Deterministic Selection of Pilot Tones for Compressive Estimation of MIMO-OFDM Channels

Zahra Shakeri and Waheed U. Bajwa
Dept. of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854
Email: {zahra.shakeri, waheed.bajwa}@rutgers.edu

***Abstract*—This paper addresses the problem of channel estimation in multiple-input and multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems for the case when the underlying multipath channel is approximately sparse in the angle-delay domain. To this end, an algorithm for deterministic selection of pilot tones over which (either random or deterministic) training vectors can be transmitted is proposed. In addition, conditions are derived for the minimum number of pilot tones needed by the proposed algorithm to ensure reliable estimation of the underlying channel using reconstruction methods from the compressed sensing literature. Finally, effectiveness of the proposed algorithm is demonstrated through numerical experiments.**

***Index Terms*—Channel estimation, compressed sensing, MIMO-OFDM systems**

I. Introduction

Orthogonal frequency division multiplexing (OFDM) is a multi-carrier modulation scheme that achieves high data rates in wireless systems. Benefits of OFDM are increased symbol duration from dividing the channel into overlapping orthogonal narrowband subchannels, thereby eliminating inter-symbol interference, and efficient implementation using fast Fourier transform (FFT) and inverse FFT. In multiple-input and multiple-output (MIMO) systems, the transmitter and/or receiver are equipped with multiple antennas, which creates multiple parallel data streams and enhances system reliability.

For coherent signal detection and low bit error rates in MIMO-OFDM systems, periodic channel estimation is essential at the receiver. Blind channel estimation methods use the statistics of unknown data in order to estimate the underlying channel from the channel output [1]–[5]. Semi-blind approaches enhance the performance of blind methods by transmitting a small amount of training data known to the receiver [6], [7]. Although these techniques have high spectral efficiency, they often require a constant channel over a large number of symbols and tend to suffer from high complexity. In training-based methods, training data known to the receiver are transmitted for channel estimation. OFDM subchannel selection (pilot tones) for transmitting training symbols and training sequence design are key aspects of this problem for minimum resource utilization and reliable estimation. Wireless industry standards require specification of deterministic pilot tones in wireless systems. Therefore, deterministic selection of pilot tones for training-based MIMO-OFDM channel estimation—the focus of this paper—is an important problem.

This work is supported by NPRP grant # 6-070-2-024 from the Qatar National Research Fund (a member of Qatar Foundation).

Most of the current training-based methods consider rich-scattering channels [8]–[17] and perform channel estimation using criteria such as maximum likelihood or minimum mean square error. In particular, in [8], a scheme is proposed for optimal selection of pilot tones and training sequence to minimize the mean square error in MIMO-OFDM systems. The selected pilot sequences are equispaced in frequency, equipowered, and phase-shift orthogonal, which result in low channel estimation error. But such techniques overuse channel resources, as they fail to exploit the fact that in wideband MIMO-OFDM systems, multipath wireless channels are usually approximately sparse in the delay-angle domain [18].

One of the main contributions of this work is providing an algorithm for deterministic pilot tone selection to estimate approximately sparse multipath channels in MIMO-OFDM systems. We also provide a lower bound on the number of selected pilot tones, N_{tr}, to ensure reliable estimation of the channel using the Dantzig selector (DS) [19]. We focus on DS for mainly theoretical reasons, since it provides some of the best guarantees in the literature for estimation of approximately sparse signals in the presence of Gaussian noise. In general, however, our selection of pilot tones results in reliable estimation of channels using any state-of-the-art reconstruction method. This is established in the paper using extensive numerical simulations.

In terms of relation to prior work, some works on compressive channel estimation in MIMO-OFDM systems include [18], [20], [21]. But none of these works address the problem of deterministic selection of pilot tones across all transmit antennas of a MIMO-OFDM system for approximately sparse channels. In [18], the selection of pilot tones and training sequence design are random, whereas [21] addresses this problem using a deterministic approach. We introduce a deterministic pilot tone selection algorithm and design random training sequences for reliable channel recovery in this paper. Note that the random sequence can also be substituted with a deterministic sequence having low autocorrelation. The main difference between our work and [21] is the selection of deterministic pilot tones across transmitting antennas. We select the same set of tones to transmit training data on all transmit antennas, while in [21] the pilot tones vary across the transmit antennas. The implementation of [21] seems an extremely difficult, if not an impossible, task. Finally, several works have studied single-input and single-output (SISO) OFDM compressive channel estimation [22]–[26]. In [26], a deterministic scheme for pilot tone and training sequence selection is provided. Our work is an extension of the pilot tone selection procedure of [26] to MIMO-OFDM systems.

Notational Convention: Lowercase letters are used for scalars and vectors, and uppercase letters are used for matrices. Also, $\|\mathbf{v}\|_1$, $\|\mathbf{v}\|_2$, and $\|\mathbf{v}\|_\infty$ denote the ℓ_1-, ℓ_2- and ℓ_∞-norms of the vector $\mathbf{v}$, respectively, while $\mathbf{v}^T$ and $\mathbf{v}^H$ denote the transpose and Hermitian of the vector $\mathbf{v}$, respectively.

II. System Model

In this section, we describe our channel model in the MIMO-OFDM setting. We consider a broadband multipath channel $\boldsymbol{\mathcal{H}}$ with $N_T \in \mathbb{N}$ and $N_R \in \mathbb{N}$ half-wavelength spaced linear arrays at the transmitter and receiver, respectively, and the time-varying frequency response matrix $\mathbf{H}(t,f) \in \mathbb{C}^{N_R \times N_T}$. We assume N_p physical propagation paths with β_n, $\theta_{R,n}, \theta_{T,n}, \tau_n$, and ν_n being the complex path gain, angle of arrival (AoA), angle of departure (AoD), delay, and Doppler shift associated with the n-th path, respectively. In this case, the frequency response matrix can be expressed as

$$\mathbf{H}(t,f) = \sum_{n=1}^{N_p} \beta_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,n}) e^{-j2\pi\tau_n f} e^{j2\pi\nu_n t}, \quad (1)$$

where $\mathbf{a}_T^H(\theta_T) \in \mathbb{C}^{N_T}$ and $\mathbf{a}_R(\theta_R) \in \mathbb{C}^{N_R}$ denote the array steering and response vectors, respectively, for transmitting/receiving in the θ_T/θ_R direction [18]. We assume the channel is maximally spread in the angle space, $(\theta_{R,n}, \theta_{T,n}) \in [-1/2, 1/2] \times [-1/2, 1/2]$, while $\tau_n \in [0, \tau_{\max}]$, and $\nu_n \in [-\nu_{\max}/2, \nu_{\max}/2]$.

Next, let $\mathbf{x}(t) \in \mathbb{C}^{N_T}$ denote the transmitted signal over $\boldsymbol{\mathcal{H}}$. We assume the symbol duration to be $[0,T]$ (i.e., $\mathbf{x}(t) = 0 \ \forall t \notin [0,T]$) and signal bandwidth to be $[-W/2, W/2]$ (i.e., $\mathbf{X}(f) = 0 \ \forall f \notin [-W/2, W/2]$), resulting in a temporal signal space of dimension $N_0 = WT$. We assume $\boldsymbol{\mathcal{H}}$ to be strictly frequency selective (i.e., $W\tau_{\max} > 1$ and $T\nu_{\max} \ll 1$), which implies $\mathbf{H}(t,f) \approx \mathbf{H}(f)$ since the $e^{j2\pi\nu_n t}$ factor in (1) can be ignored for $t \in [0,T]$. Instead of working directly with (1), which has nonlinear dependencies on the channel parameters, we will work with the canonical linear representation $\widetilde{\mathbf{H}}(f)$ of $\boldsymbol{\mathcal{H}}$, which corresponds to uniform sampling of the angle-delay space at the Nyquist rate $(1/N_T, 1/N_R, 1/W)$. That is [18],

$$\mathbf{H}(f) \approx \widetilde{\mathbf{H}}(f) = \sum_{\ell=0}^{L-1} \mathbf{A}_R \mathbf{H}_v^T(\ell) \mathbf{A}_T^H e^{-j2\pi \frac{\ell}{W} f}. \quad (2)$$

Here, $L \triangleq \lceil W\tau_{\max} \rceil + 1$, $\mathbf{A}_T \in \mathbb{C}^{N_T \times N_T}$ and $\mathbf{A}_R \in \mathbb{C}^{N_R \times N_R}$ are unitary (Fourier) matrices, and $\mathbf{H}_v(\ell) \in \mathbb{C}^{N_T \times N_R} = [\mathbf{h}_{v,1}(\ell) \ldots \mathbf{h}_{v,N_R}(\ell)]$, $\ell = 0, \ldots, L-1$, is termed the channel coefficient matrix associated with the ℓ-th discrete delay. In the following, $H_v(i,k,\ell), k = 1, \ldots, N_T$, will be used to denote the k-th entry of $\mathbf{h}_{v,i}(\ell) \in \mathbb{C}^{N_T}$. Note that the sampled channel coefficients $\{H_v(i,k,\ell)\}$ can be expressed in terms of the physical paths by first dividing the propagation paths into the following subsets:

$$\begin{aligned} S_{R,i} &\triangleq \{n : \theta_{R,n} \in (i/N_R - 1/2N_R, i/N_R + 1/2N_R]\}, \\ S_{T,k} &\triangleq \{n : \theta_{T,n} \in (i/N_T - 1/2N_T, i/N_T + 1/2N_T]\}, \\ S_{\tau,\ell} &\triangleq \{n : \tau_n \in (\ell/W - 1/2W, \ell/W + 1/2W]\}, \end{aligned} \quad (3)$$

and then noting from [18] that

$$\begin{aligned} H_v(i,k,\ell) \approx &\sum_{n \in S_{R,i} \cap S_{T,k} \cap S_{\tau,\ell}} \beta_n f_{N_R}(i/N_R - \theta_{R,n}) \\ &\times f_{N_T}^*(k/N_T - \theta_{T,n}) \operatorname{sinc}(\ell - W\tau_n), \end{aligned} \quad (4)$$

where $f_N(\theta) \triangleq 1/N \sum_{i=0}^{N-1} e^{-j2\pi i\theta}$ and $\operatorname{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$.

Our goal here is to estimate the $N_T N_R L$ channel coefficients $\{H_v(i,k,\ell)\}$ for the case when $\boldsymbol{\mathcal{H}}$ is approximately sparse, or s-compressible, per resolvable AoA.

Definition 1 (s-compressible channel)**.** Let $H_{v,i_{(j)}}, j = 1, \ldots, N_T L$, denote the j-th largest (in magnitude) coefficient associated with i-th resolvable AoA. We say that the channel $\boldsymbol{\mathcal{H}}$ is s-compressible per resolvable AoA if

$$|H_{v,i_{(j)}}| \le r j^{-\frac{1}{s}}, \ i = 1, \ldots, N_R, \quad (5)$$

for parameters $r > 0$ and $s \le 1$.

In training based methods, the transmitted signal can be expressed as the addition of the training and data components: $\mathbf{x}(t) = \mathbf{x}_{tr}(t) + \mathbf{x}_{data}(t)$. Assuming $\mathbf{x}_{tr}(t)$ to be orthogonal to $\mathbf{x}_{data}(t)$, we can focus only on the training component at the receiver for channel estimation. In OFDM systems, the transmitted training signal is of the form

$$\mathbf{x}_{tr}(t) = \sqrt{\frac{\epsilon_{tr}}{N_T}} \sum_{p=1}^{N_{tr}} \widetilde{\mathbf{x}}_{n_p} g(t) e^{j2\pi \frac{n_p}{T} t}, \ 0 \le t \le T, \quad (6)$$

where ϵ_{tr} denotes the total transmit energy for training, N_{tr} is the number of pilot tones, $n_p \in \{0, \ldots, N_0 - 1\}$ are the indices of the pilot tones, and $\{\widetilde{\mathbf{x}}_{n_p} \in \mathbb{R}^{N_T}\}_{p=1}^{N_{tr}}$ is the training sequence with total energy $\sum_{p=1}^{N_{tr}} \|\widetilde{\mathbf{x}}_{n_p}\|_2^2 = N_T$. Finally, $\{g(t) e^{j2\pi \frac{n_p}{T} t}\}_{p=1}^{N_{tr}}$ in (6) denote the OFDM orthogonal waveforms. The output associated with the training signal is

$$\mathbf{y}_{tr}(t) = \boldsymbol{\mathcal{H}}(\mathbf{x}_{tr}(t)) + \mathbf{z}_{tr}(t), \ 0 \le t \le T + \tau_{\max}, \quad (7)$$

where $\mathbf{z}_{tr}(t) \in \mathbb{C}^{N_R}$ is the additive white Gaussian noise (AWGN) at the receiver. Match filtering $\mathbf{y}_{tr}(t)$ at the receiver with the OFDM waveforms, we have

$$\widetilde{\mathbf{y}}_{n_p} = \sqrt{\frac{\epsilon_{tr}}{N_T}} \mathbf{H}_{n_p} \widetilde{\mathbf{x}}_{n_p} + \widetilde{\mathbf{z}}_{n_p}, \ p = 1, \ldots, N_{tr}, \quad (8)$$

where $\mathbf{H}_{n_p} \approx \widetilde{\mathbf{H}}(f)|_{f=\frac{n_p}{T}}$ and $\tilde{\mathbf{z}}_{n_p}$ is AWGN with distribution $\mathcal{N}(\mathbf{0}_{N_R}, \mathbf{I}_{N_R})$. Defining $\{\mathbf{y}_{n_p}^T = \widetilde{\mathbf{y}}_{n_p}^T \mathbf{A}_R^*\}_{p=1}^{N_{tr}}$, we have

$$\mathbf{y}_{n_p}^T = \sqrt{\frac{\epsilon_{tr}}{N_T}} \widetilde{\mathbf{x}}_{n_p}^T \mathbf{A}_T^* \sum_{\ell=1}^{L-1} \mathbf{H}_v(\ell) e^{-j2\pi \frac{\ell}{N_0} n_p} + \mathbf{z}_{n_p}^T, \quad (9)$$

for $p = 1, \ldots, N_{tr}$. Finally, constructing a matrix $\mathbf{Y}$ by stacking $\{\mathbf{y}_{n_p}^T\}$ as rows, we obtain the standard linear form

$$\mathbf{Y} = \sqrt{\frac{\epsilon_{tr}}{N_T}}\mathbf{X}\mathbf{H}_v + \mathbf{Z}. \tag{10}$$

Here, $\mathbf{X} \in \mathbb{C}^{N_{tr}\times N_T L}$ is a known *measurement matrix* with rows $\{\widetilde{\mathbf{x}}_{n_p}^T(\mathbf{u}_{n_p}^T \otimes \mathbf{A}_T^*),\ p = 1, \ldots, N_{tr}\}$, where $\mathbf{u}_{n_p}^T = \{e^{-j\frac{2\pi n_p}{N_0}\ell}\}_{\ell=0}^{L-1}$, while $\mathbf{H}_v = [\mathbf{h}_{v,1}, \ldots, \mathbf{h}_{v,N_R}]$ with columns $\mathbf{h}_{v,i} \in \mathbb{C}^{N_T L}$, $i = 1, \ldots, N_R$, comprising channel coefficients $\{H_v(i, k, \ell)\}$ associated with the i-th resolvable AoA.

Our goal is to specify the number of pilot tones, N_{tr}, the indices of the pilot tones, $\{n_p\}_{p=1}^{N_{tr}}$, and the training sequence, $\{\widetilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}}$, that will result in reliable estimation of $\mathbf{H}_v$ from $\mathbf{Y}$. To this end, we first need to understand conditions on the matrix $\mathbf{X}$ and the corresponding reconstruction method that lead to reliable estimation of approximately sparse channels.

III. Compressive Channel Estimation

We will estimate the columns of the unknown channel $\mathbf{H}_v$ one at a time. It is now a well-known fact that an s-compressible $\mathbf{h}$ can be estimated from noisy observations $\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w}$ with $\mathbf{X}$ having unit-norm columns and $\mathbf{w}$ distributed as $\mathcal{N}(\mathbf{0}_{N_R}, \sigma^2\mathbf{I}_{N_R})$ by solving

$$\widehat{\mathbf{h}} = \underset{\widetilde{\mathbf{h}}}{\operatorname{argmin}}\, \|\widetilde{\mathbf{h}}\|_1 \text{ s.t. } \|\mathbf{X}^H(\mathbf{y} - \mathbf{X}\widetilde{\mathbf{h}})\|_\infty \leq \lambda \tag{11}$$

for an appropriate parameter $\lambda > 0$ [19]. The reconstruction method (11) is known as the Dantzig selector and it results in near-optimal guarantees for the estimation error $\|\mathbf{h} - \widehat{\mathbf{h}}\|_2$ as long as the matrix $\mathbf{X}$ satisfies a certain property, termed the restricted isometry property (RIP) [27].

Definition 2 (Restricted isometry property)**.** The matrix $\mathbf{X} \in \mathbb{C}^{N_{tr}\times N_T L}$ with unit ℓ_2-norm columns is said to satisfy the restricted isometry property (RIP) of order S with parameter $\delta_S \in (0, 1)$ (i.e., $\mathbf{X}$ is RIP(S, δ_S)) as long as

$$(1 - \delta_S)\|\mathbf{h}\|_2^2 \leq \|\mathbf{X}\mathbf{h}\|_2^2 \leq (1 + \delta_S)\|\mathbf{h}\|_2^2 \tag{12}$$

for all $\mathbf{h} \in \mathbb{C}^{N_T L}$ having no more than S nonzero entries.

It is shown in [19] that if a measurement matrix is RIP$(2S, \delta_{2S})$ with $\delta_{2S} \leq 0.3$ then it can be used to near-optimally estimate any $\mathbf{h}$ having S nonzero entries with high probability. In addition, it is established in [19] that an s-compressible $\mathbf{h}$ can be treated as effectively having no more than $(\frac{r}{\sigma})^{1/s}$ nonzero entries [19]. Our goal in the following, therefore, reduces to specification of the pilot tones and the corresponding training sequence that result in the matrix $\mathbf{X}$ in (10) satisfying RIP$(2S_*, 0.3)$ with $S_* \geq (\frac{r}{\sigma})^{1/s}$, where $\sigma^2 \triangleq \frac{N_T}{\epsilon_{tr}}$ in our problem formulation.

IV. Deterministic Selection of Pilot Tones and Design of the Training Sequence

In this section, we provide a deterministic algorithm for selection of pilot tones across the transmit antennas and discuss training sequences that can result in measurement matrices that satisfy the RIP.

A. Selection of Pilot Tones

Our procedure for selecting pilot tones is similar to the algorithm provided in [26] for single-antenna OFDM systems. We use the same pilot tones across all the transmit antennas. We assume the number of subcarriers in the OFDM channel, N_0, is prime. An integer $R \geq 2$ is selected and integers $\{\alpha_i\}_{i=1}^R$ are chosen such that $\alpha_R \in \{1, 2, ..., N_0 - 1\}$ is relatively prime to N_0, while $\{\alpha_i\}_{i=1}^{R-1} \in \{0, 1, ..., N_0 - 1\}$. Next, an integer $M \geq 1$ is selected and an R-degree polynomial is constructed with coefficients $\{\alpha_i\}_{i=1}^R$: $Q(m) = \alpha_1 m + ... + \alpha_R m^R$, where $m = 1, \ldots, M$. Finally, a multiset $\mathcal{T} = \{Q(m) \mod N_0 : m = 1, 2, .., M\}$ is formed. The number of pilot tones, N_{tr}, is the number of unique elements of $\mathcal{T}$ and the set of pilot tones, $\{n_p\}_{p=1}^{N_{tr}}$, corresponds to the unique elements of $\mathcal{T}$. Defining the multiplicity of elements in $\mathcal{T}$ as $C_{n_p},\ p = 1, \ldots, N_{tr}$, notice that if all pilot tone multiplicities are 1 then $N_{tr} = |\mathcal{T}| = M$ and $\{n_p\}_{p=1}^{N_{tr}} = \mathcal{T}$. Note that this pilot tone selection procedure is flexible and depends on the choice of N_0 (assumed to be prime), $R \geq 2$, M, and $\{\alpha_i\}_{i=1}^R$.

B. Training Sequence Design

In contrast to the tone selection procedure, the training sequence is designed randomly. The set $\{\widetilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}}$ is taken to be a matrix of independent and identically distributed (i.i.d) Rademacher random variables in which each element takes the values $\pm\sqrt{\frac{C_{n_p}}{M}}$ with equal probability; this ensures $\sum_{p=1}^{N_{tr}} \|\widetilde{\mathbf{x}}_{n_p}\|_2^2 = N_T$. Note that the training sequence is not the same across different antennas. It is worth noting here that deterministic variants of this problem include sequences with low periodic autocorrelations such as the Alltop sequence $\{\widetilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}} = \{\frac{1}{\sqrt{N_{tr}}}e^{-j2\pi k^3/N_{tr}}\}_{k=0}^{N_{tr}-1}$ [28] that is circularly shifted across transmit antennas. Although we do not have analysis for such deterministic sequences, we demonstrate their estimation performance in numerical experiments using the circularly shifted Alltop sequence.

The entire procedure is summarized below as Algorithm 1.

Algorithm 1 Pilot tone selection and training sequence design

1: Select integer $R \geq 2$.
2: Select $\alpha_R \in \{1, 2, ..., N_0 - 1\}$ relatively prime to N_0 and $\{\alpha_i\}_{i=1}^{R-1} \in \{0, 1, ..., N_0 - 1\}$.
3: Construct $Q(m) = \alpha_1 m + ... + \alpha_R m^R$.
4: Form $\mathcal{T} = \{Q(m) \mod N_0 : m = 1, 2, .., M\}$.
5: Select N_{tr} to be the number of unique elements of $\mathcal{T}$.
6: Set $\{n_p\}_{p=1}^{N_{tr}}$ to be the unique elements of $\mathcal{T}$.
7: Set elements of $\{\widetilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}}$ to take values $\pm\sqrt{\frac{C_{n_p}}{M}}$ with equal probability (in an i.i.d fashion).

V. Main Result and Discussion

In this section, we establish that pilot tones and training sequences specified in Algorithm 1 result in matrix $\mathbf{X}$ that satisfies the RIP. This involves deriving the minimum number of polynomial evaluation points M that ensures $\mathbf{X}$ is RIP$(2S_*, 0.3)$, where $S_* \geq (\frac{r}{\sigma})^{1/s}$ and $\sigma^2 \triangleq \frac{N_T}{\epsilon_{tr}}$.

Lemma 1. *Let N_0 be a prime number and N_{tr}, $\{n_p\}_{p=1}^{N_{tr}}$, and $\{\tilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}}$ be selected according to Algorithm 1. Fix $N_0 > 2$, the polynomial degree $R \geq 2$, and parameters $\epsilon_1 \in (0,1)$, $\epsilon_2 \in (0, \epsilon_1)$, and $t \in (0,1)$. Pick the number of evaluation points M such that $N_0^{1/(R-\epsilon_1)} \leq M \leq N_0$. Then there exists a constant $c(N_0, \epsilon_2)$ such that $\mathbf{X}$ satisfies $RIP(2S, \delta_{2S})$ with probability exceeding $1 - N_0^{-1}$ as long as*

$$2S \leq tc(N_0, \epsilon_2)\delta_{2S} M^{(\epsilon_1-\epsilon_2)/2^{R-1}} \tag{13}$$

and

$$M \geq \frac{8S^2 N_T C^2_{n_{p_{max}}}}{(1-t)^2\delta^2_{2S}} \log 2N_0. \tag{14}$$

Here, $C_{n_{p_{max}}} \triangleq \max_p C_{n_p}$ is the largest multiplicity in $\mathcal{T}$.

Proof: From Gersgorin's circle theroem, a matrix $\mathbf{X}$ consisting of unit ℓ_2-norm columns satisfies RIP$(2S, \delta_{2S})$ if

$$\|\mathbf{X}^H\mathbf{X} - \mathbf{I}_{N_T L}\|_{\max} \leq \frac{\delta_{2S}}{2S}, \tag{15}$$

where $\|\cdot\|_{\max}$ denotes the maximum absolute entry of a matrix, $\mathbf{I}_{N_T L}$ is the identity matrix, and $\|\mathbf{X}^H\mathbf{X} - \mathbf{I}_{N_T L}\|_{\max}$ denotes the worst-case coherence of $\mathbf{X}$. We therefore turn our attention to $\|\mathbf{X}^H\mathbf{X} - \mathbf{I}_{N_T L}\|_{\max}$. To this end, we define the Gram matrix $\mathbf{G} = \mathbf{X}^H\mathbf{X}$ and focus on its off-diagonal entries $\mathbf{G}(j_1, j_2) = \mathbf{x}_{j_1}^H \mathbf{x}_{j_2}, j_1, j_2 = 1, \ldots, N_T L, j_1 \neq j_2$, where $\mathbf{x}_i$ denotes the i-th column of $\mathbf{X}$. Further, in order to simplify analysis, we reindex the columns of $\mathbf{X}$ in terms of an ordered pair (i,k) such that $i = 1, \ldots, N_T$ denotes the antenna index and $k = 0, \ldots, L-1$ denotes the delay index. Using this convention, we can express the off-diagonal entries of $\mathbf{G}$ as

$$\mathbf{G}\Big((i_1,k_1),(i_2,k_2)\Big) = \frac{1}{N_T}\sum_{p=1}^{N_{tr}} e^{\frac{j2\pi n_p}{N_0}(k_1-k_2)} \times \sum_{a_1=0}^{N_T-1}\sum_{a_2=0}^{N_T-1} e^{-\frac{j2\pi}{N_T}(a_1 i_1 - a_2 i_2)} \widetilde{x}^{a_1}_{n_p}\widetilde{x}^{a_2}_{n_p}, \tag{16}$$

where $\widetilde{x}^a_{n_p}$ denotes the a-th element of the training vector $\widetilde{\mathbf{x}}_{n_p}$. Notice that $\mathbf{G}\Big((i,k),(i,k)\Big) = 1$. We now rewrite (16) as $\mathbf{G}\Big((i_1,k_1),(i_2,k_2)\Big) = G_1 + G_2$, where

$$G_1 \triangleq \frac{1}{N_T}\sum_{p=1}^{N_{tr}} e^{\frac{j2\pi n_p}{N_0}(k_1-k_2)} \sum_{a_1=0}^{N_T-1} e^{-\frac{j2\pi}{N_T}a_1(i_1-i_2)} \widetilde{x}^{a_1}_{n_p}\widetilde{x}^{a_1}_{n_p},$$

$$G_2 \triangleq \frac{1}{N_T}\sum_{p=1}^{N_{tr}} e^{\frac{j2\pi n_p}{N_0}(k_1-k_2)} \times \sum_{a_1=0}^{N_T-1}\sum_{\substack{a_2=0\\a_2\neq a_1}}^{N_T-1} e^{-\frac{j2\pi}{N_T}(a_1 i_1 - a_2 i_2)} \widetilde{x}^{a_1}_{n_p}\widetilde{x}^{a_2}_{n_p}. \tag{17}$$

We now provide upper bounds on $|G_1|$ and $|G_2|$. Notice that $G_1 = 0$ for $i_1 \neq i_2$, while for $i_1 = i_2$ we have

$$|G_1| = \frac{1}{M}\left|\sum_{p=1}^{N_{tr}} C_{n_p} e^{\frac{j2\pi n_p}{N_0}(k_1-k_2)}\right|. \tag{18}$$

To bound (18), we follow the approach taken in [26] and use a lemma that takes advantage of an inequality provided by Weyl in [29].

Lemma 2. *Let $R \geq 2$ and consider the polynomial $f(m) = b_1 m + \cdots + b_R m^R$ such that $b_R = \frac{b}{N_0} + \frac{\theta}{N_0^2}$, where $gcd(b, N_0) = 1$ and $|\theta| \leq 1$. Then for any $0 < \epsilon_2 < \epsilon_1 < 1$ and M satisfying $M^{\epsilon_1} \leq N_0 \leq M^{R-\epsilon_2}$, we have*

$$\left|\sum_{m=1}^{M} e^{j2\pi f(m)}\right| \leq C(R, \epsilon_1, \epsilon_2) M^{1-\frac{\epsilon_1-\epsilon_2}{2^{R-1}}}, \tag{19}$$

where $C(R, \epsilon_1, \epsilon_2)$ is a constant independent of M.

In order to apply Lemma 2 to (18), we set $f(m) = \frac{k_1-k_2}{N_0}Q(m)$, then $b_R = (k_1 - k_2)\alpha_R$ and since $k_1, k_2 \in \{0, \ldots, L-1\}$ and $k_1 \neq k_2$, we have $\gcd(b_R, N_0) = 1$. Hence,

$$|G_1| = \frac{1}{M}\left|\sum_{p=1}^{N_{tr}} e^{\frac{k_1-k_2}{N_0}Q(m)}\right| \leq C(R, \epsilon_1, \epsilon_2) M^{(\epsilon_2-\epsilon_1)/2^{R-1}}. \tag{20}$$

Next, since G_2 is a random variable, we provide a probabilistic upper bound on $|G_2|$ using McDiarmid's inequality.

Lemma 3 (McDiarmid's Inequality)**.** *Let $\{x_i\}_{i=1}^N$ be independent random variables that take values from set $\mathcal{X} \subset \mathbb{R}$. Suppose $f : \mathcal{X}^N \to \mathbb{R}$ satisfies*

$$\sup_{\{x_i\}_{i=1}^N, x_{i'}} |f(x_1, \ldots, x_i, \ldots, x_N) - f(x_1, \ldots, x_{i'}, \ldots, x_N)| \leq c_i, \forall i \in 1, \ldots, N. \tag{21}$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(|f(x_1, \ldots, x_N) - \mathbb{E}\left[f(x_1, \ldots, x_N)\right]| \geq \epsilon\right) \leq 2\exp(-\frac{2\epsilon^2}{\sum_{i=1}^N c_i^2}). \tag{22}$$

Notice that since $\widetilde{x}^a_{n_p}$'s, $p = 1, \ldots, N_{tr}$, $a = 1, \ldots, N_T$, are i.i.d, we have $\mathbb{E}(G_2) = 0$. Further, we have that

$$\sup_{\{\widetilde{\mathbf{x}}_{n_p}\}_{p=1}^{N_{tr}}, \widetilde{x}^{a'}_{n_p}} |G_2(\widetilde{x}^0_{n_1}, \ldots, \widetilde{x}^a_{n_p}, \ldots, \widetilde{x}^{N_T-1}_{n_{N_{tr}}}) - G_2(\widetilde{x}^0_{n_1}, \ldots, \widetilde{x}^{a'}_{n_p}, \ldots, \widetilde{x}^{N_T-1}_{n_{N_{tr}}})|$$
$$= \frac{1}{N_T}\left|\sum_{\substack{a_2=0\\a_2\neq i}}^{N_T-1} e^{-\frac{j2\pi}{N_T}(ai_1 - a_2 i_2)}\left(\frac{1}{\sqrt{M}} - (-\frac{1}{\sqrt{M}})\right)\widetilde{x}^{a_2}_{n_p}\right|$$
$$\leq \frac{2C_{n_{p_{\max}}}}{M}. \tag{23}$$

It therefore follows from McDiarmid's inequality that

$$\mathbb{P}(|G_2| \geq \epsilon) \leq 2\exp\left(-\frac{M\epsilon^2}{2C^2_{n_{p_{\max}}}N_T}\right). \tag{24}$$

We can finally set $\epsilon^2 = \frac{2N_T C^2_{n_{p_{\max}}}}{M}\log(2N_0)$ in the above expression to obtain

$$P(|G_2| \geq \epsilon) \leq N_0^{-1}. \tag{25}$$

Notice now from (15) that $\mathbf{X}$ satisfies RIP$(2S, \delta_{2S})$ as long as

$$|G_1| \leq \frac{\delta_{2S}}{2S}t \text{ and } \mathbb{P}(|G_2| \geq \frac{\delta_{2S}}{2S}(1-t)) \leq N_0^{-1},\ t \in (0,1).$$

Substituting (20) and (24) in the above expression completes the proof of the lemma. ■

Using Lemma 1, we can now state the main theorem of this paper for reliable channel estimation.

Theorem 1. *Suppose $\mathbf{H}_v$ comprises the sampled channel coefficients of an s-compressible multipath channel $\boldsymbol{\mathcal{H}}$ and we are interested in estimating $\mathbf{H}_v$ from the noisy observations $\mathbf{Y} = \sqrt{\frac{\epsilon_{tr}}{N_T}}\mathbf{X}\mathbf{H}_v + \mathbf{Z}$, where $\mathbf{Z}$ is AWGN with unit variance. Pick parameters $\epsilon_1 \in (0,1)$, $\epsilon_2 \in (0,\epsilon_1)$, and $t \in (0,1)$. Suppose the number of polynomial evaluation points in Algorithm 1 satisfies $M \geq \max\{M_1, M_2, M_3\}$, where*

$$M_1 \triangleq N_0^{1/(R-\epsilon_1)},$$

$$M_2 \triangleq \left(\frac{(\epsilon_{tr}r^2)^{1/2s}}{0.15tc(N_0,\epsilon_2)N_T^{1/2s}}\right)^{\frac{2^{R-1}}{\epsilon_1-\epsilon_2}}, \text{ and}$$

$$M_3 \triangleq \frac{8N_T(\epsilon_{tr}r^2)^{1/s}C^2_{n_{p_{max}}}}{0.09(1-t)^2N_T^{1/s}}\log 2N_0. \tag{26}$$

Then the measurement matrix $\mathbf{X}$ satisfies RIP$(2S,\delta_{2S})$ with $S \geq (\frac{\epsilon_{tr}r^2}{N_T})^{1/2s}$. Further, if we set the parameter $\lambda = \sqrt{2(1+a)(\log N_R N_T L)\epsilon_{tr}/N_T}$, then for any fixed $a \geq 0$, the estimator

$$\forall i = 1,\dots,N_R, \quad \widehat{\mathbf{h}}_{v,i} = \operatorname*{argmin}_{\mathbf{h}_v \in \mathbb{C}^{N_T L}} \|\mathbf{h}_v\|_1 \quad \textit{such that}$$

$$\left\|\sqrt{\frac{\epsilon_{tr}}{N_T}}\mathbf{X}^H\left(\mathbf{y}_i - \sqrt{\frac{\epsilon_{tr}}{N_T}}\mathbf{X}\mathbf{h}_v\right)\right\|_\infty \leq \lambda \tag{27}$$

satisfies

$$\|\widehat{\mathbf{H}}_v - \mathbf{H}_v\|_2^2 \leq O\left(N_R \log(N_R N_T L)\right)\left(\frac{N_T}{\epsilon_{tr}}r^{1/d}\right)^{2d/(2d+1)} \tag{28}$$

with very high probability that exceeds the value $1 - 2\max\left\{\left(\pi(1+a)\log N_R N_T L \cdot (N_R N_T L)^{2a}\right)^{-1/2}, N_0^{-1}\right\}$. Here, $d \triangleq 1/s - 1/2$ and $\widehat{\mathbf{H}}_v \triangleq [\widehat{\mathbf{h}}_{v,1}\dots\widehat{\mathbf{h}}_{v,N_R}]$.

Notable implications of this theorem include: (i) a deterministic algorithm for the selection of pilot tones, (ii) a sufficient condition on the number of polynomial evaluation points for our algorithm to be effective (although the N_T factor in (26) is not favorable), and (iii) the estimation error primarily having a logarithmic dependence on the number of channel parameters $N_R N_T L$, which is an improvement over traditional channel estimation techniques. In the next section, we demonstrate the effectiveness of Algorithm 1 using numerical experiments.

VI. Numerical Results

In this section, we conduct numerical experiments to evaluate the performance of Algorithm 1. We use Monte Carlo experiments for random channel, additive noise, and random training sequence realizations. Our realizations of multipath channels correspond to $N_p = 6N_R\sqrt{N_T}$ point scatterers with the τ_i's uniformly distributed over $[0, 12.7\mu\text{sec}]$, the $(\theta_{R,i},\theta_{T,i})$'s uniformly distributed over $[-1/2,1/2] \times [-1/2,1/2]$, and the β_i's distributed as zero-mean Gaussian random variables. We consider the case of $N_T \in \{1,\dots,8\}$ transmit antennas and we work with a bandwidth of $W = 25.12$ MHz, which results in $L = 320$. We construct the sampled channel coefficients according to (4), followed by normalization that makes each column of the channel matrix $\mathbf{H}_v$ into a unit ℓ_2-norm vector. Finally, we consider the case of $N_0 = 1049$ tones and operate with training signal-to-noise ratio of $5000N_T$ on a linear scale.

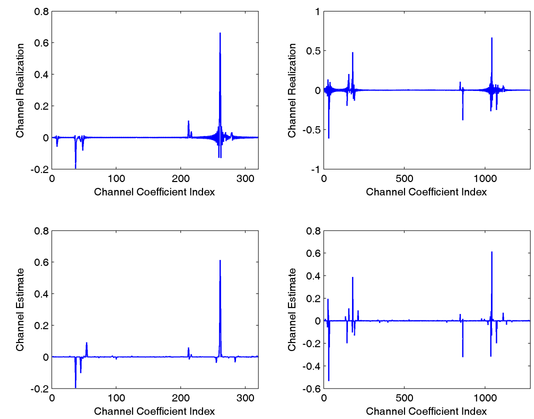
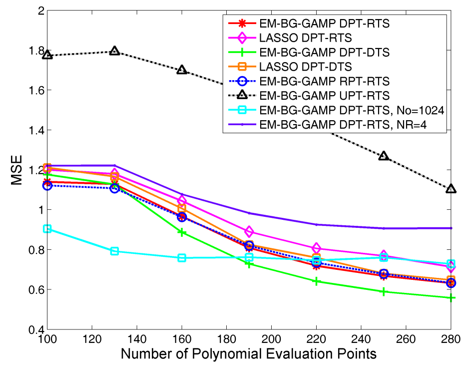


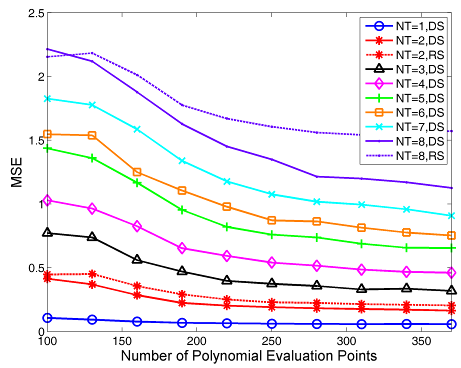
Fig. 1. A typical channel realization and its estimate using EM-BG-GAMP for the case of $N_T = 4$ and $N_R = 1$. The left panel depicts the discrete channel between one transmit and receive antenna, while the right panel shows the complete multipath discrete channel in vectorized form.

In order to implement the pilot tone selection procedure, we use $R = 2$ for the polynomial degree in Algorithm 1. To obtain the mean square error (MSE), $\|\mathbf{H}_v - \widehat{\mathbf{H}}_v\|_2^2$, of channel estimates as a function of the number of polynomial evaluation points, we average over 100 Monte Carlo experiments. We utilize compressed sensing algorithms EM-BG-GAMP [30] and LASSO (via SPGL1 [31]) for channel reconstruction.

Our first set of experiments correspond to $N_R = 1$ receive antenna. A typical channel realization with $N_T = 4$ transmit antennas in this case and its reconstruction using EM-BG-GAMP is depicted in Fig. 1. In Fig. 2(a), the MSE of channel estimates with $N_T = 4$ transmit antennas is demonstrated. The MSE obtained using deterministic pilot tone selection (DPT) along with random training sequence (RTS) and deterministic, circularly shifted Alltop sequence (DTS) is compared against random pilot tone selection (RPT) and equispaced pilot tones (UPT) with the same number of pilot tones. It is observed that DPT outperforms UPT and performs as well as RPT. Moreover, EM-BG-GAMP results in a smaller MSE than the LASSO. We also show results for both non-prime $N_0 = 1024$ and $N_R = 4$ in this figure to demonstrate effectiveness of Algorithm 1. Finally, the MSE of channel estimates obtained using EM-BG-GAMP is plotted in Fig. 2(b) as a function of



(a)



(b)

Fig. 2. Performance summary of deterministic pilot tone selection algorithm. (a) depicts the performance of proposed algorithm using various reconstruction techniques and (b) shows the MSE of EM-BG-GAMP as a function of number of polynomial evaluation points for different number of transmit antennas.

the number of polynomial evaluation points for DPT with RTS and DTS for different number of transmit antennas.

VII. Conclusion

In this paper, we addressed the problem of compressive MIMO-OFDM channel estimation using a training based technique involving deterministic selection of pilot tones. Our proposed method is quite flexible in terms of the choice of different parameters. We also provided analytical guarantees for the performance of our method. Finally, although our analysis is limited to the case of random training sequences, our numerical results suggested that deterministic sequences with low autocorrelation can be substituted for similar performance.

References

[1] H. Bolcskei, R. W. Heath, and A. J. Paulraj, “Blind channel identification and equalization in OFDM-based multiantenna systems,” *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, 2002.

[2] S. Yatawatta and A. P. Petropulu, “Blind channel estimation in MIMO OFDM systems with multiuser interference,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1054–1068, 2006.

[3] F. Gao and A. Nallanathan, “Subspace-based blind channel estimation for SISO, MISO and MIMO OFDM systems,” in *Proc. IEEE Int. Conf. Commun.*, vol. 7, 2006, pp. 3025–3030.

[4] ——, “Blind channel estimation for MIMO OFDM systems via nonredundant linear precoding,” *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 784–789, 2007.

[5] C. Shin, R. Heath, and E. Powers, “Blind channel estimation for MIMO-OFDM systems,” *IEEE Trans. Veh. Tech.*, pp. 670–685, 2007.

[6] Y. Zeng and T.-S. Ng, “A semi-blind channel estimation method for multiuser multiantenna OFDM systems,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1419–1429, 2004.

[7] F. Wan, W. Zhu, and M. N. S. Swamy, “Semiblind sparse channel estimation for MIMO-OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2569–2582, 2011.

[8] I. Barhumi, G. Leus, and M. Moonen, “Optimal training design for MIMO OFDM systems in mobile wireless channels,” *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, 2003.

[9] S. Sun, I. Wiemer, C. K. Ho, and T. T. Tjhung, “Training sequence assisted channel estimation for MIMO OFDM,” in *Wireless Commun. and Networking,*, vol. 1. IEEE, 2003, pp. 38–43.

[10] J. Yue, K. Kim, J. Gibson, and R. Iltis, “Channel estimation and data detection for MIMO-OFDM systems,” in *Proc. IEEE Global Telecommun. Conf.*, vol. 2, 2003, pp. 581–585.

[11] M. Shin, H. Lee, and C. Lee, “Enhanced channel-estimation technique for MIMO-OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 261–265, 2004.

[12] G. L. Stuber, J. R. Barry, S. W. Mclaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, “Broadband MIMO-OFDM wireless communications,” *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, 2004.

[13] D. Hu, L. Yang, Y. Shi, and L. He, “Optimal pilot sequence design for channel estimation in MIMO OFDM systems,” *Commu. Lett.*, vol. 10, no. 1, pp. 1–3, 2006.

[14] H. Minn and N. Al-Dhahir, “Optimal training signals for MIMO OFDM channel estimation,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, 2006.

[15] H. Minn, N. Al-Dhahir, and Y. Li, “Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise,” *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1754–1759, 2006.

[16] T. Qureshi, M. Zoltowski, and R. Calderbank, “MIMO-OFDM channel estimation using golay complementary sequences,” in *Proc. IEEE Int. Waveform Diversity and Design Conf.*, 2009, pp. 253–257.

[17] Y. Chi, A. Gomaa, N. Al-Dhahir, and A. Calderbank, “Training signal design and tradeoffs for spectrally-efficient multi-user MIMO-OFDM systems,” *IEEE Trans. Wireless Commun.*, pp. 2234–2245, 2011.

[18] W. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[19] E. Candes and T. Tao, “The dantzig selector: Statistical estimation when p is much larger than n,” *The Ann. Stat.*, pp. 2313–2351, 2007.

[20] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, “Channel estimation algorithms for broadband MIMO-OFDM sparse channel,” in *Proc. Personal, Indoor and Mobile Radio Comm.*, 2003, pp. 1929–1933.

[21] N. Jing, W. Bi, and L. Wang, “Deterministic pilot design for MIMO OFDM system based on compressed sensing,” in *Proc. IEEE 14th Int. Conf. Commun. Technol.*, 2012, pp. 897–903.

[22] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, “Compressed channel sensing,” in *Proc. IEEE Annu. Conf. Inf. Sci. and Syst.*, 2008, pp. 5–10.

[23] M. Sharp and A. Scaglione, “Application of sparse signal recovery to pilot-assisted channel estimation,” in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2008, pp. 3469–3472.

[24] G. Taubock 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. Acoust., Speech and Signal Process.*, 2008, pp. 2885–2888.

[25] P. Schniter, “Joint estimation and decoding for sparse channels via relaxed belief propagation,” in *Proc. IEEE Conf. Rec. Forty Fourth Asilomar Conf. Signals, Syst. and Comput.*, 2010, pp. 1055–1059.

[26] L. Applebaum, W. Bajwa, A. Calderbank, J. Haupt, and R. Nowak, “Deterministic pilot sequences for sparse channel estimation in OFDM systems,” in *Proc. Int. Conf. Digital Signal Process.*, 2011, pp. 1–7.

[27] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[28] W. O. Alltop, “Complex sequences with low periodic correlations,” *IEEE Trans. Inf. Theory*, vol. 26, pp. 350–354, 1980.

[29] H. Weyl, “Über die gleichverteilung von zahlen mod. eins,” *Mathematische Annalen*, vol. 77, no. 3, pp. 313–352, 1916.

[30] J. Vila and P. Schniter, “Expectation-maximization bernoulli-gaussian approximate message passing,” in *Proc. IEEE Conf. Rec. Asilomar Conf. on Signals, Syst. and Comput.*, 2011, pp. 799–803.

[31] E. van den Berg and M. Friedlander, “Probing the pareto frontier for basis pursuit solutions,” *SIAM J. Sci. Comput.*, pp. 890–912, 2008.