A Useful Performance Metric for Compressed Channel Sensing

Matthew Sharp, *Member, IEEE*, and Anna Scaglione, *Fellow, IEEE*

***Abstract*—Recently, new progress has been made in using basis expansion models for system identification with compressed sensing. To aid the application of these methodologies, we introduce a metric, called localized coherence, for choosing input signals that result in better estimation performance. Its definition is motivated through the analysis of the normalized mean Euclidean error of the channel estimate and its efficacy is demonstrated through numerical simulations.**

***Index Terms*—Channel estimation, compressed sensing, system identification, sparsity.**

I. Introduction

SINCE the early work of Bello [1], the estimation of doubly-selective channels characterized by both significant Doppler and delay spread has been considered a challenging problem. Basis Expansion Models (BEM) are often employed to estimate the time-varying channel impulse response by constructing a basis to model the time-variations of each tap of the discrete-time impulse response, [2], [3]. The basis used to model the time-variations of the impulse response is typically constrained by the number of observations and impulse response length. In this work we consider a specific class of channels consisting of a few dominant propagation paths, each contributing potentially large Doppler shifts and delays.

Different from typical BEMs for doubly selective channels, we consider an overcomplete basis, and thus the system identification problem is underdetermined. More specifically, samples of the observed signal are modeled as a sparse linear combination of columns from an overcomplete basis plus additive white Gaussian noise. This is similar to models found in [4]–[7] (focused on frequency-selective block fading channels) and in [8]–[12] (dedicated to doubly-selective channels). In many of these works, the *sensing matrix* resulting from the overcomplete basis expansion of the channel, combined with the training signal, is designed to satisfy certain Restricted Isometry Properties (RIP) or constraints on the mutual coherence (MC). As shown in [13], [14], these are sufficient conditions to guarantee identifiability in the absence of noise, and stability in the presence of noise for methods such as as Basis Pursuit [15] and Matching Pursuit [16]. These conditions, however, are mainly concerned with guaranteeing that every subset of S or fewer columns of the sensing matrix approximately form an orthonormal system, where S is the sparsity of the signal of interest. The RIP and MC constraints not only impose requirements on the input, but also on the class of systems that can be identified. In practice, as a result of modeling errors, there is considerable degradation in performance when one applies the restricted class of models to actual doubly selective systems. A looser condition, aimed at guaranteeing identifiability of a restricted class of models, was provided in [17]. However, the criterion in [17] provides no guarantees on the channel estimation performance in the presence of noise, which is the objective of this work.

The work in [18] relaxed the RIP and MC conditions when considering the problem of recovering signals from undersampled data, in situations where the signal is not sparse or compressible in an incoherent dictionary, but in a truly redundant dictionary. The important aspect of [18] is the acknowledgement that redundancy in the overcomplete channel basis expansion does not necessarily imply problems in estimating the channel. Our work, like [18], [19], indicates that it is not the high mutual coherence of the channel basis that drives the channel estimation performance, but rather the type of error associated with exchanging specific columns of this basis. Based on this analysis, we propose a new metric called *localized coherence* that can be used to compare different input designs, for a fixed quantization of the parameter space, in order to determine those that lead to better channel estimation performance. The derivation of this metric is based on the analysis of a combinatorial algorithm searching for the sparsest solution to a set of linear equations. In order to obtain a simple measure, the resulting metric includes only the error contribution of a sparsity model with support equal to one, which we argue should represent the overall trend of the error bound. This is confirmed by our simulation results, where estimates are formed considering support sets consisting of more than a single component. Our metric is a way of ranking input designs without restrictions on the parameterization of the system other than the minimal restrictions in [17].

II. System Model

We consider the transmission of a linearly modulated signal $x(t) = \sum_{n=-N_{cp}}^{N} x[(n)_N] g(t-nT_s)$, containing a cyclic prefix of length N_{cp} where $(n)_N$ denotes the number n modulo N. At the receiver we observe the sampled output of a multi-path channel consisting of Q propagation paths with input $x(t)$. Each path is defined by an attenuation a_q, a Doppler shift $\tilde{f}_q$ Hz, and delay $\tilde{\tau}_q$ seconds which lie in a bounded interval of frequency and time respectively. Introducing the pair of normalized Doppler and delay shifts $f_q = \tilde{f}_q T_s$ and $\tau_q = \tilde{\tau}_q / T_s$, where T_s is the sample spacing, we can express the received signal as a function of the vectors $\boldsymbol{\kappa}_q = (f_q, \tau_q) \in (-f_{\max}/2, f_{max}/2) \times (\tau_{min}, \tau_{max})$. Assuming $N_{cp} > \tau_{\max}$, and removing possible inter-block interference by taking the N samples $k = 0, \ldots, N-1$

$$\begin{aligned} y[k] &= \sum_{q=1}^{Q} e^{j2\pi f_q k} x((k-\tau_q)T_s) a_q + w[k] \\ &= \sum_{n=-N_{cp}}^{N} x[(n)_N] h[k, k-n] + w[k] \end{aligned} \tag{1}$$

where $h[k, k-n] = \sum_{q=1}^{Q} e^{j2\pi f_q k} g((k-n-\tau_q)T_s) a_q$ is the time-varying impulse response and $w[k] \overset{i.i.d.}{\sim} \mathcal{CN}(0, \sigma_w^2)$. Assuming the pulse $g(t)$ is causal with finite duration T_p sec., and defining $L-1 = \lceil \tau_{\max} + T_p/T_s \rceil$, we note $h[k,l] = 0$ for $l < 0$ and $l > L-1$, and have the following equivalent model:

$$y[k] = \sum_{l=0}^{L-1} x[(k-l)_N] h[k,l] + w[k], k = 0, \ldots, N-1. \tag{2}$$

Let $\boldsymbol{\kappa} = [\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_Q]^T$ and $\mathbf{a} = [a_1, \ldots, a_Q]^T$. Collecting these N observations in a vector

$$\mathbf{y} = \boldsymbol{\mathcal{X}} \mathbf{h} + \mathbf{w} = \boldsymbol{\mathcal{X}} \boldsymbol{G}(\boldsymbol{\kappa}) \mathbf{a} + \mathbf{w} \tag{3}$$

Manuscript received July 14, 2010; revised October 23, 2010 and February 10, 2011; accepted February 10, 2011. Date of publication March 07, 2011; date of current version May 18, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Huaiyu Dai. This work was supported by ONR contract No. N00014-05-C-0070.

M. Sharp is with Cornell University, Ithaca, NY 14853 USA (e-mail: mds72@cornell.edu).

A. Scaglione is with the University of California, Davis, CA 95616 USA(e-mail: ascaglione@ucdavis.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2123892

where the NL length vector $\{\mathbf{h}\}_{l+kL} = h[k,l]$, and the matrices[1] $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{G}(\boldsymbol{\kappa})$

$$\{\boldsymbol{\mathcal{X}}\}_{k,l+uL} = x[(k-l)_N]\delta[u-k]$$
$$\{\mathbf{G}(\boldsymbol{\kappa})\}_{l+kL,q} = e^{j2\pi f_q k} g((l-\tau_q)T_s). \tag{4}$$

In the following, we assume that the number of coefficients of the time-varying impulse response $\mathbf{h}$ is $NL \gg Q$. We further assume that each parameter vector $\boldsymbol{\kappa}_q$ lies in a bounded region of $\mathbb{R}^2$ where we have placed a finite grid of $R = R_f R_\tau$ points (f_{r_f}, τ_{r_τ}), each identified by an index $\mathbf{r} = (r_f, r_\tau)$. Suppose that for every parameter vector $\boldsymbol{\kappa}_q$ in the actual model there exists a nearby point in the grid indexed by $(r_f^{(q)}, r_\tau^{(q)})$. Assuming a fine quantization of this parameter space

$$\mathbf{y} = \underbrace{\boldsymbol{\mathcal{X}}\boldsymbol{G}}_{\mathbf{U}}\boldsymbol{\alpha}^\star + \mathbf{w} + \varepsilon_{\mathbf{y}}, \mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star + \varepsilon_{\mathbf{h}} \tag{5}$$

where $\varepsilon_{\mathbf{y}}$ and $\varepsilon_{\mathbf{h}}$ represent the modeling error that results from approximating the parameter vectors $\boldsymbol{\kappa}_q$ on the grid. The $N \times R$ matrix $\mathbf{U}$ and $NL \times R$ matrix $\boldsymbol{G}$, and the vector $\boldsymbol{\alpha}^\star$ are defined as

$$\{\mathbf{U}\}_{k,r_f+r_\tau R_f} = \sum_{l=0}^{L-1} x[(k-l)_N] e^{j2\pi k f_{r_f}} g((l-\tau_{r_\tau})T_s) \tag{6}$$

$$\{\boldsymbol{G}\}_{l+kL,r_f+r_\tau R_f} = e^{j2\pi k f_{r_f}} g((l-\tau_{r_\tau})T_s) \tag{7}$$

$$\{\boldsymbol{\alpha}^\star\}_{r_f+r_\tau R_f} = \sum_{q=1}^{Q} a_q \delta[r_f - r_f^{(q)}]\delta[r_\tau - r_\tau^{(q)}]. \tag{8}$$

If the parameters $\boldsymbol{\kappa}_q$ fall on the aforementioned grid, $\varepsilon_{\mathbf{y}} = \varepsilon_{\mathbf{h}} = \mathbf{0}$. Alternatively, one possible approximation is obtained by considering the Taylor expansion of the model with respect to $\boldsymbol{\kappa}_q$ around the closest grid point $(f_{r_f^{(q)}}, \tau_{r_\tau^{(q)}}) = (\arg\min_{f_{r_f}} |f_{r_f} - f_q|, \arg\min_{\tau_{r_\tau}} |\tau_{r_\tau} - \tau_q|), q = 1, \ldots, Q$

$$\begin{aligned} h[k,l] &= \sum_{q=1}^{Q} e^{j2\pi f_q k} g((l-\tau_q)T_s) a_q \\ &= \sum_{q=1}^{Q} e^{j2\pi f_{r_f^{(q)}} k} g((l-\tau_{r_\tau^{(q)}})T_s) a_q + \epsilon_h[k,l] \end{aligned} \tag{9}$$

$$\begin{aligned} \epsilon_h[k,l] = \sum_{q=1}^{Q} e^{j2\pi f_{r_f^{(q)}} k} & [j2\pi k g((l-\tau_{r_\tau^{(q)}})T_s)(f_{r_f^{(q)}} - f_q) \\ & + T_s \dot{g}((l-\tau_{r_\tau^{(q)}})T_s)(\tau_{r_\tau^{(q)}} - \tau_q)] \\ & + O((\tau_{r_\tau^{(q)}} - \tau_q)^2) \\ & + O((f_{r_f^{(q)}} - f_q)^2) \\ & + O((f_{r_f^{(q)}} - f_q)(\tau_{r_\tau^{(q)}} - \tau_q)). \end{aligned} \tag{10}$$

We note that the error mismatch due to the normalized frequency in (10) increases with the observation window duration N, while the normalized delay error in the second term of (10) depends on T_s times the derivatives of the pulse, whose energy scales approximately with the pulse bandwidth $B \geq 1/T_s$. Considering the impact of (9) on (5), given that $\boldsymbol{\mathcal{X}}$ is bounded, we can say that the derivatives of the columns of $\mathbf{U}$ with respect to $\boldsymbol{\kappa}_q$ are bounded. Therefore, we will say that *the grid is fine* when the error between $\boldsymbol{\kappa}_q$ and the closest grid point is very small compared to the magnitude of these derivatives, roughly $|f_{r_f^{(q)}} - f_q| < (1/N)$ and $|\tau_{r_\tau^{(q)}} - \tau_q| < 1$; furthermore, we know there exist vectors $\boldsymbol{\alpha}$ with only Q nonzero elements that result in vanishing modeling error as the grid is made very fine. The following analysis is applicable for any case where there exists a solution $\|\boldsymbol{\alpha}^\star\|_0 \leq N/2$ with negligible modeling error, though for simplicity we assume that there exists a $\|\boldsymbol{\alpha}^\star\|_0 = Q$ with $\varepsilon_{\mathbf{y}} = \varepsilon_{\mathbf{h}} = \mathbf{0}$, unless otherwise stated. We relax this assumption in our numerical analysis. The support of the vector $\boldsymbol{\alpha}$ over the basis $\mathbf{U}$ (as well as $\boldsymbol{G}$) is denoted as $\mathcal{Q} = (q_1, q_2, \ldots, q_Q)$, while the Q length vector $\boldsymbol{\alpha}_{\mathcal{Q}}$ contains these nonzero components. Similarly, the matrices $\mathbf{U}_{\mathcal{Q}}$ and $\boldsymbol{G}_{\mathcal{Q}}$ consist only of those Q columns.

Since by assumption the matrix $\boldsymbol{\mathcal{X}}$ has more columns than rows and a right null-space, there exist many solutions to the above problem, and it is problematic if our desire is to estimate $\mathbf{h}$. However, for those channels that have a unique sparsest representation $\mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star$ with $Q \ll NL$, in the absence of noise it is possible to recover the channel $\mathbf{h}$ by searching for the sparsest solution

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \text{ subj. to } \mathbf{y} = \mathbf{U}\boldsymbol{\alpha} \tag{11}$$

provided that certain conditions are met. For $\|\boldsymbol{\alpha}^\star\|_0 \leq Q$, the recovery of the channel is guaranteed if every subset of $2Q$ columns of $\mathbf{U}$ are linearly independent. These inputs are sufficiently informative (SI) for such sparse systems [17]. In the presence of noise, the above algorithm can be modified as (see [14])

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \text{ subj. to } \|\mathbf{y} - \mathbf{U}\boldsymbol{\alpha}\|_2^2 \leq \eta \tag{12}$$

where η is, with high probability, an upperbound on the noise level $\|\mathbf{w}\|^2$. Since $\mathbb{E}\{\|\mathbf{w}\|_2^2\} = N\sigma_w^2$, an appropriate choice of η should scale with the noise variance, such as[2] $\eta = C_0 N \sigma_w^2$.

We note that (12) is combinatorial in nature and many authors have studied less complex alternatives including greedy approaches such as matching pursuit (MP) and orthogonal matching pursuit (OMP) [16], [20], and the l_1 relaxation of (12) (obtained by replacing $\|(\cdot)\|_0$ with $\|(\cdot)\|_1$) termed Basis Pursuit or l_1-synthesis [15], [18]. In [18], the authors introduce a criterion and bound on the estimation error of $\mathbf{h}$ using the optimization $\hat{\mathbf{h}} = \arg\min_{\tilde{\mathbf{h}}} \|\boldsymbol{G}^H \tilde{\mathbf{h}}\|_1$ s.t.$\|\mathbf{y} - \boldsymbol{\mathcal{X}}\tilde{\mathbf{h}}\|_2^2 \leq \eta$, called l_1-analysis, mentioning in Section IV that the l_1-synthesis and analysis problems generally provide very different results. While l_1-analysis may provide reasonable results for certain channel estimation problems, the error bound derived in [18] is applicable when the matrix $\boldsymbol{\mathcal{X}}$ satisfies a form of the RIP adapted to the matrix $\boldsymbol{G}$, called the D-RIP. One could potentially reduce the error bound by reducing the D-RIP of the matrix $\boldsymbol{\mathcal{X}}$ in (4), but we are not aware of a tractable approach for computing the D-RIP of various deterministic input sequences. We have observed that the D-RIP constrained to vectors consisting of a single column of $\boldsymbol{G}$ has a similar trend to the localized coherence for those input sequences examined in Section V, though the implications of this trend are not immediately clear.

The rest of this paper is devoted to deriving and demonstrating the usefulness of the metric we call *localized coherence*, which can be used to compare different input sequence designs in terms of their affects on channel estimation performance when using an algorithm similar to (12). Specifically, the localized coherence ψ is defined as

$$\psi = \sum_{q=1}^{R} \sum_{s \neq q} \frac{d(\{q\},\{s\}) e^{\eta}}{RN\left(1 + a\left(\|\mathbf{u}_q\|_2^2 - \frac{|\mathbf{u}_q^H \mathbf{u}_s|^2}{\|\mathbf{u}_s\|^2}\right)\right)} \tag{13}$$

where $\mathbf{u}_i$ denotes the ith column of the $N \times R$ matrix $\mathbf{U}$ in (12) and $d(\{q\},\{s\})$ is the error in modeling the channel that results from exchanging the qth column of $\boldsymbol{G}$, assumed to be matching the channel model, with the sth column, when the sparsity of the channel is assumed to be one, as will be more clearly explained in the next section.

[1] In the purely time-selective channel the matrix $\boldsymbol{\mathcal{X}}$ is an $\boldsymbol{N \times N}$ diagonal matrix while in the purely frequency-selective channel $\boldsymbol{\mathcal{X}}$ is a $\boldsymbol{N \times L}$ column circulant matrix.

[2] The choice of $\boldsymbol{C}_0$ is typically determined via trial and error. We note that choosing $\boldsymbol{C}_0$ too large or small may significantly deteriorate the estimation performance of the algorithm.

III. Sparse Recovery Algorithm Error Analysis

In the following, we denote $\mathcal{R} = \{1, \ldots, R\}$ as the set of all column indices of $\mathbf{U}$ and $2^{\mathcal{R}}$ as its power set. The sparse recovery algorithm we analyze is based on (12). We assume that every subset of N columns of the $N \times R$ matrix $\mathbf{U}$ is linearly independent, which is sufficient to guarantee the identifiability of channels with $\|\boldsymbol{\alpha}^\star\|_0 \leq N/2$ in the absence of noise. For deterministic $\boldsymbol{\alpha}$ this assumption implies that, for each set $\mathcal{S}$ with $|\mathcal{S}| \leq N$, the estimate $\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = (\mathbf{U}_{\mathcal{S}}^H \mathbf{U}_{\mathcal{S}})^{-1} \mathbf{U}_{\mathcal{S}}^H \mathbf{y}$ exists since $\mathbf{U}_{\mathcal{S}}$ is full rank and it minimizes $\|\mathbf{y} - \mathbf{U}_{\mathcal{S}} \boldsymbol{\alpha}_{\mathcal{S}}\|_2^2$, which in turn means that if the constraint in (12) can be met with set $\mathcal{S}$, then $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$ must be a valid solution. We note that this would be the maximum likelihood solution if the support $\mathcal{Q}$ were known *a priori*, and for this reason $\hat{\boldsymbol{\alpha}}_{\mathcal{Q}}$ is sometimes referred to as the Oracle estimate. Replacing $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$ in the constraint defined in (12), and adding the constraint that the solution must satisfy $|\mathcal{S}| \geq 1$

$$\tilde{\mathcal{S}} = \underset{\mathcal{S}' \in 2^{\mathcal{R}}}{\operatorname{argmin}} |\mathcal{S}'| \text{ subj. to } \|\boldsymbol{\Pi}_{\mathcal{S}'}^{\perp} \mathbf{y}\|_2^2 \leq \eta, |\mathcal{S}'| \geq 1 \tag{14}$$

where the matrix $\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \triangleq \mathbf{I} - \mathbf{U}_{\mathcal{S}} (\mathbf{U}_{\mathcal{S}}^H \mathbf{U}_{\mathcal{S}})^{-1} \mathbf{U}_{\mathcal{S}}^H$ is the orthogonal complement to the column-space of $\mathbf{U}_{\mathcal{S}}$. Due to the assumption that every subset of N columns of $\mathbf{U}$ is linearly independent, the cardinality of the solution $\tilde{\mathcal{S}}$ will always be such that $|\tilde{\mathcal{S}}| \leq N$. In order to simplify the analysis, we define the sets

$$\mathcal{C}_1 = \left\{ \mathcal{S} : \|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \mathbf{y}\|_2^2 \leq \eta, |\mathcal{S}| \geq 1 \right\}$$

$$\mathcal{C}_2 = \{ \mathcal{S} : \mathcal{S} = \arg \min_{\mathcal{S}' \in \mathcal{C}_1} |\mathcal{S}'| \}$$

where $\mathcal{C}_1$ is the set of supports satisfying the reconstruction error constraint in (14) and $\mathcal{C}_2$ is the set of all solutions to (14). There may be multiple solutions $\tilde{\mathcal{S}}$ to (14) and there are several ways of resolving these ambiguities; for example, selecting an element from $\mathcal{C}_2$ at random or the one minimizing $\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \mathbf{y}\|_2^2$. In the following, we define $\xi(\cdot) : \mathcal{C}_2 \rightarrow \mathcal{C}_2$ as the function resolving this possible ambiguity, and returning a single solution $\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$, though the following analysis will be agnostic as to its choice. The basic algorithm is summarized in Alg. 1.

In summary, the algorithm starts exploring all $\mathcal{S} : |\mathcal{S}| = 1$, to determine if any such $\mathcal{S}$ belongs to $\mathcal{C}_1$, including them in a temporary set Θ. If Θ is non empty, then $\mathcal{C}_2 = \Theta$, and the estimated set $\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$. Otherwise, the algorithm starts exploring all $\mathcal{S} : |\mathcal{S}| = 2$, and continues the same process until Θ is not empty and $\mathcal{C}_2$ is found. Though this algorithm is combinatorial in nature, its analysis can provide insight into the performance of similar algorithms attempting to find the sparsest solution, while constraining the reconstruction error $\|\mathbf{y} - \mathbf{U}\hat{\boldsymbol{\alpha}}\|_2^2$.

Algorithm 1: Sparse Algorithm

$s = 1, \text{cont} = 0, \Theta = \{\emptyset\}$
while cont $= 1$ **do**
  for $m = 1$ to $\binom{R}{s}$ **do**
    if $\|\boldsymbol{\Pi}_{\mathcal{S}^{(m)}}^{\perp} \mathbf{y}\|_2^2 \leq \eta$ **then**
      $\Theta \Leftarrow \{\Theta, \mathcal{S}^{(m)}\}$
    end if
  end for
  if $\Theta \neq \{\emptyset\}$ **then**
    $\mathcal{C}_2 \Leftarrow \Theta$, cont $= 0$
  else
    $s = s + 1$
  end if
end while
$\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$, $\hat{\mathbf{h}} = \boldsymbol{G}_{\hat{\mathcal{S}}} \hat{\boldsymbol{\alpha}}_{\hat{\mathcal{S}}}$

Our primary concern is not the error $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^\star\|_2$, but the channel estimation error

$$\|\hat{\mathbf{h}} - \mathbf{h}\|_2 \leq \|(\mathbf{I} - \boldsymbol{G}_{\hat{\mathcal{S}}} \mathbf{U}_{\hat{\mathcal{S}}}^{\dagger} \boldsymbol{\mathcal{X}}) \boldsymbol{G}_{\mathcal{Q}} \boldsymbol{\alpha}_{\mathcal{Q}}\|_2 + \|\boldsymbol{G}_{\hat{\mathcal{S}}} \mathbf{U}_{\hat{\mathcal{S}}}^{\dagger} \mathbf{w}\|_2 \tag{15}$$

where $\mathbf{U}_{\hat{\mathcal{S}}}^{\dagger}$ is the pseudoinverse of $\mathbf{U}_{\hat{\mathcal{S}}}$. The first term is the error in modeling the channel due to the particular estimate of the columns $\hat{\mathcal{S}}$, while the second term is the noise entering the system. For any $\hat{\mathcal{S}} \supseteq \mathcal{Q}$, the first term is zero, however, large values of $|\hat{\mathcal{S}}|$ generally result in larger values of the second term. Clearly, replacing $\hat{\mathcal{S}}$ with any of its subsets would reduce the contribution of the second term. Thus, a good estimator should strike a balance between increasing the modeling error and decreasing the noise in the estimate by picking a small model size $|\hat{\mathcal{S}}|$, and for this reason sparse estimation methods have the added benefit of denoising the estimate.

Though the channel MSE is difficult to analyze due to the nonlinear dependence of $\hat{\mathcal{S}}$ on $\mathbf{y}$, we will use the following bound on the normalized mean Euclidean error (NMEE) to gauge how the performance is affected by the choice of the input signal. Defining $C = \|\boldsymbol{\alpha}_{\mathcal{Q}}\|_2 \sqrt{N}$, and noting that $\|\mathbf{A}\boldsymbol{\alpha}\|_2 \leq \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})} \|\boldsymbol{\alpha}\|_2$, we have

$$\frac{\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2\}}{\|\boldsymbol{\alpha}_{\mathcal{Q}}\|_2 \sqrt{N}} \leq \frac{\mathbb{E}\{d(\mathcal{Q}, \hat{\mathcal{S}})\}}{\sqrt{N}} + \frac{\mathbb{E}\{\|\boldsymbol{G}_{\hat{\mathcal{S}}} \mathbf{U}_{\hat{\mathcal{S}}}^{\dagger} \mathbf{w}\|_2\}}{C} \tag{16}$$

where $d(\mathcal{Q}, \hat{\mathcal{S}}) = \sqrt{\lambda_{\max}(\boldsymbol{G}_{\mathcal{Q}}^H \mathbf{W}_{\hat{\mathcal{S}}}^H \mathbf{W}_{\hat{\mathcal{S}}} \boldsymbol{G}_{\mathcal{Q}})}$ with $\mathbf{W}_{\hat{\mathcal{S}}} = \mathbf{I} - \boldsymbol{G}_{\hat{\mathcal{S}}} \mathbf{U}_{\hat{\mathcal{S}}}^{\dagger} \boldsymbol{\mathcal{X}}$, and the expectation is with respect to $\hat{\mathcal{S}}$ and $\mathbf{w}$. It is important to remark that a good estimate does not necessarily require resolving all of the Q paths and the corresponding columns of $\boldsymbol{G}_{\mathcal{Q}}$, but rather from selecting a $\boldsymbol{G}_{\hat{\mathcal{S}}}$ whose column space well approximates that of $\boldsymbol{G}_{\mathcal{Q}}$. In the following, we will analyze the first term[3] in (16) to find a criterion that could be used as an alternative to the RIP and/or MC requirements. Specifically

$$e(\mathcal{Q}) = \mathbb{E}\{d(\mathcal{Q}, \hat{\mathcal{S}})\} = \sum_{n=1}^{N} \sum_{\mathcal{S}:|\mathcal{S}|=n} d(\mathcal{Q}, \mathcal{S}) p_{\hat{\mathcal{S}}}(\mathcal{S}). \tag{17}$$

In defining the probability law $p_{\hat{\mathcal{S}}}(\mathcal{S}) \triangleq \Pr(\mathcal{S} = \xi(\mathcal{C}_2))$, the random experiment consists of performing Alg. 1 on noisy observations such as (5) and the sample space Ω consists of the power set $\Omega = 2^{\mathcal{R}}$. Determining the probability of the random outcome $\hat{\mathcal{S}}$ in closed form is difficult. We note that $\Pr(\mathcal{S} = \xi(\mathcal{C}_2)) = \Pr((\mathcal{S} = \xi(\mathcal{C}_2)) \cap (\mathcal{S} \in \mathcal{C}_2)) \leq \Pr(\mathcal{S} \in \mathcal{C}_2)$. In a similar fashion, as $\mathcal{C}_2 \subseteq \mathcal{C}_1$, $\Pr(\mathcal{S} \in \mathcal{C}_2) = \Pr((\mathcal{S} \in \mathcal{C}_2) \cap (\mathcal{S} \in \mathcal{C}_1)) = \Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) \Pr(\mathcal{S} \in \mathcal{C}_1)$

$$e(\mathcal{Q}) \leq \sum_{n=1}^{N} \sum_{\mathcal{S}:|\mathcal{S}|=n} d(\mathcal{Q}, \mathcal{S}) \times \Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) \Pr(\mathcal{S} \in \mathcal{C}_1) \tag{18}$$

We remark that

$$\Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) = 1 - \Pr(\mathcal{S} \notin \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) = 1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1} |\mathcal{S}'| < |\mathcal{S}|). \tag{19}$$

[3]The analysis of the second term is complicated due to the fact that $\hat{\boldsymbol{\mathcal{S}}}$ and $\mathbf{w}$ are dependent. One could bound the second term by a constant times $\boldsymbol{N}\boldsymbol{\sigma}_w^2$, but this would not provide any additional insight on the error trend and we will not consider the second term for this reason.

Thus $\Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1)$ is monotonically decreasing in $|\mathcal{S}|$ and is unity for all $|\mathcal{S}| = 1$, as we have restricted $|\hat{\mathcal{S}}| \geq 1$. Intuitively, we expect that for larger values of η in (14), this probability will decrease faster in $|\mathcal{S}|$, while for very small values of η it remains large for all $|\mathcal{S}| \leq N$.

Assuming $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\Sigma})$, the term $\Pr(\mathcal{S} \in \mathcal{C}_1)$ is easily upper-bounded using the Chernoff bound

$$\begin{aligned} \Pr(\mathcal{S} \in \mathcal{C}_1) &= \Pr(\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\mathbf{y}\|^2 \leq \eta) \\ &\leq \frac{e^{\eta} e^{-\boldsymbol{\mu}_{\mathbf{y}}^H \boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\mathbf{I}+\boldsymbol{\Sigma}\boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}\boldsymbol{\mu}_{\mathbf{y}}}}{\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\boldsymbol{\Sigma} + \mathbf{I})}. \end{aligned} \tag{20}$$

In contrast with (19), we note that (20) is increasing in η. We use (20) under two assumptions: first that $\boldsymbol{\alpha}_{\mathcal{Q}}$ and $\mathcal{Q}$ are deterministic, where $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}$ and $\boldsymbol{\Sigma} = \sigma_w^2\mathbf{I}$, and second, that $\mathcal{Q}$ is deterministic but $\boldsymbol{\alpha}_{\mathcal{Q}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}\sigma_a^2)$ and independent of $\mathbf{w}$ where $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{U}_{\mathcal{Q}}\mathbf{U}_{\mathcal{Q}}^H\sigma_a^2 + \sigma_w^2\mathbf{I}$. We will relate these two probability measures later to extract our *localized coherence* metric. Denoting $\pi^{(d)}$ and $\pi^{(r)}$ as (20) in these different cases, respectively, we have

$$\pi^{(d)} \leq e^{\eta} e^{-\boldsymbol{\alpha}_{\mathcal{Q}}^H \mathbf{U}_{\mathcal{Q}}^H \boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\mathbf{I}+\sigma_w^2\boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}\mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}}(1+\sigma_w^2)^{|\mathcal{S}|-N} \tag{21}$$

$$\pi^{(r)} \leq e^{\eta}\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\mathbf{U}_{\mathcal{Q}}\mathbf{U}_{\mathcal{Q}}^H\sigma_a^2 + \mathbf{I}\sigma_w^2) + \mathbf{I})^{-1}. \tag{22}$$

For small values of σ_w^2, the second exponential term in (21) can be approximated as $e^{-\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}\|_2^2}$ such that (21) is largest when the column space of $\mathbf{U}_{\mathcal{S}}$ spans that of $\mathbf{U}_{\mathcal{Q}}$. Likewise, in (22), where $\boldsymbol{\alpha}_{\mathcal{Q}}$ is random, for small σ_w^2 we observe that the largest values of (22) similarly correspond to those cases where the column space of $\mathbf{U}_{\mathcal{S}}$ spans that of $\mathbf{U}_{\mathcal{Q}}$.

Combining (19) and (20), we have the following upper-bounds on $e(\mathcal{Q})$. For the case of deterministic $\boldsymbol{\alpha}_{\mathcal{Q}}$ we have $e(\mathcal{Q}) \leq \sum_{n=1}^{N}\phi^{(d)}(n)$, where [see (23) at the bottom of the page]. In the case of random $\boldsymbol{\alpha}_{\mathcal{Q}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}\sigma_a^2)$ we have $e(\mathcal{Q}) \leq \sum_{n=1}^{N}\phi^{(r)}(n)$, where [see (24) at the bottom of the page].

For appropriately chosen η, the highest probability sets $\mathcal{S}$ are those where the columns $\mathbf{U}_{\mathcal{S}}$ closely span the column space of $\mathbf{U}_{\mathcal{Q}}$ while maintaining minimum cardinality. We can infer that an appropriate design of the input signal should try to ensure $d(\mathcal{Q}, \mathcal{S})$ is small for these high probability sets $\mathcal{S}$. We also note that $1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1}|\mathcal{S}'| < n)$ is decreasing in n, indicating a lower probability of selecting large support sizes $|\mathcal{S}|$. Further, this term decreases more rapidly for larger values of η, signaling a tradeoff between this term and e^{η} when choosing values of η. As it is intractable to analyze these bounds for all n, due to their combinatorial nature and the need to compute $1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1}|\mathcal{S}'| < n)$, we examine the case of $n = 1$ assuming $|\mathcal{Q}| = 1$ and for negligible noise $\sigma_w^2 \approx 0$. For $|\mathcal{S}| = 1$ and $|\mathcal{Q}| = 1$, we have $\mathcal{S} = \{s\}$ and $\mathcal{Q} = \{q\}$ and we define $\mathbf{u}_q \triangleq \mathbf{U}_{\{q\}}$, $\mathbf{u}_s \triangleq \mathbf{U}_{\{s\}}$ and $\alpha_q = \boldsymbol{\alpha}_{\{q\}}$. Noting that $e^{-|x|} \leq (1 + |x|)^{-1}$, we have the following approximation for both $\phi^{(d)}(1)$ and $\phi^{(r)}(1)$

$$\phi(1) \approx \sum_{\substack{\{s\}\in\mathcal{R} \\ \{s\}\neq\{q\}}} d(\{q\}, \{s\})e^{\eta} \times \underbrace{\left(1 + a\left(\|\mathbf{u}_q\|^2 - \frac{|\mathbf{u}_q^H\mathbf{u}_s|^2}{\|\mathbf{u}_s\|_2^2}\right)\right)^{-1}}_{p_{\phi}(\mathbf{u}_q, \mathbf{u}_s)} \tag{25}$$

where $a = \|\alpha_q\|^2$ for deterministic $\boldsymbol{\alpha}_{\mathcal{Q}}$ ($a = \sigma_a^2$ if random). Assuming each column of $\mathbf{U}$ is such that $\|\mathbf{u}_s\|_2^2 \approx c$, the $\mathbf{u}_s$ with largest $p_{\phi}(\mathbf{u}_q, \mathbf{u}_s)$ are those that are highly correlated with $\mathbf{u}_q$. For example, if $|\mathbf{u}_q^H\mathbf{u}_s|^2 \approx c^2$ then $p_{\phi}(\mathbf{u}_q, \mathbf{u}_s) \approx 1$. On the other hand, when $\mathbf{u}_q$ and $\mathbf{u}_s$ are approximately orthogonal $p_{\phi}(\mathbf{u}_q, \mathbf{u}_s) \approx (1 + ac)^{-1}$.

IV. Localized Coherence

In Section II, we assumed that the quantization of the parameter space of $\boldsymbol{\kappa}_q$ lead to a sparse system, which we later used as if it were the exact representation of the observations. The dilemma encountered is whether or not to refine the quantization when the parameters $\boldsymbol{\kappa}_q$ do not truly lie on the grid. Though a fine grid generally results in larger mutual coherence, it is easy to numerically verify that a fine grid is preferable to a coarse grid in terms of channel impulse response MSE performance, even when using suboptimal approximations to (14), such as OMP or Basis Pursuit, as demonstrated in Fig. 1. This has also been observed in [19], [21] and indicates that RIP and MC constraints should be complemented with other tools. The issue still remains open as to what constitutes a good input signal once a fine grid has been chosen and the RIP and MC constraints are no longer satisfied. We use the previous analysis of the NMEE to supply a possible guideline.

From the preceding analysis we observed that in high SNR situations and assuming an appropriate choice of η the most likely errors made in trying to find a sparse solution correspond to the selection of sets of columns $\mathcal{S}$ of $\mathbf{U}$ spanning a space very similar to that spanned by the set of columns $\mathcal{Q}$, and that have similar or smaller support. Assuming that the bounds obtained are reasonably tight, to reduce the channel estimation error, the input design should be such that if $\mathbf{U}_{\mathcal{S}}$ closely spans $\mathbf{U}_{\mathcal{Q}}$, then the corresponding values of $d(\mathcal{Q}, \mathcal{S})$ are small.

We observe that for small σ_w^2, the terms $\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\mathbf{U}_{\mathcal{Q}}\mathbf{U}_{\mathcal{Q}}^H\sigma_a^2 + \mathbf{I})^{-1}$ and $e^{-\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}\|_2^2}$ are exponentially decreasing in $|\mathcal{Q}|$. This combined with the fact that $\Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1)$ is decreasing in $|\mathcal{S}|$ seems to suggest that analyzing $\phi(1)$ is sufficient to compare the performance

$$\phi^{(d)}(n) = \begin{cases} \displaystyle\sum_{\mathcal{S}:|\mathcal{S}|=n} \frac{e^{\eta} d(\mathcal{Q}, \mathcal{S})\left(1+\sigma_w^2\right)^{n-N}}{e^{\boldsymbol{\alpha}_{\mathcal{Q}}^H\mathbf{U}_{\mathcal{Q}}^H\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\mathbf{I}+\sigma_w^2\boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}\mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}}}, & n = 1 \\ \displaystyle\sum_{\mathcal{S}:|\mathcal{S}|=n} \frac{\frac{e^{\eta} d(\mathcal{Q}, \mathcal{S})}{(1+\sigma_w^2)^{N-n}}(1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1}|\mathcal{S}'| < n))}{e^{\boldsymbol{\alpha}_{\mathcal{Q}}^H\mathbf{U}_{\mathcal{Q}}^H\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\mathbf{I}+\sigma_w^2\boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}\mathbf{U}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}}}, & n > 1 \end{cases}. \tag{23}$$

$$\phi^{(r)}(n) = \begin{cases} \displaystyle\sum_{\mathcal{S}:|\mathcal{S}|=n} \frac{e^{\eta} d(\mathcal{Q}, \mathcal{S})}{\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\left(\mathbf{U}_{\mathcal{Q}}\mathbf{U}_{\mathcal{Q}}^H\sigma_a^2 + \mathbf{I}\sigma_w^2\right) + \mathbf{I})}, & n = 1 \\ \displaystyle\sum_{\mathcal{S}:|\mathcal{S}|=n} \frac{e^{\eta} d(\mathcal{Q}, \mathcal{S})(1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1}|\mathcal{S}'| < n))}{\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\left(\mathbf{U}_{\mathcal{Q}}\mathbf{U}_{\mathcal{Q}}^H\sigma_a^2 + \mathbf{I}\sigma_w^2\right) + \mathbf{I})}, & n > 1 \end{cases}. \tag{24}$$

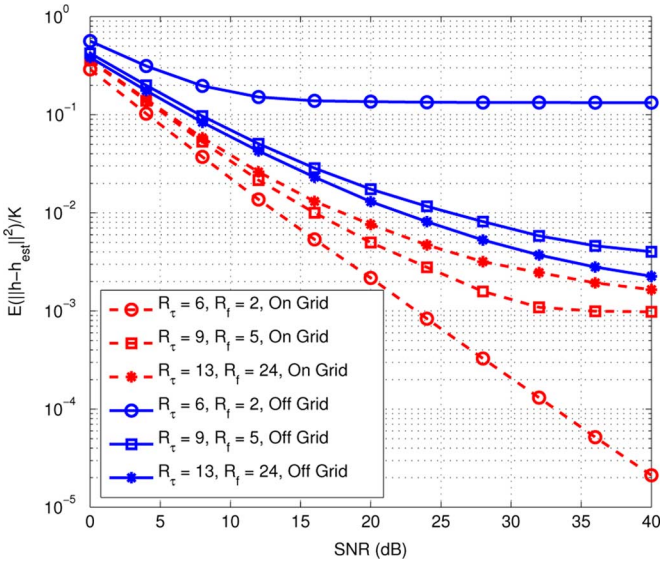

Fig. 1. Channel MSE for various grid choices.

for different input designs. In this simplest case with $|\mathcal{Q}| = 1$, we observed in (25) that the most likely candidates $\mathcal{S}$ (with $|\mathcal{S}| = 1$) to be returned by the algorithm are those with large correlation $\|\mathbf{u}_q^H \mathbf{u}_s\|^2$ and we desire for the largest values of intercolumn correlation $|\mathbf{u}_q^H \mathbf{u}_s|^2$ to correspond to small values of $d(\{q\},\{s\})$. We propose a simple measure (called localized coherence) of this property based on (25)

$$\psi = \sum_{q=1}^{R} \sum_{s \neq q} \frac{d(\{q\},\{s\}) e^{\eta}}{RN\left(1 + a(\|\mathbf{u}_q\|_2^2 - \frac{|\mathbf{u}_q^H \mathbf{u}_s|^2}{\|\mathbf{u}_s\|^2})\right)} \tag{26}$$

where $a = \|\alpha_q\|^2$ for deterministic $\boldsymbol{\alpha}_{\mathcal{Q}}$ ($a = \sigma_a^2$ if random) and $\mathbf{u}_q$ is the qth column of $\mathbf{U} = \boldsymbol{\mathcal{X}}\boldsymbol{G}$ defined in (6). We want the error to be small for all possible values of q, and this is the reason we sum over q. Localized coherence is a simple measure to aid in designing the input signal in $\boldsymbol{\mathcal{X}}$ to attain good performance in the presence of noise. Though this metric does not consider the case of more than a single column, it is obviously a starting point for designing $\boldsymbol{\mathcal{X}}$, and we examine through simulation its efficacy for cases consisting of more than a single column.

Even though most sparse recovery algorithms try to approximate the solution of (12), we have observed that this metric is applicable to predict the ranking of different input design in terms of the resulting channel estimation performance. For example greedy algorithms such as MP and OMP select at each iteration of the algorithm, the column of $\mathbf{U}$ that is most correlated with the residual signal. Other methods that relax (12) by replacing the l_0 norm with the l_1 norm and possibly modifying the reconstruction error constraint, such as Basis Pursuit and the Dantzig Selector [15], [22], appear to similarly select a set of columns $\mathbf{U}_{\mathcal{S}}$ which span a space very similar to the columns $\mathbf{U}_{\mathcal{Q}}$.

V. Numerical Experiments

In the following experiments, we will use OMP [20] to estimate the coefficients $\boldsymbol{\alpha}^{\star}$ and support $\mathcal{Q}$. We terminate the OMP algorithm when the reconstruction error $\|\mathbf{y} - \mathbf{U}\boldsymbol{\alpha}\|_2^2$ is less than $\eta = N\sigma_w^2$. In all experiments we define the received signal-to-noise ratio (SNR) as $\mathbb{E}\{\|\boldsymbol{\mathcal{X}}\mathbf{h}\|^2\}/\mathbb{E}\{\|\mathbf{w}\|\}^2$.

In the first experiment, we compare the channel estimation performance for two different scenarios; the first assumes that the parameters of the channel $\boldsymbol{\kappa}_q$ lie on the grid of the quantized parameter space while the second scenario assumes the parameters $\boldsymbol{\kappa}_q$ do not lie on the grid points but in a bounded region of $\mathbb{R}^2$. We simulate the doubly selective system model considered in Section II with $L = 6$. We observe the signal over $N = 17$ observations where the N length input sequence consists of two complementary chirp sequences (as suggested in [17]), $x[n] = e^{-j2\pi n^2\beta/N}$ for $n = 0,\ldots 8$ and $x[n] = e^{j2\pi n^2\beta/N}$ for $n = 9,\ldots,16$ preceded by a cyclic prefix of length $N_{cp} = L-1$. The parameter $\beta = 1.9$ and the pulse $g(t) = \mathrm{sinc}(\pi t)$.

The channel parameters $\boldsymbol{\kappa}_q = (f_q, \tau_q) \in (-0.025, 0.025) \times (0, L-1)$. We construct a uniform grid for the parameter space such that for a given R_τ and R_f the spacing between delays $\epsilon_\tau = (L-1)/(R_\tau - 1)$ while the spacing between frequencies $\epsilon_f = .05/(R_f - 1)$. In both scenarios, the path attenuations $\mathbf{a} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}Q^{-1})$ with $Q = 2$ paths. In the first scenario, since we assume that the parameters $\boldsymbol{\kappa}_q$ are generated on the grid, we randomly choose Q points in the grid. In the second scenario, the frequencies $f_q \overset{iid}{\sim} \mathcal{U}(-0.025, 0.025)$ while the delays $\tau_q \overset{iid}{\sim} \mathcal{U}(0, L-1)$. The system is generated according to (1).

In Fig. 1 we plot the channel MSE $\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|^2\}/N$ averaged over 10 000 channel realizations. As expected, when the parameters $\boldsymbol{\kappa}_q$ are generated on the grid and the grid becomes coarser, the columns of $\mathbf{U}$ (as well as $\boldsymbol{G}$) become less correlated and the estimation performance improves. However, when the parameters are not assumed to lie on the grid, we observe the opposite trend where the performance improves as the quantization of the parameter space is made finer. The grid with the best performance in the first scenario corresponds to the worst performance in the second scenario and vice versa. In practice we fall into the second scenario, and is precisely why we consider a fine quantization of the parameter space and propose the use of localized coherence to compare designs of the input sequence for a fixed quantization of the parameter space. Though not shown, we observe the same trend when simulating the NMEE.

We observe in Fig. 1 that as the grid size is increased, the gain in performance appears to saturate while the computational complexity increases. Indeed, we do expect diminishing returns in the second scenario when increasing the grid size. Localized coherence is not quite able to capture this behavior due to its assumption that a sparse solution exists with zero-modeling error.

We expect this behavior to mainly result from the use of suboptimal methods like OMP compared to (14), as done in this experiment.

In the second experiment, for a fixed grid, we compare the localized coherence ψ (shown in Fig. 2) for several different designs of the sequence $x[n]$. The first sequence is the chirp sequence defined in the previous experiment, the second sequence is the Alltop (see [12]) defined as $x[n] = e^{j\pi n^3/N}$ where $N = 17$, and the last two sequences are two different binary sequences

$$\mathbf{x}_{\mathrm{Bin}(1)} = (1,1,1,1,-1,-1,-1,-1, 1,1,1,1,-1,-1,-1,-1,-1)^T$$

$$\mathbf{x}_{\mathrm{Bin}(2)} = (1,-1,-1,1,1,-1,-1,-1, 1,1,1,1,-1,-1,-1,1,1)^T.$$

The number of observations $N = 17$, and each input sequence also has a cyclic prefix. The channel parameters are the same as the previous experiment, and we generate the parameters uniformly at random in the respective intervals for $Q = 3$ paths. The grid parameters $R_\tau = L + 7$ and $R_f = 24$, while ϵ_τ and ϵ_f are defined as in the previous experiment. In Fig. 2, we plot the channel MSE, while in Fig. 3 we plot the channel NMEE $(\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2/\|\mathbf{a}\|_2\}/\sqrt{N})$ as well as the first term of (16) (denoted as $\mathbb{E}(d)$ and replacing $\boldsymbol{G}_{\mathcal{Q}}$ with $\mathbf{G}(\boldsymbol{\kappa})$ since our parameters are not necessarily on the grid) averaged over 10 000 channel realizations.

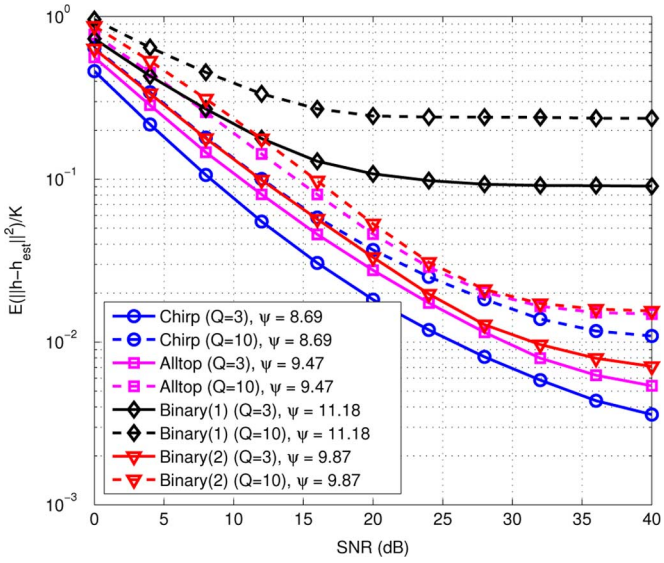

Fig. 2. Channel MSE for various training sequence designs.

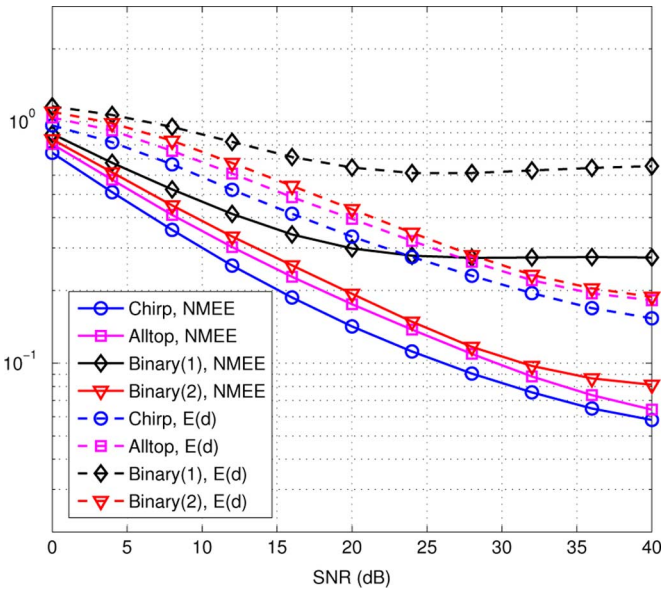

Fig. 3. Channel NMEE and $\mathbb{E}\{d(\mathcal{Q},\mathcal{S})\}$ for various training sequence designs.

For the specified choice of grid, the localized coherence values are $\psi = (8.69, 9.47, 11.18, 9.87)$ for the chirp, Alltop, Binary(1), and Binary(2) sequences, respectively. In Figs. 2 and 3 the sequences having the best channel MSE, NMEE, and $\mathbb{E}(d)$ performance have the smallest localized coherence ψ as expected. The chirp sequence and the first binary sequences appear to have similar performance, while the second binary sequence and Alltop sequence are slightly worse. We also observe that the trend of $\mathbb{E}(d)$ is similar to that of the channel MSE and NMEE, lending further support to the argument that one should try to decrease the contribution from the term $\mathbb{E}(d)$.

In Fig. 2, we have also included the case of $Q = 10$ paths, demonstrating similar trends for larger values of Q as well. We again observe that the MSE performance (as well as NMEE and $\mathbb{E}(d)$) tends to saturate at high SNR when using the OMP algorithm. We note that these input sequences coupled with grid parameters generally do not satisfy the sufficient conditions (based on MC) for identification of the channel using OMP and we expect some residual channel estimation error to be present even in the absence of noise.

VI. Conclusion

We have analyzed the performance of a sparse recovery algorithm applied to the problem of channel estimation in doubly selective channels. To estimate the channel impulse response, we constructed an overcomplete basis for the channel impulse response by quantizing the parameter space of delays and Doppler shifts. For an appropriate choice of the allowable reconstruction error in the sparse recovery algorithm, we observed the most probable sets of basis elements returned by the algorithm correspond to those minimal size sets most closely spanning the channel space. Based on this observation, for a fixed quantization of the parameter space, we introduced a measure called *localized coherence* that can be used to compare input signal designs in terms of their channel estimation performance.

We observed through simulation that in practical situations where the delays and Doppler shifts of the channel do not lie on the grid corresponding to the quantized parameter space, a fine grid is preferable over a coarse grid. We further observed through simulation that the metric of localized coherence provides a way to compare input sequence designs in terms of channel estimation performance even when using OMP in place of the ideal sparse recovery algorithm analyzed in this paper.

References

[1] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
[2] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, pp. 862–873, Jun. 1999.
[3] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, pp. 1969–1986, Oct. 1998.
[4] J. J. Fuchs, "Multipath time-delay estimation," in *Proc. IEEE ICASSP*, 1997, pp. 527–530.
[5] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. CISS*, Mar. 2008.
[6] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
[7] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. IEEE ICASSP*, Mar. 2008, pp. 3469–3472.
[8] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. Allerton'08*, Sep. 2008.
[9] W. Li and J. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.
[10] 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 ICASSP*, 2008.
[11] T. Liu and D. K. Borah, "Estimation of time-varying frequency-selective channels using a matching pursuit technique," in *Proc. IEEE WCNC*, 2003, pp. 941–946.
[12] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.
[13] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[14] D. 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.
[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scient. Comput.*, vol. 20, pp. 33–61, 1998.
[16] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, Dec. 1993.
[17] M. Sharp and A. Scaglione, "On time varying channel estimation using sparse models," in *Proc. IEEE ICC'10*, Cape Town, South Africa, May 2010.
[18] E. Candes, Y. Eldar, and D. Needell, Compressed Sensing With Coherent and Redundant Dictionaries May 2010, preprint.

[19] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 560–570, 2010.
[20] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 1993, vol. 1, pp. 40–44.
[21] M. Sharp and A. Scaglione, "Estimation of sparse multipath channels," in *Proc. MILCOM'08*, Nov. 2008.
[22] E. J. Candes and T. Tao, "The dantzig selector: Statistical estimation when $\boldsymbol{p}$ is much larger than $\boldsymbol{n}$," *Ann. Stat.*, vol. 35, pp. 2313–2351, 2007.

Precoder Design for Amplify and Forward Relaying With Complex Field Network Coding

Yanwu Ding, Hyuck M. Kwon, and Kanghee Lee

***Abstract*—In this correspondence, multiuser relay systems with amplify and forward relaying protocol are considered. To make transmission delays as short as possible, the concept of complex field network coding is applied. The relays transmit for all users simultaneously instead of using conventional relaying transmissions over orthogonal channels such as time-division multiplexing for each user. Three transmission schemes with two or three time-slot delays and symbol rates $1/2$, $2/3$, and 1 are examined. Based on the derived pairwise error probability expressions, optimal precoders achieving the maximum diversity and best coding gains are designed without changing the original ergodic capacity regions. Simulation results are provided to verify the proposed designs.**

***Index Terms*—Amplify-and-forward, coding gain, diversity gain, pairwise error probability, throughput.**

I. Introduction

To realize diversity gains and other desired properties of multiple-input multiple-output (MIMO) techniques, space-time codes and precoders have been designed in a distributed fashion for relay communication systems. Different designs, depending on the criteria applied, often have introduced different delays and symbol rates. For instance, the space-time code designed to achieve the diversity-multiplexing tradeoff transmits eight symbols in eight time-slot for a relay system with two users and one relay node [1]. The symbol rate is, therefore, one symbol per channel-use per user. A disadvantage of longer delays is that more storage and higher computational cost are required at the receiver, especially for the maximum-likelihood (ML) receiver. As a result, there is an increasing interest in relay system designs with shorter delays [2]–[4]. In [2], Wang *et al.* proposed transmission schemes with symbol rate $1/2$ using complex field network code for multiuser relaying systems. The delay for a two-user-one-relay system is only two time-slots (with transmission of one symbol per user), and a lower complexity can be achieved even if the ML detection is applied at receiver. The design is shown to achieve full diversity for the protocol of link adaptive regenerative relaying in which the transmitted bits are decoded and scaled in power at the relay node before being forwarded to the destination [2]. The scaler (in power) is adapted to the signal-to-noise-ratio (SNR) in the source-relay and relay-destination links. However, the advantage of coding gain is left unexplored.

In this correspondence, we focus on an amplify-and-forward (AF) relaying protocol. The relay retransmits a scaled version of the received signal without decoding the message. An AF protocol often has lower complexity than a decode-and-forward (DF) protocol, as it does not need decoding and encoding the messages. The idea of complex field network coding is combined with AF relaying to make transmission delays as short as possible. The relay forwards for all users simultaneously instead of using conventional relaying transmission over orthogonal channels such as time division multiplexing for each user. In particular, this correspondence examines transmission schemes with delay of two or three time-slots and symbol rates higher than $1/2$, namely, $2/3$ or 1 symbols per channel-use for each user. While higher symbol rates often provide potentials to further improvement in throughput, the error performance may be deteriorated at a price. Our goal is to design complex precoders for the source nodes to improve the error performance. Based on the derived diversity and coding gain expressions in the pairwise error probabilities, optimal precoders are designed to achieve maximum diversity and coding gains.

Simulation results demonstrate that the proposed designs, when compared with the Golden code which is often regarded as one of the best space-time codes for MIMO systems, have similar or better bit error performance. In addition, the designs provide about 1 dB gain in SNR over the space-time code in [5] at bit error rate (BER) 10^{-4}. It is also observed that by using the proposed designs, rate $2/3$ scheme provides similar bit error performance as rate $1/2$ and yet achieves 30% higher throughput. The throughput for both rate $2/3$ and $1/2$ can be further improved by adaptive modulations. As for the effect of relay locations, it is demonstrated that if the relays are much closer to the destination, lower bit error rates can be obtained.

Notation: Matrices and column vectors are denoted by uppercase and lowercase boldface characters, respectively. The transpose and Hermitian of matrix $\mathbf{A}$ are denoted by $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$, respectively. An $n \times n$ identity matrix is denoted by $\mathbf{I}_n$, and $\mathbf{0}$ stands for an all-zero matrix of appropriate dimensions. $\mathbb{E}[\cdot]$ is the expectation operator. Notation $f(x) \triangleq O(g(x))$, $g(x) > 0$ denotes that there exists a positive constant c such that $|f(x)| \leq cg(x)$, when x is large.

II. System Model

This section describes the system model for a simple system with two users and a single relay. Discussion for systems with more users and relays is presented in Section IV. As shown in Fig. 1, users S_1 and S_2 transmit to the destination D, and relay node R assists the transmission. All nodes are equipped with one antenna and they either transmit or receive signals, but do not do both at the same time. The channel coefficients from S_1 and S_2 to the destination are denoted by $h_{\mathrm{S_1D}}$ and $h_{\mathrm{S_2D}}$, respectively, whereas those from S_1 and S_2 to the relay and from the relay to the destination are denoted by $h_{\mathrm{S_1R}}$, $h_{\mathrm{S_2R}}$, and h_{RD}, respectively. The channel coefficients are modeled as independent and identically distributed (i.i.d) zero-mean unit variance circular Gaussian and quasi-static Rayleigh fading, i.e., they do not change within one period of observation and after one period, they change to independent values.

Manuscript received September 13, 2010; revised January 05, 2011; accepted February 21, 2011. Date of publication March 14, 2011; date of current version May 18, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ta-Sung Lee.

The authors are with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260 USA (e-mail: yanwu.ding@wichita.edu; hyuck.kwon@wichita.edu; kxlee1@wichita.edu).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2128315