MAXIMUM-DIVERSITY AFFINE PRECODING FOR THE NONCOHERENT DOUBLY DISPERSIVE CHANNEL

Sung-Jun Hwang and Philip Schniter

Dept. ECE, The Ohio State University, 2015 Neil Avenue, Columbus, OH 43210.
E-mail: {hwangsu, schniter}@ece.osu.edu

ABSTRACT

In this paper we characterize the maximally achievable diversity order for noncoherent block communication over the doubly dispersive channel, and propose affine precoders which facilitate such maximum-diversity reception. In fact, we show that, under mild channel conditions, almost any affine precoder is sufficient to facilitate maximum-diversity reception, regardless of precoding rate. By "noncoherent," we mean that the channel realization is unknown to both transmitter and receiver, and by "doubly dispersive," we mean that the channel exhibits both delay and Doppler spreading (i.e., the channel has a time-varying nontrivial impulse response).

1. INTRODUCTION

In this paper, we consider reliable communication over doubly dispersive (DD) channels, i.e., fading channels that exhibit significant simultaneous delay and Doppler spread. We are especially interested in the high-SNR regime, where the performance is strongly dependent on the diversity order, i.e., the negative slope of the log-error-rate versus log-SNR curve.

For the case where the receiver has channel state information (CSI) and that the channel follows a complex-exponential basis expansion model (CE-BEM), Ma and Giannakis [1] characterized the maximum achievable diversity order and proposed a linear precoding scheme that facilitates maximum-diversity reception. The assumptions of perfect receiver CSI and a CE-BEM channel are quite restrictive, however, limiting the practical impact of [1]. For example, CSI is not easy to acquire and maintain in the doubly dispersive case, where channel parameters can be multitudinous and quickly varying.

In response, we consider the more difficult problem of *noncoherent communication* over the DD channel, where neither the transmitter nor the receiver is assumed to have CSI. In this case, the receiver must exploit (a priori known) structure in the transmitted signal in order to decode reliably in the presence of channel uncertainty. Note that training-based, blind, and semi-blind schemes all fall under the category of non-coherent communication. Similarly, the term "joint channel/symbol estimation" sometimes refers to noncoherent decoding, even though explicit channel estimates are not strictly needed for data decoding.

For noncoherent communication over the DD channel, there exists a large body of work on optimal and suboptimal noncoherent reception strategies (e.g., [2–12]). For this case, there also exist several articles on training sequence design (e.g., [13–16]) with the aim of improving explicit channel estimates. But we are not aware of work addressing the general problem of transmitter design (i.e., joint design of data and training sequences) to improve the reliability of communication over the noncoherent DD channel.

In response, we first characterize the maximum achievable diversity order for noncoherent communication over the DD channel, and find (for wide-sense stationary uncorrelated scattering (WSSUS) channels with limited time-frequency spread) that the diversity order equals the product of temporal and spectral diversity orders, thereby coinciding with the maximum diversity order for coherent communication over the DD channel [1, 17]. For our analysis, we leverage certain asymptotic results from the noncoherent pairwise error probability (PWEP) analysis in [18, 19]. Next, we show that (under mild channel conditions) *almost any* affine precoder facilitates maximum diversity reception. We also show that linear precoding [20, 21] does not facilitate maximum diversity reception for commonly used symbol alphabets (e.g., uncoded QAM or PSK). Recall that affine precoding [22] refers to the general class of schemes which combine linear processing of the information symbols with additive training. It is interesting to note that, while the maximum-diversity precoder proposed for the coherent case in [1] led to a high degree of transmit-signal redundancy, the affine precoders considered here are not rate-constrained in any way. Furthermore, while the coherent results in [1] apply only to the subclass of DD channels for which the CE-BEM holds, our noncoherent results apply to a much broader class of DD channels.

Notation: We denote the transpose by $(\cdot)^T$, the conjugate transpose by $(\cdot)^H$, the determinant by $\det(\cdot)$, and the null space of matrix $\boldsymbol{A}$ by $\mathcal{N}(\boldsymbol{A})$. We denote the $M \times M$ identity matrix by $\boldsymbol{I}_M$, the $M \times 1$ zero-valued column vector by $\mathbf{0}_M$, and the $M \times N$ zero-valued matrix by $\mathbf{0}_{M\times N}$. Finally, we

This work was supported by the National Science Foundation CAREER grant CCR-0237037 and the Office of Naval Research.

abbreviate “with probability one” as “w.p.1”.

2. SYSTEM MODEL

We consider block transmission of a codeword $\boldsymbol{c} = [c_{N-1}, c_{N-2}, \ldots, c_0]^T \in \boldsymbol{\mathcal{C}}$, where $\boldsymbol{\mathcal{C}} \subset \mathbb{C}^N$ is a finite set of candidate codewords, through a doubly dispersive (DD) channel. The DD channel is characterized by a time-varying discrete impulse response $h_{n,\ell}$, such that the received sample at time n can be described as

$$r_n \;=\; \sum_{\ell=0}^{N_h-1} h_{n,\ell} c_{n-\ell} + w_n. \qquad (1)$$

In (1), N_h denotes the channel length and w_n denotes a sample of a zero-mean circular white Gaussian noise (CWGN) process with variance σ^2.

We assume that the channel is Rayleigh fading and wide-sense stationary (WSS). Thus, $\boldsymbol{h}_\ell := [h_{N-1,\ell}, h_{N-2,\ell}, \ldots, h_{0,\ell}]^T$, the random vector defined by the N-sample trajectory of the ℓ^{th} channel tap, can be expressed (without loss of generality) using its Karhunen-Lòeve (KL) expansion as $\boldsymbol{h}_\ell = \boldsymbol{B}_\ell \boldsymbol{\theta}_\ell$, where $\boldsymbol{B}_\ell \in \mathbb{C}^{N \times N_b}$ is a fixed basis matrix such that $\boldsymbol{B}_\ell^H \boldsymbol{B}_\ell = \boldsymbol{I}_{N_b}$, and where $\boldsymbol{\theta}_\ell \in \mathbb{C}^{N_b}$ is a zero-mean circular Gaussian random vector. The parameter $N_b \leq N$ quantifies the degrees-of-freedom in the tap's time-variation. In cases of practical interest, the channel varies slowly enough that $N_b \ll N$. For evidence of this claim, Fig. 1 plots the effective[1] degrees-of-freedom for the commonly assumed “Jakes' channel,” i.e., $\mathrm{E}\{h_{n,\ell} h^*_{n+m,\ell}\} = J_0(2\pi f_{\mathsf{D}} T_s m)$, where $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, f_{D} denotes the single-sided Doppler spread in Hz and T_s denotes the channel-use interval in seconds. We furthermore assume that our channel exhibits WSS uncorrelated scattering (WSSUS), so that $\boldsymbol{\theta} := [\boldsymbol{\theta}_0^T, \ldots, \boldsymbol{\theta}_{N_h-1}^T]^T \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_\theta)$, where $\boldsymbol{R}_\theta$ has full rank $N_h N_b$. In addition, we assume that each tap has the same Doppler profile, so that $\boldsymbol{B}_\ell = \boldsymbol{B}\ \forall \ell$.

Using $\boldsymbol{b}_n^H$ to denote the row of $\boldsymbol{B}$ such that $h_{n,\ell} = \boldsymbol{b}_n^H \boldsymbol{\theta}_\ell$, the model (1) can be rewritten, for $n \in \{0, \ldots, N-1\}$, as

$$r_n \;=\; \boldsymbol{b}_n^H \sum_{\ell=0}^{N_h-1} c_{n-\ell} \boldsymbol{\theta}_\ell + w_n. \qquad (2)$$

The vector $\boldsymbol{r} := [r_{N-1}, \ldots, r_0]^T$ can then be written as

$$\boldsymbol{r} \;=\; \boldsymbol{C}\boldsymbol{\theta} + \boldsymbol{w}, \qquad (3)$$

where

$$\boldsymbol{w} \;=\; [w_{N-1}, \ldots, w_0]^T \qquad (4)$$

$$\boldsymbol{C} \;=\; \begin{bmatrix} c_{N-1}\boldsymbol{b}_{N-1}^H & \cdots & c_{N-N_h}\boldsymbol{b}_{N-1}^H \\ \vdots & & \vdots \\ c_1 \boldsymbol{b}_1^H & \cdots & c_{-N_h+2}\boldsymbol{b}_1^H \\ c_0 \boldsymbol{b}_0^H & \cdots & c_{-N_h+1}\boldsymbol{b}_0^H \end{bmatrix} \qquad (5)$$

[1] We define the “effective degrees-of-freedom” as the number of eigenvalues in $\mathrm{E}\{\boldsymbol{h}_\ell \boldsymbol{h}_\ell^H\}$ which are larger than $1/1000$ of the principle eigenvalue.

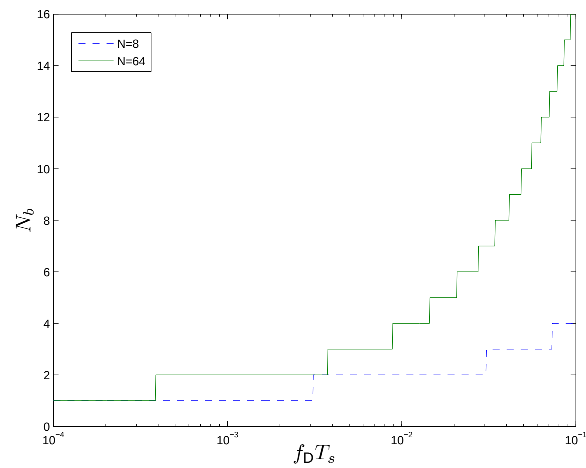

Fig. 1. Effective degrees-of-freedom versus normalized single-sided Doppler spread $f_{\mathsf{D}} T_s$ for Jakes' channel at different block lengths N.

For simplicity, we assume that $c_n = 0$ for $n < 0$, as occurs when block transmissions are separated by zero-valued guards with duration $\geq N_h - 1$. However, we note that such guards may not be needed in the high-SNR regime, where good estimates of $\{c_n\}_{n<0}$ are available from previously detected blocks and thus do not pose a problem when detecting the unknown codeword $\boldsymbol{c}$.

We assume that the receiver knows the channel statistics, i.e., $\boldsymbol{B}$ and $\boldsymbol{R}_\theta$, but not the channel realization. In this case, the (noncoherent) ML estimate of $\boldsymbol{c} \in \boldsymbol{\mathcal{C}}$ has the well known form [12, 18]

$$\hat{\boldsymbol{c}}_{\mathsf{ML}} \;=\; \arg\min_{\boldsymbol{c} \in \boldsymbol{\mathcal{C}}} \boldsymbol{r}^H \boldsymbol{\Phi} \boldsymbol{r} - \log\det(\sigma^{-2}\boldsymbol{C}^H\boldsymbol{C} + \boldsymbol{R}_\theta^{-1})$$

$$\boldsymbol{\Phi} \;:=\; \left(\boldsymbol{C}\boldsymbol{R}_\theta\boldsymbol{C}^H + \sigma^2 \boldsymbol{I}_N\right)^{-1}.$$

3. DIVERSITY-ORDER ANALYSIS

3.1. Pairwise Error Probability Analysis

In this section, we quantify the diversity order attained by the noncoherent ML detector over the doubly dispersive (DD) channel via pairwise error probability (PWEP) analysis, leveraging the work of Brehler and Varanasi [18] and Siwamogsatham, Fitz, and Grimm [19].

Let $\boldsymbol{c}_k$ denote the k^{th} codeword in $\boldsymbol{\mathcal{C}}$, and let the corresponding versions of $\boldsymbol{C}$, $\boldsymbol{\Phi}$, and $Q := \det(\sigma^{-2}\boldsymbol{C}^H\boldsymbol{C} + \boldsymbol{R}_\theta^{-1})$ be denoted by $\boldsymbol{C}_k$, $\boldsymbol{\Phi}_k$, and Q_k, respectively. Then E_{kl}, the event that $\boldsymbol{c}_k$ is transmitted and $\boldsymbol{c}_{l \neq k}$ is chosen by the ML detector, becomes

$$E_{kl} = \{\boldsymbol{r}^H \boldsymbol{\Phi}_k \boldsymbol{r} - \log Q_k \;>\; \boldsymbol{r}^H \boldsymbol{\Phi}_l \boldsymbol{r} - \log Q_l\}. \qquad (6)$$

A closed-form expression for the PWEP $\Pr\{E_{kl}\}$ has been derived [18, 19] for the high-SNR asymptotic case, i.e., $\sigma^2 \rightarrow$

0. Adapted to the specifics of our model, the result can be summarized as follows:

Lemma 1 (High-SNR PWEP [18,19]) *If the matrix*

$$\boldsymbol{M}_{kl} \;:=\; \boldsymbol{C}_k^H(\boldsymbol{I}_N - \boldsymbol{C}_l(\boldsymbol{C}_l^H\boldsymbol{C}_l)^{-1}\boldsymbol{C}_l^H)\boldsymbol{C}_k \qquad (7)$$

has full rank $N_h N_b$*, then, as* $\sigma^2 \to 0$*,*

$$\Pr\{E_{kl}\} \to \left(\frac{1}{\sigma^2}\right)^{-N_h N_b} \det(\boldsymbol{R}_\theta \boldsymbol{M}_{kl})^{-1}\binom{2N_hN_b - 1}{N_hN_b}. \qquad (8)$$

Lemma 1 establishes that the maximum achievable diversity order equals N_hN_b, and that achieving this maximum diversity order requires that $\boldsymbol{M}_{kl}$ be full rank for all k and all $l \neq k$.

3.2. Maximum-Diversity Conditions

We now translate the full-rank condition on $\boldsymbol{M}_{kl}$ to a more convenient form.

Lemma 2 $\boldsymbol{M}_{kl}$ *has full rank* N_hN_b *if and only if* $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$ *has full rank* $2N_hN_b$*, where* $\boldsymbol{D}_{lk} := \boldsymbol{C}_l - \boldsymbol{C}_k$*.*

Proof : From (7), we see that $\boldsymbol{M}_{kl}$ shares the rank of $\boldsymbol{\Pi}_l^{\perp}\boldsymbol{C}_k\boldsymbol{C}_k^H$, where $\boldsymbol{\Pi}_l^{\perp} := \boldsymbol{I}_N - \boldsymbol{C}_l(\boldsymbol{C}_l^H\boldsymbol{C}_l)^{-1}\boldsymbol{C}_l^H$ accomplishes projection onto the null space of $\boldsymbol{C}_l$. Since $\boldsymbol{C}_k \in \mathbb{C}^{N\times N_bN_h}$, full rank $\boldsymbol{M}_{kl}$ occurs iff the following two conditions are satisfied: $\boldsymbol{C}_k$ has full rank N_bN_h, and the column space of $\boldsymbol{C}_k$ is contained in the null space of $\boldsymbol{C}_l$, i.e., the column spaces of $\boldsymbol{C}_k$ and $\boldsymbol{C}_l$ share no common subspace. In other words, $\boldsymbol{M}_{kl}$ has full rank iff $[\boldsymbol{C}_k, \boldsymbol{C}_l]$ has full rank $2N_hN_b$. Furthermore, since rank is not affected by subtracting the first N_hN_b columns from the last, the rank of $[\boldsymbol{C}_k, \boldsymbol{C}_l]$ equals the rank of $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$. ■

Lemma 2 states that, for full diversity noncoherent detection, the following must hold for all k and $l \neq k$: both the codeword matrix $\boldsymbol{C}_k$ and the codeword-difference matrix $\boldsymbol{D}_{lk}$ must be full rank, and their column spaces must not intersect. Notice that the full-rank condition requires that $N \geq 2N_hN_b$. This latter condition specifies the maximum degree of time-frequency spreading for which maximum-diversity reception is possible. Notice that the condition $N \geq 2N_hN_b$ is stronger than $N > N_hN_b$, the condition for an "underspread" channel.

3.3. Linear Precoding

We refer to the class of schemes in which the codewords are generated according to

$$\boldsymbol{c} \;=\; \boldsymbol{P}\boldsymbol{s}, \qquad (9)$$

for general $\boldsymbol{P} \in \mathbb{C}^{N\times N_s}$, as *linear precoders* [20,21]. In this case, we associate the k^{th} codeword $\boldsymbol{c}_k$ with the k^{th} symbol vector $\boldsymbol{s}_k \in \boldsymbol{\mathcal{S}}$, where $\boldsymbol{\mathcal{S}} \subset \mathbb{C}^{N_s}$ is a finite set.

Lemma 3 *Linear precoding does not facilitate maximum-diversity detection when* $\exists \boldsymbol{s}_k, \boldsymbol{s}_l \in \boldsymbol{\mathcal{S}}$ *and* $a \in \mathbb{C}$ *such that* $\boldsymbol{s}_k = a\boldsymbol{s}_l$*, i.e., when* $\boldsymbol{\mathcal{S}}$ *contains symbol vectors which differ only by a scale factor.*

Proof : With linear precoding, $\boldsymbol{s}_k = a\boldsymbol{s}_l$ implies $\boldsymbol{C}_k = a\boldsymbol{C}_l$, and hence $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}] = [\boldsymbol{C}_k, (1-a)\boldsymbol{C}_k]$. Since this $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$ has rank of at most N_hN_b, Lemmas 1 and 2 establish that this rank is insufficient for maximum-diversity detection. ■

The situation described in Lemma 3 is common and arises, e.g., when $\boldsymbol{s}$ is composed of uncoded QAM or PSK symbols.

3.4. Affine Precoding

We refer to the class of schemes in which the codewords are generated according to

$$\boldsymbol{c} = \boldsymbol{P}\boldsymbol{s} + \boldsymbol{t}, \qquad (10)$$

for general $\boldsymbol{P} \in \mathbb{C}^{N\times N_s}$ and $\boldsymbol{t} \in \mathbb{C}^N$, as *affine precoders* [22]. Here again, we associate the k^{th} codeword $\boldsymbol{c}_k$ with the k^{th} symbol vector $\boldsymbol{s}_k \in \boldsymbol{\mathcal{S}}$, where $\boldsymbol{\mathcal{S}} \subset \mathbb{C}^{N_s}$ is a finite set. The affine precoder described in (10) is parameterized by a precoding matrix $\boldsymbol{P}$ and a (superimposed) training vector $\boldsymbol{t}$. In this section, we demonstrate that *almost any* choice of $\{\boldsymbol{P}, \boldsymbol{t}\}$ is sufficient to facilitate maximum-diversity detection under some mild channel conditions. Before stating our result, we define $\tilde{\boldsymbol{B}}$ as the matrix created from the top $N - N_h + 1$ rows of $\boldsymbol{B}$, i.e.,

$$\tilde{\boldsymbol{B}} \;:=\; \begin{bmatrix} \boldsymbol{b}_{N-1}^H \\ \boldsymbol{b}_{N-2}^H \\ \vdots \\ \boldsymbol{b}_{N_h-1}^H \end{bmatrix}. \qquad (11)$$

Lemma 4 *If* $N \geq 2N_hN_b$*, if* $\tilde{\boldsymbol{B}}$ *is full rank, and if* $[\boldsymbol{P}, \boldsymbol{t}]$ *is chosen randomly from a distribution whose support contains an open ball in* $\mathbb{C}^{N\times(N_s+1)}$*, then* $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$ *is full rank w.p.1.* $\forall k$ *and* $\forall l \neq k$*.*

Proof : See the appendix. ■

We now make some observations. First, Lemma 4 holds for general N_s, i.e., for precoders of arbitrary rate. Second, the rank condition on $\tilde{\boldsymbol{B}}$ is quite mild, and states that the first $N_h - 1$ samples (out of $N \geq 2N_hN_b$) of each tap trajectory are not essential to experiencing the N_b degrees-of-freedom in tap time-variation. This is expected behavior for WSS channels. (Recall that $\boldsymbol{B}$ satisfied $\boldsymbol{B}^H\boldsymbol{B} = \boldsymbol{I}_{N_b}$.)

4. NUMERICAL EXAMPLES

Figure 2 plots average PWEP versus SNR (σ^{-2}) for a randomly chosen affine precoder for $\boldsymbol{\mathcal{S}}$ = BPSK assuming an energy-preserving two-tap (i.e., $N_h = 2$) channel whose time evolution is governed by Jakes' model[2] with $f_{\mathsf{D}}T_s = 0.003$.

[2]Jakes' model was described in Section 2.

By "average" PWEP, we mean that the PWEP is averaged across symbol pairs. Our experiments assumed $N = 8$, for which the channel model yields $N_b = 2$ (see Fig. 1). To demonstrate that the results hold for general N_s, Fig. 2 investigates $N_s \in \{6, 8, 10\}$, which covers the cases that $N_s > N$, $N_s = N$, and $N_s < N$. In all cases, it can be seen that the asymptotic slope of the average PWEP equals $-N_b N_h = -4$, which confirms full-diversity reception.

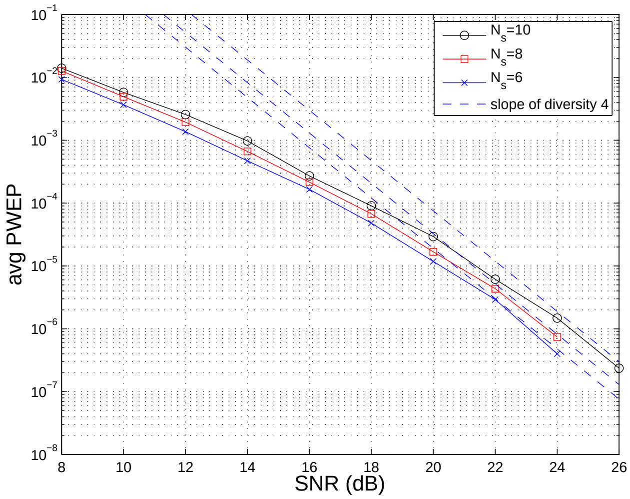

Fig. 2. Average PWEP versus SNR for $\boldsymbol{\mathcal{S}}$ = BPSK, $N = 8$, $N_h = N_b = 2$, and various N_s. The dashed line confirms the asymptotic slope of -4.

5. CONCLUSION

In this paper, we have characterized the maximum diversity-order that can be attained for noncoherent detection of doubly dispersed block transmissions, and we have provided a set of sufficient conditions under which this maximum diversity-order can be attained. Specifically, we have shown that, when the channel spreading is gentle enough to ensure $N \geq 2N_b N_h$ (and when certain other mild channel conditions are satisfied), *almost any* affine precoder will facilitate maximum-diversity noncoherent ML detection. In addition, we have shown that linear precoding does not facilitate maximum-diversity detection for certain commonly used symbol alphabets.

In the future, we plan to investigate the effect of various constrained affine precoders, such as those with orthogonal training (i.e., $\boldsymbol{t}^H \boldsymbol{P} = \boldsymbol{0}$) and those with systematic precoding matrices (i.e., $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}' \\ \boldsymbol{I}_{N_s} \end{bmatrix}$). The latter would facilitate near-ML sequential detection at very low complexity (e.g., $\mathcal{O}(N^2)$ in [12]). We also plan to investigate the design of full-diversity precoders with good finite-SNR performance (i.e., good coding gain).

6. APPENDIX

Our strategy is to characterize the $[\boldsymbol{P}, \boldsymbol{t}]$ which cause $[\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$ to be rank deficient, and show that these problematic $[\boldsymbol{P}, \boldsymbol{t}]$ are avoided w.p.1. In the sequel, we consider arbitrary k and arbitrary $l \neq k$, and we use the abbreviations $\boldsymbol{s} = \boldsymbol{s}_k$, $\boldsymbol{\delta} = \boldsymbol{s}_l - \boldsymbol{s}_k$, and $[\boldsymbol{C}, \boldsymbol{D}] = [\boldsymbol{C}_k, \boldsymbol{D}_{lk}]$.

Rank deficiency occurs when $\exists \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right] \neq \boldsymbol{0}$ such that $[\boldsymbol{C}, \boldsymbol{D}] \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right] = \boldsymbol{0}_N$. We would like to rewrite $[\boldsymbol{C}, \boldsymbol{D}] \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right]$ so that the role of $[\boldsymbol{P}, \boldsymbol{t}]$ is explicit. From the construction of $\boldsymbol{C}$, and from the partitions $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_{N_h-1}^T]^T$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{N_h-1}^T]^T$ where $\boldsymbol{\alpha}_\ell, \boldsymbol{\beta}_\ell \in \mathbb{C}^{N_b}$, we rewrite $[\boldsymbol{C}, \boldsymbol{D}] \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right] = [\boldsymbol{F}, \boldsymbol{G}] \left[\begin{smallmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{smallmatrix}\right]$ with

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{b}_{N-1}^H \boldsymbol{\alpha}_0 \cdots \boldsymbol{b}_{N-1}^H \boldsymbol{\alpha}_{N_h-1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \boldsymbol{b}_{N_h-1}^H \boldsymbol{\alpha}_0 \cdots \boldsymbol{b}_{N_h-1}^H \boldsymbol{\alpha}_{N_h-1} \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \boldsymbol{b}_0^H \boldsymbol{\alpha}_0 \end{bmatrix}$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{b}_{N-1}^H \boldsymbol{\beta}_0 \cdots \boldsymbol{b}_{N-1}^H \boldsymbol{\beta}_{N_h-1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \boldsymbol{b}_{N_h-1}^H \boldsymbol{\beta}_0 \cdots \boldsymbol{b}_{N_h-1}^H \boldsymbol{\beta}_{N_h-1} \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \boldsymbol{b}_0^H \boldsymbol{\beta}_0 \end{bmatrix}.$$

for $\boldsymbol{c}$ defined in (10) and $\boldsymbol{d} := \boldsymbol{P}\boldsymbol{\delta}$. Here we used the fact that $\{d_n = 0\}_{n<0}$ and $\{c_n = 0\}_{n<0}$. Using $\boldsymbol{p}_n^H$ to denote the row of $\boldsymbol{P}$ such that $c_n = \boldsymbol{p}_n^H \boldsymbol{s}$, we can then write

$$\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \left[\begin{array}{ccc|ccc} \boldsymbol{s}^T & & & 1 & & \\ & \ddots & & & \ddots & \\ & & \boldsymbol{s}^T & & & 1 \\ \hline \boldsymbol{\delta}^T & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \boldsymbol{\delta}^T & & & 0 \end{array}\right] \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{t} \end{bmatrix}$$

$$\boldsymbol{p} = [\boldsymbol{p}_{N-1}^H, \boldsymbol{p}_{N-2}^H, \ldots, \boldsymbol{p}_0^H]^T$$

$$\boldsymbol{t} = [t_{N-1}, t_{N-2}, \ldots, t_0]^T.$$

Putting these together, we have $[\boldsymbol{C}, \boldsymbol{D}] \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right] = [\boldsymbol{H}, \boldsymbol{F}] \left[\begin{smallmatrix} \boldsymbol{p} \\ \boldsymbol{t} \end{smallmatrix}\right]$ with

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{b}_{N-1}^H(\boldsymbol{\alpha}_0 \boldsymbol{s}^T + \boldsymbol{\beta}_0 \boldsymbol{\delta}^T) & \cdots \boldsymbol{b}_{N-1}^H(\boldsymbol{\alpha}_{N_h-1} \boldsymbol{s}^T + \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \boldsymbol{b}_{N_h-1}^H(\boldsymbol{\alpha}_0 \boldsymbol{s}^T + \boldsymbol{\beta}_0 \boldsymbol{\delta}^T) & \cdots & \boldsymbol{b}_{N_h-1}^H(\boldsymbol{\alpha}_{N_h-1} \boldsymbol{s}^T + \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T) \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \boldsymbol{b}_0^H(\boldsymbol{\alpha}_0 \boldsymbol{s}^T + \boldsymbol{\beta}_0 \boldsymbol{\delta}^T) \end{bmatrix}$$

and with $\boldsymbol{F}$ as defined earlier. Thus, $[\boldsymbol{C}, \boldsymbol{D}] \left[\begin{smallmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{smallmatrix}\right] = \boldsymbol{0}_N$ becomes equivalent to $\left[\begin{smallmatrix} \boldsymbol{p} \\ \boldsymbol{t} \end{smallmatrix}\right] \in \mathcal{N}([\boldsymbol{H}, \boldsymbol{F}])$.

Notice that, if $[\boldsymbol{H}, \boldsymbol{F}] \neq \mathbf{0}_{N \times N(N_s+1)}$, then $\mathcal{N}([\boldsymbol{H}, \boldsymbol{F}])$ is a strict subspace of $\mathbb{C}^{N(N_s+1)}$. In this case, our assumptions on the distribution of $\left[\begin{smallmatrix}\boldsymbol{p}\\ \boldsymbol{t}\end{smallmatrix}\right]$ imply that the set $\mathcal{N}([\boldsymbol{H}, \boldsymbol{F}])$ has measure zero, so that $\left[\begin{smallmatrix}\boldsymbol{p}\\ \boldsymbol{t}\end{smallmatrix}\right] \notin \mathcal{N}([\boldsymbol{H}, \boldsymbol{F}])$ w.p.1. Thus, we need to show that $[\boldsymbol{H}, \boldsymbol{F}] \neq \mathbf{0}$ for all $\boldsymbol{s}$, for all nonzero $\boldsymbol{\delta}$, and for all nonzero $\left[\begin{smallmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{smallmatrix}\right]$. To do this, we consider two cases.

Case 1) $\boldsymbol{\alpha} \neq \mathbf{0}$: Here we show that $[\boldsymbol{H}, \boldsymbol{F}] \neq \mathbf{0}$ by showing that $\boldsymbol{F} \neq \mathbf{0}$. Since $\boldsymbol{\alpha} \neq \mathbf{0}$, we know that $\boldsymbol{\alpha}_\ell \neq \mathbf{0}$ for some ℓ. The assumption of full rank $\tilde{\boldsymbol{B}}$ then implies that $\tilde{\boldsymbol{B}}\boldsymbol{\alpha}_\ell \neq \mathbf{0}$ for some ℓ, which ensures that $\boldsymbol{b}_n^H \boldsymbol{\alpha}_\ell \neq 0$ for some $n \in \{N_h - 1, \dots, N-1\}$. The latter condition implies $\boldsymbol{F} \neq \mathbf{0}$. Clearly, this occurs for any $\{\boldsymbol{s}, \boldsymbol{\delta}\}$.

Case 2) $\boldsymbol{\alpha} = \mathbf{0}$: Here it is evident that $\boldsymbol{\beta} \neq \mathbf{0}$, $\boldsymbol{F} = \mathbf{0}$, and

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{b}_{N-1}^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T & \cdots \boldsymbol{b}_{N-1}^H \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \boldsymbol{b}_{N_h-1}^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T & \cdots & \boldsymbol{b}_{N_h-1}^H \boldsymbol{\beta}_{N_h-1} \boldsymbol{\delta}^T \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \boldsymbol{b}_0^H \boldsymbol{\beta}_0 \boldsymbol{\delta}^T \end{bmatrix}$$

Thus, we need to show that there is no combination of $\boldsymbol{s}$, nonzero $\boldsymbol{\delta}$, and nonzero $\boldsymbol{\beta}$ that yields $\boldsymbol{H} = \mathbf{0}$. But, since $\boldsymbol{\delta} \neq \mathbf{0}$, the condition $\boldsymbol{H} = \mathbf{0}$ is equivalent to $\boldsymbol{G} = \mathbf{0}$. Now, since $\tilde{\boldsymbol{B}}$ is full rank and $\boldsymbol{\beta}_\ell \neq \mathbf{0}$ for some ℓ, we know that $\boldsymbol{b}_n^H \boldsymbol{\beta}_\ell \neq 0$ for some $n \in \{N_h - 1, \dots, N-1\}$, which ensures that $\boldsymbol{G} \neq \mathbf{0}$. Clearly, this occurs for any $\boldsymbol{s}$ and any nonzero $\boldsymbol{\delta}$.

7. REFERENCES

[1] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 1832–1840, July 2003.

[2] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE Journal on Selected Areas In Communications*, vol. 10, pp. 579–588, Apr. 1992.

[3] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. on Communications*, vol. 42, pp. 941–950, Feb./Mar./Apr. 1994.

[4] H. Kubo, K. Murakami, and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels," *IEEE Trans. on Communications*, vol. 42, pp. 1872–1880, Feb./Mar./Apr. 1994.

[5] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. on Communications*, vol. 43, pp. 1534–1544, Feb./Mar./April 1995.

[6] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. on Communications*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[7] D. K. Borah and B. D. Hart, "Receiver structures for time-varying frequency-selective fading channels," *IEEE Journal on Selected Areas In Communications*, vol. 17, pp. 1863–1875, Nov. 1999.

[8] B. D. Hart, "Maximum likelihood sequence detection using a pilot tone," *IEEE Trans. on Vehicular Technology*, vol. 49, pp. 550–560, Mar. 2000.

[9] H. Chen, K. Buckley, and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 1005–1009, 2000.

[10] H. Chen, R. Perry, and K. Buckley, "On MLSE algorithms for unknown fast time-varying channels," *IEEE Trans. on Communications*, vol. 51, pp. 730–734, May 2003.

[11] A. E.-S. El-Mahdy, "Adaptive channel estimation and equalization for rapidly mobile communication channels," *IEEE Trans. on Communications*, vol. 52, pp. 1126–1135, July 2004.

[12] S.-J. Hwang and P. Schniter, "Near-optimal noncoherent sequence detection for doubly dispersive channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 134–138, Nov. 2006.

[13] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1351–1366, May 2003.

[14] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, 2005.

[15] A. P. Kannu and P. Schniter, "Capacity analysis of MMSE pilot patterns for doubly selective channels," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, 2005.

[16] J. K. Tugnait, X. Meng, and S. He, "Doubly selective channel estimation using superimposed training and exponential bases models," *EURASIP Journal on Applied Signal Processing*, pp. Article ID 85303, 11 pages, 2006.

[17] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. on Communications*, vol. 47, pp. 123–132, Jan. 1999.

[18] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. on Information Theory*, vol. 47, pp. 2383–2399, Sept. 2001.

[19] S. S. M. Fitz and J. Grimm, "A new view of performance analysis of transmit diversity schemes in correlated Rayleigh fading," *IEEE Trans. on Information Theory*, vol. 48, pp. 950–956, April 2002.

[20] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. on Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[21] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part II: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. on Signal Processing*, vol. 47, pp. 2007–2022, July 1999.

[22] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2749–2752, June 2000.