On the Spectral Efficiency of Noncoherent Doubly Selective Channels

Arun P. Kannu and Philip Schniter
Dept. ECE, The Ohio State University, Columbus, OH 43210.

***Abstract*— This paper considers block transmissions over single-antenna doubly selective channels that obey a complex-exponential basis expansion model. We consider the noncoherent case, when the channel fading coefficients are not known at both transmitter and receiver. We characterize the spectral efficiency of the channel when the inputs are chosen from continuous distributions. Then, we study several pilot aided transmissions (PAT) over the doubly selective channels. We establish that all the PAT schemes designed to minimize the channel estimation error variance are spectrally inefficient. We also design novel spectrally efficient PAT schemes.**

Index Terms— Noncoherent channels, doubly selective channels, spectral efficiency, channel capacity, achievable rates, pilot symbols, channel estimation, optimal training, minimum mean squared error.

I. Introduction

Wireless channels which exhibit multipath fading are typically modeled as a linear transformation parameterized by random fading coefficients. The scenario when the knowledge of the fading coefficients is available at the transmitter and/or at the receiver is commonly referred as coherent scenario. The capacity of the coherent channels are well studied and several efficient coherent coding and decoding techniques have been developed in terms of complexity and performance. In many practical scenarios, however, neither the transmitter nor the receiver has this channel state information (CSI), which is commonly referred as noncoherent scenario. In this paper, we study the capacity limits of noncoherent doubly selective channels and develop simple and efficient communication techniques for those channels.

For wireless channels, the delay spread (in seconds) τ, caused by the multiple propagation paths between the transmitter and the receiver, governs the frequency selective nature of the channel. Similarly, the Doppler spread (in Hz) $2f_{\mathsf{D}}$, caused by the mobility between the transmitter and the receiver, governs the time selectivity. The product $2f_{\mathsf{D}}\tau$ quantifies the channel variation across time and frequency. The channels for which $2f_{\mathsf{D}}\tau < 1$ are referred as underspread channels.

The capacities of noncoherent multiple-input multiple-output (MIMO) flat, single-input single-output (SISO) time-selective, and SISO frequency-selective block-fading channels have been obtained in the high signal-to-noise ratio (SNR) regime in [1], [2], and [3], respectively. Referring to the pre-log factor of the high-SNR capacity expression as the *achievable spectral efficiency*, it has been shown that the achievable spectral efficiency of the noncoherent case is less than that of the perfect-receiver-CSI case [1]–[3], henceforth referred to as the *coherent* case. In fact the capacity of the overspread channel (for which $2f_{\mathsf{D}}\tau \geq 1$) has been shown to grow only double-logarithmically with SNR [4]. For MIMO time-varying channels, the relation between channel capacity, rate of channel variation, and SNR was recently analyzed in [5].

In this paper, we study the achievable spectral efficiency of block transmissions over noncoherent underspread doubly selective channels (DSC) which can be characterized by a complex-exponential basis expansion model (CE-BEM). Due to the limitations on the velocity of the moving objects, the time variation of the DSC is band-limited. For large enough block size, a band-limited random sequence can be approximated to any degree of accuracy by its Fourier expansion using the coefficients within the band, usually referred as truncated Fourier series. Using this CE-BEM model, we establish that, for an underspread DSC the capacity grows logarithmically with SNR. In particular, we show that, with a continuous input distribution, the DSC's spectral efficiency is approximately equal to $1 - 2f_{\mathsf{D}}\tau$. For such channels, the rapid variation in time and frequency can be seen to limit the achievable spectral efficiency.

Pilot-aided transmission (PAT) is a well-known and practical noncoherent communication strategy, whereby the transmitter embeds known pilot (i.e., training) signals that the receiver can use to estimate the channel. The estimated channel coefficients can then be used by coherent coding and decoding techniques. Cavers [6] authored one of the first analytical studies of PAT. Since then, there has been a growing interest in PAT design. See [7] for a recent comprehensive overview. We say that a given noncoherent scheme is *spectrally efficient* if it attains the achievable spectral efficiency of the channel. PAT schemes designed to minimize channel estimation error variance are often referred to as minimum mean-squared error (MMSE) PAT schemes. Spectrally efficient MMSE-PAT schemes have been established for MIMO flat, SISO frequency-selective, and SISO time-selective fading channels [1]–[3]. In this paper, we study MMSE-PAT schemes for the CE-BEM DSC and establish that *none* of them are spectrally efficient. We also design novel spectrally efficient (non-MMSE) PAT schemes for the DSC.

The paper is organized as follows. Section II outlines the modeling assumptions, Section III analyzes the spectral efficiency of the DSC, Section IV examines several PAT schemes for the DSC, and Section V concludes.

Notation: Hermitian, transpose and conjugate are denoted

2015 Neil Avenue, Columbus, OH 43210. E-mail: {kannu,schniter}@ece.osu.edu
This work supported by NSF CAREER grant 237037.

by $(\cdot)^{\mathsf{H}}$, $(\cdot)^{\top}$, and $(\cdot)^{*}$ respectively. The notation $[\cdot]_{n,m}$ extracts the $(n,m)^{th}$ element of a matrix, where the indices n,m begin with 0. The expectation, trace, Kronecker delta, Kronecker product and modulo-N operations are denoted by $\mathrm{E}\{\cdot\}$, $\mathrm{tr}\{\cdot\}$, $\delta(\cdot)$, $\otimes$ and $\langle\cdot\rangle_N$, respectively. The $N\times N$ identity matrix and unitary discrete Fourier transform matrices are denoted by $\boldsymbol{I}_N$ and $\boldsymbol{F}_N$ respectively.

II. System Model

A. Block Transmission Model

Equation (1) relates the sampled complex-baseband DSC output $\{y(i)\}$ to the input signal $\{x(i)\}$, where $\{v(i)\}$ is a zero-mean unit-variance circular white Gaussian noise (CWGN) process and $h(i,\ell)$ is the time-i channel response to an impulse applied at time $i-\ell$.

$$y(i) \;=\; \sqrt{\rho}\sum_{\ell=0}^{N_t-1} h(i,\ell)x(i-\ell)+v(i), \tag{1}$$

Here, N_t denotes the discrete channel length, i.e., the channel's delay spread τ normalized by the sampling interval T_s; by convention, we assume that $h(i,\ell)=0$ for $\ell\notin\{0,\dots,N_t-1\}$. To avoid interference from the preceding block, each length-N transmission block[1] is separated from it's neighbors by a guard interval of length N_t-1. In cyclic prefix (CP) systems, $x\big(i+(N+N_t-1)z\big)=x\big(i+N+(N+N_t-1)z\big)$ for $i\in\{-N_t+1,\dots,-1\}$ and all $z\in\mathbb{Z}$, while in zero prefixed (ZPr) systems, $x\big(i+(N+N_t-1)z\big)=0$ for $i\in\{-N_t+1,\dots,-1\}$ and all $z\in\mathbb{Z}$. Assuming a stationary channel, we can, without loss of generality, focus on the block with index $z=0$, whose input samples (excluding the guard portion) are given by $\{x(i)\}_{i=0}^{N-1}$. We assume that this block is demodulated from the received vector $\boldsymbol{y}=[y(0),\dots,y(N-1)]^{\top}$, from which it is evident that the guard samples have been discarded.

With $\boldsymbol{x}=[x(0),...,x(N-1)]^{\top}$ and $\boldsymbol{v}=[v(0),...,v(N-1)]^{\top}$, the block transmission structure applied to (1) implies that

$$\boldsymbol{y} \;=\; \sqrt{\rho}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v}, \tag{2}$$

where $\boldsymbol{H}\in\mathbb{C}^{N\times N}$ is the matrix of channel coefficients. For CP systems, $\boldsymbol{H}$ is given element-wise as $[\boldsymbol{H}]_{n,m}=h(n,\langle n-m\rangle_N)$ and for ZPr systems, $\boldsymbol{H}$ is a lower triangular matrix with $[\boldsymbol{H}]_{n,m}=h(n,\langle n-m\rangle_N)$, for $m\le n$. The inputs are assumed to obey the power constraint

$$\frac{1}{N}\,\mathrm{E}\{\|\boldsymbol{x}\|^2\} \;=\; 1. \tag{3}$$

B. Doubly Selective Channel Model

The following CE-BEM [8, p. 65], [9] will be used to describe the channel response over the N-length block duration. For $i\in\{0,\dots,N-1\}$ and $\ell\in\{0,\dots,N_t-1\}$, we assume that

$$h(i,\ell) \;=\; \frac{1}{\sqrt{N}}\sum_{k=-(N_f-1)/2}^{(N_f-1)/2}\lambda(k,\ell)e^{j\frac{2\pi}{N}ki}, \tag{4}$$

where the CE-BEM coefficients $\{\lambda(k,\ell)\}$ are uncorrelated zero-mean circular Gaussian with positive variance. The CE-BEM (4) has been widely used to model time-varying communication channels (e.g., [2], [8], [10]) and can be interpreted as an N_f-term truncated Fourier-series approximation of each of the N_t coefficient trajectories $\{h(0,\ell),\dots,h(N-1,\ell)\}_{\ell=0}^{N_t-1}$. The application of truncated Fourier series can be motivated by the bandlimited nature of coefficient trajectories that results from finite mobile velocities. Specifically, path lengths which vary by at most $v_{\max}$ meters per second imply a maximum single-sided Doppler spread of $f_{\mathsf{D}}=2v_{\max}/\lambda_c$ Hz, where λ_c denotes the carrier wavelength [8]. Since the use of $N_f=2\lfloor f_{\mathsf{D}}T_sN\rfloor+1$ terms in the Fourier series yields a reasonably accurate approximation to each trajectory, we assume this value of N_f throughout. We allow CE-BEM coefficients with possibly unequal variances in order to model arbitrary delay profiles and Doppler spectra. See [2] for a thorough discussion on the validity of the CE-BEM.

Defining the $N\times N_f$ matrix $\bar{\boldsymbol{F}}$ element-wise as $[\bar{\boldsymbol{F}}]_{n,m}=\frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}n(m-(N_f-1)/2)}$, and noticing that $\bar{\boldsymbol{F}}^{\mathsf{H}}\bar{\boldsymbol{F}}=\boldsymbol{I}_{N_f}$, the definitions $\boldsymbol{U}=\boldsymbol{I}_{N_t}\otimes\bar{\boldsymbol{F}}$, $\boldsymbol{\lambda}_\ell=[\lambda(-\frac{N_f-1}{2},\ell),\dots,\lambda(\frac{N_f-1}{2},\ell)]^{\top}$, $\boldsymbol{\lambda}=[\boldsymbol{\lambda}_0^{\top}\ \cdots\ \boldsymbol{\lambda}_{N_t-1}^{\top}]^{\top}\in\mathbb{C}^{N_fN_t}$, allow (4) to be written compactly as

$$\boldsymbol{h}=\boldsymbol{U}\boldsymbol{\lambda}. \tag{5}$$

Note that, because $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U}=\boldsymbol{I}_{N_fN_t}$ and because $\boldsymbol{R}_\lambda=\mathrm{E}\{\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathsf{H}}\}$ is diagonal and positive definite, (5) gives the Karhunen-Loeve (KL) expansion of $\boldsymbol{h}$. We assume that the DSC has average energy gain of unity, so that $\frac{1}{N}\mathrm{E}\{\|\boldsymbol{h}\|^2\}=\frac{1}{N}\mathrm{tr}\{\boldsymbol{R}_\lambda\}=1$. For this energy preserving channel, we define the SNR $\mathrm{E}\{\|\sqrt{\rho}\boldsymbol{x}\|^2\}/\mathrm{E}\{\|\boldsymbol{v}\|^2\}=\rho$.

In the sequel, we refer to N_t as the "discrete time spread" and to N_f as the "discrete frequency spread." In addition, we refer to $f_{\mathsf{D}}T_s$ as the "normalized Doppler spread" and to $\gamma=2f_{\mathsf{D}}\tau\approx N_tN_f/N$ as the channel's "spreading index." We restrict our focus to the underspread case, i.e., $\gamma<1$, so that N_tN_f, the number of independent channel parameters per block, is less than the block length N.

III. Spectral Efficiency of Doubly Selective Channels

We study the ergodic capacity of the channel (2), assuming the channel fading coefficients $\boldsymbol{\lambda}$ fade independently from block to block, which can be attained by interleaving of blocks. The ergodic capacity per-channel-use of the DSC (2) is expressed as [11]

$$\mathcal{C}(\rho) \;=\; \sup_{\boldsymbol{x}:\mathrm{E}\{\|\boldsymbol{x}\|^2\}=N}\frac{1}{N}\mathrm{I}(\boldsymbol{y};\boldsymbol{x}), \tag{6}$$

where $\mathrm{I}(\boldsymbol{y};\boldsymbol{x})$ is the mutual information between random vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ and the supremum is taken over all the input random vectors satisfying the power constraint.

We define the *achievable spectral efficiency* η of the channel as the pre-log factor in the high-SNR expression for the

[1]When we refer to a "transmission block of length N," we do not include the contribution from the guard portion.

channel capacity. Precisely, we have

$$\eta = \lim_{\rho \to \infty} \frac{\mathcal{C}(\rho)}{\log \rho}. \quad (7)$$

For Rayleigh-fading SISO channels (flat, frequency-selective, or time-selective), in the coherent case, i.e., perfect receiver CSI, the achievable spectral efficiency is unity. But in the non-coherent case, the achievable spectral efficiency is generally less than unity. The loss in achievable spectral efficiency has been shown to be proportional to the channel's time spread N_t and frequency spread N_f for frequency-selective and time-selective channels, respectively [2], [3]. In this paper, we establish that, for a DSC with inputs generated according to a continuous distribution, the loss in achievable spectral efficiency is proportional to the channel's spreading index $\gamma \approx \frac{N_f N_t}{N}$.

In general, the random vector $\boldsymbol{x}$ which achieves the capacity (6) is a function of ρ. The following theorem characterizes the achievable spectral efficiency of the DSC when the channel input is a continuous random vector.

Theorem 1 (Achievable Spectral Efficiency)**.** *For transmissions obeying (2)-(3) over the CE-BEM DSC, any sequence of continuous random input vectors* $\{\boldsymbol{x}^{(\rho)}\}$ *indexed by SNR* ρ*, and converging in distribution to a continuous random vector* $\boldsymbol{x}^{(\infty)}$*, yields*

$$\limsup_{\rho \to \infty} \frac{\frac{1}{N}\mathrm{I}(\boldsymbol{y}; \boldsymbol{x}^{(\rho)})}{\log \rho} \le \frac{N - N_f N_t}{N}. \quad (8)$$

Proof of Theorem 1 for CP systems appears in Appendix I and for ZPr systems the proof follows a similar approach. The following lemma specifies a fixed input distribution which achieves equality in (8).

Lemma 1 (Achievability)**.** *For transmissions obeying (2)-(3) over the CE-BEM DSC, i.i.d. inputs chosen from the zero-mean circular Gaussian distribution, i.e.,* $\boldsymbol{x} \sim \mathsf{CN}(\boldsymbol{0}, \boldsymbol{I})$*, yield*

$$\lim_{\rho \to \infty} \frac{\frac{1}{N}\mathrm{I}(\boldsymbol{y}; \boldsymbol{x})}{\log \rho} = \frac{N - N_f N_t}{N}. \quad (9)$$

See Appendix II for proof of the above Lemma.

From Theorem 1 and Lemma 1, we see that, for continuous inputs, the achievable spectral efficiency of the noncoherent DSC is (approximately) equal to $1 - \gamma$, where γ denotes the channel's spreading index. Since $\gamma = 2 f_{\mathsf{D}} \tau$, larger γ implies more channel dispersion in time and/or frequency. Our result, which shows that channel dispersion limits the achievable spectral efficiency, is intuitively satisfying. For relatively small γ, the achievable spectral efficiency will be close to unity, i.e., that of the coherent case. A channel with small γ could be interpreted as one with few unknown parameters, and thus one which does not demand much training overhead. Similar observations were made for flat and singly selective fading channels [1], [2], [12].

A note on the ZPr systems: We ignored the observations during the guard interval portion $\boldsymbol{y}_z = [y(N), ..., y(N + N_t - 2)]^\top$. Let $\boldsymbol{y}_z = \sqrt{\rho}\boldsymbol{H}_z \boldsymbol{x} + \boldsymbol{v}_z$, for suitably constructed channel coefficient matrix $\boldsymbol{H}_z$ and the CWGN vector $\boldsymbol{v}_z$. Intuitively, these "additional" observations of the "same" input $\boldsymbol{x}$ can not increase the prelog factor of the capacity since they can only increase the effective SNR through noise averaging, which affects the term inside the logarithm. Mathematically, we have $\mathrm{I}(\boldsymbol{y}, \boldsymbol{y}_z; \boldsymbol{x}) = \mathrm{I}(\boldsymbol{y}; \boldsymbol{x}) + \mathrm{I}(\boldsymbol{y}_z; \boldsymbol{x}|\boldsymbol{y})$. Using the coherent scenario to bound the mutual information, we have $\mathrm{I}(\boldsymbol{y}_z; \boldsymbol{x}|\boldsymbol{y}) \le \mathrm{I}(\boldsymbol{y}_z; \boldsymbol{x}|\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{H}_z)$. Since $\boldsymbol{H}$ is full rank with probability 1, we have $\boldsymbol{y}_z = \boldsymbol{H}_z \boldsymbol{H}^{-1}\boldsymbol{y} - \boldsymbol{H}_z \boldsymbol{H}^{-1}\boldsymbol{v} + \boldsymbol{v}_z$. Denoting the differential entropy by $\mathrm{h}(\cdot)$, $\mathrm{I}(\boldsymbol{y}_z; \boldsymbol{x}|\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{H}_z) = \mathrm{h}(\boldsymbol{y}_z|\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{H}_z) - \mathrm{h}(\boldsymbol{y}_z|\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{H}_z, \boldsymbol{x}) = \mathrm{h}(\boldsymbol{v}_z - \boldsymbol{H}_z \boldsymbol{H}^{-1}\boldsymbol{v}) - \mathrm{h}(\boldsymbol{v}_z) = O(1)$, as $\rho \to \infty$. Since $\boldsymbol{y}_z$ does not change the spectral efficiency, we ignore it in the rest of the manuscript.

IV. Pilot Aided Transmissions

A. PAT System Model

In pilot aided transmission (PAT), the transmitter embeds known pilot (i.e., training) signals that the receiver can use to estimate the channel. The estimated channel coefficients can then be used by coherent coding and decoding techniques. In PAT, the channel input $\boldsymbol{x}$ in (2) is generated as

$$\boldsymbol{x} = \boldsymbol{p} + \boldsymbol{d}, \quad (10)$$

where $\boldsymbol{p}$ is a deterministic pilot vector and $\boldsymbol{d}$ is zero-mean data. We assume that the data vector $\boldsymbol{d}$ results from linearly modulating $N_s (\le N)$ information bearing symbols $\boldsymbol{s} = [s(0), \dots, s(N_s - 1)]^\top$ according to

$$\boldsymbol{d} = \boldsymbol{B}\boldsymbol{s}, \quad (11)$$

using a "data modulation matrix" $\boldsymbol{B} \in \mathbb{C}^{N \times N_s}$ with orthonormal columns. We refer to $N_s = \mathrm{rank}(\boldsymbol{B})$ as the "data dimension" of the PAT scheme. Defining $E_p = \|\boldsymbol{p}\|^2$ and $E_s = \mathrm{E}\{\|\boldsymbol{d}\|^2\} = \mathrm{E}\{\|\boldsymbol{s}\|^2\}$, we require that $E_p \ge 0$, $E_s > 0$, and $\frac{1}{N}(E_p + E_s) = 1$.

We use the linear-MMSE (LMMSE) estimate of $\boldsymbol{h}$ given the knowledge of $\{\boldsymbol{y}, \boldsymbol{p}\}$ and the knowledge of the second-order statistics of $\{\boldsymbol{h}, \boldsymbol{d}, \boldsymbol{v}\}$ [13],

$$\hat{\boldsymbol{h}} = \boldsymbol{R}_{y,h}^{\mathsf{H}} \boldsymbol{R}_y^{-1} \boldsymbol{y}, \quad (12)$$

where $\boldsymbol{R}_{y,h} = \mathrm{E}\{\boldsymbol{y}\boldsymbol{h}^{\mathsf{H}}\}$ and $\boldsymbol{R}_y = \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}\}$. The channel estimation error $\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}}$ has variance $\sigma_e^2 = \mathrm{E}\{\|\tilde{\boldsymbol{h}}\|^2\}$. Notice that we consider only non-data-aided estimators which use only the second-order statistics of data.

There have been many studies on developing powerful encoding and decoding techniques for the coherent case of perfect receiver CSI. In general, for coherent channels, the capacity-achieving input distribution is continuous. In fact, for coherent channels, when the additive noise is Gaussian, the capacity-optimal input distribution is also Gaussian. One of the main advantages of PAT schemes is that they enable the use of coherent communication techniques in noncoherent scenarios. So, in our study of PAT schemes, we restrict our focus to continuously distributed $\boldsymbol{s}$. Motivated by Theorem 1, the characterization of achievable spectral efficiency for the CE-BEM DSC with continuous inputs, we make the following definition.

Definition 1. *A PAT scheme is called* spectrally efficient *if its achievable rate* $\mathcal{R}(\rho)$ *over the CE-BEM DSC satisfies* $\lim_{\rho \to \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = \frac{N - N_f N_t}{N}$.

Recall that, a rate $\mathcal{R}(\rho)$ is said to be *achievable* by a PAT scheme if the probability of decoding error, with some encoding and decoding strategy, can be made arbitrarily small when communicating at that rate. For the case of flat or frequency-selective channels, PAT schemes designed to minimize the channel estimation error variance have been shown to be spectrally efficient [1], [3], [14]. Here we study MMSE-PAT schemes for CE-BEM DSC and establish that they are *not* spectrally efficient. We also design spectrally efficient PAT schemes for the CE-BEM DSC.

B. Spectral efficiency of MMSE-PAT

In our study of MMSE-PAT schemes, we restrict our attention to CP transmissions since they render the design and analysis of MMSE-PAT for DSC tractable. CP-MMSE-PAT schemes for the CE-BEM DSC have been characterized and their achievable rates studied in [15]. We briefly review those results and present new results about the spectral efficiency of CP-MMSE-PAT.

A PAT parameter pair $(\boldsymbol{p}, \boldsymbol{B})$ which jointly minimizes the channel estimate MSE σ_e^2 is referred as an MMSE-PAT scheme. We now recall the design requirements of CP-MMSE-PAT for DSC [15].

Lemma 2 ([15])**.** *For CP-PAT over the DSC described in Section II-B, with the non-data-aided estimator (12), the total channel estimation error variance σ_e^2 obeys*

$$\sigma_e^2 \geq \operatorname{tr}\left\{\left(\boldsymbol{R}_\lambda^{-1} + \frac{\rho E_p}{N}\boldsymbol{I}_{N_f N_t}\right)^{-1}\right\}, \tag{13}$$

where equality in (13) occurs if and only if the following conditions hold: $\forall k \in \{-N_t+1, ..., N_t-1\}$, $\forall m \in \{-N_f+1, ..., N_f-1\}$, we require

$$\sum_{i=0}^{N-1} p(i)p^*(i+k)e^{-j\frac{2\pi}{N}mi} = E_p\delta(k)\delta(m) \tag{14}$$

$$\sum_{i=0}^{N-1} b_q(i)p^*(i+k)e^{-j\frac{2\pi}{N}mi} = 0, \forall q \in \{0, ..., N_s-1\} \tag{15}$$

where $\boldsymbol{b}_q$ denotes q^{th} column of $\boldsymbol{B}$.

The requirement (15) has the interpretation that the pilots and data should be multiplexed in a way that ensures their orthogonality at the channel output, and the requirement (14) has the interpretation that the pilots should be constructed so that all the channel modes are excited with equal energy [15]. To design an MMSE-PAT, first we need to get a pilot vector $\boldsymbol{p}$ which satisfies (14). Then the data basis are chosen to meet the requirement (15). A general procedure to design $(\boldsymbol{p}, \boldsymbol{B})$ which meet the above requirements has been given in [15].

CP-MMSE-PAT examples from [15] are given below, using the $(\boldsymbol{p}, \boldsymbol{B})$ parameterization.

Example 1 (TDKD)**.** *Assuming $\frac{N}{N_f} \in \mathbb{Z}$, with the pilot index set $\mathcal{P}_t^{(l)} = \{l, l+\frac{N}{N_f}, ..., l+\frac{(N_f-1)N}{N_f}\}$ and the guard index set $\mathcal{G}_t^{(l)} = \bigcup_{k\in\mathcal{P}_t^{(l)}}\{-N_t+1+k, ..., N_t-1+k\}$, an N-block CP-MMSE-PAT scheme for the CE-BEM DSC is given by*

$$p(k) = \begin{cases}\sqrt{\frac{E_p}{N_f}}e^{j\theta(k)} & k \in \mathcal{P}_t^{(l)} \\ 0 & k \notin \mathcal{P}_t^{(l)}\end{cases} \tag{16}$$

and by $\boldsymbol{B}$ constructed from the columns of $\boldsymbol{I}_N$ with indices not *in the set $\mathcal{G}_t^{(l)}$. Both $l \in \{0, \dots, \frac{N}{N_f}-1\}$ and $\theta(k) \in \mathbb{R}$ are arbitrary. The corresponding data dimension is $N_s = N - N_f(2N_t-1)$.*

Example 2 (FDKD)**.** *Assuming $\frac{N}{N_t} \in \mathbb{Z}$, with the pilot index set $\mathcal{P}_f^{(l)} = \{l, l+\frac{N}{N_t}, ..., l+\frac{(N_t-1)N}{N_t}\}$ and the guard index set $\mathcal{G}_f^{(l)} = \bigcup_{k\in\mathcal{P}_f^{(l)}}\{-N_f+1+k, ..., N_f-1+k\}$, an N-block CP-MMSE-PAT scheme for the CE-BEM DSC is given by $\boldsymbol{p} = \boldsymbol{F}_N^{\mathsf{H}}\breve{\boldsymbol{p}}$, with*

$$\breve{p}(k) = \begin{cases}\sqrt{\frac{E_p}{N_t}}e^{j\theta(k)} & k \in \mathcal{P}_f^{(l)} \\ 0 & k \notin \mathcal{P}_f^{(l)}\end{cases} \tag{17}$$

and by $\boldsymbol{B}$ constructed from the columns of the IDFT matrix $\boldsymbol{F}_N^{\mathsf{H}}$ with indices not *in the set $\mathcal{G}_f^{(l)}$. Both $l \in \{0, \dots, \frac{N}{N_t}-1\}$ and $\theta(k) \in \mathbb{R}$, are arbitrary. The corresponding data dimension is $N_s = N - N_t(2N_f-1)$.*

Example 3 (Superimposed Chirps)**.** *Assuming even N, an N-block CP-MMSE-PAT scheme for the CE-BEM DSC is given by*

$$p(k) = \sqrt{\frac{E_p}{N}}e^{j\frac{2\pi}{N}\frac{N_f}{2}k^2} \tag{18}$$

$$[\boldsymbol{B}]_{k,q} = \frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}(q+N_fN_t)k}e^{j\frac{2\pi}{N}\frac{N_f}{2}k^2}, \tag{19}$$

for $k \in \{0, \dots, N-1\}$ and $q \in \{0, \dots, N_s-1\}$, where the data dimension $N_s = N - 2N_fN_t + 1$.

From the achievable rate analysis of CP-MMSE-PAT in [15], we have the following result about the spectral efficiency of CP-MMSE-PAT.

Lemma 3 ([15])**.** *For an N-block CP-MMSE-PAT $(\boldsymbol{p}, \boldsymbol{B})$ over the CE-BEM DSC with the data dimension $\operatorname{rank}(\boldsymbol{B}) = N_s$, the achievable rate $\mathcal{R}(\rho)$ satisfies,*

$$\limsup_{\rho\to\infty} \frac{\mathcal{R}(\rho)}{\log\rho} \leq \frac{N_s}{N}. \tag{20}$$

Also, equality in (20) is achieved with an i.i.d. Gaussian distribution for $\boldsymbol{s}$.

For strictly doubly selective channels (i.e., $N_t > 1$ and $N_f > 1$), the three MMSE-PAT Examples 1-3 yield $N_s < N - N_fN_t$, and are clearly not spectrally efficient. But does there exist some other CP-MMSE-PAT scheme which *is* spectrally efficient over the CE-BEM DSC? The answer is given in the following theorem.

Theorem 2. *In CE-BEM DSC with $N_t > 1$ and $N_f > 1$, the data dimension N_s of any CP-MMSE-PAT scheme $(\boldsymbol{p}, \boldsymbol{B})$ satisfying the necessary requirements given in Lemma 2, is strictly bounded as $N_s < N - N_fN_t$.*

Proof. See Appendix III. □

Combining the results from Theorem 2 and Lemma 3, we have the following result.

Corollary 1 (Spectral Inefficiency)**.** *No CP-MMSE-PAT scheme is spectrally efficient over the CE-BEM DSC with $N_t > 1$ and $N_f > 1$.*

C. Spectrally Efficient PAT

Note that there are $N_f N_t$ independent unknown BEM coefficients in each N-length block. All the CP-MMSE-PAT schemes are shown to sacrifice more than $N_f N_t$ dimensions for pilots, leaving $N_s < N - N_f N_t$ dimensions for data symbols, and hence are spectrally inefficient. Hence, to get a spectrally efficient PAT, we need to relax the MMSE requirements in Lemma 2. Since we are considering non-data-aided estimators, the pilot-data orthogonality at the channel output (15) is desirable, otherwise the channel estimation will suffer interference from data and the channel estimates will not be perfect even in the absence of noise (i.e., asymptotically as $\rho \to \infty$). Thus, a PAT scheme which preserves pilot-data orthogonality at the channel output, and which can yield perfect channel estimates in the absence of noise using only $N_f N_t$ pilot dimensions, is a candidate for spectrally efficient PAT. We give one such example and establish its spectral efficiency.

Example 4 (SEKD)**.** *With the pilot index set $\mathcal{P}_s = \{0, N_t, ..., (N_f - 1)N_t\}$ and the guard index set $\mathcal{G}_s = \{0, ..., N_f N_t - 1\}$, an N-block ZPr MMSE-PAT scheme for the CE-BEM DSC is given by*

$$p(k) = \begin{cases} \sqrt{\frac{E_p}{N_f}} e^{j\theta(k)} & k \in \mathcal{P}_s \\ 0 & k \notin \mathcal{P}_s \end{cases} \tag{21}$$

and by $\boldsymbol{B}$ constructed from the columns of $\boldsymbol{I}_N$ with indices not *in the set $\mathcal{G}_s$. $\theta(k) \in \mathbb{R}$ is arbitrary.*

In the above PAT, the first $N_f N_t$ time slots are used by the pilots and the remaining time slots are used for data transmission. The spectral efficiency of the PAT is established in the following lemma.

Lemma 4. *For the SEKD PAT from Example 4, achievable rate $\mathcal{R}(\rho)$ at SNR level ρ satisfies $\lim_{\rho\to\infty} \frac{\mathcal{R}(\rho)}{\log \rho} = \frac{N - N_f N_t}{N}$ and hence the PAT scheme is spectrally efficient.*

In general, any PAT scheme which preserves pilot-data orthogonality and which can perfectly estimate the BEM coefficients in the absence of noise while sacrificing only $N_f N_t$ pilot dimensions can be shown to be spectrally efficient. Spectrally efficient PAT schemes are guaranteed to yield higher achievable rate than CP-MMSE-PAT schemes in the high-SNR regime.

V. Conclusion

In this paper, the achievable spectral efficiency of the noncoherent CE-BEM DSC with continuous input distributions and was shown to be approximately $1 - 2f_{\mathsf{D}}\tau$, where f_{D} denotes single-sided Doppler spread and τ denotes delay spread of the channel. In addition, this paper established that CP-MMSE-PAT schemes are spectrally inefficient, and discussed the design of spectrally efficient PAT schemes.

Appendix I
Proof for Theorem 1

Defining the two vectors $\boldsymbol{y}_s = [y(0), ..., y(N_f N_t - 1)]^\top$ and $\boldsymbol{y}_r = [y(N_f N_t), ..., y(N-1)]^\top$, and using the chain rule for mutual information [11], we have $\mathrm{I}(\boldsymbol{y}; \boldsymbol{x}^{(\rho)}) = \mathrm{I}(\boldsymbol{y}_s; \boldsymbol{x}^{(\rho)}) + \mathrm{I}(\boldsymbol{y}_r; \boldsymbol{x}^{(\rho)} | \boldsymbol{y}_s)$. Using the coherent capacity to bound the second term, we have $\mathrm{I}(\boldsymbol{y}_r; \boldsymbol{x}^{(\rho)} | \boldsymbol{y}_s) \le \mathrm{I}(\boldsymbol{y}_r; \boldsymbol{x}^{(\rho)} | \boldsymbol{y}_s, \boldsymbol{H}) = (N - N_f N_t) \log \rho + O(1)$.

What remains to be shown is that $\lim_{\rho\to\infty} \frac{\mathrm{I}(\boldsymbol{y}_s; \boldsymbol{x}^{(\rho)})}{\log \rho} = 0$. Using the chain rule for mutual information again, we have

$$\begin{aligned} \mathrm{I}(\boldsymbol{y}_s; \boldsymbol{x}^{(\rho)}) &= \mathrm{I}(y(0); \boldsymbol{x}^{(\rho)}) \\ &\quad + \sum_{i=1}^{N_f N_t - 1} \mathrm{I}(y(i); \boldsymbol{x}^{(\rho)} | y(0), ..., y(i-1)) \quad (22) \\ &\le \mathrm{I}(y(0); \boldsymbol{x}^{(\rho)}) \\ &\quad + \sum_{i=1}^{N_f N_t - 1} \mathrm{I}(y(i); \boldsymbol{x}^{(\rho)}, y(0), ..., y(i-1)) \quad (23) \end{aligned}$$

We shall analyze each term in (23) separately. We define the vectors $\boldsymbol{x}_i^{(\rho)} = [x^{(\rho)}(i), ..., x^{(\rho)}(i - N_t + 1)]^\top$ and their “complements” $\bar{\boldsymbol{x}}_i^{(\rho)}$, which are composed of elements of $\boldsymbol{x}^{(\rho)}$ not in $\boldsymbol{x}_i^{(\rho)}$. With these definitions, the first term in (23) can be written $\mathrm{I}(y(0); \boldsymbol{x}^{(\rho)}) = \mathrm{I}(y(0); \boldsymbol{x}_0^{(\rho)}) + \mathrm{I}(y(0); \bar{\boldsymbol{x}}_0^{(\rho)} | \boldsymbol{x}_0^{(\rho)})$. Conditioned on $\boldsymbol{x}_0^{(\rho)}$, the uncertainty in $y(0)$ is due to channel coefficients and additive noise which are independent of $\bar{\boldsymbol{x}}_0^{(\rho)}$. Hence, $\mathrm{I}(y(0); \bar{\boldsymbol{x}}_0^{(\rho)} | \boldsymbol{x}_0^{(\rho)}) = 0$. Now, $\mathrm{I}(y(0); \boldsymbol{x}_0^{(\rho)})$ corresponds to a overspread channel (i.e., one observation with N_t unknown channel coefficients) and, using the result from [4], we have $\mathrm{I}(y(0); \boldsymbol{x}_0^{(\rho)}) \le \log\log \rho + O(1)$. Hence $\lim_{\rho\to\infty} \frac{\mathrm{I}(y(0); \boldsymbol{x}^{(\rho)})}{\log \rho} = 0$. Now considering the general term inside the summation of (23),

$$\begin{aligned} &\mathrm{I}(y(i); \boldsymbol{x}^{(\rho)}, y(0), ..., y(i-1)) \\ &\quad = \underbrace{\mathrm{I}(y(i); \boldsymbol{x}_i^{(\rho)})}_{\le \log\log\rho + O(1)} \\ &\qquad + \underbrace{\mathrm{I}(y(i); \bar{\boldsymbol{x}}_i^{(\rho)} | \boldsymbol{x}_i^{(\rho)})}_{=0} \\ &\qquad + \underbrace{\mathrm{I}(y(i); y(0), ..., y(i-1) | \boldsymbol{x}^{(\rho)})}_{T_i}, \end{aligned}$$

it remains to be shown that $\lim_{\rho\to\infty} \frac{T_i}{\log \rho} = 0$.

Recall that $\boldsymbol{y}$ and $\boldsymbol{h}$ are jointly Gaussian conditioned on $\boldsymbol{x}^{(\rho)}$. In terms of differential entropies, $\mathrm{I}(y(i); y(0), ..., y(i-1) | \boldsymbol{x}^{(\rho)}) = \mathrm{h}(y(i) | \boldsymbol{x}^{(\rho)}) - \mathrm{h}(y(i) | \boldsymbol{x}^{(\rho)}, y(0), ..., y(i-1))$. It easily follows that

$$\mathrm{h}(y(i) | \boldsymbol{x}^{(\rho)}) = \mathrm{E}\{\log(1 + \rho \sum_{\ell=0}^{N_t - 1} \mathrm{E}\{|h(i, \ell)|^2\} |x^{(\rho)}(i - \ell)|^2)\}, \tag{24}$$

where the expectation is with respect to $\boldsymbol{x}^{(\rho)}$. Now, given $\{y(0), ..., y(i-1)\}$, we split $y(i)$ into MMSE estimate and error as $y(i) = \mathrm{E}\{y(i) | y(0), ..., y(i-1)\} + \tilde{y}(i)$, and we have

$\mathrm{h}(y(i)|y(0),...,y(i-1),\boldsymbol{x}^{(\rho)}) = \mathrm{E}\log(\mathrm{E}\,|\tilde{y}(i)|^2)$. Defining $\boldsymbol{h}_i = [h(i,0),...,h(i,N_t-1)]^\top$ and denoting the covariance of $\boldsymbol{h}_i - \mathrm{E}\{\boldsymbol{h}_i|y(0),...,y(i-1)\}$ by $\tilde{\boldsymbol{R}}_i$, we have $\mathrm{E}\,|\tilde{y}(i)|^2 = 1+\rho\boldsymbol{x}_i^{(\rho)\mathsf{H}}\tilde{\boldsymbol{R}}_i\boldsymbol{x}_i^{(\rho)}$. Let $\mu_{\max,i}$ denote the maximum eigenvalue of $\tilde{\boldsymbol{R}}_i$ and $\boldsymbol{q}_i$ denote the corresponding eigenvector. Now define $\kappa_{\max,i} = \inf_{\boldsymbol{x}^{(\rho)}\in\mathbb{C}^N}\mu_{\max,i}$. For $i \in \{1,...,N_fN_t-1\}$, all the elements of $\boldsymbol{h}_i$ can not be estimated perfectly, even in the absence of noise ($\rho=\infty$), since $\{y(0),...,y(i-1)\}$ correspond to a projection of $\boldsymbol{\lambda}$ onto a subspace of smaller dimension, and hence $\kappa_{\max,i} > 0$. Now, $\mathrm{E}\,|\tilde{y}(i)|^2 \geq 1+\rho\kappa_{\max,i}|\sum_{k=0}^{N_t-1} q_i(k)x^{(\rho)}(i-k)|^2$, and hence

$$\mathrm{h}(y(i)|\boldsymbol{x}^{(\rho)},y(0),...,y(i-1)) \geq \mathrm{E}\{\log(1+\rho\kappa_{\max,i}|\sum_{k=0}^{N_t-1} q_i(k)x^{(\rho)}(i-k)|^2)\} \tag{25}$$

Combining (24) and (25), we have $T_i \leq \mathrm{E}\log\frac{1+\rho\sum_{\ell=0}^{N_t-1}\mathrm{E}\{|h(i,\ell)|^2\}|x^{(\rho)}(i-\ell)|^2}{1+\rho\kappa_{\max,i}|\sum_{k=0}^{N_t-1}q_i(k)x^{(\rho)}(i-k)|^2}$. Since $\boldsymbol{x}^{(\rho)}$ is a sequence of continuous random vectors with $\|\boldsymbol{x}^{(\rho)}\|^2 = N$ converging to a continuous random vector, $\lim_{\rho\to\infty}|\sum_{k=0}^{N_t-1}q_i(k)x^{(\rho)}(i-k)|^2 > 0$ with probability 1, and $\lim_{\rho\to\infty}\frac{T_i}{\log\rho} = 0$.

Appendix II
Proof for Lemma 1

Using the chain rule for mutual information, we have

$$\mathrm{I}(\boldsymbol{y};\boldsymbol{x}) = \mathrm{I}(\boldsymbol{y};\boldsymbol{x},\boldsymbol{H}) - \mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}) \tag{26}$$
$$\geq \mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H}) - \mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}). \tag{27}$$

Now, $\mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H})$ corresponds to coherent case of perfect receiver CSI and since $\boldsymbol{H}$ is full rank with probability 1, we have,

$$I(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H}) = \mathrm{E}\{\log\det[\boldsymbol{I}_N + \rho\boldsymbol{H}^H\boldsymbol{H}]\} \tag{28}$$
$$= N\log(\rho) + O(1). \tag{29}$$

Now, for appropriately constructed matrix $\boldsymbol{X}$ using the input samples $\{x(i)\}_{i=0}^{N-1}$, (2) can be written as

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{X}\boldsymbol{h} + \boldsymbol{v}.$$

Using the BEM model (5), we have $\boldsymbol{y} = \sqrt{\rho}\boldsymbol{X}\boldsymbol{U}\boldsymbol{\lambda}+\boldsymbol{v}$. Since $\boldsymbol{\lambda}$ captures all the degrees of freedom of DSC over a block, we have $\mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}) = \mathrm{I}(\boldsymbol{y};\boldsymbol{\lambda}|\boldsymbol{x}) = \mathrm{I}(\boldsymbol{y};\boldsymbol{\lambda}|\boldsymbol{X})$. Conditioned on $\boldsymbol{X}$, $\boldsymbol{y}$ and $\boldsymbol{\lambda}$ are jointly Gaussian and hence using the statistics of $\boldsymbol{\lambda}$, we have

$$\mathrm{I}(\boldsymbol{y};\boldsymbol{\lambda}|\boldsymbol{X}) = \mathrm{E}\log\det[\boldsymbol{I} + \frac{\rho N}{N_fN_t}(\boldsymbol{X}\boldsymbol{U})^\mathsf{H}\boldsymbol{X}\boldsymbol{U}] \tag{30}$$

Let $\{\alpha_i\}_0^{N_fN_t-1}$ be eigen values of $(\boldsymbol{X}\boldsymbol{U})^H\boldsymbol{X}\boldsymbol{U}$. For both CP and ZPr systems, from the structure of $\boldsymbol{X}$ and $\boldsymbol{U}$, with $\varphi_N = \sum_{n=0}^{N-1}|x(n)|^2$, we have

$$\sum_{i=0}^{N_fN_t-1}\alpha_i \leq \frac{N_fN_t}{N}\varphi_N. \tag{31}$$

Notice that φ_N is chi-squared distributed with $2N$ degrees of freedom. Now,

$$\log\det[\boldsymbol{I} + \frac{\rho N}{N_fN_t}(\boldsymbol{X}\boldsymbol{U})^H\boldsymbol{X}\boldsymbol{U}] = \log\prod_{i=0}^{N_fN_t-1}(1+\frac{\rho N}{N_fN_t}\alpha_i) \tag{32}$$
$$\leq \log\prod_{i=0}^{N_fN_t-1}(1+\rho\frac{\varphi_N}{N_fN_t}) \tag{33}$$

where the above inequality follows from maximizing right hand side of (32) using the method of Lagrange multipliers with the constraint (31). So,

$$I(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}) \leq N_fN_t\,\mathrm{E}\log(1+\rho\frac{\varphi_N}{N_tN_f})\} \tag{34}$$
$$= N_fN_t\log(\rho) + O(1). \tag{35}$$

Using (29) and (35) in (27), we have the desired result.

Appendix III
Proof of Theorem 2

We use modulo-N indexing throughout this proof. First define $\boldsymbol{e}_{(k,m)} = \frac{1}{\sqrt{E_p}}[p(k)e^{j\frac{2\pi}{N}m.0}, p(k+1)e^{j\frac{2\pi}{N}m.1}, ...p(k+N-1)e^{j\frac{2\pi}{N}m(N-1)}]^\top$, which are normalized to have unit norm, for convenience. We also define the sets, $\mathcal{N}_t = \{-N_t+1,...,N_t-1\}$ and $\mathcal{N}_f = \{-N_f+1,...,N_f-1\}$. Let $\boldsymbol{W}$ be a matrix whose columns are constructed from the set $\{\boldsymbol{e}_{(k,m)},\ k\in\mathcal{N}_t,\ m\in\mathcal{N}_f\}$. Now, the orthogonality requirement (15) can be written as $\boldsymbol{W}^\mathsf{H}\boldsymbol{B} = \boldsymbol{0}$ and hence the number of information symbols in each block $N_s = \mathrm{rank}(\boldsymbol{B})$ is equal to the dimension of the null space of $\boldsymbol{W}^\mathsf{H}$.

We proof the theorem by contradiction. We assume there are MMSE-PAT schemes for which $\mathrm{rank}(\boldsymbol{B}) = N - N_fN_t$ and find the necessary requirements on their pilot vectors. Then we establish that the pilot vectors satisfying these requirements does not yield $\mathrm{rank}(\boldsymbol{B}) = N - N_fN_t$.

Let $(\boldsymbol{p},\boldsymbol{B})$ correspond to a MMSE-PAT with $\mathrm{rank}(\boldsymbol{B}) = N-N_fN_t$. We proceed to establish the necessary requirements for $\boldsymbol{p}$. To start with, optimal excitation (14) is necessary for MMSE-PAT and let $\boldsymbol{p}$ be any vector which satisfies (14). For convenience, define $D = \frac{N_f-1}{2}$. Figure 1 gives a pictorial representation of the elements of the set $\{\boldsymbol{e}_{(k,m)},\ k\in\mathcal{N}_t,\ m\in\mathcal{N}_f\}$ arranged in a grid.

We define the quantity

$$r_{(k,m)} := \frac{1}{E_p}\sum_{i=0}^{N-1}p(i)p^*(i+k)e^{-j\frac{2\pi}{N}mi} \tag{36}$$
$$= <\boldsymbol{e}_{(0,0)},\boldsymbol{e}_{(k,m)}> \tag{37}$$

where $<\boldsymbol{x},\boldsymbol{y}> = \boldsymbol{y}^H\boldsymbol{x}$ denotes the inner product between $\boldsymbol{x}$ and $\boldsymbol{y}$. From (14), note that

$$r_{(k,m)} = \delta(k)\delta(m) \text{ for } k\in\mathcal{N}_t, \text{ and } m\in\mathcal{N}_f. \tag{38}$$

It easily follows that

$$<\boldsymbol{e}_{(k_1,m_1)},\boldsymbol{e}_{(k_2,m_2)}> = e^{j\frac{2\pi}{N}(m_2-m_1)k_1}r_{(k_2-k_1,m_2-m_1)}. \tag{39}$$

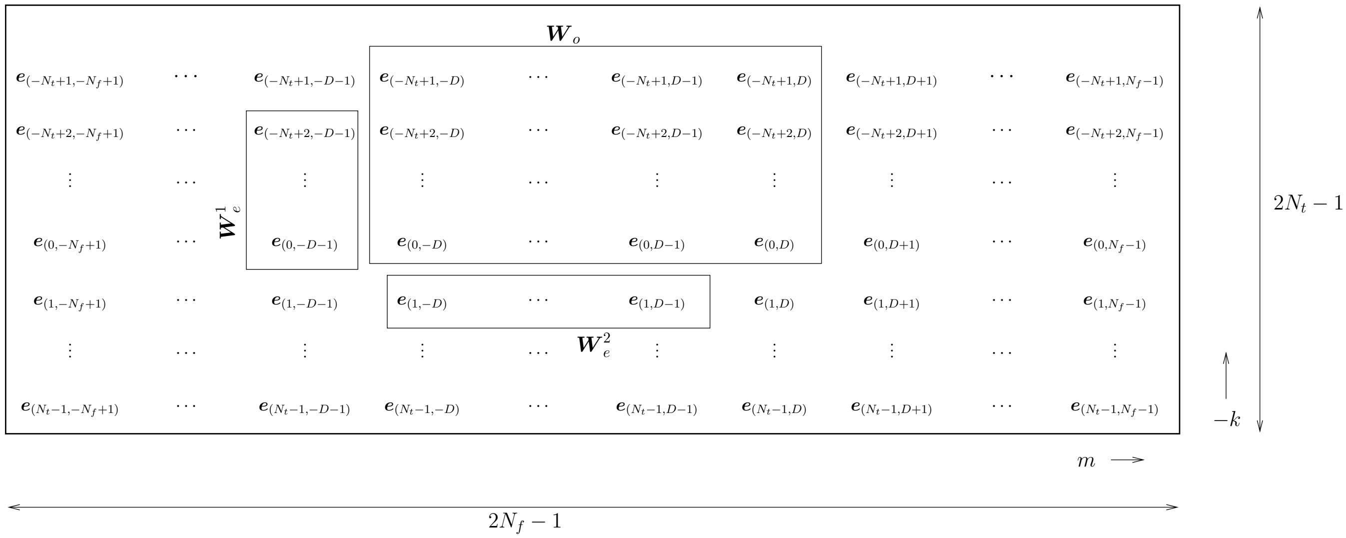

Fig. 1. Elements of the set $\{\boldsymbol{e}_{(k,m)},\ k \in \mathcal{N}_t,\ m \in \mathcal{N}_f\}$ arranged in a grid

From (38) and (39), all the elements in Fig. 1 *within any rectangle of height N_t and width N_f are orthonormal*. We also have,

$$r^*_{(k,m)} = <\boldsymbol{e}_{(k,m)}, \boldsymbol{e}_{(0,0)}> = e^{-j\frac{2\pi}{N}mk} r_{(-k,-m)}. \quad (40)$$

We use the following intermediate result.

Lemma 5. *When $\boldsymbol{B}$ is of rank $N - N_f N_t$, we have $|r_{(0,N_f)}| = 1$ or $|r_{(N_t,0)}| = 1$.*

Proof. Since $\mathrm{rank}(\boldsymbol{B}) = N - N_f N_t$ is the null space dimension of $\boldsymbol{W}^{\mathsf{H}}$, it follows that $\mathrm{rank}(\boldsymbol{W}) = N_f N_t$. Let $\boldsymbol{W}_o$ be a matrix whose columns are from the set $\{\boldsymbol{e}_{(k,m)} : k \in \{0, -1, ..., -N_t + 1\},\ m \in \{-D, ..., D\}\}$, and let $\boldsymbol{W}_e^1 = [\boldsymbol{e}_{(-1,-D-1)}, ..., \boldsymbol{e}_{(-N_t+2,-D-1)}]$ and $\boldsymbol{W}_e^2 = [\boldsymbol{e}_{(1,-D)}, ..., \boldsymbol{e}_{(1,D-1)}]$. (See Fig. 1.) From (14),(38) and (39), the $N_f N_t$ columns of $\boldsymbol{W}_o$ are orthonormal and, since $\boldsymbol{W}$ is of rank $N_f N_t$, we have the following basis expansion: $\forall k \in \mathcal{N}_t, \forall m \in \mathcal{N}_f$,

$$\boldsymbol{e}_{(k,m)} = \sum_{i=0}^{N_t-1} \sum_{j=-D}^{D} <\boldsymbol{e}_{(k,m)}, \boldsymbol{e}_{(-i,j)}> \boldsymbol{e}_{(-i,j)}, \quad (41)$$

$$= \sum_{i=0}^{N_t-1} \sum_{j=-D}^{D} e^{j\frac{2\pi}{N}(j-m)k} r_{(-i-k,j-m)} \boldsymbol{e}_{(-i,j)}. \quad (42)$$

Since any two elements inside the rectangle of height N_t and width N_f are orthogonal ((38),(39)), for the columns of $\boldsymbol{W}_e^1$, we have

$$[\boldsymbol{e}_{(0,-D-1)}, \boldsymbol{e}_{(-1,-D-1)}, ..., \boldsymbol{e}_{(-N_t+2,-D-1)}] = [\boldsymbol{e}_{(0,D)}, \boldsymbol{e}_{(-1,D)}, ..., \boldsymbol{e}_{(-N_t+1,D)}]\boldsymbol{M}_1$$

where $\boldsymbol{M}_1 \in \mathbb{C}^{N_t \times N_t - 1}$ is given by

$$\boldsymbol{M}_1 = \begin{bmatrix} r_{(0,N_f)} & \cdots & e^{-j\frac{2\pi}{N}N_f(N_t-2)} r_{(N_t-2,N_f)} \\ r_{(-1,N_f)} & \cdots & e^{-j\frac{2\pi}{N}N_f(N_t-2)} r_{(N_t-3,N_f)} \\ \vdots & \cdots & \vdots \\ r_{(-N_t+1,N_f)} & \cdots & e^{-j\frac{2\pi}{N}N_f(N_t-2)} r_{(-1,N_f)} \end{bmatrix}.$$

Similarly, we have the following expansion for the columns of $\boldsymbol{W}_e^2$,

$$[\boldsymbol{e}_{(1,-D)}, \boldsymbol{e}_{(1,-D+1)}, ..., \boldsymbol{e}_{(1,D-1)}] = [\boldsymbol{e}_{(-N_t+1,-D)}, \boldsymbol{e}_{(-N_t+1,-D+1)}, ..., \boldsymbol{e}_{(-N_t+1,D)}]\boldsymbol{M}_2$$

where $\boldsymbol{M}_2 \in \mathbb{C}^{N_f \times N_f - 1}$ is equal to

$$\begin{bmatrix} r_{(-N_t,0)} & \cdots & e^{-j\frac{2\pi}{N}(N_f-2)} r_{(-N_t,-N_f+2)} \\ e^{j\frac{2\pi}{N}} r_{(-N_t,1)} & \cdots & e^{-j\frac{2\pi}{N}(N_f-3)} r_{(-N_t,-N_f+3)} \\ \vdots & \cdots & \vdots \\ e^{j\frac{2\pi}{N}(N_f-1)} r_{(-N_t,N_f-1)} & \cdots & e^{j\frac{2\pi}{N}} r_{(-N_t,1)} \end{bmatrix}.$$

Since each column of $\boldsymbol{W}_e^1$ is orthogonal to each column of $\boldsymbol{W}_e^2$, from their basis expansions, we see that they have only one common basis vector $\boldsymbol{e}_{(-N_t+1,D)}$. So, to meet the orthogonality requirement, we have

$$r(-1, N_f) = \cdots = r(-N_t + 1, N_f) = 0 \quad (43)$$

or

$$r(-N_t, 1) = \cdots = r(-N_t, N_f - 1) = 0. \quad (44)$$

Using (43) in the basis expansion of $\boldsymbol{e}_{(0,-D-1)}$, we have

$$\boldsymbol{e}_{(0,-D-1)} = r_{(0,N_f)} \boldsymbol{e}_{(0,D)}. \quad (45)$$

Since both $\boldsymbol{e}_{(0,-D-1)}$ and $\boldsymbol{e}_{(0,D)}$ have unit norm, we have

$$|r_{(0,N_f)}| = 1 \Rightarrow r_{(0,N_f)} = e^{j\theta} \text{ for some } \theta \in \mathbb{R}. \quad (46)$$

Similarly, when the condition (44) is met, we have

$$\boldsymbol{e}_{(1,-D)} = r_{(-N_t,0)} \boldsymbol{e}_{(-N_t+1,-D)} \quad (47)$$

and $|r_{(-N_t,0)}| = 1$. So, from (40) we have

$$r_{(N_t,0)} = e^{j\bar{\theta}} \text{ for some } \bar{\theta} \in \mathbb{R}. \quad (48)$$

□

Now we study the pilot vectors $\boldsymbol{p}$ which satisfy (14) with the additional constraint that $|r_{(N_t,0)}| = 1$ or $|r_{(0,N_f)}| = 1$. Considering these two cases separately, we establish that there is no such $\boldsymbol{p}$ for which $\mathrm{rank}(\boldsymbol{B}) = N - N_f N_t$.

A. Case I: $r_{(0,N_f)} = e^{j\theta}$

From (45), we have

$$p(i)(e^{-j\frac{2\pi}{N}N_f i} - e^{j\theta}) = 0. \tag{49}$$

Now, if $\theta \neq \frac{2\pi}{N}N_f q$ for some $q \in \mathbb{Z}$, then $p(i) = 0 \forall i$, which clearly does not satisfy the MMSE-PAT requirement (14), and hence ruled out from consideration. Now, if $\theta = \frac{2\pi}{N}N_f q$ for some $q \in \mathbb{Z}$, from (49), $p(i)$ may be non-zero only if $i = q + \frac{kN}{N_f}$ for $k \in \mathbb{Z}$ such that $\frac{kN}{N_f} \in \mathbb{Z}$. Now, for $k \in \mathbb{Z}$, defining

$$a_q(k) = \begin{cases} |p(q + \frac{kN}{N_f})|^2 & \text{if } \frac{kN}{N_f} \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \tag{50}$$

then from the requirement (38), it follows that $\sum_{i=0}^{N_f-1} a_q(i) e^{j\frac{2\pi}{N_f}mi} = E_p \delta(m), \ \forall m \in \mathcal{N}_f$, which can be met if and only if

$$a_q(i) = \frac{E_p}{N_f}, \forall \ i \in \{0, ..., N_f - 1\}. \tag{51}$$

From the definition (50), it follows that, the above requirement can be met if and only if $\frac{N}{N_f} \in \mathbb{Z}$. If $\frac{N}{N_f} \notin \mathbb{Z}$, there is no training sequence which satisfies both (14) and (46). Now, if $\frac{N}{N_f} \in \mathbb{Z}$, from (50) and (51), the sequence $p(i)$ is of the form given in Example 1. For Example 1, as noted earlier $\mathrm{rank}(\boldsymbol{B}) = N_s = N - (2N_t - 1)N_f < N - N_f N_t$. This contradicts the initial assumption that $\boldsymbol{B}$ is of rank $N_f N_t$.

B. Case II: $r_{(N_t,0)} = e^{j\bar{\theta}}$

From (40), (47) and (48), it follows that

$$p(i) = e^{j\bar{\theta}} p(i + Nt). \tag{52}$$

Because of the circular symmetry $p(i+N) = p(i)$, using (52), we can find the largest integer $L \in \{1, ..., N_t\}$ so that $\frac{N}{L} \in \mathbb{Z}$ and $p(i) = e^{j\theta} p(i + L)$ for some $\theta \in \mathbb{R}$. Note that, if $\frac{N}{N_t} \in \mathbb{Z}$ then $L = N_t$ else $L < N_t$. Again from the circular symmetry, $\theta = \frac{2\pi}{N}Lq$ for some $q \in \mathbb{Z}$. Let $\breve{\boldsymbol{p}}$ denote the N-point unitary discrete Fourier transform (DFT) of $\boldsymbol{p}$. For the sequence $\boldsymbol{p}$ with the given "periodic" structure, we have

$$\breve{p}(k) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} p(i) e^{-j\frac{2\pi}{N}ik} \tag{53}$$

$$= \frac{1}{\sqrt{N}} \sum_{i=0}^{L-1} p(i) e^{-j\frac{2\pi}{N}ik} \sum_{m=0}^{\frac{N}{L}-1} e^{-j\frac{2\pi}{N}L(k-q)m} \tag{54}$$

and hence $\breve{p}(k) = 0 \ \forall \ k \notin \{q, q + \frac{N}{L}, ..., q + \frac{N(L-1)}{L}\}$. Now, the optimal excitation requirement (14) can be written in terms of $\breve{\boldsymbol{p}}$ as [15], $\forall k \in \mathcal{N}_f, \ \forall m \in \mathcal{N}_t$,

$$\sum_{i=0}^{N-1} \breve{p}(i) \breve{p}^*(i-k) e^{-j\frac{2\pi}{N}mi} = E_p \delta(k)\delta(m). \tag{55}$$

Defining $|\breve{p}(q + \frac{iN}{L})|^2 = \breve{a}_q(i), i \in \{0, ..., L-1\}$ and using the MMSE-PAT requirements in the frequency domain, we require

$$g(m) = \sum_{i=0}^{L-1} \breve{a}_q(i) e^{-j\frac{2\pi}{N}(i\frac{N}{L}+q)m} = E_p \delta(m), \ \forall \ m \in \mathcal{N}_t. \tag{56}$$

If $L < N_t$, then (56) can not be satisfied since $g(L) = g(0)e^{j\frac{2\pi}{N}qL}$. So, if $\frac{N}{N_t} \notin \mathbb{Z}$, there is no MMSE-PAT with $\dim(\boldsymbol{W}) = N_f N_t$. Now, if $\frac{N}{N_t} \in \mathbb{Z}$, then $L = N_t$ and the only sequence $\{\breve{a}_q(i)\}$ satisfying the requirement (56) is $\breve{a}_q(i) = c$ for some constant c, $\forall \ i$. This corresponds to the equi-spaced, equi-powered frequency domain pilot sequence of FDKD in Example 2. Again, for this pilot sequence, from Example 2, we have $\mathrm{rank}(\boldsymbol{B}) < N - N_f N_t$. Again, we reach a contradiction on the initial assumption that $\boldsymbol{B}$ is of rank $N_f N_t$.

APPENDIX IV
PROOF OF LEMMA 4

The pilot observations of the PAT scheme in Example 4, $\boldsymbol{y}_p = [y(0), ..., y(N_f N_t - 1)]^\top$, can be written as $\boldsymbol{y}_p = \sqrt{\rho}\boldsymbol{G}\boldsymbol{\lambda} + \boldsymbol{v}_p$, where $\boldsymbol{v}_p = [v(0), ..., v(N_f N_t - 1)]^\top$, for some $\boldsymbol{G}$. It can be verified that $\boldsymbol{G}$ is full rank and hence the minimum eigen value of $\boldsymbol{G}^\mathsf{H}\boldsymbol{G}$ denoted by β_{min} is positive. The covariance matrix of the estimation error $\boldsymbol{R}_{\tilde{h}} = \mathrm{E}\{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^\mathsf{H}\} = \boldsymbol{U}(\boldsymbol{R}_\lambda^{-1} + \rho\boldsymbol{G}^\mathsf{H}\boldsymbol{G})^{-1}\boldsymbol{U}^\mathsf{H}$. Denoting the smallest eigen value of $\boldsymbol{R}_\lambda^{-1}$ by α_{min}, we have $\boldsymbol{R}_{\tilde{h}} \leq (\alpha_{min} + \rho\beta_{min})^{-1}\boldsymbol{U}\boldsymbol{U}^\mathsf{H}$, in the positive semi-definite sense. Now, constructing $\boldsymbol{B}_d \in \mathbb{C}^{N \times N - N_f N_t}$ using the last columns $N - N_f N_t$ columns of identity matrix, the data observations $\boldsymbol{y}_d = [y(N_f N_t), ..., y(N-1)]^\top$ are given by $\boldsymbol{y}_d = \boldsymbol{B}_d^\mathsf{H}\boldsymbol{H}\boldsymbol{B}\boldsymbol{s} + \boldsymbol{v}_d$, where $\boldsymbol{v}_d = [v(N_f N_t), ..., v(N-1)]^\top$. The effective channel between the observations and the data is $\boldsymbol{H}_e = \boldsymbol{B}_d^\mathsf{H}\boldsymbol{H}\boldsymbol{B}$. Splitting $\boldsymbol{H}$ into estimate $\hat{\boldsymbol{H}}$ and error $\tilde{\boldsymbol{H}}$ components, we have

$$\boldsymbol{y}_d = \underbrace{\boldsymbol{B}_d^\mathsf{H}\hat{\boldsymbol{H}}\boldsymbol{B}}_{\hat{\boldsymbol{H}}_e} \boldsymbol{s} + \underbrace{\boldsymbol{B}_d^\mathsf{H}\tilde{\boldsymbol{H}}\boldsymbol{B}\boldsymbol{s} + \boldsymbol{v}_d}_{\boldsymbol{v}_e}. \tag{57}$$

Orthogonality principle of MMSE estimator guarantees that $\hat{\boldsymbol{H}}_e\boldsymbol{s}$ and $\boldsymbol{v}_e$ are uncorrelated. In this case, the worst case distribution for $\boldsymbol{v}_e$ (from a mutual information perspective) is Gaussian [14]. With i.i.d. Gaussian distribution for the information symbols $\boldsymbol{s}$ satisfying the power constraint, with $\sigma_s^2 = \frac{E_s}{N - N_f N_t}$, we have

$$\mathrm{I}(\boldsymbol{y}_d; \boldsymbol{s}) \geq \mathrm{E} \log\det[\boldsymbol{I} + \rho\sigma_s^2 \hat{\boldsymbol{H}}_e \boldsymbol{R}_e^{-1} \hat{\boldsymbol{H}}_e^\mathsf{H}] \tag{58}$$

where $\boldsymbol{R}_e = \mathrm{E}\{\boldsymbol{v}_e\boldsymbol{v}_e^\mathsf{H}\}$. Since $\boldsymbol{B}$ and $\boldsymbol{B}_d$ have orthonormal columns, it easily follows that $\boldsymbol{R}_e \leq (1 + \frac{\rho N_t \sigma_s^2}{\alpha_{\min} + \rho\beta_{\min}})\boldsymbol{I}$, in the positive definite sense. Using this in (58), the achievable rate of the system obeys

$$\mathcal{R}(\rho) \geq \frac{1}{N} \mathrm{E} \log\det[\boldsymbol{I} + \rho_e \hat{\boldsymbol{H}}_e \hat{\boldsymbol{H}}_e^\mathsf{H}] \tag{59}$$

where $\rho_e = \frac{\rho\sigma_s^2(\alpha_{\min} + \rho\beta_{\min})}{\alpha_{\min} + \rho\beta_{\min} + \rho N_t \sigma_s^2}$. We have $\rho_e \geq k\rho, \forall \rho > 1$ for some constant k and $\lim_{\rho\to\infty} \hat{\boldsymbol{H}}_e = \boldsymbol{H}_e$ almost surely. Using Fatou's lemma, taking the limit inside the expectation, we have

$$\lim_{\rho\to\infty} \mathcal{R}(\rho) \geq \frac{1}{N} \mathrm{E} \lim_{\rho\to\infty} \log\det[\boldsymbol{I} + \rho_e \hat{\boldsymbol{H}}_e \hat{\boldsymbol{H}}_e^\mathsf{H}] \tag{60}$$

$$\geq \frac{1}{N} \mathrm{E} \log\det \lim_{\rho\to\infty} [\boldsymbol{I} + k\rho \boldsymbol{H}_e \boldsymbol{H}_e^\mathsf{H}] \tag{61}$$

$$= \frac{N - N_f N_t}{N} \log\rho + O(1) \tag{62}$$

since $\boldsymbol{H}_e$ is full rank with probability 1.

REFERENCES

[1] L. Zheng and D. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.

[2] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 3095–3110, Dec. 2004.

[3] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2572–2583, Sep. 2004.

[4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple antenna systems on flat fading channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2426–2467, Oct. 2003.

[5] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread mimo fading channels," *IEEE Trans. on Information Theory*, vol. 52, pp. 1576–1608, Apr. 2006.

[6] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. on Vehicular Technology*, vol. 40, pp. 686–693, Nov. 1991.

[7] L. Tong, B. M. Sadler, and M. Dong, "Pilot assisted wireless transmissions: general model, design criteria and signal processing," *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov. 2004.

[8] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.

[9] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Internat. Journal of Adaptive Control & Signal Processing*, vol. 10, pp. 156–176, Mar. 1996.

[10] 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.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[12] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, Jan. 1999.

[13] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[14] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.

[15] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. on Signal Processing*. (in submission).