# Low Coherence Compressed Channel Estimation for High Mobility MIMO OFDM Systems

Xiang Ren, Wen Chen, Zijian Wang
Department of Electronic Engineering
Shanghai Jiao Tong University, China
Email: {renx; wenchen; wangzijian1786}@sjtu.edu.cn

***Abstract*—In this paper, a low coherence compressed channel estimation method is proposed for high mobility MIMO OFDM systems. High mobility always causes large Doppler frequency spread which costs large spectrum and time resources to obtain the accurate channel state information (CSI). As numerous recent experimental studies have shown that high mobility broadband wireless channels tend to have some inherent sparsity, compressed sensing (CS) has been introduced to utilize the inherent sparsity and reduce the CSI estimation complexity. In this paper, the coherence of CS is studied and we prove that lower coherence leads to better CS performance. An iterative algorithm is proposed to reduce the coherence by designing pilots with the known channel model before transmitted. Numerical results confirm that the proposed method has satisfied channel estimation performance in high mobility environments.**

***Index Terms*—Compressed sensing, channel estimation, low coherence, high mobility, MIMO OFDM.**

## I. Introduction

To achieve high data rates and high spectral efficiency, multiple input multiple output (MIMO) and orthogonal frequency division multiplexing (OFDM) techniques have been considered to be the the most promising technologies for the 4G wireless communication systems [1]. In high mobility environments, the effects of Doppler frequency spread destroy the orthogonality among subcarriers in OFDM, and lead to intercarrier interference (ICI) at the receiver. Channel estimation in high mobility environments has already been considered in a number of recent papers. In [2]-[4], different kinds of methods are proposed to overcome the ICI caused by high mobility. However, these methods are based on the implicit assumption of a rich underlying multipath environment and focused on decreasing the influences caused by ICI. They introduced several complicated iterative algorithms at the receiver part and estimated CSI with costing many iterations.

Recently, growing experimental evidences have shown that the high mobility channels in broadband wireless communication systems tend to exhibit a sparse structure at some high dimensional signal space such as the delay-Doppler domain, and can be characterized with significantly few parameters in those domains. To utilize the inherent channel sparsity, many researchers have studied the application of compressed sensing (CS) methods in the doubly-selective channel [5][6] which well reflects the natures of the high mobility channel. However, these methods have seldom considered the coherence of CS, which degrades the CS reconstruction performance directly. Fundamental research [7] shows that the coherence plays an important role in CS, and [7] made a conclusion that lower coherence between the measurement matrix and the dictionary matrix leads to better CS performance. Therefore, how to reduce the coherence in a certain environment, e.g. high mobility environment in this paper, is a very interesting and valuable problem.

In this paper, a low coherence compressed channel estimation method is proposed for a MIMO OFDM system in the high mobility environment, in which we study the coherence between the pilot matrix and the channel model dictionary matrix. An low coherence pilot design algorithm is proposed to reduce the coherence with the fixed channel model. Numerical results show that the proposed method has satisfied estimation performance in high mobility environments.

Through out this paper, $diag(\cdot)$ denotes a diagonal matrix, $\lceil x \rceil$ denotes the maximum integer that is no larger than $x$, $\|\cdot\|_{\ell_0}$ counts the number of nonzero entries in a matrix, $\|\cdot\|_{\ell_2}$ is the Euclidean norm, $(\cdot)^T$ denotes the transposition of a matrix, $\otimes$ denotes the Kronecker product, and $\mathbf{a} = vec(\mathbf{A})$ denotes the vector obtained by stacking columns of $\mathbf{A}$.

## II. System Model

In this section, we consider a MIMO OFDM system with $N_t$ transmit antennas and $N_r$ receive antennas in a high mobility environment. And we suppose that there are $K$ subcarriers in each OFDM symbol. In the $n$th OFDM symbol, the information signal $X_t^n(k)$ is transmitted over the frequency $k$ at the $t$th transmit antenna, in which $n = 1, ..., N$, $k = 1, ..., K$ is the subcarrier, and $t = 1, ..., N_T$ is the transmit antenna. Each transmit antenna performs the inverse discrete Fourier transform (IDFT) and sends independent OFDM symbols. After the parallel to serial module, the cyclic prefix (CP) is inserted into the transmitted signals to avoid the intersymbol interference (ISI). Let $H_{rt}^n(k)$ be the channel state information (CSI) between the $t$th transmit antenna and the $r$th receive antenna at the $k$th subcarrier, for $r = 1, ..., N_R$. Since there are $N_T$ transmit antennas emitting different signals, the received signal $Y_r^n(k)$ is the sum of $N_T$ transmitted signals passing all subchannels.

In this paper, pilot-assisted estimation is used to reduce the estimation complexity and ensure the CSI accuracy. Assume

This work is supported by the National 973 Project #2012CB316106, and by NSF China #61161130529.

that there are $P$ pilots placed at subcarriers $k_1, k_2, ..., k_P$ at each transmit antenna, and $P \leq K$. Let $\mathbf{H}_{rt}^n = [H_{rt}^n(k_1), H_{rt}^n(k_2), ..., H_{rt}^n(k_P)]^T$ be the high mobility channel frequency response vector from the $t$th transmit antenna to the $r$th receive antenna at the pilot subcarriers, and let $W_r^n(k)$ be the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma_r^2$. Then the received pilot vector at the $r$th receive antenna can be represented as the matrix form:

$$\mathbf{Y}_r^n = \sum_{t=1}^{N_T} \mathbf{X}_t^n \mathbf{H}_{rt}^n + \sum_{t=1}^{N_T} \mathbf{H}_{rt_{ICI}}^n \mathbf{X}_{t_{vec}}^n + \mathbf{W}_r^n, \quad (1)$$

$$= \sum_{t=1}^{N_T} \mathbf{X}_t^n \mathbf{H}_{rt}^n + \mathbf{N}_r^n, \quad (2)$$

where $\mathbf{Y}_r^n$ is the received pilot vector of the $r$th receive antenna, $\mathbf{X}_t^n = diag([X_t^n(k_1), X_t^n(k_2), ..., X_t^n(k_P)]^T)$ is the transmitted pilot matrix from the $t$th transmit antenna at the pilot subcarriers, and $\mathbf{X}_{t_{vec}}^n = vec(\mathbf{X}_t^n)$. $\mathbf{H}_{rt_{ICI}}^n$ is a $P \times P$ channel matrix with zero diagonal entries and whose off-diagonal entries represent the ICI caused by time-variant channels. $\mathbf{N}_r^n = [N_r^n(k_1), N_r^n(k_2), ..., N_r^n(k_P)]^T$ is the sum vector of $\mathbf{W}_r^n$ and the ICI part (here we consider the ICI as noise so that we can use CS easily).

### *A. High Mobility Channels*

In high mobility environment, e.g. 350km/h or more, the wireless channel changes rapidly and causes large Doppler frequency spread leading to time selective fading. In addition, frequency selective fading caused by the multipath effect is also unavoidable. So the high mobility channel can be considered as a time and frequency doubly-selective channel [5]. In this paper, we assume that the channel between a transmit and a receive antenna is independent and identically distributed (i.i.d.). Let $v_{\max}$ be the maximum Doppler frequency spread, $\tau_{\max}$ be the delay spread, $T_d$ be the packet duration and $W$ be the bandwidth. Therefore, we can give the representation of the high mobility channel between the $t$th transmit antenna and the $r$th receive antenna in delay-Doppler domain as

$$H_{rt}(n, f) = \sum_{l=0}^{L-1} \sum_{m=-M}^{M} \beta_{l,m,r,t} e^{j2\pi \frac{m}{T_d} n} e^{-j2\pi \frac{l}{W} f}, \quad (3)$$

where $L = \lceil W\tau_{\max} + 1 \rceil$ represents the maximum number of resolvable delays and $M = \lceil T_d v_{\max}/2 \rceil$ represents the maximum Doppler shifts within the delay-Doppler spread of the high mobility channel, $f$ is the subcarrier frequency and $n$ is the time slot. Let $\mathbf{u}_f = \left[1, e^{-j2\pi \frac{1}{W} f}, ..., e^{-j2\pi \frac{(L-1)}{W} f}\right]$ and $\mathbf{u}_n = \left[e^{j2\pi \frac{-M}{T_d} n}, e^{j2\pi \frac{(-M+1)}{T_d} n}, ..., e^{j2\pi \frac{M}{T_d} n}\right]$. Then we have

$$H_{rt}(n, f) = \mathbf{u}_f \mathbf{B}_{rt} \mathbf{u}_n^T = (\mathbf{u}_n \otimes \mathbf{u}_f)\, \mathbf{b}_{rt}, \quad (4)$$

where $\mathbf{B}_{rt}$ is an $L \times (2M+1)$ channel coefficient matrix in the delay-Doppler domain of the high mobility channel between the $r$th receive antenna and the $t$th transmit antenna, i.e.,

$$\mathbf{B}_{rt} = \begin{bmatrix} \beta_{0,-M,r,t} & \cdots & \beta_{0,M,r,t} \\ \vdots & \ddots & \vdots \\ \beta_{L-1,-M,r,t} & \cdots & \beta_{L-1,M,r,t} \end{bmatrix}. \quad (5)$$

In this paper we assume that the coefficients are constant. Let $\mathbf{b}_{rt} \triangleq vec(\mathbf{B}_{rt})$ be the stacking vector of the channel coefficient matrix. As $M$ increases with the speed of system, $\mathbf{B}_{rt}$ can still model the high mobility precisely with more coefficients. Coefficients $\{\beta_{l,m,r,t}\}$ are approximately equal to the sum of the complex gains of all physical paths at the sampling points in the delay-Doppler space. Thereby, the task of estimating the high mobility channel in frequency domain is reduced to estimating the channel coefficients $\{\beta_{l,m,r,t}\}$, which are fortunately sparse in practice [5][6].

### *B. Channel Estimation*

Let $\psi^n = [\mathbf{u}_n \otimes \mathbf{u}_{k_1}, \mathbf{u}_n \otimes \mathbf{u}_{k_2}, \cdots, \mathbf{u}_n \otimes \mathbf{u}_{k_P}]^T$ be the $P \times L(2M+1)$ subchannel dictionary matrix in the $n$th OFDM symbol, in which $P < L(2M+1)$. Substitute the high mobility channel model (4) into (2) to derive the matrix form

$$\mathbf{Y}_r^n = \sum_{t=1}^{N_T} \mathbf{X}_t^n \mathbf{H}_{rt}^n + \mathbf{N}_r^n = \mathbf{X}^n \mathbf{\Psi}^n \mathbf{b}_r + \mathbf{N}_r^n, \quad (6)$$

where $\mathbf{X}^n = \left[\mathbf{X}_1^n, \cdots, \mathbf{X}_{N_T}^n\right]$ is the $P \times N_T P$ pilot matrix of all transmit antennas, $\mathbf{b}_r = \left[\mathbf{b}_{r1}, \cdots, \mathbf{b}_{rN_T}\right]^T$ is the $N_T L(2M+1) \times 1$ channel coefficient matrix, and $\mathbf{\Psi}^n = diag([\psi^n, \cdots, \psi^n]^T)$ is the $N_T P \times N_T L(2M+1)$ block-diagonal channel model dictionary matrix. Here we define the dominant non-zero coefficients in $\mathbf{b}_r$ as those contributing significant channel coefficients, i.e. $|\beta_{l,m,r,t}|^2 > \gamma$, where $\gamma$ is an appropriately chosen threshold whose value depends upon the design accuracy. For some appropriate threshold $\gamma > 0$, the channel is said to be $S$-sparse, if $\|\mathbf{b}_r\|_{\ell_0} = S < P \ll N_T L(2M+1)$. Recent researchers in [5] and [6] have shown that the doubly-selective channels have a certain inherent sparsity and can be represented sparsely in the delay-Doppler domain, which means that $\mathbf{b}_r$ is sparse in practice. Therefore CS is introduced in this paper to utilize the inherent channel sparsity.

## III. Coherence Optimized Compressed Estimation

Different to the conventional channel estimation method based on linear estimators such as least-squares (LS), CS can utilize the inherent sparsity of the wireless channel, which is known as the compressed channel estimation [5]-[7]. For better explanation, here we review CS simply. Let signal $\mathbf{x} \in \mathbb{R}^m$ be an $m \times 1$ vector. Assuming that signal $\mathbf{x}$ has the sparsity of $S$ under the dictionary basis $\mathbf{D} \in \mathbb{R}^{m \times U}$ ($m < U$), that is $\mathbf{x} = \mathbf{Da}$, and there are only $S$ non-zero elements in vector $\mathbf{a}$ with $\|\mathbf{a}\|_{\ell_0} = S \ll m < U$. By using a measurement matrix $\mathbf{P} \in \mathbb{R}^{p \times m}$, which is not related to the dictionary basis $\mathbf{D}$, it projects the signal $\mathbf{x}$ to $\mathbf{y}$: $\mathbf{y} = \mathbf{Px} = \mathbf{PDa}$, where $\mathbf{PD}$ should satisfy the Restricted Isometric Property (RIP) [7]. Then the CS reconstruction methods such as the Basis Pursuit (BP) [10] and Orthogonal Matching Pursuit (OMP) [12] can be used to reconstruct $\mathbf{x}$ by $\mathbf{y}$.

### *A. Coherence*

Here we recall the definition of the coherence [7][15], which is a fundamental concept in CS.

*Definition 1:* For a matrix $\mathbf{M}$ with the $i$th column of $\mathbf{d}_i$, its coherence is defined as the largest absolute value of the normalized inner product between different columns in $\mathbf{M}$ and can be written as follows:

$$\mu\{\mathbf{M}\} = \max_{i \neq j} \frac{\left|\mathbf{d}_i^T \mathbf{d}_j\right|}{\|\mathbf{d}_i\| \cdot \|\mathbf{d}_j\|}. \tag{7}$$

The coherence provides a measure of the worst similarity between the dictionary columns. It is a value that exposes the matrix's vulnerability, as two closely related columns may confuse the reconstruction techniques. The concept of the coherence of the dictionary $\mathbf{D}$ plays a major role in CS. Previous work [14]-[16] established that both BP and orthogonal greedy algorithms (OGA) (including OMP) approaches can be used if the condition in the following theorem is satisfied.

*Theorem 1:* For a matrix $\mathbf{D}$, if some representation of the signal $\mathbf{x} = \mathbf{Da}$ satisfies

$$S = \|\mathbf{a}\|_{\ell_0} < \frac{1}{2}\left(1 + \frac{1}{\mu\{\mathbf{D}\}}\right), \tag{8}$$

then a) $\mathbf{a}$ is the unique sparsest such representation of $\mathbf{x}$; b) the deviation of the reconstructed $\hat{\mathbf{a}}$ from $\mathbf{a}$ by BP or OGA can be bounded by

$$\|\hat{\mathbf{a}} - \mathbf{a}\|_{\ell_2}^2 \leq \frac{\epsilon^2}{1 - \mu\{\mathbf{D}\}(2S-1)}, \tag{9}$$

for some constant $\epsilon > 0$.

*Proof:* The proof is given in [11]. ■

Theorem 1 implies that if the sparsity $S$ determined by the system is constant, which in our system is determined by the channel model, then the sparse signal $\mathbf{a}$ can be recovered with a given error bound related to $\mu$. It is easy to found that lower $\mu\{\mathbf{D}\}$ will result in higher upperbound of $S$, which means that $\mathbf{a}$ can contain more nonzero atoms, and introduces lower reconstruction error bound leading to better CS recovery performance.

Based on the previous point, suppose that the measurement matrix $\mathbf{P}$ has been chosen independent of the dictionary basis $\mathbf{D}$. Then we can solve the vector $\mathbf{a}$ in $\mathbf{y} = \mathbf{Px} = \mathbf{PDa}$ by BP or OMP as described in the following corollary.

*Corollary 1:* For a dictionary matrix $\mathbf{D}$ and measurement matrix $\mathbf{P}$, assume that $\mathbf{PD}$ satisfies the RIP, if the representation $\mathbf{y} = \mathbf{Px} = \mathbf{PDa}$ satisfies the requirement

$$S = \|\mathbf{a}\|_{\ell_0} < \frac{1}{2}\left(1 + \frac{1}{\mu\{\mathbf{PD}\}}\right), \tag{10}$$

then a) $\mathbf{a}$ is the unique sparsest such representation of $\mathbf{x}$; b) the deviation of the reconstructed $\hat{\mathbf{a}}$ from $\mathbf{a}$ by BP or OGA can be bounded by

$$\|\hat{\mathbf{a}} - \mathbf{a}\|_{\ell_2}^2 \leq \frac{\epsilon^2}{1 - \mu\{\mathbf{PD}\}(2S-1)}, \tag{11}$$

for some constant $\epsilon > 0$.

*Proof:* The proof is similar to the proof of Theorem 1 by proving that $\mathbf{PD}$ satisfies the RIP. ■

Corollary 1 implies that if $\mathbf{P}$ is designed with a fixed $\mathbf{D}$ such that $\mu\{\mathbf{PD}\}$ is as small as possible, a large number of candidate signals are able to reside under the umbrella of successful CS behavior and lead to better CS performance.

While these conclusions are true from a worst-case standpoint, it turns out that the coherence as defined in Definition 1 does not justify the actual behavior of the practical system. Considering the performance of the practical CS reconstruction, an average measure of coherence is more likely to describe the true behavior. So we consider the average coherence [14] in our system to reflect the practical system behavior. The previous theorem and corollary are still valid to the average coherence $\mu_\delta$ as described in Definition 2.

*Definition 2:* For a matrix $\mathbf{M}$ with the $i$th column of $\mathbf{d}_i$, its average coherence is defined as the average of all absolute inner products between the different normalized columns in $\mathbf{M}$ that are beyond $\delta$, where $\delta$ is a threshold and $0 < \delta < 1$. Put formally

$$\mu_\delta\{\mathbf{M}\} = \frac{\sum\limits_{i \neq j} (|g_{ij}| \geq \delta) \cdot |g_{ij}|}{\sum\limits_{i \neq j} (|g_{ij}| \geq \delta)}, \tag{12}$$

where $g_{ij} = \tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_j$ and $\tilde{\mathbf{d}}_i = \mathbf{d}_i / \|\mathbf{d}_i\|_{\ell_2}$.

### B. Coherence Optimization

As we have already known that lower $\mu_\delta$ leads to better CS performance, we are going to reduce the coherence $\mu_\delta\{\mathbf{PD}\}$ in our system to get better estimation performance. In this paper, we design the pilot entries of the pilot matrix $\mathbf{X}^n$ (i.e. the measurement matrix in CS) with given pilot locations and known channel model dictionary $\mathbf{\Psi}^n$ (i.e. the dictionary basis matrix in CS) to minimize the average coherence $\mu_\delta\{\mathbf{X}^n \mathbf{\Psi}^n\}$. The influence of the pilot location is not considered in this paper since considering both of the two factors is a difficult joint optimization problem, which will be discussed in the future work. Here we assume that the pilot locations are fixed and are the same for each antenna. Therefore, the problem can be formulated as following optimization problem

$$\min_{\mathbf{X}^n} \mu_\delta\{\mathbf{X}^n \mathbf{\Psi}^n\}. \tag{13}$$

Hence the coherence optimized pilot matrix is given as

$$\hat{\mathbf{X}}^n = \arg\min_{\mathbf{X}^n} \mu_\delta\{\mathbf{X}^n \mathbf{\Psi}^n\}. \tag{14}$$

### C. Low Coherence Pilot Design Algorithm

In this subsection, we propose an algorithm to get the optimized pilot $\hat{\mathbf{X}}^n$ in (14). According to the system model (6) and Definition 2, the average coherence can be calculated as $\mu_\delta\{\mathbf{X}^n \mathbf{\Psi}^n\} = \mu_\delta\left\{\left[\mathbf{X}_1^n \psi^n, \cdots, \mathbf{X}_{N_T}^n \psi^n\right]\right\}$ with columns $\mathbf{d}_i$, and $1 \leq i \leq N_T L(2M+1)$. As $\mathbf{X}^n$ consists of $N_T$ diagonal matrices, it is difficult to get the optimal $\hat{\mathbf{X}}^n$ directly. So we assume that the pilots at each transmit antenna are the same, which means $\mathbf{X}_1^n = \cdots = \mathbf{X}_{N_T}^n$. From Definition 2, we can know that the average coherence only care about the columns with $|g_{ij}| \geq \delta$. Therefore the problem of minimizing $\mu_\delta\{\mathbf{X}^n \mathbf{\Psi}^n\}$ is equivalent to minimizing each $\mu_\delta\{\mathbf{X}_t^n \psi^n\}$ respectively. An iterative algorithm is proposed here to reduce $\mu_\delta\{\mathbf{X}_t^n \psi^n\}$ and get the optimized pilot matrix $\hat{\mathbf{X}}_t^n$ with fixed subchannel basis model $\psi^n$. The algorithm is given in Algorithm 1.

**Algorithm 1 :** Low Coherence Pilot Design Algorithm

1: Initialization: Set the initial pilot matrix $\mathbf{X}_t^n$ as the complex Gaussian random variables sequence with zero mean and variance $\frac{1}{N_T}$; set the threshold $\delta$, the shrink factor $\lambda$ and the maximum iteration time $Iter$.
2: Obtain the effective dictionary $\hat{\mathbf{D}}$ by normalizing each column in the matrix $\mathbf{X}_t^n \psi^n$.
3: Compute Gram matrix: $\mathbf{G} = \hat{\mathbf{D}}^T \hat{\mathbf{D}}$.
4: Update the Gram matrix and obtain $\hat{\mathbf{G}}$ with

$$\hat{g}_{ij} = \begin{cases} \lambda g_{ij}, & |g_{ij}| \geq \delta, \\ \lambda\delta \cdot sign(g_{ij}), & \delta \geq |g_{ij}| \geq \lambda\delta, \\ g_{ij}, & \lambda\delta \geq |g_{ij}|, \end{cases}$$

where $sign(x) = 1$ if $x \geq 0$, and $-1$ otherwise.
5: Apply singular value decomposition (SVD) to $\hat{\mathbf{G}}$ with the diagonal entries in decreasing order. Reduce the rank of $\hat{\mathbf{G}}$ to $P$ by keeping only the first $P$ diagonal entries of the diagonal matrix.
6: Build the square root of $\hat{\mathbf{G}}$: $\mathbf{S}^T\mathbf{S} = \hat{\mathbf{G}}$.
7: Find the new diagonal $\mathbf{X}_t^n$ that minimizes the error $\|\mathbf{S} - \mathbf{X}_t^n \psi^n\|_F^2$ and goto $Step$ 2 if the iteration is less than $Iter$.
8: Output the optimized pilot matrix $\hat{\mathbf{X}}_t^n$ after $Iter$ iterations.

After $Iter$ iterations, we can get the optimized pilot matrix $\hat{\mathbf{X}}_t^n$ and hence the optimized $\hat{\mathbf{X}}^n$ is obtained since the pilots at each transmit antenna are the same. Therefore, $\hat{\mathbf{X}}^n$ can be used to estimate the CSI by CS estimators. The numerical results in the next section will show that Algorithm 1 will converge to a stationary pilot matrix. As afore algorithm analysis is not restricted to any specific channel, Algorithm 1 is also effective to other channels with known channel model dictionary.

### *D. Low Coherence Compressed Channel Estimation*

The flow chart of the low coherence compressed channel estimation in practical system is shown in Fig. 1.

Firstly, the high mobility wireless channel is modeled as (4) and known at both the transmit and receive sides. $\mathbf{\Psi}^n$ reflects the channel properties and the mobility is represented by the Doppler frequency shifts. In this paper, we assume that the channel coefficients $\mathbf{b}$ is sparse in the delay-Doppler domain. Secondly, the random pilot matrix $\mathbf{X}^n$ is initialized at the transmit side as Algorithm 1. Since $\mathbf{X}^n$ does not always have the lowest coherence with the high mobility channel model $\mathbf{\Psi}^n$, Algorithm 1 is operated to get the coherence reduced pilot $\hat{\mathbf{X}}^n$. And $\hat{\mathbf{X}}^n$ is also known at both sides for pilot-assisted channel estimation. After $\hat{\mathbf{X}}^n$ is transmitted and passed the high mobility wireless channel, $\mathbf{Y}^n$ is received at the receive side for all antennas with AWG noise. Then, CS reconstruction algorithms such as BP and OMP can reconstruct the estimated channel coefficients $\hat{\mathbf{b}}$ by $\mathbf{Y}^n = \hat{\mathbf{X}}^n\mathbf{\Psi}^n\mathbf{b} + \mathbf{N}^n$. Finally, the reconstructed CSI is obtained by $\hat{\mathbf{H}}^n = \mathbf{\Psi}^n\hat{\mathbf{b}}$ at the receive side. The mean square error (MSE) performance between the reconstructed $\hat{\mathbf{H}}^n$ and the actual channel $\mathbf{H}^n$ is introduced in this paper to measure the performances of different channel estimators and pilots.

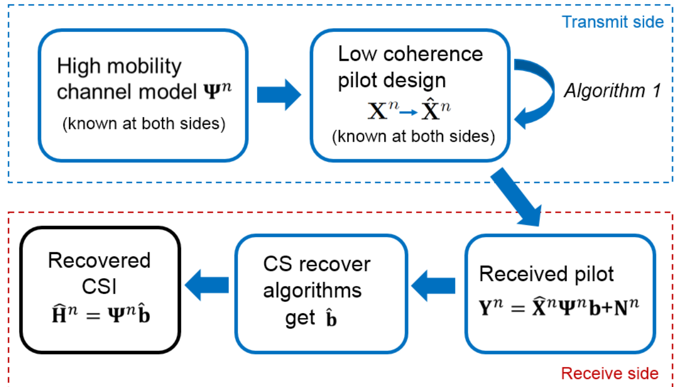


Fig. 1. Flow chart of the low coherence compressed channel estimation

## IV. Numerical Results

In this section, under the high mobility environment, we compare the MSE performances of the LS estimator, the linear minimum mean square error (LMMSE) [8] estimator and the best linear unbiased estimator (BLUE) [9] with two compressed channel estimators BP [10] and OMP [12]. LS, LMMSE and BLUE estimators are all considering the ICI as described in [8]. BP and OMP are both based on the random pilot and the optimal pilot. Both BP and OMP are shown to benefit from the coherence optimized pilot.

Here we consider a MIMO OFDM system with 4 transmit antennas and 4 receive antennas in the high mobility environment. Assumed that there are 512 subcarriers in each OFDM system and $12.5\%$ are pilot subcarriers. All pilots are allocated with equal intervals. The bandwidth is 5MHz, the packet duration is $T_d = 0.5$ms and the carrier frequency is operated at $f_c = 2.6$GHz, according to the LTE standard. The additive noise is a Gaussian and white random process. The high mobility channel is modeled as (4). The maximum delay spread is $\tau_{max} = 40\mu s$ and the maximum Doppler frequency spread is $v_{max} = 1.204$KHz, which means that the maximum velocity of the receiver is 500km/h. In our experiment, we assumed that there are only $10\%$ of the channel coefficients are nonzero, which means that we only care about the largest $10\%$ coefficients as dominant coefficients and set others to zero. Simulations are carried out with the random pilot matrix $\mathbf{X}_t^n$ at each transmit antenna and the optimized pilot matrix $\hat{\mathbf{X}}_t^n$. $\mathbf{X}_t^n$ consists of the complex Gaussian random variables sequence with zero mean and variance $\frac{1}{N_T}$ as Algorithm 1. $\hat{\mathbf{X}}_t^n$ is generated by Algorithm 1 with $\delta = 0.2$, $\lambda = 0.8$, $Iter = 8$.

Fig. 2 presents the comparison of the MSE performances of LS, LMMSE and BLUE channel estimators with BP and OMP compressed channel estimators versus the SNR at 500km/h. It can be observed that the CS channel estimators significantly improves the MSE performances by utilizing the inherent sparsity of high mobility channels. On the other hand, LS and LMMSE need more pilots to obtain enough channel informations and reconstruct the CSI accurately. Therefore, CS estimators save more spectrum resources than linear estimators and get better performance. As expected, both BP and OMP with optimized pilots (BP-OP and OMP-OP with $Iter = 8$) offer better performances than those with random pilots. It

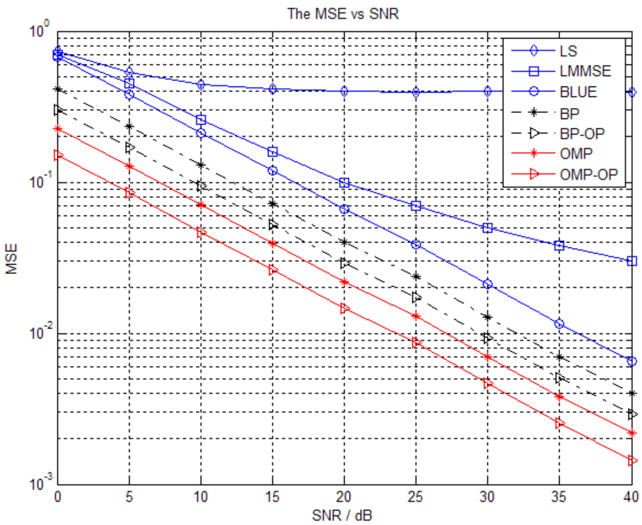


Fig. 2. MSE performance of estimators and BP OMP with random and optimized pilot matrix in a $4 \times 4$ MIMO OFDM system at 500km/h. 512 subcarriers in each OFDM system, and 12.5% are pilot subcarriers.

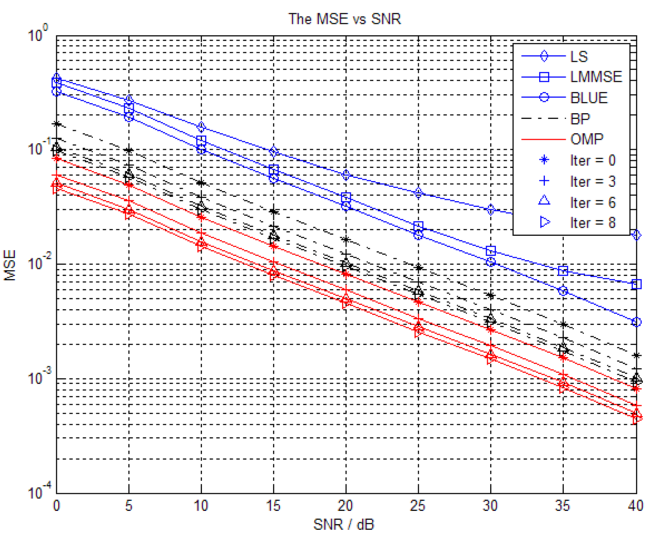


Fig. 3. MSE performance of estimators and BP OMP with random and different optimized pilot matrices in a $4\times4$ MIMO OFDM system at 100km/h. 512 subcarriers in each OFDM system, and 12.5% are pilot subcarriers.

means that the proposed pilot design algorithm reduces the coherence between the pilot matrix and the high mobility channel dictionary matrix effectively and improves the system performance directly.

Fig. 3 presents the comparison of the MSE performances of different estimators at 100km/h, that generates the maximum Doppler shift frequency spread as $v_{max} = 241$Hz. As can be seen, all estimators get better performances at lower speed because there are more sparsity in channel coefficients, but LS and LMMSE still need more pilots to get the accurate CSI. The MSE performances of BP and OMP estimators with optimized pilots by different iterations are given in this figure. $Iter$ is the iterations of Algorithm 1 and $Iter = 0$ means the initial random pilot without optimization. It can be found that the gains become smaller and smaller with $Iter$ growing, which shows the convergence property of Algorithm 1. Numerical results show that the optimized pilot with $Iter = 8$ is good enough for our system. As can be seen, the coherence optimized pilots also improve the performances of both BP and OMP estimators.

## V. Conclusion

In this paper a compressed channel estimation method with coherence optimized pilot is proposed for MIMO OFDM systems in high mobility environments. The proposed algorithm reduces the the coherence between the pilot matrix and the high mobility channel model effectively. Numerical results show that the proposed method get better performance than the conventional estimators with the same number of pilots in high mobility environments and also show the convergence property. In addition, as the optimized pilots can be operated to the known channel model and known at both the transmit side and the receive side before transmitted, the proposed scheme does not cost more time or frequency resources than the conventional pilot-assisted channel estimation methods and can be realized easily in practical communication systems.

## References

[1] W. U. Bajwa, A. M. Sayeed, R. Nowak. "Sparse multipath channels: Modeling and estimation," *IEEE Digital Signal Processing Workshop*, 2009, 1:1-6.

[2] E. Panayirci, H. Senol, and H. V. Poor, "Joint channel estimation, equalization and data detection for OFDM systems in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4225-4238, Aug. 2010

[3] T. Y. Al-Naffouri, K. M. Z. Islam, N. Al-Dhahir, and S. Lu, "A model reduction approach for OFDM channel estimation under high mobility conditions," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2181-2193, Apr. 2010

[4] N. Aboutora, W. Hardjawana, B. Vucetic, "A new iterative doppler-assisted channel estimation joint with parallel ICI cancellation for high-mobility MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1577-1589, May. 2012

[5] W. U. Bajwa, J. Haupt, A. M. Sayeed, R. Nowak, "Compressed channel sensing: a new approach to estimating sparse multipath channels," *IEEE Signal Processing Mag.*, vol. 98, no. 6, pp. 1058-1076, June. 2010

[6] W. U. Bajwa, A. M. Sayeed, R. Nowak, "Learning sparse doubly-selective channels," *Conference on Communication Control and Computing*, 46th, pp.575-582. Sept. 2008

[7] J. C. Emmanuel, and B. W. Michael, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, Mar. 2008.

[8] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol.55, pp. 2226-2238, 2007.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey, USA, 1993.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decompostiion by basis pursuit," *SIAM Review*, vol.43, pp. 129-159, 2001.

[11] D.L.Donoho, M.Elad, and V.Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no.1, pp. 6-18, Jan. 2006.

[12] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993

[13] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol.59, no.8, pp.1207-1223, 2006.

[14] M. Elad, "Optimized projections for compressed sensing," *IEEE Transcation on Signal Processing*, vol.55, no.12, pp.5695-5702, Dec. 2007.

[15] R. Gribonval and M. Nielsen, "Sparse representationsin unions of bases," *IEEE Trans. Inf. Theory, vol.* 49, no.12, pp.3320-3325, Dec. 2004

[16] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol.50, no.10, pp.2231-2242, Oct. 2004