DETERMINISTIC PILOT SEQUENCES FOR SPARSE CHANNEL ESTIMATION IN OFDM SYSTEMS

Lorne Applebaum[1], *Waheed U. Bajwa*[2], *A. Robert Calderbank*[1,2], *Jarvis Haupt*[3], *and Robert Nowak*[4]

[1]Princeton University, [2]Duke University, [3]University of Minnesota,
and [4]University of Wisconsin–Madison

ABSTRACT

This paper examines the problem of multipath channel estimation in single-antenna orthogonal frequency division multiplexing (OFDM) systems. In particular, we study the problem of pilot assisted channel estimation in wideband OFDM systems, where the time-domain (discrete) channel is approximately sparse. Existing works on this topic established that techniques from the compressed sensing literature can yield accurate channel estimates using a relatively small number of pilot tones, provided the pilots are selected *randomly*. Here, we describe a general purpose procedure for *deterministic* selection of pilot tones to be used for channel estimation, and establish guarantees for channel estimation accuracy using these sequences along with recovery techniques from the compressed sensing literature. Simulation results are presented to demonstrate the effectiveness of the proposed procedure in practice.

Index Terms— Channel estimation, compressed sensing, deterministic RIP matrices, OFDM systems

1. INTRODUCTION

Two key metrics used to evaluate the performance of wireless systems include: (i) the bit-error rate (BER) and (ii) the spectral efficiency (i.e., bits transmitted per second per Hz). It is generally recognized that the BER in wideband wireless systems can be significantly reduced if the receiver has knowledge of the underlying multipath channel response; the so-called coherent communications [1]. In practice, however, the channel response is seldom (if ever) known to the receiver. Instead, it needs to be periodically estimated at the receiver in order to reap the benefits of coherent communications.

Our focus in this paper is on multipath channel estimation for single-antenna *orthogonal frequency-division multiplexing* (OFDM) wireless systems [2]. There are two classes of methods that are commonly employed for channel estimation in OFDM systems, namely, training-based methods and blind methods. Blind methods attempt to estimate the channel by making use of the statistics of the unknown data only. Therefore, blind channel estimation has the potential to yield a lower BER without affecting the systems's spectral efficiency. Blind methods, however, tend to be effective only if the underlying multipath channel remains constant over a large number of OFDM symbols. This is clearly a disadvantage in the case of a mobile system where the underlying channel can change from one symbol to the next symbol.

Training-based methods, on the other hand, try to estimate the channel by transmitting the unknown data multiplexed with some *training data* already known to the receiver. Such methods are preferred for channel estimation in mobile OFDM systems since they yield reliable estimates even if the channel changes from one symbol to the next symbol. The most prevalent form of training-based channel estimation in OFDM systems involves dedicating a few of the OFDM subcarriers (tones) solely for transmitting the training data (pilots) [3]. The key questions that arise in such *pilot-assisted channel estimation* (PACE) methods include: (i) which, and how many, OFDM tones should be used as pilot tones? and (ii) how does the choice of pilot tones affect the channel estimation error? Note that the former question directly impacts the spectral efficiency, while the latter question directly impacts the BER in wireless systems.

To the best of the authors' knowledge, Rinne and Renfors [4] and Negi and Cioffi [5] made some of the first attempts to concretely answer the above questions for PACE methods in OFDM systems. In particular, it has been argued in [5] that: (i) the best set of pilot tones corresponds to a set of cardinality equal to the length, L, of the underlying (discrete) multipath channel, with the pilot tones equally spaced within the OFDM subcarriers, and (ii) this set of equally spaced pilot tones results in a channel estimation error of $\frac{L\sigma^2}{\mathcal{E}_{tr}}$, where σ^2 denotes the receiver noise variance and $\mathcal{E}_{tr}$ denotes the training energy. The channel estimation results of [5] are based on the maximum likelihood (ML) criterion and are in fact optimal for narrowband OFDM systems. Many wideband OFDM systems of recent interest, such as underwater acoustic systems [6], digital television systems [7] and residential ultrawideband systems [8], however, correspond to underlying (discrete) multipath channels in which a large number of time-

LA, WUB, and RC are supported in part by NSF under grants CNS-1011962 and DMS-0914892, by ONR under grant N00014-08-1-1110, and by AFOSR under grants FA9550-09-1-0643 and FA9550-09-1- 0551. JH is supported in part by DARPA/ONR grant N66001-10-1-4090.

domain channel coefficients tend to have very small magnitudes; see [9, 10] for a mathematical justification of this phenomenon from the channel modeling perspective.

Conventional *linear* PACE methods based on the ML criterion fail to capitalize on this anticipated structure (i.e., *approximate sparsity* or *compressibility*) of the underlying multipath channels in wideband OFDM systems. In contrast, the main contribution of this paper is a *nonlinear* PACE scheme that makes use of a deterministically chosen set of pilot tones having cardinality much smaller than L and a convex optimization-based reconstruction method, known as the *Dantzig selector* [11], to result in a channel estimation error that is significantly smaller than that achievable using the traditional PACE techniques. Stated differently, the PACE scheme proposed in here is significantly superior to the ones proposed in [4, 5] in terms of both the spectral efficiency (number of pilot tones is smaller than L) and the BER (channel estimation error is smaller than $\frac{L\sigma^2}{\mathcal{E}_{tr}}$) in wideband OFDM systems.

In terms of relationship to previous work, note that the results reported in this paper leverage some of the recent advances in the field of compressed sensing (CS) [12]. This makes our approach to PACE in OFDM systems somewhat similar to the ones proposed independently in [13–15]. However, note that [13–15] require using *randomly* chosen sets of pilot tones in order to provide concrete guarantees for finite training energy. In contrast, to the best of our knowledge this is the first paper in the literature that provides rigorous guarantees for CS-based PACE methods for OFDM systems using deterministically chosen sets of pilot tones (deterministic probe designs for multi-user multi-antenna OFDM systems were recently examined in [16], for systems employing *linear least-squares* channel estimation).

Finally, note that recently Schniter [17] has also proposed a novel channel estimation scheme for wideband OFDM systems. The scheme proposed in [17] can be termed as *semi-blind* since it makes use of both the training data and the statistics of the unknown data to carry out joint channel estimation and data decoding using belief-propagation ideas. The primary difference between [17] and our work is that [17] assumes the underlying multipath channel to have no more than $S \ll L$ nonzero time-domain channel coefficients (i.e., S-sparse channels), whereas we assume the multipath channel to be approximately sparse (i.e., compressible) but not necessarily exactly sparse.

2. SYSTEM MODEL

In this section, we describe the problem setup and accompanying assumptions for PACE in wideband OFDM systems. For the sake of this exposition, we restrict ourselves to the canonical discrete channel and system model; we refer the reader to [10] for the relationship between the discrete-time mathematical model and the continuous-time physical setup (see also the simulation setup and (8) in Section 6 of this paper).

To begin, we assume that the transmitter communicates with the receiver over a discrete multipath channel of length L, $h = \begin{bmatrix} h_0 & h_1 & \dots & h_{L-1} \end{bmatrix}^T \in \mathbb{C}^L$, that remains fixed for a period of $N + L$ with $N \gg L$. The main assumption that we make here concerns the structure of h in wideband wireless systems. Specifically, we assume that the j-th largest (in magnitude) entry, $h_{(j)}$, of h obeys

$$|h_{(j)}| \leq B \cdot j^{-\alpha - 1/2} \tag{1}$$

for some $B > 0$ and $\alpha > 0$. The parameter α here controls the rate of decay of the magnitudes of the ordered entries of h and we term any h that satisfies (1) as α-compressible.

Next, we assume that the (unknown and training) data is transmitted using an OFDM symbol that consists of a total of N subcarriers and has an L-length cyclic prefix. Using $d = \begin{bmatrix} d_0 & d_1 & \dots & d_{N-1} \end{bmatrix}^T \in \mathbb{C}^N$ in this case to denote the data transmitted over each of the N OFDM tones, it can then be easily shown that the received data vector $y \in \mathbb{C}^N$ at the receiver is related to the transmitted data vector d as follows [1]:

$$y = Hd + w. \tag{2}$$

Here, $w \in \mathbb{C}^N$ represents a zero-mean additive white Gaussian noise (AWGN) vector of variance σ^2, while $H \in \mathbb{C}^{N \times N}$ is a diagonal matrix comprising of the N *OFDM channel coefficients* that are related to h as $\{H_{kk} = \sum_{\ell=0}^{L-1} F_{k\ell} h_\ell\}$ with $F_{k\ell} = e^{-j\frac{2\pi}{N}k\ell}$ denoting the (k, ℓ)th element of the N-point discrete Fourier transform (DFT) matrix.

In order to carry out PACE under this setup, we can now proceed as follows. First, we select a set of indices $\mathcal{P} \subset \{0, 1, \dots, N-1\}$, corresponding to the set of pilot tones, of cardinality $N_{tr} = |\mathcal{P}|$. Next, we construct a training data vector $d_{tr} \in \mathbb{C}^{Ntr}$ having energy $\|d_{tr}\|_2^2 = \mathcal{E}_{tr}$ and transmit this vector using the pilot tones specified by $\mathcal{P}$; in other words, we have that $d_{|\mathcal{P}} = d_{tr}$, where $d_{|\mathcal{P}}$ denotes the restriction of d to the indices in $\mathcal{P}$. Then defining $y_{tr} = y_{|\mathcal{P}}$, it is easy to see because of the diagonal nature of the OFDM channel matrix H [cf. (2)] that

$$y_{tr} = H_{|\mathcal{P}\times\mathcal{P}} d_{tr} + w_{|\mathcal{P}} \tag{3}$$

where $H_{|\mathcal{P}\times\mathcal{P}}$ denotes the restriction of H to the ordered pairs in $\mathcal{P} \times \mathcal{P}$. Further, (3) can be easily expressed in terms of the (time-domain) multipath channel h by noting that

$$H_{|\mathcal{P}\times\mathcal{P}} d_{tr} = D_{tr} A h$$

where $D_{tr} = \mathsf{diag}(d_{tr})$ and A is an $N_{tr} \times L$ matrix that comprises $\{\begin{bmatrix} F_{p0} & F_{p1} & \dots & F_{p(L-1)} \end{bmatrix} : p \in \mathcal{P}\}$ as its rows. Therefore, defining $X = D_{tr} A$, the goal of any PACE scheme in OFDM systems is to (i) specify the pair $(\mathcal{P}, d_{tr})$ such that

$N_{tr} = |\mathcal{P}|$ and $\|d_{tr}\|_2^2 = \mathcal{E}_{tr}$, and (ii) provide a reliable estimate of the underlying multipath channel h from the corresponding received training data vector

$$y_{tr} = Xh + w_{|\mathcal{P}}.$$

3. COMPRESSED SENSING BACKGROUND

The PACE scheme proposed in this paper leverages existing concepts from the CS literature. In particular, we will employ a reconstruction method whose success relies upon a normalized version of the *measurement matrix* $X = D_{tr}A$ described above satisfying a property known as the Restricted Isometry Property (RIP) [18].

Definition 3.1 (Restricted Isometry Property)**.** *An $N_{tr} \times L$ matrix Z with unit norm columns satisfies the restricted isometry property of order $S \in \mathbb{N}$ with parameter $\delta_S \in [0, 1)$—or, in shorthand, Z satisfies RIP(S,δ_S)—if*

$$(1 - \delta_S)\|v\|_2^2 \leq \|Zv\|_2^2 \leq (1 + \delta_S)\|v\|_2^2 \tag{4}$$

holds for all $v \in \mathbb{C}^L$ having no more than S nonzero entries.

In words, the RIP of order S says that the matrix Z acts like an *almost isometry* on all S-sparse vectors v. For example, if Z is the identity matrix (say, of dimension L), then it satisfies the RIP trivially for any $S = 1, 2, \ldots, L$ with $\delta_S = 0$.

The cases of particular interest in the CS literature, however, correspond to when Z has fewer rows than columns. Indeed, the RIP has been widely adopted in the CS literature, and many results have been established for procedures which guarantee reliable and efficient reconstruction of sparse and compressible signals from a relatively small number of linear measurements obtained via a measurement matrix satisfying the RIP. In the following, we will leverage the results for one such reconstruction method, known as the Dantzig selector, which was proposed in [11] and which is particularly well-suited for measurements corrupted by stochastic noise. The following result is a complex-valued variant of the result originally reported in [11]; see [19, Th. 2.13] for further details.

Lemma 3.2 (The Dantzig Selector)**.** *Let Z be a measurement matrix satisfying RIP($2S$,δ_{2S}) with $\delta_{2S} < 1/3$ for some $S \in \mathbb{N}$. Let $\gamma = Z\beta + \eta$ be a vector of noisy measurements of $\beta \in \mathbb{C}^L$, where η is an AWGN vector with variance σ^2. Choose $\lambda_L = \sqrt{2(1 + a)\log L}$ for any $a \geq 0$. Then the estimator*

$$\widehat{\beta} = \arg\min_{v \in \mathbb{C}^L} \|v\|_1 \text{ subject to } \|Z^H(\gamma - Zv)\|_\infty \leq \sigma\lambda_L,$$

where the notation Z^H denotes the Hermitian, or conjugate transpose, of Z, satisfies

$$\|\widehat{\beta} - \beta\|_2^2 \leq c_0 \min_{1 \leq k \leq S} \left(\sigma\lambda_L\sqrt{k} + \frac{\|\beta_k - \beta\|_1}{\sqrt{k}} \right)^2,$$

with probability at least $1 - 2\left(\sqrt{\pi(1 + a)\log L} \cdot L^a\right)^{-1}$, where β_k is the best k-term approximation of β, formed by setting all but the k largest entries (in magnitude) of β to zero, and the constant $c_0 = 16/\left(1 - 3\delta_{2S}\right)^2$.

4. DETERMINISTIC PILOT SEQUENCE AND TRAINING DATA SELECTION

In this section, we describe our proposed procedure for deterministic selection of the pair $(\mathcal{P}, d_{tr})$ corresponding to the set of pilot tones and training data. Our procedure, which is based on the method outlined in [20], is quite general and in fact gives rise to a *family* of pilot tone/training data selection procedures, each of which is fully-described by a small number of integer-valued parameters. We will see in the following section that each selection $(\mathcal{P}, d_{tr})$ results in a set of pilot tones and corresponding training data whose performance for estimating compressible multipath channels using the Dantzig selector can be rigorously quantified.

In order to describe our selection procedure, we first assume that the number of subcarriers, N, is prime. Under this condition, our procedure can be described as follows. Begin by first selecting an integer $R \geq 2$. Next, choose a set of R integers denoted $\{a_i\}_{i=1}^R$ such that a_R is relatively prime to N (i.e., $a_R \in \{1, 2, \ldots, N-1\}$), while $a_i \in \{0, 1, 2, \ldots, N-1\}$ for the remaining $i \neq R$. The integers $\{a_i\}_{i=1}^R$ become the coefficients of a degree-R polynomial $Q(m) = a_1 m + \cdots + a_R m^R$. Finally, choose an integer $M \geq 1$ and construct a multiset $\mathcal{T}$ by evaluating $Q(m) \mod N$ for integers $m = 1, 2, \ldots, M$. Formally, we have that $\mathcal{T} = \{Q(m) \mod N : m = 1, 2, \ldots, M\}$ with multiplicities.

Now, the set of pilot tones $\mathcal{P}$ is the (sub)set of unique elements in $\mathcal{T}$. Note that if each element of $\mathcal{T}$ appears with multiplicity 1 then $\mathcal{P} = \mathcal{T}$, otherwise $\mathcal{P} \subset \mathcal{T}$. The entries of the training data vector d_{tr} corresponding to the pilots $p \in \mathcal{P}$, $\{d_{tr}(p)\}_{p \in \mathcal{P}}$, are functions of the multiplicity of the elements $p \in \mathcal{P}$ in the multiset $\mathcal{T}$. Specifically, let C_p denote the number of times the element $p \in \mathcal{P}$ appears in $\mathcal{T}$ and note that the cardinality of $\mathcal{T}$ is equal to M. Then, we select the training data associated with the pilot $p \in \mathcal{P}$ as

$$d_{tr}(p) = \sqrt{\frac{C_p\, \mathcal{E}_{tr}}{M}}$$

where $\mathcal{E}_{tr}$ is the training data energy as described earlier. Notice that, by construction, we have $\sum_{p \in \mathcal{P}} C_p = M$, and so the selection in (5) ensures $\|d_{tr}\|_2^2 = \sum_{p \in \mathcal{P}} d_{tr}^2(p) = \mathcal{E}_{tr}$, as required. This entire pilot sequence and training data selection procedure is summarized as Procedure 1.

Notice that this procedure provides considerable flexibility in selecting a pair $(\mathcal{P}, d_{tr})$, and the selection of each pair is fully-parameterized by the polynomial degree R (≥ 2), the number of polynomial evaluation points M, the coefficients

Procedure 1 : Deterministic Pilot/Training Data Selection

1. Select an integer $R \geq 2$.
2. Choose integers $a_R \in \{1, 2, \dots, N-1\}$ and $a_i \in \{0, 1, 2, \dots, N-1\}$ for $i = 1, \dots, R-1$.
3. Construct the polynomial $Q(m) = a_1 m + \cdots + a_R m^R$.
4. Choose an integer $M \geq 1$ and form the multiset of integers $\mathcal{T} = \{Q(m) \mod N : m = 1, 2, \dots, M\}$.
5. Select the set of pilots to be the unique elements of $\mathcal{T}$.
6. Select the training data vector entries according to $d_{tr}(p) = \sqrt{C_p \, \mathcal{E}_{tr}/M}$ for $p \in \mathcal{P}$, where C_p denotes the multiplicity of each $p \in \mathcal{P}$ in the multiset $\mathcal{T}$.

$\{a_i\}_{i=1}^R$, and the number of subcarriers N (which is assumed to be prime). The next section establishes conditions under which these deterministic selections of $(\mathcal{P}, d_{tr})$ enable provably accurate estimation of compressible multipath channels using the Dantzig selector.

5. MAIN RESULTS

In the previous section, we specified a deterministic procedure for selecting the pair $(\mathcal{P}, d_{tr})$ for PACE in OFDM systems. We now claim that Procedure 1 can result in a measurement matrix $X = D_{tr} A$ that facilitates reconstruction of compressible multipath channels using the Dantzig selector. To that end, we follow the approach proposed in [20] to obtain the following lemma, which outlines conditions under which our deterministic selection of $(\mathcal{P}, d_{tr})$ corresponds to a normalized matrix $\Psi = (\mathcal{E}_{tr})^{-1/2} X$ satisfying the RIP. The proof of this lemma is provided in the appendix.

Lemma 5.1. *Suppose $N > 2$ is prime, and let $(\mathcal{P}, d_{tr})$ be selected according to Procedure 1, with parameters R (the polynomial degree), M (the number of polynomial evaluation points), and $\{a_i\}_{i=0}^R$ (the polynomial coefficients). Let*

$$\Psi = (\mathcal{E}_{tr})^{-1/2} D_{tr} A \tag{5}$$

where (as above) $D_{tr} = \textsf{diag}(d_{tr})$ and A is an $N_{tr} \times L$ matrix that comprises $\{\begin{bmatrix} F_{p0} & F_{p1} & \dots & F_{p(L-1)} \end{bmatrix} : p \in \mathcal{P}\}$ as its rows. Choose $R \geq 2$, and any $\epsilon_1 \in (0,1)$ and $\epsilon_2 \in (0, \epsilon_1)$. There exists a constant $c = c(N, \epsilon_2)$ such that whenever the number of evaluation points M satisfies, $N^{1/(R-\epsilon_1)} \leq M \leq N$, the matrix Ψ satisfies RIP(S,δ_S) for any $\delta_S \in (0,1)$ and

$$S \leq c(N, \epsilon_2) \delta_S M^{(\epsilon_1 - \epsilon_2)/2^{R-1}}. \tag{6}$$

Now, taken together with the results of Lemma 3.2, the results of Lemma 5.1 allow us to obtain the following theorem.

Theorem 5.2. *Suppose that the time-domain multipath channel $h \in \mathbb{C}^L$ obeys (1). Define $y'_{tr} = (\mathcal{E}_{tr})^{-1/2} y_{tr}$; in other words, $y'_{tr} = \Psi h + w'$, where w' is an AWGN vector with variance $\mathsf{SNR}^{-1} = \frac{\sigma^2}{\mathcal{E}_{tr}}$.*

Select $(\mathcal{P}, d_{tr})$ according to Procedure 1, such that for any choice of $\epsilon_1 \in (0,1)$ and $\epsilon_2 \in (0, \epsilon_1)$ the number of polynomial evaluation points satisfies

$$M > \max\Bigg\{ N^{1/(R-\epsilon_1)}, \left(\frac{6}{c(N,\epsilon_2)}\right)^{\frac{2^{R-1}}{\epsilon_1 - \epsilon_2}} B^{\frac{2^{R-1}(\alpha+1/2)}{\epsilon_1-\epsilon_2}} \cdot \mathsf{SNR}^{\frac{2^{R-1}(2\alpha+1)}{4(\epsilon_1-\epsilon_2)}} \Bigg\} \tag{7}$$

where $c(N, \epsilon_2)$ is the same constant as in Lemma 5.1. Finally, choose $\lambda_L = \sqrt{2(1+a)\log L}$ for any $a \geq 0$. Then the reconstruction error of the Dantzig selector channel estimate

$$\widehat{h} = \arg\min_{v \in \mathbb{C}^L} \|v\|_1 \text{ subject to } \|\Psi^H(y'_{tr} - \Psi v)\|_\infty \leq \frac{\lambda_L}{\sqrt{\mathsf{SNR}}},$$

satisfies

$$\|\widehat{h} - h\|_2^2 \leq c'_0 \, B^{\frac{2}{2\alpha+1}} \mathsf{SNR}^{-\frac{2\alpha}{2\alpha+1}} \log L,$$

with probability at least $1 - 2\left(\sqrt{\pi(1+a)\log L} \cdot L^a\right)^{-1}$, where c'_0 is an absolute constant.

The proof of Theorem 5.2 follows from the fact that the condition (7) implies that the matrix Ψ described in Lemma 5.1 satisfies RIP($2S$,δ_{2S}) with $S \geq B^{\alpha+1/2} \cdot \mathsf{SNR}^{\frac{2\alpha+1}{4}}$ and $\delta_{2S} < 1/3$. Thus, we may apply Lemma 3.2 to obtain the stated estimation error bounds.

A few comments are in order regarding the result of the theorem. First, as our pilot tone selection procedure is entirely deterministic, the probabilistic statement is with respect to the randomness in the additive noise vector w'. Second, note that ignoring constants, our estimation error bounds scale like $\|\widehat{h} - h\|_2^2 \precsim \mathsf{SNR}^{-\frac{2\alpha}{2\alpha+1}} \log L$; in other words, our error guarantees exhibit only a logarithmic dependence on the channel length L. Compare this with the estimation error bounds using ML estimation, which are on the order of L/SNR. When α is large the error bound obtained here is $\sim \log(L)/\mathsf{SNR}$, which may represent a significant scaling improvement over the ML-based estimation methods.

Finally, note that for our results to hold M must exceed some minimum value as specified in (7). Without specifying a relationship between N and L, and given that the modular arithmetic nature of Procedure 1 makes it difficult to precisely quantify $|\mathcal{P}|$ (though we have $|\mathcal{P}| \leq M$ trivially), the spectral efficiency of the proposed method is difficult to quantify analytically. In Section 6 we examine the estimation error as a function of the number of pilot tones $|\mathcal{P}|$ via simulation.

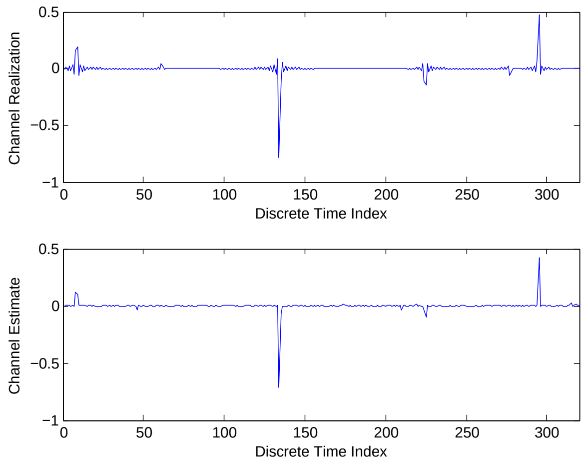

Fig. 1. A typical channel realization and estimate. The channel was generated with $S = 6$ point scatterers with a discrete channel length of $L = 320$. Estimation was performed with pilots selected from a degree $R = 2$ polynomial.

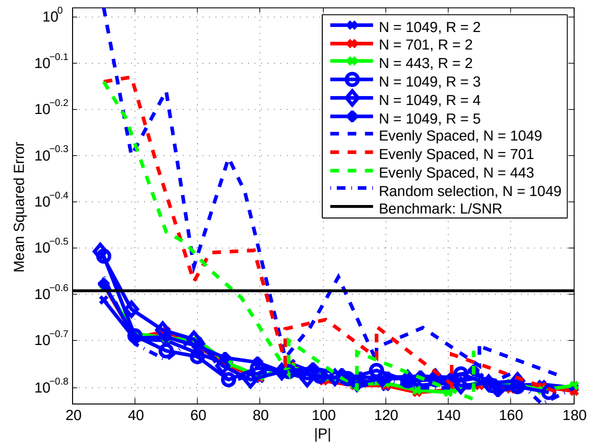

Fig. 2. Numerical experiments plotting mean squared-error versus number of pilot tones for various Fourier matrix subselections. Line color designates the primes N while line style designates pilot selection method.

6. NUMERICAL EXPERIMENTS AND DISCUSSION

In this section, we present results from numerical experiments which illustrate and verify our results. Though our procedure deterministically selects the set of pilot tones, we use Monte Carlo experiments with random channel realizations and noise.

An important and novel feature of our results is their applicability to compressible channels (compared to the strictly sparse channels addressed in [17]). This is reflected in the model used for h. In the numerical experiments, we generate h as the convolution of the response from S point scatterers with that of a low pass filter. This is given by

$$h_j = \sum_{i=1}^{S} \beta_i \operatorname{sinc}(j - W\tau_i) \tag{8}$$

where β_i and τ_i are the coefficients and continuous-time delays of the multipath scatterers, while W is the two-sided bandwidth of the system. Since the τ_i's are not taken to be multiples of $1/W$, h is not strictly sparse. We again refer the reader to [10] for further discussion of the relationship between the discrete-time mathematical model and the continuous-time physical setup.

More particularly, for our simulations we take $W = 25.12$MHz with $S = 6$ scatterers and τ_i distributed uniformly on $[0, 12.7\mu\text{sec}]$. These parameters are based on the "Brazil B" channel reported in [21]. Sampled at $1/W$, this results in a discrete channel length of $L = 320$. Further, we generate β_i as zero-mean Gaussian and normalize h such that $\|h\|_2 = 1$. A typical channel response is shown in Figure 1. Outside of the generation of h, we use $\sigma = 0.02\sqrt{2}$ and take the training data to have energy $\mathcal{E}_{tr} = 1$.

Results from our numerical experiments are shown in Figure 2 where we plot the mean squared-error (MSE) $\|\hat{h} - h\|_2^2$ of the channel estimate, averaged over 100 Monte Carlo trials, as a function of the number of pilot tones $|\mathcal{P}|$. We include results using various primes N and polynomials $Q(m)$. For comparison with previous work, we also include recovery using the Dantzig selector from both equally spaced and randomly selected pilot tones.

To place the results in context, note that the classic results of [5] require $|\mathcal{P}| = L$ pilot tones to achieve an MSE of $\frac{L}{\text{SNR}} = 0.256$. Figure 2 shows that this MSE is achieved by our procedure using approximately 30 rather than 320 pilots while using additional pilots reduces the MSE further. Moreover, as a function of $|\mathcal{P}|$ the MSE of our procedure matches that of randomly selected tones. Random subselections of Fourier matrices, studied in the context of channel estimation in [13], are known to have near-optimal RIP guarantees. In contrast, our selection performs as well while having the advantage of being deterministically constructed.

Though the bounds of (6) suggest that recovery can be a function of the prime N and polynomial degree $R \geq 2$, a strong dependence was not found empirically. This is apparent since the plots for various N and polynomials roughly coincide. This, however, may be due to our theory providing worst-case guarantees over possible channel responses and polynomials which requires exhaustive search to validate empirically. The polynomials chosen for the experiments displayed in Figure 2 have the coefficients $a_i = 1$ for $i = 1, \ldots, R$.

An examination of the numerical experiments with evenly spaced pilots shows the utility of selecting pilots using the non-linear polynomials $Q(m)$. When $|\mathcal{P}| = L$, [5] finds the optimal set of pilots to be evenly spaced within the N subcarriers. Though the estimation technique of [5] becomes ill-posed, one might expect that evenly spaced tones may also be effective when $|\mathcal{P}| < L$ and the Dantzig selector is used to estimate h. We test this idea numerically and find this is not the case. Figure 2 shows that pilot tones selected as our procedure prescribes outperform evenly spaced tones, allowing superior estimation using fewer pilots.

Finally, the MSE of our procedure can be further reduced by "debiasing" the estimate. As noted in [11], the estimate of the Danztig selector is improved when, after the initial estimate, a least-squares step is performed to fit the data on the estimate's support. For example, with 180 pilot tones, Figure 2 shows our procedure results in a MSE of 0.16. But with the additional debiasing step, the MSE is reduced to 0.10. In the interest of space, we have chosen not to display the debiased results in this exposition.

7. CONCLUSIONS

In this paper we proposed a general purpose procedure for deterministic selection of pilot tones for estimation of approximately sparse (or compressible) multipath channels in single-antenna OFDM systems. Our approach utilized estimation techniques from the compressed sensing literature, and was based on establishing the RIP for certain deterministically subsampled discrete Fourier transform matrices, as in [20].

It is interesting to note that our pilot tone selection procedure provides considerable flexibility when selecting parameters, such as the order R of the polynomial $Q(m)$ and its integer coefficients $\{a_i\}_{i=1}^{R}$. That said, our guarantees apply to *all* selections of these parameters which satisfy the conditions outlined in Procedure 1. It would be illustrative to perform a more comprehensive evaluation of the performance of our pilot tone selection and channel estimation procedure to determine whether there is a strong dependence on the selection of these parameters over a wide range of possible choices. We defer this to a future effort.

8. APPENDIX

8.1. Proof of Lemma 5.1

Assume that a matrix Z has unit-norm columns, then the *worst-case coherence* $\mu(Z)$ of Z is defined to be the largest (in magnitude) inner product between unique columns of Z. Formally, if Z_i denotes the ith column of Z, then

$$\mu(Z) = \max_{i,j,\ i\neq j} \left| Z_i^H Z_j \right|.$$

A general procedure for parlaying the coherence of a matrix into a statement of RIP was described, for example, in [22, 23]. In particular, Geršgorin's theorem [24] can be applied to bound the extremal eigenvalues of the Gram matrix $G = Z^H Z$. It follows that for a specified δ_S a matrix Z with coherence $\mu(Z)$ (and unit-norm columns) satisfies RIP(S, δ_S) for $S \leq \delta_S / \mu(Z)$.

Our goal, then, is to obtain an upper bound on the coherence of the matrix

$$\Psi = (\mathcal{E}_{tr})^{-1/2} D_{tr} A,$$

where $D_{tr} = \mathsf{diag}(d_{tr})$ and A is an $N_{tr} \times L$ matrix that comprises $\{\begin{bmatrix} F_{p0} & F_{p1} & \ldots & F_{p(L-1)} \end{bmatrix} : p \in \mathcal{P}\}$ as its rows. Let Ψ_i, $i = 1, 2, \ldots, L$ denote the columns of Ψ. Then, the entries of the Gram matrix G of Ψ are given by the expression

$$\begin{aligned} G(k,\ell) &= \sum_{p\in\mathcal{P}} \frac{C_p}{M} F_{pk}^* F_{p\ell} \\ &= M^{-1} \sum_{p\in\mathcal{P}} C_p \exp\left(-j\frac{2\pi}{N} p(\ell - k)\right). \end{aligned}$$

Now, notice that because of how we defined the multiplicity terms $\{C_p\}_{p\in\mathcal{P}}$, we can write the expression for $G(k, \ell)$ equivalently as a sum over the indices $m = 1, 2, \ldots, M$, in terms of the elements of the multiset $\mathcal{T}$. That is,

$$G(k,\ell) = M^{-1} \sum_{m=1}^{M} \exp\left(-j\frac{2\pi}{N}(\ell - k)Q(m)\right).$$

Note that $G(k, k) = 1$, implying that the columns of Ψ are unit norm.

It remains to bound the coherence of Ψ, which is just the largest value of $|G(k, \ell)|$ when $k \neq \ell$. For that, we follow the approach described in [20]. We make use of the following lemma which is credited to H. Weyl [25], and appears in its present form in [26].

Lemma 8.1 (Weyl). *Let $R \geq 2$, and let $P(m) = b_1 m + \cdots + b_R m^R$, where $b_R = \alpha/N + \theta/N^2$, $|\theta| \leq 1$, and $\gcd(\alpha, N) = 1$. If, for $0 < \epsilon_2 < \epsilon_1 < 1$, the condition $M^{\epsilon_1} \leq N \leq M^{R-\epsilon_1}$ holds for some integer M, then*

$$M^{-1}\left|\sum_{m=1}^{M} \exp\left(2\pi j P(m)\right)\right| \leq \gamma(R,\epsilon_2)\cdot M^{(\epsilon_2-\epsilon_1)/2^{R-1}}.$$

For completeness, we note the constant $\gamma(R, \epsilon_2)$ is given by

$$\gamma(R,\epsilon_2) = 2\left[\left(\frac{64R}{\epsilon_2}\right)\left(\frac{R^2}{\epsilon_2\log 2}\right)^{\exp\left(R^2/\epsilon_2\right)} R!\right]^{1/2^{R-1}}.$$

Let $P(m) = (\ell - k)Q(m)/N$, then $b_R = (\ell - k)a_R$. For $\ell, k \in \{0, 1, 2, \ldots, N - 1\}$ and $\ell \neq k$, it follows that $\gcd(b_R, N) = 1$ since a_R was selected to be relatively prime to N. If $M \geq N^{1/(R-\epsilon_1)}$ then Lemma 8.1 implies that the worst-case coherence of Ψ is no more than $\gamma(R, \epsilon_2) \cdot M^{(\epsilon_2-\epsilon_1)/2^{R-1}}$. The stated results follow from Geršgorin's theorem.

9. REFERENCES

[1] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, NY, 2005.

[2] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[3] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[4] J. Rinne and M. Renfors, "Pilot spacing in orthogonal frequency division multiplexing systems on practical channels," *IEEE Trans. Consumer Electron.*, vol. 42, no. 4, pp. 959–962, Nov. 1996.

[5] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.

[6] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, vol. 25, no. 1, pp. 4–27, Jan. 2000.

[7] ATSC, "Recommended practices for digital television: Receiver performance guidelines," Tech. Rep., 2004.

[8] A. F. Molisch, "Ultrawideband propagation channels-Theory, measurement, and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1528–1545, Sept. 2005.

[9] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Proc. 13th IEEE Digital Signal Processing Workshop*, Marco Island, FL, Jan. 2009, pp. 320–325.

[10] W. U. 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, June 2010.

[11] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[12] "IEEE Signal Processing Mag. Special Issue on Compressive Sampling," Mar. 2008.

[13] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. Annu. Conf. Information Sciences and Systems*, March 2008, pp. 5–10.

[14] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, April 2008, pp. 3469–3472.

[15] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, April 2008, pp. 2885–2888.

[16] Y. Chi, A. Gomaa, N. Al-Dhahir, and A. R. Calderbank, "Training signal design and tradeoffs for spectrally-efficient multi-user MIMO-OFDM systems," *IEEE Transactions on Wireless Communication*, 2010, to appear.

[17] P. Schniter, "Joint estimation and decoding for sparse channels via relaxed belief propagation," in *Proc. 44th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2010.

[18] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, December 2005.

[19] W. U. Bajwa, *New Information Processing Theory and Methods for Exploiting Sparsity in wireless Systems*, Ph.D. thesis, University of Wisconsin-Madison, Madison, WI, 2009.

[20] J. Haupt, L. Applebaum, and R. Nowak, "On the restricted isometry of deterministically subsampled Fourier matrices," in *Proc. 44th Annual Conf. on Information Sciences and Systems*, Princeton, NJ, March 2010.

[21] ABERT/SET Group, "Digital television systems - Brazilian testes - Final report," Tech. Rep., Sociedade Brasileira de Engenharia de Televisão.

[22] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inform. Theory*, vol. 56, no. 11, pp. 5862–5875, November 2010.

[23] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Proc.*, vol. 58, no. 10, pp. 5030–5043, October 2010.

[24] R. S. Varga, *Geršgorin and His Circles*, Number 36 in Springer Series in Computational Mathematics. Springer-Verlag, Berlin, Germany, 2004.

[25] H. Weyl, "Über die Gleichverteilung von Zahlen mod. Eins," *Math. Ann.*, vol. 77, pp. 312–352, 1916.

[26] N. Korobov, *Exponential sums and their applications*, Kluwer Academic Publishers, 1992.