

Pilot Reduction Techniques for Sparse Channel Estimation in Massive MIMO Systems

Farzana Kulsoom*, Anna Vizziello*, Hassan Nazeer Chaudhry†, Pietro Savazzi*

* Department of Electrical, Computer and Biomedical Engineering
University of Pavia

Via Ferrata 5, 27100 Pavia, Italy
Email: farzana.kulsoom01@universitadipavia.it,
anna.vizziello@unipv.it, pietro.savazzi@unipv.it

† Department of Electrical, Information and Bioengineering
Politecnico di Milano

Via Camillo Golgi, 42, 20133 Milano, Italy
Email: hassannazeer.chaudhry@polimi.it

Abstract—The current high gain frequency division duplex (FDD) Massive multiple-input, multiple-output (MIMO) systems pose several challenges to carry out the downlink beamforming. Specifically, downlink beamforming requires a channel estimation that usually needs long training and feedback overhead, scaling with the number of antennas at the base station (BS). We exploit compressive sensing (CS) techniques to accurately estimate the channel, while assuring overhead reduction which is proportional to the sparsity level of the channel. The sparse virtual channel representation is obtained through the proposed dictionary design, which is more flexible, robust and able to estimate the cell characteristics. We specifically focus on massive MIMO-Orthogonal Frequency-Division Multiplexing (OFDM) systems that show more robustness to multipath fading, and analyze several CS algorithms to select among them the best technique with the proposed dictionary design. Numerical results demonstrate that greedy solutions approach the basic pursuit bound with lower complexity and consequent shorter training period. The normalized hard thresholding pursuit (NHTP) technique is the greedy algorithm with the best performance complexity trade-off.

I. INTRODUCTION

Next-generation massive multiple-input multiple-output (MIMO) systems allow fine beamforming with narrow beams, thanks to the large number of antennas at the base station (BS) [1] [2]. This turns out in interference avoidance among users and higher throughput compared to current architectures.

However, to achieve effective downlink beamforming, precise channel state information (CSI) is required. In time-division duplexing (TDD) systems, CSI can be estimated exploiting channel reciprocity [3]. Specifically, the uplink channel information, obtained easily due to the low number of antennas at the mobile station, can be exploited for downlink beamforming.

On the contrary, frequency-division duplexing (FDD) architectures do not allow the use of channel reciprocity since different frequency bands are used for uplink and downlink with consequent different channels. Since FDD systems are more suitable for delay-sensitive and symmetric traffic systems, new solutions for downlink training overhead reduction are needed. On this purpose, the compressive sensing (CS) framework can be exploited to reduce the downlink training

period, by sparsely representing the channel response with some dictionary or basis.

The discrete Fourier transform (DFT) matrix has been already employed to represent the channel in a sparse manner [4] [5]. However, such representation is not valid for all the antenna geometries, but only for uniform linear array (ULA) with a sufficient number of antennas and limited scattering, which reduces its real feasibility.

New dictionary based virtual channel model have been proposed to sparsely represent the channel, which are able to adjust to the cell characteristics, with no restriction to the array geometry. We exploit dictionary for sparse channel model and, once sparse channel representation is obtained, we apply compressive sensing techniques for channel estimation with a reduced downlink training period.

We specifically focus on MIMO-orthogonal frequency division multiplexing (OFDM) systems. In order to mitigate the frequency selective fading, OFDM is used along with MIMO effectively converting it into flat fading channel. Each subcarrier is transmitted over a narrow band hence simplifying equalizer design at the receiver. We compare different compressive sensing techniques, i. e., convex optimization based approaches (basic pursuit (BP)) and greedy methods (matching pursuit (MP)). Among the greedy algorithms, we employ the well known orthogonal matching pursuit (OMP), as well as the iterative hard thresholding (IHT) and the normalized hard thresholding pursuit (NHTP) methods.

The main contributions of this work are:

- 1) a compressive sensing based solution is proposed for channel estimation in massive MIMO-OFDM FDD systems with reduced downlink training overhead;
- 2) we implement different CS algorithms and choose the optimal one with the proposed dictionary, while [3] only demonstrates the usefulness and potential of dictionary learning based channel modeling applied to an OMP CS solution, without analyzing the problem of optimal CS algorithm selection for massive-MIMO using learned dictionary;
- 3) a comparison of DFT basis and proposed dictionary is presented varying the pilot spacing, which shows that

proposed dictionary based compressed sensing is more robust and reliable even when increasing pilot spacing.

- 4) simulation results show that the basic pursuit gives the best performance at the cost of high complexity. Greedy solutions approach the basic pursuit bound with a lower complexity and consequent shorter training period. Specifically, the NHTP approach is the greedy algorithm with the best performance complexity trade-off.

The paper is organized as follows: Sec. II-A describes sparse channel estimation, while Sec. III the compressive sensing methods. Sec. IV shows some interesting simulations results, and, finally, some concluding remarks wrap up and close the paper in Sec. V.

II. SYSTEM MODEL

Wireless channel involves multiple paths of varying delays and gains. The transmission at high data rates generally implies a small symbol duration, so that the received signal is affected by intersymbol interference (ISI) of successive modulation symbols, due to several paths in the propagation environment. A time varying channel with L_t paths can therefore be represented as

$$h(t) = \sum_{k=0}^{L_t-1} \alpha_k \delta(t - \tau_k) \quad (1)$$

Where τ_k and α_k denote the k^{th} channel path delay and complex gain respectively. In order to combat multipath fading, the well known multicarrier scheme OFDM may be used, by employing narrow spaced subcarriers at the transmit side. An OFDM signal consists of N_o subcarriers, equally frequency spaced at $\Delta f = \frac{1}{T_s}$, with T_s the sampling time. An OFDM discrete time baseband signal can be expressed as

$$S_i(n) = \frac{1}{\sqrt{N_o}} \sum_{k=0}^{N_o-1} x_i(k) e^{j \frac{2\pi kn}{N_o}} \quad (2)$$

where $S_i(n)$ represents the n^{th} sample of the i^{th} OFDM symbol, and $x_i(k)$ represents the data transmitted over the k^{th} subcarrier in the i^{th} symbol interval, while $\frac{1}{\sqrt{N_o}}$ is a normalization factor. The k^{th} subcarrier in the equivalent lowpass domain is described by the signal $g_k(t)$ as

$$g_k(t) = e^{j2\pi\Delta f kt} \cdot \text{rect}\left(\frac{t}{T_s}\right) \quad (3)$$

Despite the fact that these subcarriers overlap each other, they do not interfere since they are orthogonal. The maximum channel delay $\tau_{max} = \tau_{L_t-1}$ may cause intercarrier interference (ICI) which is usually mitigated by appending a cyclic extension of each OFDM symbol at its end, for a number of samples $G_i \geq \tau_{max}$, i.e. the guard interval. Therefore, the OFDM total symbol duration is $T_{sym} = (N_o + G_i) \cdot T_s$. A perfect channel knowledge is a fundamental requirement for channel equalization. The channel may be usually estimated by blind or pilot based techniques. In OFDM pilot based channel estimation, pilots are arranged in a time frequency grid over time (block based) or frequency (combo based). The receiver can directly

estimate channel behavior at the frequencies where pilots are transmitted using conventional least square based technique

$$H_{i,k} = \frac{R_{i,k}}{S_{i,k}} \quad (4)$$

Where $S_{i,k}$ is k^{th} position of i^{th} OFDM symbol transmitted and $R_{i,k}$ is k^{th} position of i^{th} OFDM symbol received. Recently, research has been focused on exploiting communication channel sparsity, considering the fact that channel exhibit only few dominant propagation paths which may be approximated as a linear combination over a known basis or dictionary, resulting in a sparse Channel impulse response (CIR). Traditionally DFT basis are used for sparse channel representation.

$$\phi = \begin{bmatrix} e^{-j2\pi k_{0,1}} & \dots & e^{-j2\pi k_{0,N_o-1}} \\ e^{-j2\pi k_{1,1}} & \dots & e^{-j2\pi k_{1,N_o-1}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi k_{N_o-1,1}} & \dots & e^{-j2\pi k_{N_o-1,N_o-1}} \end{bmatrix}$$

The use of the DFT base is compliant with the theoretical results of signals estimation in CS [6], and it has been used to represent sparse channel models. However, DFT basis may represent the channel only in a few, although orthogonal, directions. In order to better estimate the channel, more robust and refined basis may be defined.

A. Sparse Channel Estimation

In this work, we will focus on a downlink massive MIMO-OFDM system, with a number of transmit antennas equal to N , and a single antenna at the receiver. Multicarrier OFDM modulation is used in order to mitigate frequency selective fading and aid channel estimation by means of pilot subcarriers. Data frames are represented by \mathbf{d}_j with $j = 1, \dots, N-L$, while the number of training pilots is L .

$$\mathbf{x} = \{\mathbf{P}_1, \mathbf{d}_1, \mathbf{P}_2, \mathbf{d}_2, \dots, \mathbf{P}_L, \dots, \mathbf{d}_{N-L}\} \quad (5)$$

Pilots \mathbf{P} are built according to a Rademacher distribution (RD), comprising equally likely symbols belonging to $\{+1, -1\}$ and multiplied by a phase rotation as in the following equation.

$$\mathbf{P} = [\mathbf{p}_1 e^{-j\pi l_1}, \mathbf{p}_2 e^{-j\pi l_2}, \dots, \mathbf{p}_L e^{-j\pi l_L}] \quad (6)$$

where l_j are uniformly distributed random numbers belonging to the interval $[0, 1]$, while the channel delay profile can be formulated using OFDM sample time and guard interval for each dictionary point as follows

$$\tau_h = [0, \alpha, 2\alpha, \dots, (M-1)\alpha] \quad (7)$$

where M is the dictionary length and $\alpha = G_i \cdot T_s - G_i \cdot T_s / M$ is the minimum channel spacing which is calculated by considering the constraint that τ_{max} must not exceed G_i . The downlink channel vector $\mathbf{h} \in \mathbb{C}^{N \times 1}$ is estimated at the mobile receiver by using pilot symbols, and this channel state information is sent back to the base station [3]. In more details, the transmitted pilot symbols may be represented by the matrix $\mathbf{A} \in \mathbb{C}^{L \times N_o}$, for DFT based sensing matrix and $\mathbf{A} \in \mathbb{C}^{L \times M}$ for dictionary based sensing matrix. Block based pilot arrangement is used as it is most suitable for multipath fading channel. For example, for N_o equals to 2048, 32 pilots are transmitted per OFDM symbol

with pilot interval (PI) equal to 64, while the remaining 2016 sub carriers are used for data. After the OFDM DFT matrix multiplication, the received signal at each single user can be expressed as in the following:

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} \quad (8)$$

where the elements of $\mathbf{n} \in \mathbb{C}^{L \times 1}$ are additive white Gaussian Noise (AWGN) samples with variance $\sigma_n^2 = N_0/2$. Least square (LS) channel estimation requires a number of pilot symbols $L \geq N$, needed to compute the pseudoinverse of the pilot matrix \mathbf{A} . For massive MIMO the large number of Antennas N can make this unfeasible, considering also that the estimated channel information shall be sent back to the base station.

B. Sparse Channel Representation

If $L < N$ the compressed sensing paradigm may be taken under consideration, since channel estimation may be viewed as measuring a high dimensional signal with a very limited number of measurements, assuming that the original signal, i.e. the channel estimate \mathbf{h} , may be sparse in a some suitable basis.

In order to construct the dictionary \mathbf{D} basis, we are only considering the useful OFDM symbol duration, without G_i , and taking the channel delay profile into account

$$\mathbf{D} = \begin{bmatrix} e^{-j2\pi \cdot 1 \tau_1} & e^{-j2\pi \cdot 1 \tau_2} & \dots & e^{-j2\pi \cdot 1 \tau_M} \\ e^{-j2\pi \cdot 2 \tau_1} & e^{-j2\pi \cdot 2 \tau_2} & \dots & e^{-j2\pi \cdot 2 \tau_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi \cdot N_o \tau_1} & e^{-j2\pi \cdot N_o \tau_2} & \dots & e^{-j2\pi \cdot N_o \tau_M} \end{bmatrix}$$

where $\tau(m) = 1/T_{\text{sym}} \cdot \tau_h(m)$ where $m = \{0, 1, \dots, M-1\}$

Following this reasoning, to effectively represent channel sparsity, if the basis $\mathbf{D} \in \mathbb{C}^{M \times N_o}$ may be redefined such that $\mathcal{D} = \mathbf{D}_{L \times M}$ is a sub-matrix of \mathbf{D} , we can then write

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} = \mathcal{D} \cdot \mathbf{P}\boldsymbol{\beta} + \mathbf{n} \quad (9)$$

where $\boldsymbol{\beta} \in \mathbb{C}^{L \times 1}$ is a representation compressed vector, and $\mathbf{y} \in \mathbb{C}^{M \times 1}$ the estimated one. The design of dictionary \mathbf{D} and sensing matrix \mathbf{A} is always critical when using compressed sensing. It can be seen that $\mathbf{D} \in \mathbb{C}^{N_o \times M}$, and $\mathbf{A} \in \mathbb{C}^{L \times M}$ is basically consisting of some specific rows of \mathbf{D} related to pilot locations as in [7]:

$$\mathbf{A} = \begin{bmatrix} e^{-j2\pi l_1 \tau_1} & \dots & e^{-j2\pi \cdot l_1 \tau_M} \\ e^{-j2\pi l_2 \tau_1} & \dots & e^{-j2\pi \cdot l_2 \tau_M} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi l_L \tau_1} & \dots & e^{-j2\pi \cdot l_L \tau_M} \end{bmatrix} \begin{bmatrix} [p_1 e^{-j\pi l_1}] \\ p_2 e^{-j\pi l_2} \\ \vdots \\ p_L e^{-j\pi l_L}] \end{bmatrix} \quad (10)$$

For comparison purposes, the sensing matrix \mathbf{A} with dft basis ϕ is computed in a similar manner from \mathbf{D} , i.e. only some specific rows of the basis function related to the training subcarriers are considered:

$$\mathbf{A}_\phi = \begin{bmatrix} e^{-j2\pi k_{l_1,1}} & \dots & e^{-j2\pi k_{l_1, N_o}} \\ e^{-j2\pi k_{l_2,1}} & \dots & e^{-j2\pi k_{l_2, N_o}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi k_{l_L,1}} & \dots & e^{-j2\pi k_{l_L, N_o}} \end{bmatrix} \begin{bmatrix} [p_1 e^{-j\pi l_1}] \\ p_2 e^{-j\pi l_2} \\ \vdots \\ p_L e^{-j\pi l_L}] \end{bmatrix} \quad (11)$$

The equivalent channel representation in (9) is sparse if $\|\boldsymbol{\beta}\|_0 = s \ll N$. Further, if we are able to solve for compressed vector $\boldsymbol{\beta}$, then the channel estimate can be obtained. Additionally, sparse signal reconstruction is based on the assumption that \mathbf{h} is a s -sparse CIR, i.e. its channel energy is uniformly distributed among few dominant taps. The exact position of these dominant taps is not a priori known, and it must be estimated for effective channel sensing [7].

Sensing matrix \mathbf{A} can be represented in the form of a DFT basis $\phi \in \mathbb{C}^{N_o \times N_o}$ or in the form of an overcomplete dictionary $\mathbf{D} \in \mathbb{C}^{N_o \times M}$, to solve a channel estimation problem with $M > N_o$.

The main objective of any dictionary is to sparsely represent data in terms of some basic elements, called atoms, that are not required to be orthogonal and they may be seen as an over-complete spanning set. The basic philosophy of these algorithms is that a dictionary may be inferred from input data, inspired by the fact that we need to represent each data set by using as few samples as possible.

To solve the sparse channel estimation problem, each iteration alternatively minimizes the CS error with respect to either \mathbf{D} or $\boldsymbol{\beta}$, while keeping the other fixed [3]. The algorithm convergence depends on the specific sparse recovery and dictionary update algorithms. Our work aims to optimize $\boldsymbol{\beta}$, starting from an overcomplete dictionary representation. Finally, a channel estimate is built at all location merely from dictionary $\hat{\mathbf{x}} = \mathbf{D}\mathbf{y}$.

III. COMPRESSIVE SENSING METHODS

We apply different CS algorithms to the proposed dictionary based sparse channel representation, in order to select the best CS solution able to reconstruct the estimated channel from few measurements. CS methods can be categorized broadly into convex relaxation iterative algorithms [8], [6], and greedy ones [9], [10]. Other methods have been proposed as a combination of these techniques [11], [12], [13].

A. Convex relaxation iterative algorithms

The convex relaxation type exploits linear programming to solve undetermined systems. Some convex techniques are formulated as \mathcal{L}_0 , \mathcal{L}_1 , and \mathcal{L}_2 norm minimizations. In the following, we denote with \mathbf{x} the estimated channel response $\hat{\mathbf{h}}$. Considering the basic pursuit algorithm, the solution of a \mathcal{L}_1 norm minimization can be formulated as a cost function minimization problem. BS is an iterative algorithm that initializes the first guess ($\mathbf{x} = \mathbf{x}_0 = \mathbf{A}^T \mathbf{y}$), which represents the minimal signal energy, and then computes the cost function by minimizing $\|\mathbf{y} - \mathbf{A}\mathbf{x}_0\|_2$. For the next i iterations, the measurement matrix is calculated selecting the s required measurements, and the signal coefficients are adjusted according to the minimized form. The iterations stop when enough signal coefficients are obtained, less than its sparsity level [14].

Convex relaxation methods needs only a small number of measurements to reconstruct the signal; anyway, they exhibit a greater complexity in terms of computation and time consuming [14].

B. Greedy algorithms

Differently from the convex relaxation algorithms, greedy techniques are easy to implement and able to quickly recover a

sparse signal, although their solution is not optimal [14]. Such methods are in line with our objective to reduce the downlink training overhead for channel estimation, however we should also strictly take into account their performance.

Let us call \mathbf{x} the large dimensional signal, that is the channel to be estimated in our case, with a high number M of samples, which is sparse in some domain. Greedy algorithms pick one position of a non-zero element of \mathbf{x} , which corresponds to select one column from the measurement matrix.

Some techniques have been proposed under this category: orthogonal matching pursuit (OMP) [15], compressive sampling matching pursuit (CoSaMP) [12], iterative hard thresholding (IHT) [16], and normalized hard thresholding pursuit (NHTP) [17]. Matching pursuit selects a column from the measurement matrix that maximizes the inner product of the current residuals. As a first step, the algorithm selects from the dictionary the vector corresponding to the longest projection of \mathbf{x} . In the second step, the signal \mathbf{x} is orthogonalized by removing any element of the selected vector from \mathbf{x} in order to obtain its residual with the lowest energy. Finally, such two steps are iteratively processed for the remaining part of the dictionary until the residual norm is lower than a certain threshold [14]. Orthogonal matching pursuit method is a variant of the matching pursuit algorithm by discharging both the elements of the selected vector from \mathbf{x} , and from the basis, before doing again the process. Orthogonal matching pursuit shows better results than matching pursuit, at the cost of a higher computational complexity.

Given the high-dimensional signal \mathbf{x} , observed via low-dimensional measurements \mathbf{y} , the aim is to reconstruct the signal from the measurements. Iterative Hard Thresholding (IHT) is built on the insights of [18]:

- solving the rectangular system $\mathbf{A}\mathbf{x} = \mathbf{y}$ is equivalent to solve the square system $\mathbf{A} \cdot \mathbf{A}\mathbf{x} = \mathbf{A} \cdot \mathbf{y}$;
- a sequence \mathbf{x}^i can be defined as a recursion $\mathbf{x}^{i+1} = (\mathbf{I} - \mathbf{A} \cdot \mathbf{A})\mathbf{x}^i$;
- given the objective of sparse vectors, each step includes the hard thresholding operator that keeps s nonzero largest components of a vector and fixes the other ones to zero.

Compressive Sampling Matching Pursuit algorithm pursues first a suitable candidate for the support, and then finds the vector, with this support, that fits well the measurements [18].

The Hard Threshold Pursuit (HTP) algorithm is a combination of the IHT and the CoSaMP or Subspace Pursuit algorithms [18], so that it selects the s largest component of $\mathbf{x}^i + \mathbf{A} \cdot \mathbf{A}(\mathbf{x} - \mathbf{x}^i) \approx \mathbf{x}$. The HTP main feature is its high speed.

Finally, Normalized Hard Thresholding Pursuit is a variant of the HTP method.

IV. PERFORMANCE EVALUATION

In this section, the performance of the proposed scheme are evaluated for different CS techniques.

A. Simulation Setup

We consider a MIMO-OFDM system, with 100 transmit antennas and one receive one. The base station uses 256 OFDM subcarriers for downlink transmission. Some subcarriers are employed for pilot sequences and others for data. The multipath channel length L_{taps} is equal to 6, whose consequent interference is directly solved by the OFDM guard interval,

TABLE I
SIMULATION PARAMETERS

| Parameter | Symbol | Value |
|--------------------------|--------|-------|
| Transmit antennas | N | 100 |
| Receive antennas | R | 1 |
| OFDM subcarriers | N_o | 256 |
| OFDM guard interval | G_i | 16 |
| QAM Modulations | QAM | 32 |
| Multipath channel length | L_t | 6 |

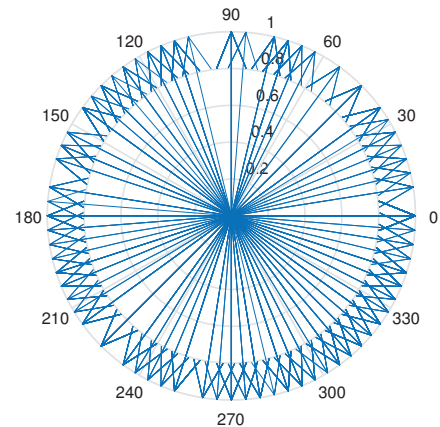


Fig. 1. Analysis of angles of dictionary basis 16x20

whose length is set equal to 16. Table I summarizes the system parameters.

B. Performance Analysis

Wireless channel in the angular domain can be represented by sparse DFT basis as reported in [7], [19]. Although DFT basis are orthonormal, only a few angles or directions may be defined in it. In practical systems, signals may arrive from any arbitrary directions, so predefined patterns of dft bins do not represent a good choice [3]. On the contrary, in the proposed solution, a dictionary with more refined bins is used in order to estimate the sparse channel model. In Fig. 1 and 2, the angular representation of DFT and Dictionary basis are shown, for a $N_o \times N_o$ DFT matrix and a $N_o \times M$ dictionary one. If $N_o = M$ the dictionary is defined as complete, while when $M > N_o$ the dictionary is said over-complete, providing more redundant basis which is more flexible to estimate the channel.

The effects of Different pilot spacing and sparsity levels on on system performance are analyzed. In massive MIMO, pilot training overhead is one of the main concern, since it is proportional to the number of transmit antennas, in order to acquire good CSI at the base station.

In Fig. 3, we analyze the effects of the OMP algorithm on BER for the following different values of the sparsity level $s = 15, 30, 45, 60, 65$. We also observe the system performance for pilot interval (PI) values equal to 4, 14. Similar analysis can be also conducted for the other CS techniques. Fig. 3 shows that the proposed dictionary performs better than the DFT based one, and it is more robust to pilot spacing changes. Indeed, when $PI = 14$, the DFT basis performance behaves worse than $PI = 4$, while the dictionary based algorithm almost

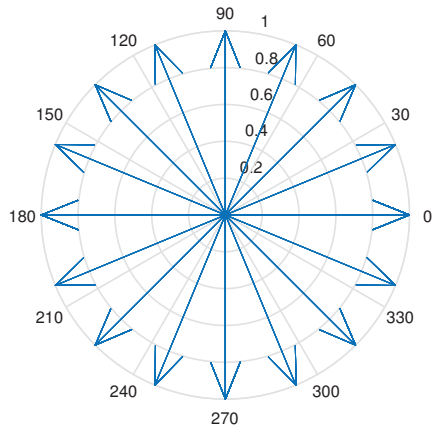
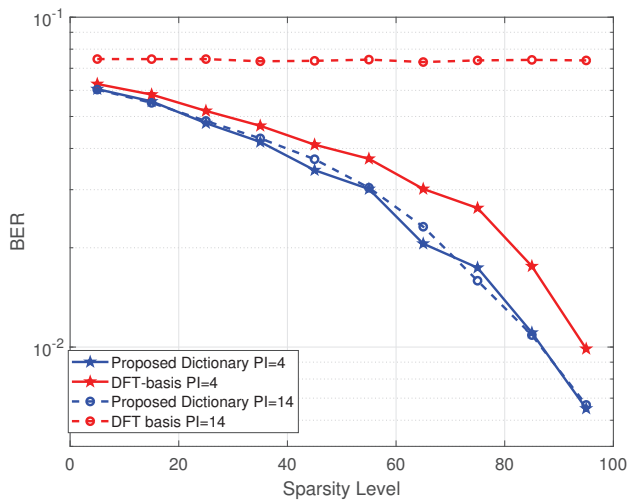


Fig. 2. Analysis of angles of dft basis 16x16


 Fig. 3. Analysis of DFT basis and proposed dictionary with different sparsity levels s and pilot spacing for OMP algorithm

maintains the same performance. Further, when the sparsity level s is increasing, all the channel measurements are more accurately represented, indeed the bit error rate is decreasing.

Although there is a huge list of compressed sensing algorithms, having different benefits and drawbacks that are suitable for different systems/models, and with distinct convergence and complexity constraints, in the following figures we focus on some prominent greedy CS techniques, i. e., OMP, CoSaMP, IHT, NHTP with dictionary, compared to convex optimization based basic pursuit algorithms.

Fig. 4 illustrates the performance of CS algorithms with the proposed dictionary, showing that the basic pursuit method outperforms all the considered greedy algorithms in terms of BER, while the least square estimation (LSE) approach gives the worst performance. However, the curve of the greedy NHTP is really close the BP one, with the advantage of high computational complexity reduction. Indeed, there is a trade-off between performance and computational complexity that we need to take into account, as shown in Sec. IV-C.

Finally, in Fig. 5 we analyze the effect of multipath channel for the NHTP with the proposed dictionary by varying the

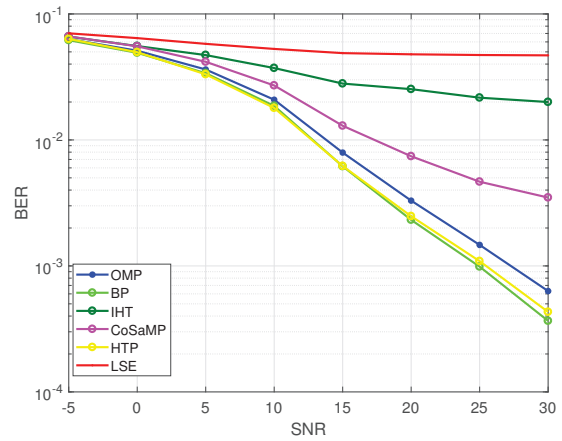
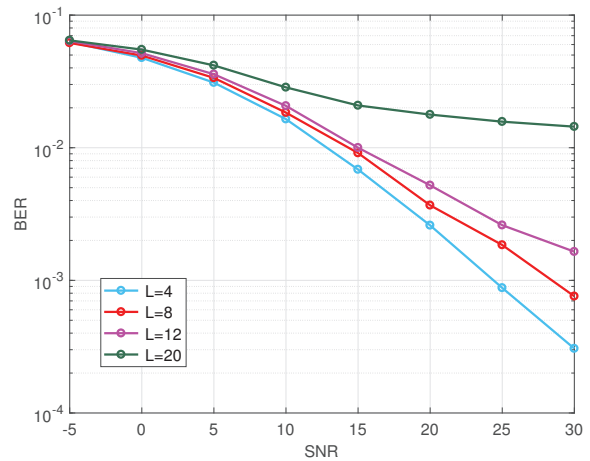

 Fig. 4. Performance of CS techniques with dictionary basis: BER vs SNR (dB) for transmit antennas $N = 100$, receive antenna $R = 1$, number of channel taps $L_t = 6$


Fig. 5. BER vs SNR (dB) for NHTP algorithm with dictionary changing the number of channel taps

number of channel taps. We focus only on this algorithm since it is the best choice in terms of performance complexity trade off, as detailed in the following.

Pilot overhead reduction has been compared among different compressed sensing based algorithms including convex optimization based BP and other greedy algorithms. It has been observed that the performance curve rapidly move down with decreasing pilot interval. The previous work [19] investigated pilot overhead reduction using CS based algorithm by random pilot allocation at different subcarriers for each antenna. They have used random sending matrix with DFT basis for sparse channel representation. In this section, we have assumed the same parameter setting for comparison purposes: transmit antennas = 128, number of subcarriers = 2048, SNR equal to 20 dB, 10 percent pilots overhead, and MSE equal to -25 db. As shown in figure 6, the proposed dictionary based algorithm exhibits better performance with pilot overhead less

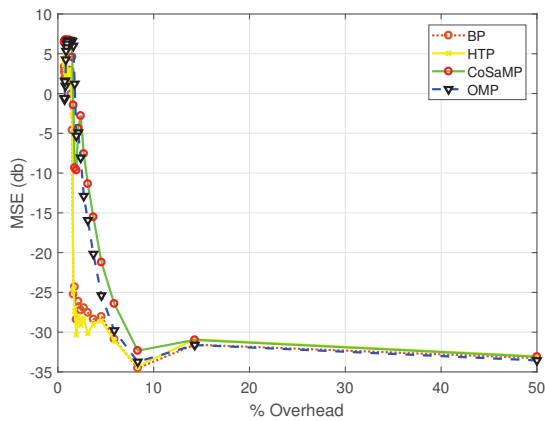


Fig. 6. Percentage pilot overhead

 TABLE II
 COMPUTATIONAL COMPLEXITY

| BP | OMP | CoSaMP | IHT | NHTP |
|----------|------------|----------|----------|----------|
| $O(n^3)$ | $O(s m n)$ | $O(m n)$ | $O(m n)$ | $O(m n)$ |

than 5 percent at -30 db MSE.

C. Complexity Analysis

Tables II and Fig. 7 refer to the computational complexity and the required simulation time, respectively.

Considering Fig. 4, 7 and Table II, the greedy orthogonal matching pursuit and normalized hard thresholding pursuit show the best performance trade off in terms of BER and computational complexity/simulation time. Note that, although OMP shows the lowest simulation time in Fig. 7, the NHTP BER performance in Fig. 4 greatly outperforms the OMP one, with a 2.5 dB gain at 10^{-3} BER. All the simulations are performed on an intel core i5 (7th Generation) architecture.

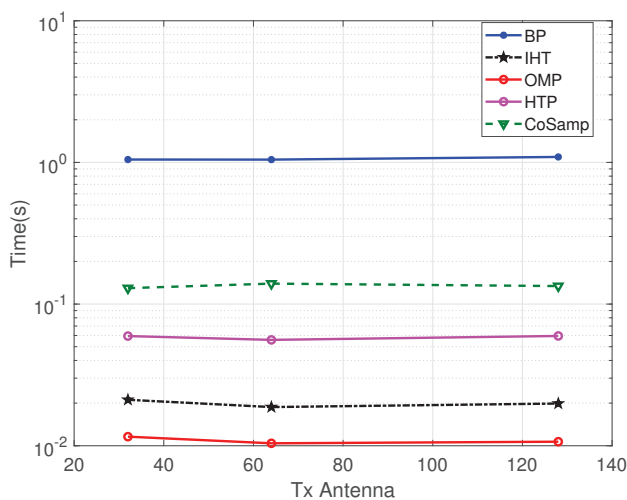


Fig. 7. Simulation time in seconds for CS techniques with dictionary

V. CONCLUSION

We have proposed a CS based channel estimation with a suitable dictionary for downlink beamforming in massive MIMO-OFDM FDD systems. We have exploited CS techniques to reduce training overhead, that is proportional to the sparsity level of the channel. The sparse channel representation is obtained through a proposed dictionary able to self-adapt to the cell characteristics. We have analyzed several CS algorithms in order to select among them the best technique with respect to the proposed dictionary. Numerical results demonstrate that basic pursuit shows the best performance at the cost of a higher complexity, while greedy solutions give good results with a lower complexity and hence shorter training period. Among them, the NHTP approach is the greedy algorithm with the best performance complexity trade-off.

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [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 Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [3] Y. Ding and B. D. Rao, "Compressed downlink channel estimation based on dictionary learning in fdd massive mimo systems," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.
- [4] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [5] X. Rao and V. K. N. Lau, "Distributed compressive csit estimation and feedback for fdd multi-user massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, June 2014.
- [6] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.
- [7] A. N. Uwaechia and N. M. Mahyuddin, "A review on sparse channel estimation in ofdm system using compressed sensing," *IETE Technical Review*, vol. 34, no. 5, pp. 514–531, 2017.
- [8] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [9] M. Masood and T. Y. Al-Naffouri, "Sparse reconstruction using distribution agnostic bayesian matching pursuit," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5298–5309, Nov 2013.
- [10] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [11] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, Feb 2012.
- [12] J. T. D. Needell, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," pp. 301–321 *Anal.* 26 (3), 2009.
- [13] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6202–6216, Dec 2012.
- [14] F. Salahdine, N. Kaabouch, and H. E. Ghazi, "A survey on compressive sensing techniques for cognitive radio networks," *Physical Communication*, vol. 20, pp. 61 – 73, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1874490716300386>
- [15] R. Jani and Y. Sharma, "A review on spectrum sensing for cognitive radio networks," in *Int. J. Adv. Eng. Res. Dev. I (5)*, 2014, pp. 5–8.
- [16] M. D. T. Blumensath, "Iterative hard thresholding for compressed sensing," pp. 265–274 *Anal.* 27 (3), 2009.
- [17] R. V. N. Deanna, J. Tropp, "Greedy signal recovery review," in *IEEE Conf. Signals Syst. Computers*, 2008.
- [18] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," pp. 2543–2563 *Volume 49 Issue 6*, November 2011.
- [19] J. W. Choi, B. Shim, and S. H. Chang, "Downlink pilot reduction for massive mimo systems via compressed sensing," *IEEE Communications Letters*, vol. 19, no. 11, pp. 1889–1892, Nov 2015.