding operation is not required in the last step as no sharing takes place after that. Therefore, to avoid unnecessary computation and the resulting information loss, the round up operation is not performed on the average scores. At the end of the D sharing and averaging steps each node computes a belief metric given by

$$b(\alpha_i^r) = \psi(\alpha_i^r) / T_{\max}, \quad (30)$$

where $b(\alpha_i^r)$ is the estimated belief that the i th tap of receiver r is active. Each node uses the beliefs as the Bernoulli priors to re-estimate the channels using RS1. We call this algorithm the *Integer-based Algorithm*. The steps followed by this algorithm are summarized in Algorithm 2. This algorithm has the following advantages over the marginal-based algorithm:

- 1) Reduced communication cost since it totally avoids communicating floating point numbers and,
- 2) Lower computational complexity since it does not compute marginal probabilities.

We now move on to suggest another level of refinement for the marginal probability/scores vectors by selecting reliable data carriers to perform channel estimation.

V. DATA-AIDED CHANNEL ESTIMATION

By virtue of the channel sparsity property we can perform channel estimation using a small number of pilots K compared to the channel length L as discussed in the last two sections. We can enhance the channel estimate by increasing the number of pilots. Alternatively, we take a data-aided approach as

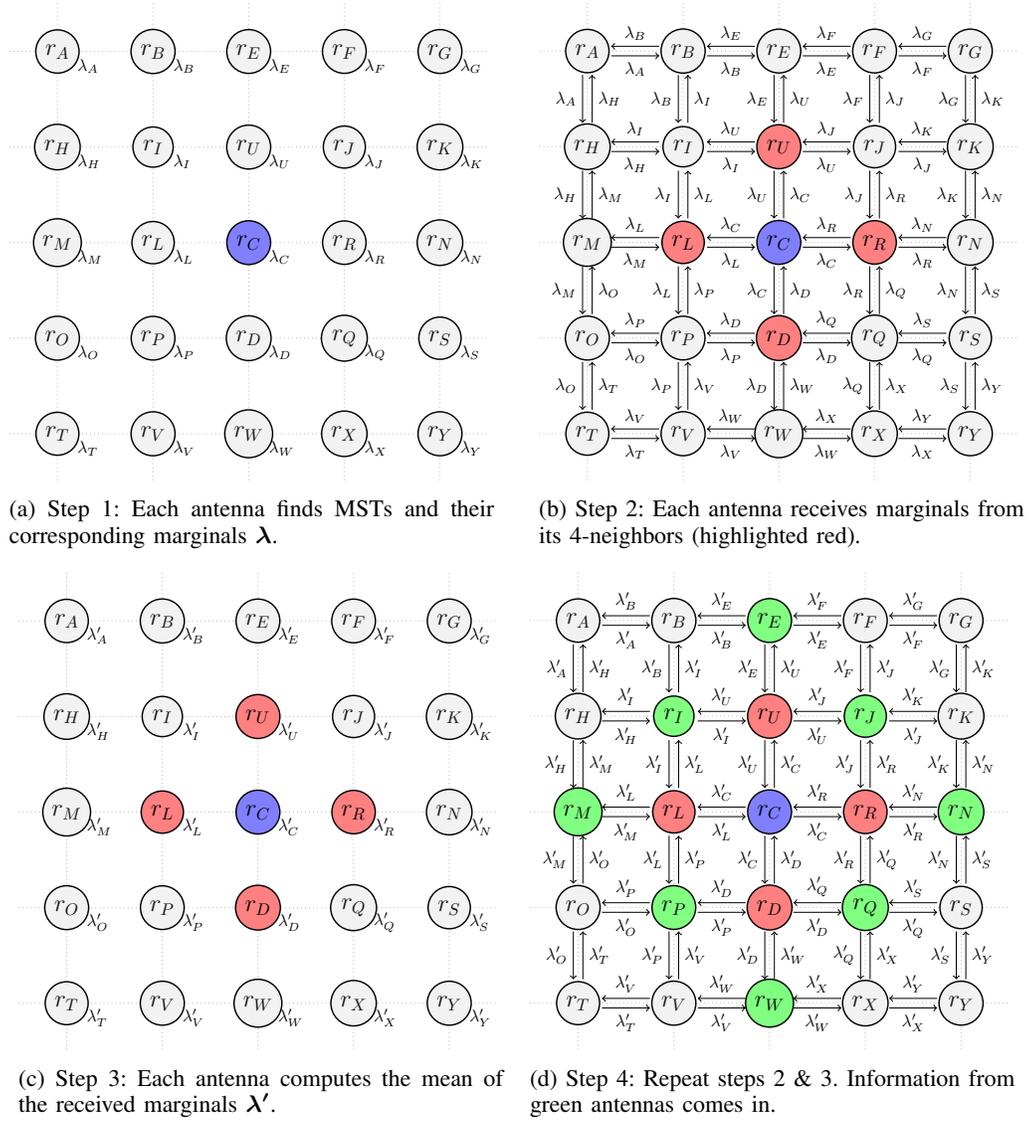


Fig. 5: Description of the steps followed by Algorithm 1 when $D = 2$. Although these steps are followed by all antennas in parallel, the process is highlighted only for the blue antenna.

it is more spectrally efficient. Here, the pilot-based channel estimate is used for data detection which along with the pilots is used to enhance the channel estimate further. Note however that we do not need to use all the detected data for channel estimation thanks to the channel sparsity; a few additional observations would enhance the channel estimate significantly. Therefore, we can be selective and use only the samples which are reliable. So each antenna could independently determine which carrier is reliable by assigning a reliability measure $\mathfrak{R}(i)$, $i \in \{1, \dots, N\} \setminus \mathcal{P}$ to each of the $N - |\mathcal{P}|$ data carriers. That said, we recognize that there are two sources of error in data detection that play important role in determining the reliable data carriers, namely, *a*) noise, and *b*) error in

channel estimation. Their combined distortion effect could be expressed by substituting the estimated channel $\hat{\mathbf{h}}_{\text{AMMSE}}$ from (18) into the system model (9) as follows

$$\mathbf{y} = \mathbf{A}(\mathbf{h} + \tilde{\mathbf{h}}) + \mathbf{W} = \mathbf{A}\mathbf{h} + \mathbf{Z},$$

where, $\mathbf{Z} = \mathbf{A}\tilde{\mathbf{h}} + \mathbf{W}$ is the combined distortion which is assumed to be Gaussian with zero mean and covariance $\mathbf{R}_{\mathbf{Z}}$, where $\mathbf{R}_{\mathbf{Z}}$ is represented in terms of the error covariance $\mathbf{R}_{\tilde{\mathbf{h}}}$, calculated in (20), as

$$\begin{aligned} \mathbf{R}_{\mathbf{Z}} &= \mathbb{E}[\mathbf{Z}\mathbf{Z}^H] = \mathbb{E}[(\mathbf{A}\tilde{\mathbf{h}} + \mathbf{W})(\mathbf{A}\tilde{\mathbf{h}} + \mathbf{W})^H] \\ &= \mathbb{E}[\mathbf{A}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{A}^H + \mathbf{W}\mathbf{W}^H] = \mathbf{A}\mathbf{R}_{\tilde{\mathbf{h}}}\mathbf{A}^H + \sigma_w^2\mathbf{I}. \end{aligned} \quad (31)$$

Here we have assumed that noise \mathcal{W} and error $\tilde{\mathbf{h}}$ are uncorrelated. Note that \mathcal{Z} includes the effect of both the channel estimation error and the noise and plays the central role in the calculation of reliability measure. Specifically, we use the reliability criterion proposed in [43] which takes into consideration the fact that for some carrier i , the distortion $\mathcal{Z}(i)$ might be strong enough to take the estimated data symbol $\hat{\mathcal{X}}(i)$ out of its correct decision region, while for some other carriers the distortion is not strong enough and the data is decoded correctly. All those data carriers i which satisfy this condition $\langle \hat{\mathcal{X}}(i) \rangle = \mathcal{X}(i)$, where $\langle \cdot \rangle$ represents rounding to the nearest constellation point, are termed *reliable* carriers and the following metric is used to compute the reliability of carrier i

$$\mathfrak{R}(i) = \frac{p(\mathcal{Z}(i) = \mathcal{X}(i) - \langle \hat{\mathcal{X}}(i) \rangle)}{\sum_{k=0, \mathcal{A}_k \neq \langle \hat{\mathcal{X}}(i) \rangle}^{Q-1} p(\mathcal{Z}(i) = \mathcal{X}(i) - \mathcal{A}_k)} \quad (32)$$

where $p(\cdot)$ represents the pdf of \mathcal{Z} . The numerator in (32) is the probability that $\mathcal{Z}(i)$ does not take $\mathcal{X}(i)$ beyond its correct decision region and the denominator sums the probabilities of all possible incorrect decisions that $\mathcal{Z}(i)$ can cause (i.e., $\mathcal{Z}(i)$ takes $\langle \hat{\mathcal{X}}(i) \rangle$ to a QAM constellation point \mathcal{A}_k different from $\mathcal{X}(i)$). The idea of reliability calculation is shown graphically in Fig. 6. In this figure, although both $\hat{\mathcal{X}}(1)$ and $\hat{\mathcal{X}}(2)$ are equiprobable to be decoded as \mathcal{X} (numerator of (32) will have same value), $\hat{\mathcal{X}}(2)$ is less likely to be decoded as any other constellation point (denominator of (32) for $\hat{\mathcal{X}}(2)$ will be smaller) and thus more reliable. Thus, it is obvious that the higher the value of \mathfrak{R} the higher the probability of staying in correct decision region and hence the higher the reliability of the carrier. Note that our reliability calculations of (32) require the error covariance which is already available at the output of the RS1 algorithm as mentioned in Sec. III-B. Here we would like to point out that the general approach of using reliable data carriers for enhanced channel estimation is not new and techniques employing reliable carriers exist [44]–[46]. Specifically, the reliability measure \mathfrak{R} in (32) is similar to log-likelihood ratios (LLRs) commonly used in joint channel estimation and data detection methods similar to turbo-equalizers (for example, see [47] and the references therein).

Using (32) each antenna determines the reliability of all data carriers and then select the U carriers with highest reliability values. Let \mathcal{R}^r denote the index set of these U reliable carriers for antenna r . One possible approach could be that each receiver uses these reliable carriers to enhance the CIR estimate by using Algorithm 1 or 2. However, the antennas can collaborate to enhance the reliability even further. First, each antenna r_C acting as a central antenna collects the indices of the reliable carriers from its 4-neighbors and returns the indices of the reliable carriers common to all antennas, i.e., $\mathcal{R} = \bigcap_{r \in \mathcal{N}^+} \mathcal{R}^r$. The central antenna can go one step further and ask its neighbors to share their equalized data on the common carriers. The central antenna in turn prunes the set \mathcal{R} further and only retains those carriers \mathcal{R}^* on which there is agreement among the neighbors on the value of the transmitted data. The central antenna can now use the enlarged set of pilots

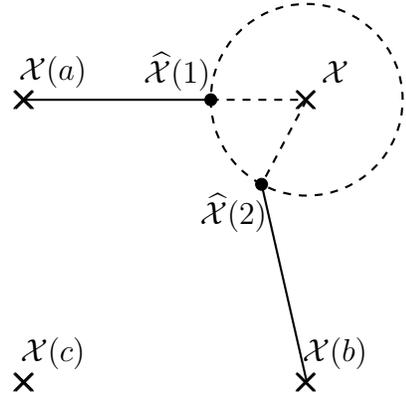


Fig. 6: Geometrical representation of the reliability measurement. $\hat{\mathcal{X}}(2)$ is more reliable than $\hat{\mathcal{X}}(1)$ as it is less probable to be confused with other constellation points.

Algorithm 3 Channel Estimation using Pilots + Reliable Carriers

- 1) Run Algorithm 1 or 2 to get CIR estimates
 - 2) Each antenna r uses its estimated channel to find top U reliable carriers \mathcal{R}^r and sends \mathcal{R}^r to its central antenna
 - 3) Each antenna finds the intersection of received reliable carriers $\mathcal{R} = \bigcap_{r \in \mathcal{N}^+} \mathcal{R}^r$ and sends it back to its neighbors
 - 4) Each antenna sends back data corresponding to \mathcal{R} to its central antenna
 - 5) Each antenna further refines the reliable carriers by selecting only those with same data. Call this list \mathcal{R}^* .
 - 6) Each antenna uses the carriers \mathcal{R}^* and the pilots to perform SABMP recovery
-

plus reliable carriers $\mathcal{P} \cup \mathcal{R}^*$ to revisit the channel estimation problem starting from the system of equations

$$\mathbf{y}^r(\mathcal{P} \cup \mathcal{R}^*) = \mathbf{A}(\mathcal{P} \cup \mathcal{R}^*)\mathbf{h}^r + \mathcal{W}^r(\mathcal{P} \cup \mathcal{R}^*) \quad (33)$$

and estimate channel \mathbf{h}^r . The resulting algorithm is presented in Algorithm 3.

At this stage we would like to point out that implementation of all three algorithms is independent of the antenna array configuration. This is due to the fact that each antenna only deals with its direct neighbors. Therefore, as far as the antennas have the knowledge of their neighbors these algorithms can be implemented on any one-, two-, or three- dimensional arrays with arbitrary topology. Furthermore, the development of algorithms assumed single-antenna UE's; however, the techniques could be easily extended to multiple-antenna UE's such as the LTE UE's which are often equipped with two or four closely located antennas. Since the antennas are closely located, their channels will exhibit the approximately common support property. Therefore, the algorithms explained above could be used to exploit correlation among the channel tap locations to further improve the channel estimates. The only difference is that the effective number of collaborating antennas in each tier

will scale with the number of antennas on UE. For example, if there are two transmit antennas on a UE, the number of collaborating antennas in each tier will double.

VI. SIMULATION RESULTS

A. System Setup

In this section we will present extensive simulation results to demonstrate the performance of our proposed channel estimation algorithms. Specifically, we consider a MIMO-OFDM system with the simulation parameters given in Table IV. For

Parameters	Value
Uniform Rectangular Array ($M \times G$)	20×20
Number of carriers (N)	512
Number of pilots (K)	8, 16
QAM modulation order (Q)	4, 16
Channel length (L)	32, 64
Channel sparsity (n)	$\approx 3, 5, 7$
Collaboration parameter (D)	3

TABLE IV: System Parameters for simulation

simulations, sparse Rayleigh channels are generated where the channel statistics are assumed to be unknown at the receivers. For space-invariant arrays the active tap locations remain fixed across the array. However, for the space-variant case the active tap locations vary slowly across the array. Specifically, we use the `IlmProp` channel modeling tool [48], [49] for channel generation. It is important to note that there is a general lack of channel models for massive MIMO scenarios and currently `IlmProp` seems to be one of the best options available to the research community for channel generation. Please refer to Appendix A for relevant discussion. The channels are generated by placing point-like scatterers and the transmitter randomly in the environment and make sure that the line-of-sight is obstructed. Moreover, the number of scatterers is set according to the desired sparsity i.e., n . Since the resulting CIR contains many small non-zero components along with the dominant ones, we discard the small ones and keep just the top n components. Further, the center frequency and signal bandwidth are chosen to be 2.6 GHz and 20 MHz respectively as specified in the 3GPP-LTE standard. Moreover, to generate the SIA and SVA behavior the distance between antennas was adjusted accordingly.

B. Methods for Performance Comparisons

The channel vectors \mathbf{h}^r are estimated using *a*) least-squares method with known true MST locations (oracle-LS), *b*) block-sparse recovery method (BR), *c*) proposed Marginal-based channel estimation using pilots (MB-P), *d*) proposed Integer-based channel estimation using pilots (IB-P), and *e*) proposed Marginal- or Integer-based channel estimation using pilots and reliable carriers (MB-R / IB-R),

The first two methods are used to benchmark the performance of our algorithm. Oracle-LS knows the channel support at each antenna and hence the only burden is tap estimation

using the available pilots. The block sparse recovery method (BR) works in the space-invariant case and uses the fact that the channel support is the same across the array. It casts the problem as several block sparse problems where each receiver collects all observations from its neighbors to estimate the channels. We use the block sparse Bayesian learning algorithm (BSBL) proposed in [50] for block sparse vector estimation as it has been shown to be superior to other methods.

C. Evaluation Criteria

To evaluate channel estimation performance we use:

- 1) Normalized mean-squared error (NMSE) between true and estimated channel vectors.

$$\text{NMSE} = 10 \log_{10} \left(\frac{1}{\Theta} \sum_{\theta=1}^{\Theta} \frac{\|\hat{\mathbf{h}}_{\theta} - \mathbf{h}_{\theta}\|^2}{\|\mathbf{h}_{\theta}\|^2} \right), \quad (34)$$

where Θ is the number of trials. \mathbf{h}_{θ} and $\hat{\mathbf{h}}_{\theta}$ are the original and estimated CIR at the θ th iteration respectively.

- 2) Bit-error-rate (BER) between the transmitted data and the recovered data at receivers using the estimated channels. We use zero-forcing equalization to recover the data passed through the channels. In all of the experiments we average the NMSE and BER over $\Theta = 100$ trials.

D. Experiments

1) *Experiment 1 - How many pilots?:* In this experiment, we are interested in finding the required number of pilots for successful recovery of channels of length $L = 64$. The graphs in Fig. 7 show the channel recovery success rate vs varying number of pilots for both SIA and SVA. Note that both pilot-based and data-aided versions of MB and IB algorithms were simulated. Here, success rate is defined as the ratio of the number of successful trials to the number of total trials, where a trial was considered successful when the NMSE was better than -10 dB. The SNR was fixed at 10 dB and the number of pilots K was varied from 2 to 42 while $\Theta = 100$ trials were conducted for each value of K . Channel sparsity was assumed to be $n = 3$ and QAM signals of order $Q = 4$ were passed through the channels. It is evident from the graphs that for SIA just 6 pilots are needed by both MB-R and IB-R to achieve a success rate $> 50\%$ and only 12 pilots to achieve a 100% success rate. This is a small fraction of the channel length $L = 64$ (i.e., 9.37% and 18.75% respectively).

2) *Experiment 2 - Comparison between MB and IB:* In this experiment, we compare the performance of the proposed MB and IB channel estimation algorithms. Channels of length $L = 64$ and sparsity $n = 3$ were estimated using $K = 16$ pilots. The top row of Fig. 8 shows the NMSE in estimated CIRs while the bottom row shows the BER of the recovered data using these CIR estimates. Moreover, as apparent from the labels, this experiment was run using three different choices of parameters, namely, ($Q = 4$, SIA), ($Q = 4$, SVA), and ($Q = 16$, SIA) respectively. The figure shows that incorporating reliable carriers results in significant performance gains.

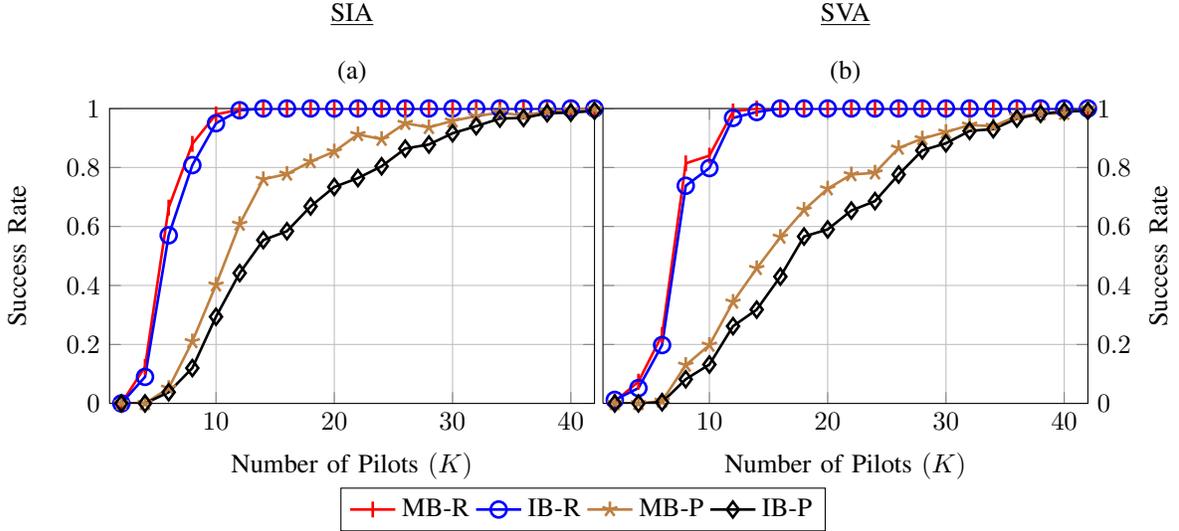


Fig. 7: Experiment 1: How many pilots are needed to successfully recover the CIR?

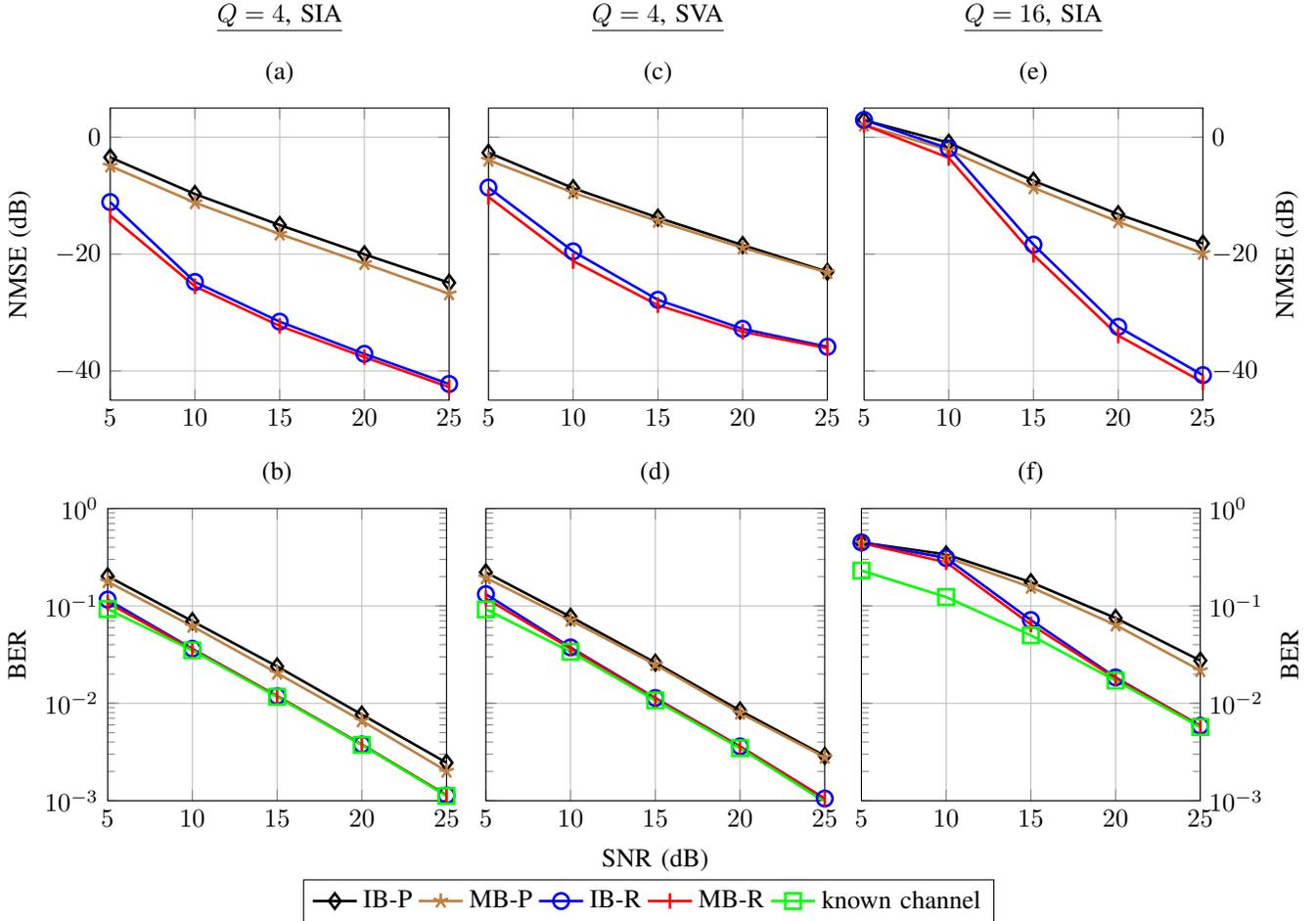


Fig. 8: Experiment 2: Performance comparison between marginal-based (MB) and integer-based (IB) algorithms.

The figure also shows that the algorithms perform equally well for both SIA and SVA configurations and, hence are robust to how accurate property 2 is.

An important observation is that there is little advantage of MB algorithms over IB algorithms in this setting. However in some scenarios the improvement could be better as it uses more (and accurate) information to estimate the channels. For example, the difference between the performance of MB and IB algorithms is more evident when the number of pilots is further reduced, as could be seen in the success rate curves of Fig. 7. Also note that the gain of MB algorithms is much more noticeable for their pilot-based versions. However, this advantage is at the expense of a relatively high computational and communication cost as depicted in Table V which compares the runtime for the two algorithms.

Algorithm	Run time (sec)
MB-P	0.3897
IB-P	0.3095

TABLE V: Run time comparison of the MB-P and IB-P.

3) *Experiment 3 - Comparison with BR and oracle-LS algorithms:* In this experiment, we benchmark the performance of the proposed algorithms against BR and oracle-LS. Here we use QAM signals of order $Q = 4$ and $Q = 16$ and pass them through a channel of length $L = 32$ and sparsity $n = 3$. We use $K = 8$ pilots and confine our attention to the space-invariant case as this is an essential requirement for block sparsity algorithm to work.

It is obvious from the graphs of Fig. 9 that the proposed MB-R algorithm has the best performance among all algorithms. The gain in the performance of the proposed algorithm over others is more prominent for higher values of SNR. Specifically, note that there is a difference of nearly two orders of magnitude in the BER of MB-R and BR when $\text{SNR} = 35$ dB and $Q = 4$.

4) *Experiment 4 - Effect of sparsity rate:* In this experiment, we study the performance of the proposed algorithms under different sparsity rates. Channels of length $L = 64$ were generated having $n = 3, 5$ and 7 non-zero taps corresponding to sparsity rate of 4.7% – 11%. In this experiment, QAM signals of order $Q = 4$ were passed through the channels and $K = 16$ number of pilots were used. Fig. 10 shows the NMSE performance of the proposed IB-R algorithm. It is evident from the graphs that the performance of the proposed algorithm degrades gracefully as CIR gets denser. A similar performance is achieved for the MB-R algorithm. The degradation in reconstruction accuracy with increased number of non-zeros is a common trend in all sparse recovery algorithms.

5) *Experiment 5 - Effect of D :* In this experiment, we study the effect of the number of collaborating antennas on channel estimation for both the SIA and SVA cases. In this experiment we compute the BER in the recovered data for various values of SNR and parameter D . Fig. 11 and 12 show the BER vs SNR graphs for SIA and SVA respectively. Specifically, we plot for $D = 1, 2, 3, 4$ and 5 . Note that when

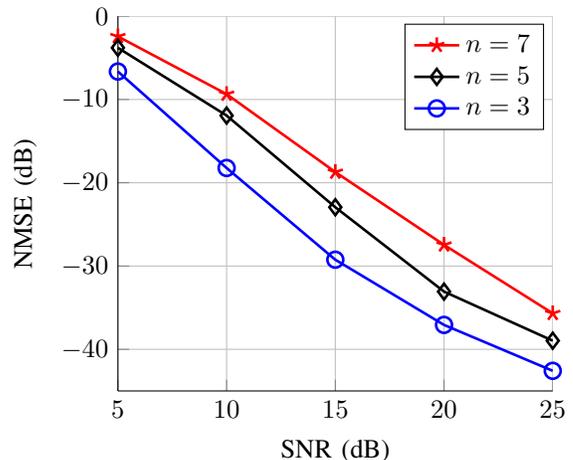


Fig. 10: Experiment 4: Effect of channel sparsity on its recovery.

$D = 1$, information from only the direct 4 neighbors is taken into consideration and for $D = 2$ information belonging to the neighbors of neighbors is also incorporated in the computations.

In the SIA case (Fig. 11) QAM signals of order $Q = 4$ were passed through channels of length $L = 32$ and sparsity $n = 3$ and the corresponding CIRs were estimated using IB-P with the help of $K = 8$ pilots. Fig. 11 shows that sharing improves the BER performance. Specifically, as we increase the scope of sharing from the first tier of neighbors ($D = 1$) to the fifth tier ($D = 5$), we observe a drop in BER. However, the improvement in BER is not significant beyond $D = 3$. This behavior is dependent on several factors such as the length of channel (L) to be estimated, number of pilots (K) and the channel sparsity (n). For instance, in this example, if the number of pilots is reduced (i.e., $K < 8$) the estimated CIRs will be more erroneous. However, the effect of this error is compensated by using more neighbors to average the marginals/scores (i.e., higher D). Therefore, in this reduced number of pilots scenario we might observe significant improvement beyond $D = 3$ as well.

In the SVA case (Fig. 12) QAM signals of order $Q = 4$ were passed through channels of length $L = 64$ and sparsity $n = 3$ and the corresponding CIRs were estimated using IB-P with the help of $K = 16$ pilots. Fig. 12 shows that we do not gain anything for higher values of D . This is due to the space-variant nature of the impulse response that adding more information does not help in the improvement of estimation accuracy. Therefore, setting $D = 1$ or 2 might be sufficient in this scenario.

VII. CONCLUSION AND FUTURE WORK

Massive MIMO systems provide substantial performance gains as compared to the traditional MIMO systems. However, these gains come with a huge requirement of estimating a large number of channels. In this paper we have shown that these

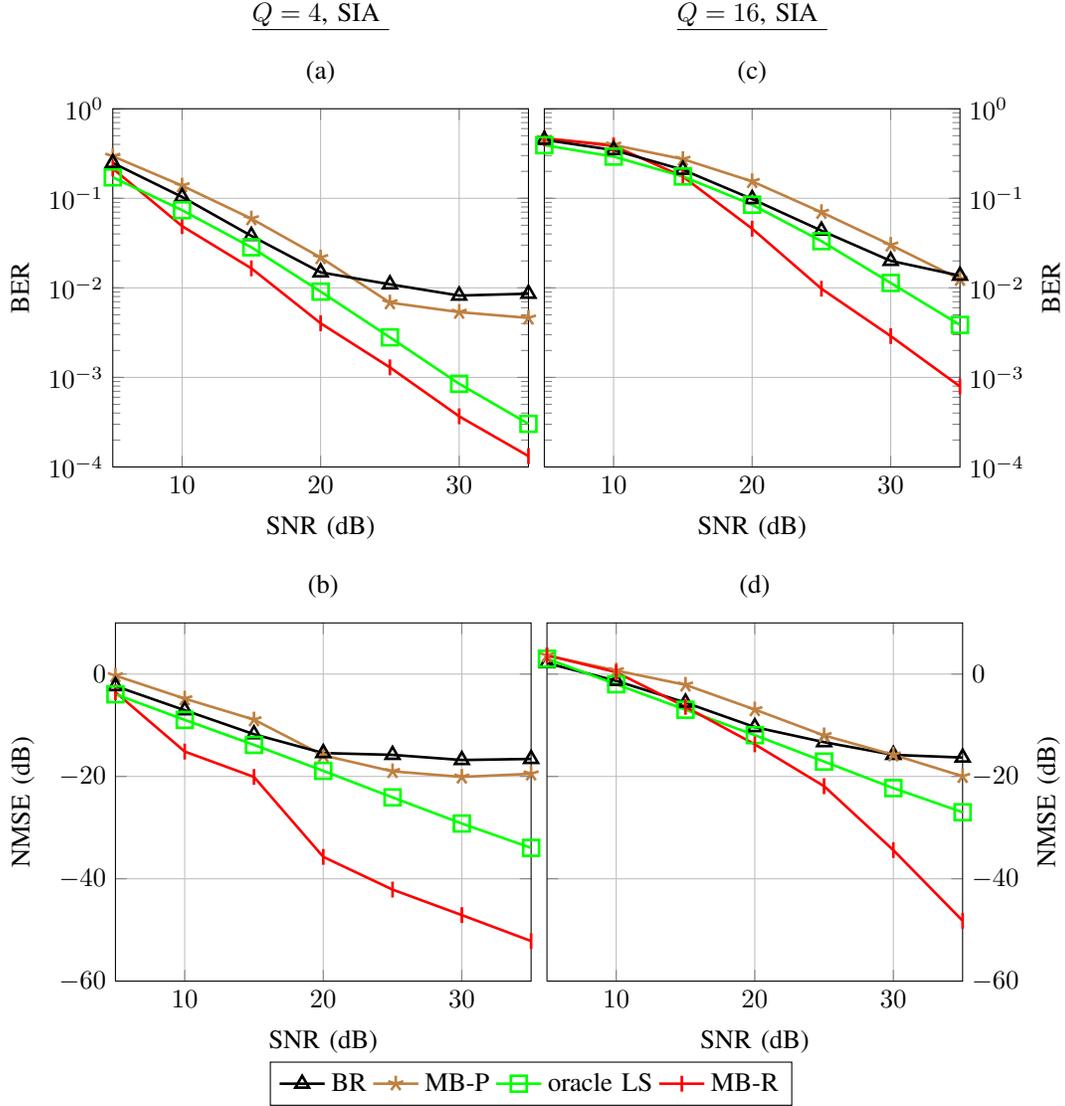


Fig. 9: Experiment 3: Performance comparison between the proposed and the BR and oracle-LS algorithms.

channels could be estimated in a collaborative manner where the antennas collaborate with their neighboring antennas. Three algorithms based on this collaborative method have been presented. The algorithms show good performance under different scenarios and that too while using a relatively small number of pilots.

APPENDIX A CHANNEL MODELS

Numerous good/accurate models exist for MIMO wireless channels. However, almost all consider the uniform linear array configuration of the antennas [51], [52]. It should be noted that from a practical point of view, assuming a uniform linear array (ULA) becomes unfeasible for the purpose of modeling a large scale antenna array. Therefore, the use of two- or

three-dimensional antenna arrays can be more appropriate in massive MIMO systems. There have been a few attempts in developing channel models for the two- and three-dimensional antenna array configurations. For example, the IlmProp tool available online [49] allows to generate CIR for 2D antenna array configuration. A similar proposal was put forward in [53], [54], to extend the spatial channel model (SCM) standard [52]. Similar models have also been considered in Winner II [55] and Winner+ [56] initiatives. However, these models come with their limitations. For example, in IlmProp the parameters have been estimated using much lower spatial resolution. Indeed, there is a general lack of channel models for the massive MIMO scenarios and there is a need of more work in this direction.

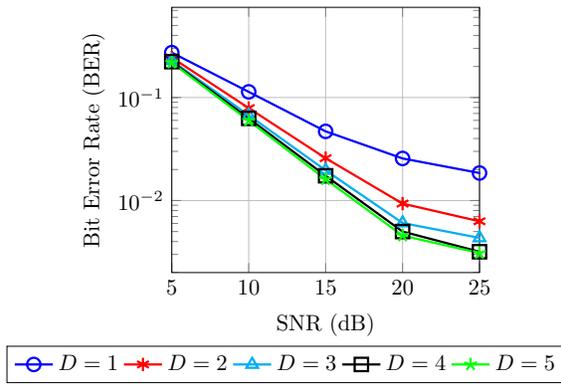


Fig. 11: Experiment 5: Information sharing among antennas belonging to different neighbor levels ($D = 1 - 5$) adds to CIR estimation accuracy in the SIA case.

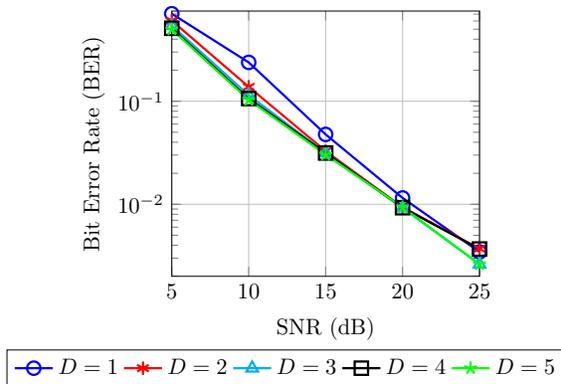


Fig. 12: Experiment 5: In the SVA case information sharing does not help in the improvement of CIR estimation accuracy.

REFERENCES

- [1] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*. Wiley, 2012.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, pp. 40–60, Jan. 2013.
- [3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 160–171, Feb. 2013.
- [4] T. L. Marzetta, "How much training is required for multiuser MIMO," in *Proc. Asilomar Conf. on Signals and Systems and Computers*, Oct. 2006.
- [5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [8] 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 *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 3464–3467.
- [9] A. Medles, D. T. M. Slock, and E. D. Carvalho, "Linear prediction based semiblind estimation of MIMO FIR channels," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2001, pp. 58–61.
- [10] F. Wan, W. Zhu, and M. Swamy, "Semiblind channel estimation of MIMO-OFDM systems with pulse shaping," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, 2008, pp. 125–128.
- [11] —, "Semiblind sparse channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2569–2582, Jul. 2010.
- [12] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.
- [13] D. Eiuwen, G. Tauböck, F. Hlawatsch, and H. Feichtinger, "Group sparsity methods for compressive channel estimation in doubly dispersive multicarrier systems," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2010, pp. 1–5.
- [14] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *ArXiv e-prints*, Jan. 2014.
- [15] H. Yin, 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, no. 2, pp. 264–273, Feb. 2013.
- [16] S. L. H. Nguyen and A. Ghayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2013, pp. 2890–2895.
- [17] W. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, Jun 1995.
- [18] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3715, Dec. 2012.
- [19] I. Fevrier, S. Gelfand, and M. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927–937, Jun 1999.
- [20] S.-S. Sadough, M. Ichir, P. Duhamel, and E. Jaffrot, "Wavelet-based semiblind channel estimation for ultrawideband OFDM systems," *IEEE Trans. Veh. Technol.*, pp. 1302–1314, 2009.
- [21] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1743–1763, May 2007.
- [22] J.-K. Hwang, R.-L. Chung, and M.-F. Tsai, "Highly efficient sparse multipath channel estimator with optimal chu-sequence prefilter for mimo frequency-domain dfe receiver," in *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, Aug 2006, pp. 755–759.
- [23] T. S. Rappaport, *Wireless Communications: Principles and Practice, 2nd edition*. Prentice Hall, 2002.
- [24] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. Annu. Conference Information Sciences and Systems (CISS)*, Mar. 2008.
- [25] M. Masood and T. Y. Al-Naffouri, "Sparse reconstruction using distribution agnostic Bayesian matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5298–5309, Nov. 2013.
- [26] W. Bajwa, A. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Proc. Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, 2009, pp. 320–325.

- [27] H. Minn and V. Bhargava, "An investigation into time-domain approach for OFDM channel estimation," *IEEE Trans. Broadcast.*, pp. 240–248, 2000.
- [28] A. A. Quadeer and T. Y. Al-Naffouri, "Structure-based Bayesian sparse reconstruction," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6354–6367, Dec 2012.
- [29] Y. Xiong and Y. M. Lu, "Blind estimation and low-rate sampling of sparse MIMO systems with common support," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 3893–3896.
- [30] D. Eiwen, G. Tauböck, F. Hlawatsch, H. Rauhut, and N. Czink, "Multichannel-compressive estimation of doubly selective channels in MIMO-OFDM systems: Exploiting and enhancing joint sparsity," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, 2010, pp. 3082–3085.
- [31] S. Nguyen, "Compressive sensing for multi-channel and large-scale MIMO networks," August 2013. [Online]. Available: <http://spectrum.library.concordia.ca/977759/>
- [32] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
- [33] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [34] K. M. Z. Islam, T. Y. Al-Naffouri, and N. Al-Dhahir, "On optimum pilot design for comb-type OFDM transmission over doubly-selective channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 930–935, Apr. 2011.
- [35] —, "Asymptotically mmse-optimum pilot design for comb-type OFDM channel estimation in high-mobility scenarios," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 3776–3779.
- [36] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.
- [37] —, "Optimal pilot placement for time-varying channels," in *Proc. IEEE Workshop Signal Process Adv. Wireless Commun. (SPAWC)*, Jun. 2003, pp. 219–223.
- [38] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems," *IEEE Signal Process. Lett.*, vol. 18, no. 12, pp. 749–752, Dec. 2011.
- [39] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2008, pp. 2885–2888.
- [40] Tareq Y. Al-Naffouri and Mudassir Masood, "Distribution agnostic structured sparsity recovery algorithms," in *Proc. International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, 2013, pp. 283–290.
- [41] H. V. Poor, *An Introduction to Signal Detection and Estimation*, ser. A Dowden & Culver book. Springer, 1994.
- [42] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, July 2005.
- [43] E. Al-Safadi and T. Y. Al-Naffouri, "Pilotless recovery of clipped OFDM signals by compressive sensing over reliable data carriers," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2012, pp. 580–584.
- [44] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 255–271, April 2010.
- [45] S. Das and P. Schniter, "Max-SINR ISI/ICI-shaping multicarrier communication over the doubly dispersive channel," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5782–5795, Dec 2007.
- [46] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, April 2004.
- [47] R. Koetter, A. C. Singer, and M. Tüchler, "Turbo equalization," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 67–80, Jan 2004.
- [48] G. Del Galdo, M. Haardt, and C. Schneider, "Geometry-based channel modelling of MIMO channels in comparison with channel sounder measurements," *Advances in Radio Science - Kleinheubacher Berichte*, vol. 50, no. 5, pp. 117–126, October 2003.
- [49] "IlmProp," <http://www2.tu-ilmeneau.de/nt/en/ilmprop/>, [Online; accessed January 10, 2014].
- [50] Z. Zhang and B. Rao, "Recovery of block sparse signals using the framework of BSBL," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 3345–3348.
- [51] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, no. 12, pp. 1926–1934, Mar. 2002.
- [52] "3GPP TR 25.996 V11.0.0, Spatial channel model for multiple input multiple output (MIMO) simulations," <http://www.3gpp.org/ftp/Specs/html-info/25996.htm>, [Online; accessed January 10, 2014].
- [53] B. L. Ng, Y. Kim, J. Lee, Y. Li, Y.-H. Nam, J. Zhang, and K. Sayana, "Fulfilling the promise of massive MIMO with 2D active antenna array," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2012, pp. 691–696.
- [54] "3GPP TSG RAN WG1 Meeting 74, Channel Coefficient Generation in 3D Channel Modelling," http://www.3gpp.org/ftp/tsg_ran/wg1_r11/TSGR1_74/Docs/R1-133114.zip, [Online; accessed January 10, 2014].
- [55] "WINNER II," <http://www.ist-winner.org/deliverables.html>, [Online; accessed January 10, 2014].
- [56] "WINNER+," <http://projects.celtic-initiative.org/winner+/>, [Online; accessed January 10, 2014].