On the Spectral Efficiency of Noncoherent Doubly Selective Block-Fading Channels

Arun Pachai Kannu and Philip Schniter, *Senior Member, IEEE*

***Abstract*—In this paper, we consider noncoherent single-antenna communication over doubly selective block-fading channels with discrete block-fading interval N. In our noncoherent setup, neither the transmitter nor the receiver know the channel fading coefficients, though both know the channel statistics. In particular, we consider discrete-time channels whose impulse-response trajectories obey a complex-exponential basis expansion model with uncorrelated coefficients, and we show that such a model holds in the limit $N \to \infty$ for pulse-shaped transmission/reception over certain wide-sense stationary uncorrelated scattering channels. First, we show that, when the inputs are chosen from continuous distributions, the channel's multiplexing gain (i.e., capacity pre-log factor) equals $\max(0, 1 - N_{\mathsf{delay}} N_{\mathsf{Dopp}}/N)$, for discrete delay spread N_{delay} and discrete Doppler spread N_{Dopp}. Next, for the case of strictly doubly selective fading (i.e., $N_{\mathsf{Dopp}} > 1$ and $N_{\mathsf{delay}} > 1$), we establish that, for cyclic-prefixed affine pilot-aided transmission (PAT) schemes designed to minimize the mean-squared error (MSE) attained by pilot-aided minimum-MSE channel estimation, the pre-log factor of the achievable rate is less than the channel's multiplexing gain. We then provide guidelines for the design of PAT schemes whose achievable-rate pre-log factor equals the channel's multiplexing gain and construct an example.**

***Index Terms*—Achievable rate, channel capacity, channel estimation, doubly dispersive, doubly selective, multiplexing gain, noncoherent, pilots, spectral efficiency, training.**

I. Introduction

RECENTLY, there has been great interest in characterizing the capacity of wireless multipath channels under the practical assumption that neither the transmitter nor the receiver has channel state information (CSI). In this paper, we focus on channels that are simultaneously time- and frequency-selective, which pertain to applications with simultaneously high signaling bandwidth and mobility. The high-SNR capacity of the noncoherent Gaussian flat-fading channel was characterized in the MIMO case by Zheng and Tse [1] using the block-fading approximation, whereby the channel coefficients are assumed to remain constant over a block of N symbols and change independently from block to block. Later, Vikalo *et al.* [2] characterized the high-SNR capacity of the noncoherent Gaussian frequency-selective block-fading SISO channel under the assumption that the discrete block-length N exceeds the discrete channel delay spread N_{delay}. Liang and Veeravalli [3] characterized the high-SNR capacity of the SISO Gaussian time-selective block-fading channel, assuming that, within the block, the channel coefficients vary according to a finite-term Fourier series with $N_{\mathsf{Dopp}} \leq N$ expansion coefficients that have a full-rank covariance matrix,[1] but change independently from block to block. In [3], they also find the asymptotic capacity of a MIMO sub-block correlated time-selective fading model, in which the channel remains constant within a sub-block. For the aforementioned noncoherent block-fading Gaussian channels, it has been shown that the capacity C as a function of SNR ρ obeys $\lim_{\rho\to\infty} C(\rho)/\log(\rho) = \eta$, where the *multiplexing gain* η is given by $\eta = \frac{N-1}{N}$ in the SISO flat-fading case, $\eta = \frac{N-N_{\mathsf{delay}}}{N}$ in the SISO frequency-selective case,[2] and $\eta = \frac{N-N_{\mathsf{Dopp}}}{N}$ in the SISO time-selective case. While the aforementioned works focus on block-fading channels, there exists other work by Lapidoth [4], [5] on the capacity of stationary fading channels.

In this paper, we consider a SISO channel that combines the frequency-selectivity of [2] with the time-selectivity of [3], henceforth referred to as the block-fading doubly selective channel (DSC). More precisely, this discrete-time channel uses a finite-length impulse response whose N_{delay} Gaussian coefficients vary according to an N_{Dopp}-term Fourier series within the block, but change independently from block to block. When the fading coefficients are uncorrelated in both time and frequency, we show that, under continuous input distributions, the channel's multiplexing gain obeys $\eta = \max(0, \frac{N-N_{\mathsf{delay}}N_{\mathsf{Dopp}}}{N})$.

Next, we study pilot-aided transmission (PAT) over this block-fading DSC. In PAT, the transmitter embeds a known pilot (i.e., training) signal that aids the receiver in data decoding under channel uncertainty. Often, PAT enables the receiver to compute an explicit channel estimate, thereby facilitating the use of coherent decoding strategies (see [6] for a recent comprehensive PAT overview). We are interested in finding PAT schemes for which the pre-log factor of the asymptotic achievable-rate expression equals the channel's multiplexing gain, i.e., $\eta = \max(0, \frac{N-N_{\mathsf{Dopp}}N_{\mathsf{delay}}}{N})$. Throughout the paper, we refer to such PAT schemes as "spectrally efficient." For the design of such PAT schemes, we consider the only the case that $N > N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ since, in the case that $N \leq N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, it would be trivial to achieve a multiplexing gain of $\eta = 0$.

When linear minimum mean-squared error (LMMSE) pilot-aided channel estimation is performed at the receiver, the resulting mean-squared error (MSE) remains dependent on the PAT scheme in use. Thus, PAT schemes have been proposed to minimize this MSE—under fixed levels of pilot and data

Manuscript received March 19, 2007; revised February 03, 2010. Current version published May 19, 2010. This work supported in part by NSF CAREER Grant 237037 and in part by the Office of Naval Research Grant N00014-07-1-0209.

The authors are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: arunpachai@ee.iitm.ac.in; schniter@ece.osu.edu).

Communicated by H. Bölcskei, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2046202

[1]Note that in the case of $N_{\mathsf{Dopp}} < N$, the N-length vector of coefficients has a rank-deficient correlation matrix.

[2]Assuming uncorrelated intersymbol interference (ISI) coefficients.

power—in [2], [7]–[9]. Henceforth, we refer to these optimized PAT schemes as "MMSE-PAT" schemes. Previous studies have established that MMSE-PAT schemes are spectrally efficient for flat [1], [7]; frequency-selective [2]; and time-selective [9], [10] block-fading channels. We establish here, however, that cyclic-prefixed MMSE-PAT schemes are *not* spectrally efficient for *strictly* doubly selective (i.e., $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$) block-fading channels. For these channels, we then develop guidelines for the design of spectrally efficient PAT schemes and propose one such scheme.

Before continuing, a few comments are in order.

1) Our work relies on the block-fading assumption, which can be justified in systems that employ block interleaving or frequency hopping. Other investigations have circumvented the block-fading assumption through the use of time-selective channel models whose coefficients vary from symbol to symbol in a stationary manner. For these stationary models, it is necessary to make a distinction between *nonregular*[3] (e.g., bandlimited) fading processes and *regular* (e.g., Gauss–Markov) fading processes. While nonregular fading channels have been shown to behave similarly to time-selective block-fading channels, regular fading channels behave quite differently [11]. Recent results on stationary doubly selective channels have a similar flavor [5]. The details, however, lie outside the scope of this work.
2) Our work relies on the assumption that intrablock time-variation can be accurately modeled by a finite-term Fourier series with uncorrelated coefficients. Though we provide a detailed justification in the sequel, the key idea is that, due to velocity limitations on the communicating terminals and the scattering surfaces, the channel fading processes will be *bandlimited*. It is well known that bandlimited random sequences can be well approximated by finite-term Fourier series, where the approximation error decreases with block size.
3) Some authors (e.g., [12]) have studied the capacity of noncoherent underspread doubly selective channels by first claiming that there exists a fixed set of *approximate* channel eigenfunctions (as motivated by [13]) but then later ignoring the resulting approximation error. The approximation error, which—if not ignored—would contribute signal- and channel-dependent additive interference to the observation, can be very small when the spreading is very mild, i.e., when $\frac{N_{\mathsf{delay}}N_{\mathsf{Dopp}}}{N} \ll 1$, but can become large as the spreading gets more severe, i.e., as $\frac{N_{\mathsf{delay}}N_{\mathsf{Dopp}}}{N} \to 1$ (see the interference lower bounds in [14]). Note that $\frac{N_{\mathsf{delay}}N_{\mathsf{Dopp}}}{N} \to 1$ for some underwater acoustic channels [15]. We avoid the approximate eigenfunction approach since we do not assume very mild spreading. We also note the existence of very recent results [16] that account for the approximation error.

The paper is organized as follows. Section II details the modeling assumptions, Section III analyzes the high-SNR capacity of the noncoherent doubly selective block-fading channel, Section IV details the PAT setup for this channel, and Sections V–VI analyze several PAT schemes.

[3]Regular processes allow for perfect prediction of the future samples from (a possibly infinite number of) past samples while nonregular processes do not. For more details, see [11].

A. Notation

Matrices (column vectors) are denoted by upper (lower) boldface letters. The Hermitian is denoted by $(\cdot)^{\mathsf{H}}$, the transpose by $(\cdot)^{\top}$, the conjugate by $(\cdot)^*$, the determinant by $\det(\cdot)$, and the Frobenius norm by $\|\cdot\|_F$. The Loewner partial order is denoted by $\preceq$, i.e., $\boldsymbol{B} \succeq \boldsymbol{A}$ means that $\boldsymbol{B}-\boldsymbol{A}$ is positive semidefinite. The expectation is denoted by $\mathrm{E}\{\cdot\}$, the trace by $\mathrm{tr}\{\cdot\}$, the Dirac delta by $\delta(\cdot)$, the Kronecker delta by $\delta[\cdot]$, the Kronecker product by $\otimes$, the modulo-N operation by $\langle\cdot\rangle_N$, and the integer ceiling operation by $\lceil\cdot\rceil$. The null space of a matrix is denoted by $\mathrm{null}(\cdot)$, the column space by $\mathrm{col}(\cdot)$, and the dimension of a vector space by $\dim(\cdot)$. The operation $[\cdot]_{n,m}$ extracts the (n,m)th element of a matrix, where the indices n, m begin with 0, and $\mathrm{diag}(\cdot)$ constructs a diagonal matrix from its vector-valued argument. Appropriately dimensioned identity and all-zero matrices are denoted by $\boldsymbol{I}$ and $\boldsymbol{0}$, respectively, while the $N \times N$ identity matrix is denoted by $\boldsymbol{I}_N$. The set-union operation is denoted by $\cup$, set-intersection by $\cap$, set-minus by $\setminus$, and the empty set by $\emptyset$. The integers are denoted by $\mathbb{Z}$, reals by $\mathbb{R}$, positive reals by $\mathbb{R}^+$, and complex numbers by $\mathbb{C}$.

II. System Model

In Section II-A, we describe the baseband-equivalent discrete-time block-fading doubly selective channel model assumed for the analysis in Sections III–IV. One of our key assumptions is that the channel can be parameterized using a discrete Fourier complex exponential (CE) basis with uncorrelated coefficients.

To lend credence to our discrete-time CE basis expansion model (BEM) and to establish links with physical channel descriptors like Doppler spread and delay spread, we show in Section II-B that, when pulse-shaped transmission/reception is used to communicate over a continuous-time channel that exhibits wide-sense stationary uncorrelated scattering (WSSUS) with limited Doppler and delay spreads, and when the combined transmission/reception pulse has a width of at most one symbol interval, the resulting system yields a discrete-time channel parameterization whose discrete Fourier coefficients become uncorrelated in the large-block limit (i.e., $N \to \infty$).

A. Block-Fading CE-BEM Doubly Selective Channel

Our discrete-time block-fading DSC model is now summarized. Within a fading block of length N, we assume that the channel output can be described as

$$y[n] = \sqrt{\rho} \sum_{l=0}^{N_{\mathsf{delay}}-1} h[n;l]x[n-l] + v[n] \qquad n \in \{0,\ldots,N-1\} \tag{1}$$

where $\{x[n]\}$ is the channel input, $\{h[n;l]\}_{l=0}^{N_{\mathsf{delay}}-1}$ is the time-n channel impulse response, and $\{v[n]\}$ is circular white Gaussian noise (CWGN) of unit variance. Here, N_{delay} refers to the discrete delay spread. We assume an energy-preserving channel, i.e., $\sum_{l=0}^{N_{\mathsf{delay}}-1} \mathrm{E}\{|h[n;l]|^2\} = 1$, so that ρ describes the signal-to-noise ratio (SNR), as well as the input

power constraint $\frac{1}{N+N_{\mathsf{delay}}-1}\mathrm{E}\left\{\sum_{i=-N_{\mathsf{delay}}+1}^{N-1}|x[i]|^2\right\} \leq 1$. Defining $\boldsymbol{y}\triangleq[y[0],\ldots,y[N-1]]^\top$, $\boldsymbol{v}\triangleq[v[0],\ldots,v[N-1]]^\top$ and $\boldsymbol{x}\triangleq[x[-N_{\mathsf{delay}}+1],\ldots,x[N-1]]^\top$, we obtain the vector model

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

where $\boldsymbol{H}\in\mathbb{C}^{N\times(N+N_{\mathsf{delay}}-1)}$ is given element-wise as $[\boldsymbol{H}]_{p,q}\triangleq h[p;p+N_{\mathsf{delay}}-1-q]$.

Using a discrete Fourier expansion of the lth tap trajectory $\{h[n;l]\}_{n=0}^{N-1}$

$$h[n;l]=\frac{1}{\sqrt{N}}\sum_{k=-(N_{\mathsf{Dopp}}-1)/2}^{(N_{\mathsf{Dopp}}-1)/2}\lambda[k;l]e^{j\frac{2\pi}{N}nk}$$
$$n\in\{0,\ldots,N-1\} \tag{3}$$

where $N_{\mathsf{Dopp}}\leq N$ denotes the discrete Doppler spread, we will assume that the Fourier coefficients are uncorrelated, i.e., $\mathrm{E}\{\lambda[k;l]\lambda^*[k-p;l-q]\}=0$ when either $p\neq 0$ or $q\neq 0$, as well as Gaussian. Using the definitions $\boldsymbol{h_l}\triangleq[h[0;l],\ldots,h[N-1;l]]^\top$, $\boldsymbol{h}\triangleq[\boldsymbol{h}_0^\top,\ldots,\boldsymbol{h}_{N_{\mathsf{delay}}-1}^\top]^\top$, $\boldsymbol{\lambda_l}\triangleq\left[\lambda[-\frac{N_{\mathsf{Dopp}}-1}{2};l],\ldots,\lambda[\frac{N_{\mathsf{Dopp}}-1}{2};l]\right]^\top$, and $\boldsymbol{\lambda}\triangleq[\lambda_0^\top,\ldots,\lambda_{N_{\mathsf{delay}}-1}^\top]^\top$, (3) can be written in vector form as

$$\boldsymbol{h}=\boldsymbol{U}\boldsymbol{\lambda} \tag{4}$$

where $\boldsymbol{U}\triangleq\boldsymbol{I}_{N_{\mathsf{delay}}}\otimes\bar{\boldsymbol{F}}$ and where $\bar{\boldsymbol{F}}\in\mathbb{C}^{N\times N_{\mathsf{Dopp}}}$ is given element-wise as $[\bar{\boldsymbol{F}}]_{n,m}\triangleq\frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}n(m-(N_{\mathsf{Dopp}}-1)/2)}$. Given our previous assumptions, $\boldsymbol{\lambda}$ is zero-mean Gaussian with diagonal positive-definite covariance matrix $\boldsymbol{R}_\lambda=E\{\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathsf{H}}\}$ such that $\mathrm{tr}\{\boldsymbol{R}_\lambda\}=N$. The discrete Fourier expansion model (3) is sometimes referred to as a complex-exponential basis expansion model (CE-BEM) after [17]. For later use, we define the spreading index $\gamma\triangleq N_{\mathsf{Dopp}}N_{\mathsf{delay}}/N$. The role of γ on the multiplexing gain η of the CE-BEM DSC will become evident in Section III.

Across blocks, we assume that the channel coefficients are independent and identical distributed. This assumption can be justified for block-interleaved systems or for time-division or frequency-hopped systems where blocks are sufficiently separated across time and/or frequency. Finally, we assume that there is no interblock interference, as when a suitable guard interval has been placed between blocks.

B. Connection to Pulse-Shaped Communication Across WSSUS Channels

In an effort to justify the block-fading CE-BEM DSC model described in Section II-A and assumed for the analyses in Sections III–IV, we now draw parallels to continuous-time pulse-shaped communication. In particular, we show that pulse-shaped communication can, in certain cases, yield a discrete-time channel parameterization whose discrete Fourier coefficients become uncorrelated as $N\to\infty$.

Consider a baseband-equivalent wireless multipath channel that can be modeled as a linear time-variant (LTV) distortion plus an additive noise

$$y(t)=\int h(t;\tau)x(t-\tau)d\tau+v(t). \tag{5}$$

We assume that, over a small time duration of $\mathcal{T}_{\mathsf{small}}$ seconds, the channel $h(t;\tau)$ obeys the following wide-sense stationary uncorrelated scattering (WSSUS) [18] model

$$\mathrm{E}\{h(t;\tau)h^*(t-t_o;\tau-\tau_o)\}=R_{\mathsf{lag;delay}}(t_o;\tau)\delta(\tau_o). \tag{6}$$

(we refer interested readers to the discussion of stationarity over $\mathcal{T}_{\mathsf{small}}$ in [3, p.3097]). If we define

$$R_{\mathsf{Dopp;delay}}(f;\tau)=\int R_{\mathsf{lag;delay}}(t;\tau)e^{-j2\pi ft}dt \tag{7}$$

then the practical assumptions of finite path-length differences and finite rates of path-length variation imply that

$$R_{\mathsf{Dopp;delay}}(f;\tau)=0 \text{ for } \begin{cases} f\notin[-\mathcal{B}_{\mathsf{Dopp}},\mathcal{B}_{\mathsf{Dopp}}] \\ \tau\notin[0,\mathcal{T}_{\mathsf{delay}}]\end{cases} \tag{8}$$

where $\mathcal{T}_{\mathsf{delay}}$ denotes causal delay spread (in seconds) and $\mathcal{B}_{\mathsf{Dopp}}$ single-sided Doppler spread (in Hz).

Now consider baseband-equivalent modulation, as described by $x(t)=\sum_n x[n]\psi(t-n\mathcal{T}_{\mathsf{s}})$, where $\mathcal{T}_{\mathsf{s}}$ is the sampling interval in seconds and where $\psi(t)$ is a unit-energy pulse, and baseband-equivalent demodulation, as described by the received samples $y[n]=\int y(t)\psi^*(t-n\mathcal{T}_{\mathsf{s}})dt$ for $n\in\mathbb{Z}$. We will assume that the baud rate is larger than the Doppler spread, i.e., $\frac{1}{\mathcal{T}_{\mathsf{s}}}>2\mathcal{B}_{\mathsf{Dopp}}$. From (5), one can write

$$y[n]=\sum_{l=-\infty}^{\infty}h[n;l]x[n-l]+v[n] \tag{9}$$

with $v[n]=\int v(t)\psi^*(t-n\mathcal{T}_{\mathsf{s}})dt$ and

$$h[n;l]=\int\int\psi^*(t)h(t+n\mathcal{T}_{\mathsf{s}};\tau+l\mathcal{T}_{\mathsf{s}})\psi(t-\tau)dtd\tau. \tag{10}$$

Parsing the received signal $\{y[n]\}$ into length-N blocks, we obtain a discrete-time block-fading model akin to (1), but with possibly infinite discrete delay spread. We will assume that the block duration $\mathcal{T}_{\mathsf{burst}}\approx N\mathcal{T}_{\mathsf{s}}$ is less than the small-scale fading duration $\mathcal{T}_{\mathsf{small}}$, so that the WSSUS property holds within each block. Note that the $\frac{1}{\mathcal{T}_{\mathsf{s}}}$ Hz sampling rate implicit in pulse-shaped transmission/reception may limit the capacity of the discrete-time channel relative to the continuous-time channel from which it is derived.

Consider now the block $\{y[n]\}_{n=0}^{N-1}$, for which channel response is characterized by $h[n;l]$ for $n\in\{0,\ldots,N-1\}$ and $l\in\mathbb{Z}$. The lth channel tap trajectory can be parameterized w.l.o.g. using the N-term discrete Fourier expansion

$$h[n;l]=\frac{1}{\sqrt{N}}\sum_{k=-\lfloor N/2\rfloor}^{\lceil N/2\rceil-1}\lambda[k;l]e^{j\frac{2\pi}{N}nk}$$
$$n\in\{0,\ldots,N-1\}. \tag{11}$$

Using the pulse ambiguity function

$$A(\tau,f)\triangleq\int\psi(t)\psi^*(t-\tau)e^{-j2\pi ft}dt \tag{12}$$

we can state the following lemma.

Lemma 1 (Statistics of Discrete Fourier Coefficients): Say that the support of $\psi(t)$ is $\left(-\frac{\mathcal{T}_\psi}{2},\frac{\mathcal{T}_\psi}{2}\right]$ with $\mathcal{T}_\psi\le\frac{\mathcal{T}_\mathsf{s}}{2}$. Then, for any $\zeta\in\left[-\frac{1}{2},\frac{1}{2}\right)$

$$\begin{aligned}\lim_{N\to\infty}\mathrm{E}\left\{\lambda\left[\lfloor\zeta N\rfloor;l\right]\lambda^*\left[\lfloor\zeta N\rfloor-p;l-q\right]\right\}&=\delta[p]\delta[q]\int\left|A\left(\tau,\frac{\zeta}{\mathcal{T}_\mathsf{s}}\right)\right|^2\\&\quad R_{\mathsf{Dopp;delay}}\left(\frac{\zeta}{\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s}\right)d\tau.\end{aligned} \tag{13}$$

Furthermore, $\lim_{N\to\infty}\mathrm{E}\left\{\left|\lambda\left[\lfloor\zeta N\rfloor;l\right]\right|^2\right\}=0$ when either $|\zeta|>\mathcal{B}_\mathsf{Dopp}\mathcal{T}_\mathsf{s}$ or $l\notin[0,\mathcal{T}_\mathsf{delay}/\mathcal{T}_\mathsf{s}+\frac{1}{2}]$.

Proof: See Appendix A ∎

Notice that, in (13), $\lim_{N\to\infty}\mathrm{E}\left\{\left|\lambda\left[\lfloor\zeta N\rfloor;l\right]\right|^2\right\}$ is a local average of $R_{\mathsf{Dopp;delay}}(\frac{\zeta}{\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s})$ over the interval $\tau\in(-\mathcal{T}_\psi,\mathcal{T}_\psi]\subset\left(-\frac{\mathcal{T}_\mathsf{s}}{2},\frac{\mathcal{T}_\mathsf{s}}{2}\right]$. Due to the support of $R_{\mathsf{Dopp;delay}}(f;\tau)$ specified by (8), it follows that $\lim_{N\to\infty}\mathrm{E}\left\{\left|\lambda\left[\lfloor\zeta N\rfloor;l\right]\right|^2\right\}=0$ when either $|\zeta|>\mathcal{B}_\mathsf{Dopp}\mathcal{T}_\mathsf{s}$ or $l\notin\{0,\dots,\mathcal{T}_\mathsf{delay}/\mathcal{T}_\mathsf{s}+\frac{1}{2}\}$.

When using a pulse that satisfies the condition in Lemma 1 and a block size N that is finite but large, Lemma 1 suggests that it is reasonable to approximate the discrete Fourier coefficients $\{\lambda[k;l]\}$, for $k\in\{-\frac{N}{2},\dots,\frac{N}{2}-1\}$ and $l\in\mathbb{Z}$, as uncorrelated with variance

$$\begin{aligned}\mathrm{E}\left\{|\lambda[k;l]|^2\right\}=\int\left|A\left(\tau,\frac{k}{N\mathcal{T}_\mathsf{s}}\right)\right|^2&\\R_{\mathsf{Dopp;delay}}\left(\frac{k}{N\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s}\right)d\tau.&\end{aligned} \tag{14}$$

In this case, $\mathrm{E}\left\{|\lambda[k;l]|^2\right\}$ is zero-valued when either $k\notin\left\{-\frac{N_\mathsf{Dopp}-1}{2},\dots,\frac{N_\mathsf{Dopp}-1}{2}\right\}$ or $l\notin\{0,\dots,N_\mathsf{delay}-1\}$, where

$$N_\mathsf{Dopp}=2\lceil\mathcal{B}_\mathsf{Dopp}\mathcal{T}_\mathsf{s}N\rceil+1 \tag{15}$$

$$N_\mathsf{delay}=\lfloor\mathcal{T}_\mathsf{delay}/\mathcal{T}_\mathsf{s}+\frac{1}{2}\rfloor+1. \tag{16}$$

Furthermore, the N_delay-sample delay spread, in combination with (9), implies that $\{y[n]\}_{n=0}^{N-1}$ depends only on the input samples $\{x[n]\}_{n=-N_\mathsf{delay}+1}^{N-1}$.

In summary, if we apply the large-N approximation (14) to the pulse-shaped continuous-time WSSUS model (5)–(8), and use a pulse $\psi(t)$ with maximum width $\mathcal{T}_\mathsf{s}/2$, then we obtain a discrete-time CE-BEM DSC model that satisfies the conditions of Section II-A with discrete Doppler and delay spreads given by (15)–(16). We note that $\psi(t)$ of width $\mathcal{T}_\mathsf{s}/2$ corresponds to a combined transmission/reception pulse $\int\psi(\tau)\psi^*(t-\tau)d\tau$ of width $\mathcal{T}_\mathsf{s}$. The discrete-time channel model of Section II-A will be assumed for the remainder of the paper.

III. Capacity Analysis

For the noncoherent block-fading CE-BEM DSC described in Section II-A, we now analyze the per-channel-use ergodic capacity, which can be expressed as [19]

$$\mathcal{C}(\rho)=\sup_{\boldsymbol{x}:\mathrm{E}\{\|\boldsymbol{x}\|^2\}\le N+N_\mathsf{delay}-1}\frac{1}{N}\mathrm{I}(\boldsymbol{y};\boldsymbol{x}) \tag{17}$$

where $\mathrm{I}(\boldsymbol{y};\boldsymbol{x})$ denotes mutual information between the channel output and input, and where the supremum is taken over all random input distributions satisfying the power constraint. It is known that all rates below the ergodic capacity can be achieved by coding over a large number of block-fading intervals [19], [20].

We define η, the channel's *multiplexing gain*, as the pre-log factor in the high-SNR expression for the channel capacity

$$\eta=\lim_{\rho\to\infty}\frac{\mathcal{C}(\rho)}{\log\rho}. \tag{18}$$

For the block-fading DSC, the *coherent* ergodic capacity (i.e., when $\boldsymbol{H}$ is known to the receiver), is given by [20]

$$\mathcal{C}_\mathsf{coh}(\rho)=\frac{1}{N}\sup_{\boldsymbol{R}_x\succeq\mathbf{0},\mathrm{tr}\{\boldsymbol{R}_x\}\le N+N_\mathsf{delay}-1}\mathrm{E}\{\log\det[\boldsymbol{I}_N+\rho\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^\mathsf{H}]\} \tag{19}$$

where $\boldsymbol{R}_x=E\{\boldsymbol{x}\boldsymbol{x}^\mathsf{H}\}$ and the expectation is taken over the random matrix $\boldsymbol{H}$. Using $\boldsymbol{R}_x=\boldsymbol{I}$ gives a lower bound on $\mathcal{C}_\mathsf{coh}(\rho)$. Also, any $\boldsymbol{R}_x$ meeting the constraint in (19) satisfies $\boldsymbol{R}_x\preceq(N+N_\mathsf{delay}-1)\boldsymbol{I}$. Thus, we have[4]

$$\begin{aligned}&\frac{1}{N}\mathrm{E}\{\log\det[\boldsymbol{I}_N+\rho\boldsymbol{H}\boldsymbol{H}^\mathsf{H}]\}\\&\quad\le\mathcal{C}_\mathsf{coh}(\rho)\\&\quad\le\frac{1}{N}\mathrm{E}\{\log\det[\boldsymbol{I}_N+\rho(N+N_\mathsf{delay}-1)\boldsymbol{H}\boldsymbol{H}^\mathsf{H}]\}.\end{aligned} \tag{20}$$

Denoting the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^\mathsf{H}$ by $\{\nu_i\}_{i=0}^{N-1}$, we have

$$\begin{aligned}&\frac{1}{N}\sum_{i=0}^{N-1}\mathrm{E}\log(1+\rho\nu_i)\\&\quad\le\mathcal{C}_\mathsf{coh}(\rho)\\&\quad\le\frac{1}{N}\sum_{i=0}^{N-1}\mathrm{E}\log(1+(N+N_\mathsf{delay}-1)\rho\nu_i).\end{aligned} \tag{21}$$

Since the random fading matrix $\boldsymbol{H}\boldsymbol{H}^\mathsf{H}$ is full rank (almost surely)[5] the eigenvalues are positive and $\lim_{\rho\to\infty}\frac{\mathcal{C}_\mathsf{coh}(\rho)}{\log\rho}=1$. Thus, in the coherent case, the multiplexing gain of the doubly selective channel is unity. But, in the noncoherent case, the multiplexing gain is generally less than unity. In particular, we claim that the multiplexing gain of the noncoherent block-fading DSC, in the case of continuously distributed inputs, is $\max(0,1-\frac{N_\mathsf{delay}N_\mathsf{Dopp}}{N})$. To prove this claim, we first derive an upper bound on the pre-log factor of mutual information between the input and output of the block-fading DSC, and later establish the achievability of this bound. Since

[4]Since $\boldsymbol{A}\succeq\boldsymbol{B}\succeq\mathbf{0}$ implies $\log\det\boldsymbol{A}\ge\log\det\boldsymbol{B}$.

[5]This property follows from the fact that first N columns of $\boldsymbol{H}$ form an upper triangular matrix and the diagonal elements $\{h[n;N_\mathsf{delay}-1],0\le n<N\}$ are almost surely nonzero, each being Gaussian with nonzero variance.

the optimal input distribution in terms of mutual information may depend on the SNR ρ, we allow the input distribution to change with respect to ρ to find upper bound on the asymptotic mutual information.

Theorem 1 (Achievable Spectral Efficiency): For the block-fading CE-BEM DSC, any sequence of continuous random input vectors $\{\boldsymbol{x}^\rho\}$ indexed by SNR ρ, satisfying the power constraint $\mathrm{E}\{\|\boldsymbol{x}^\rho\|^2\} \leq N + N_{\mathsf{delay}} - 1$, and converging in distribution to a continuous random vector $\boldsymbol{x}^\infty$, yields

$$\limsup_{\rho\to\infty} \frac{\frac{1}{N}\mathrm{I}(\boldsymbol{y};\boldsymbol{x}^\rho)}{\log\rho} \leq \max\left(0, \frac{N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}}{N}\right). \quad (22)$$

Proof: See Appendix B. ∎

The following lemma specifies a fixed input distribution that achieves the mutual information upper bound given in (22).

Lemma 2 (Achievability): For the block-fading CE-BEM DSC, when the input $\boldsymbol{x}$ is i.i.d. zero-mean unit-variance circular Gaussian

$$\lim_{\rho\to\infty} \frac{\frac{1}{N}\mathrm{I}(\boldsymbol{y};\boldsymbol{x})}{\log\rho} = \max\left(0, \frac{N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}}{N}\right). \quad (23)$$

Proof: See Appendix C. ∎

It can be seen, from (23), that the loss in multiplexing gain, relative to the coherent case, increases with the *spreading index* γ. Since $\gamma \approx 2\mathcal{B}_{\mathsf{Dopp}}\mathcal{T}_{\mathsf{delay}}$, larger values of γ correspond to higher levels of time-frequency dispersion. Thus, our findings, which imply that channel dispersion limits multiplexing gain, are intuitively satisfying. For $\gamma \ll 1$, the multiplexing gain will be close to unity, i.e., that of the coherent case. Such channels have relatively few unknown parameters and thus are not expected to incur much "training overhead." For general $\gamma < 1$, the multiplexing gain of the block-fading DSC, under continuously distributed inputs, coincides with previous results on special cases of this channel: flat fading (i.e., $N_{\mathsf{delay}} = 1, N_{\mathsf{Dopp}} = 1$) [1], [21]; time-selective fading (i.e., $N_{\mathsf{delay}} = 1$) [3]; and frequency-selective fading (i.e., $N_{\mathsf{Dopp}} = 1$) [2].

For $\gamma \geq 1$, Theorem 1 and Lemma 2 establish that the pre-log factor of mutual information with continuous inputs is zero. DSCs for which $\gamma > 1$ can be interpreted as "overspread" channels [22]. As noted by Kailath [23], time and frequency variations of overspread channels are impossible to track even in the absence of noise since they imply that the number of unknown channel parameters $(N_{\mathsf{Dopp}}N_{\mathsf{delay}})$ will be more than the number of received observations (N). Our $\gamma \geq 1$ result can be compared with a related result from Lapidoth [24] that shows that the noncoherent channel capacity grows only *double-logarithmically* when the differential entropy (denoted by $\mathrm{h}(\cdot)$) of the channel matrix satisfies $\mathrm{h}(\boldsymbol{H}) > -\infty$. Intuitively, if $\mathrm{h}(\boldsymbol{H}) > -\infty$, no element of $\boldsymbol{H}$ can be perfectly estimated with the full knowledge of other elements of $\boldsymbol{H}$, so that there are more unknowns than observations. In fact, we make use of this result in our proof.

Note that, because Theorem 1 restricts the input distribution to be continuous, it does not characterize the pre-log factor of the *capacity*[6] of the DSC.

[6]We have not established that the capacity achieving input distribution for our DSC model is a continuous one.

IV. Pilot-Aided Transmission

In this section, we detail the encoding and decoding techniques assumed for the PAT schemes analyzed in this paper. Since a primary advantage of using PAT for noncoherent channels is the application of communication techniques developed for coherent channels, we focus on the use of Gaussian coding and (weighted) minimum-distance decoding via pilot-aided linear MMSE (LMMSE) channel estimates. We are mainly interested in designing PAT schemes that achieve the pre-log factors promised by the mutual information bounds in Theorem 1 and Lemma 2. We restrict our attention to the case where $\gamma < 1$, which allows a nonzero pre-log factor.

A. PAT Encoder

We assume either cyclic-prefixed (CP) or zero-prefixed (ZP) block-transmission, so that

$$[x[-N_{\mathsf{delay}}+1],\ldots,x[-1]] = \begin{cases} \boldsymbol{0} & \text{if } ZP \\ [x[N-N_{\mathsf{delay}}+1],\ldots,x[N-1]] & \text{if } CP. \end{cases} \quad (24)$$

Since, for both CP and ZP, the vector $\boldsymbol{x}' \triangleq [x[0],\ldots,x[N-1]]^\top$ completely specifies the transmission vector $\boldsymbol{x}$ defined in Section Section II-A, we focus our attention on the structure of $\boldsymbol{x}'$. We consider $\boldsymbol{x}'$ generated by the general class of *affine precoding* schemes [25]

$$\boldsymbol{x}' = \boldsymbol{p} + \boldsymbol{B}\boldsymbol{s} \quad (25)$$

where $\boldsymbol{p}$ is a fixed pilot vector, $\boldsymbol{B} \in \mathbb{C}^{N\times N_s}$ is a fixed full-rank linear precoding matrix, and $\boldsymbol{s} \in \mathbb{C}^{N_s}$ is a zero-mean information-bearing symbol vector and we refer to its dimension N_s as "data dimension." For the purpose of achievable-rate analysis, we can assume w.l.o.g. that the columns of $\boldsymbol{B}$ are orthonormal, since the mutual information between $\boldsymbol{s}$ and $\boldsymbol{y}$ remains unaffected by invertible transformations of $\boldsymbol{s}$. Denoting the CP/ZP precoding matrix by $\boldsymbol{M} \in \mathbb{C}^{(N+N_{\mathsf{delay}}-1)\times N}$, so that $\boldsymbol{x} = \boldsymbol{M}\boldsymbol{x}'$, the DSC model (2) becomes

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{H}\boldsymbol{M}(\boldsymbol{p} + \boldsymbol{B}\boldsymbol{s}) + \boldsymbol{v}. \quad (26)$$

The transmitted power constraint $E\{\|\boldsymbol{x}\|^2\} \leq N + N_{\mathsf{delay}} - 1$ will be enforced via constraints on $E_p = \|\boldsymbol{p}\|^2 > 0$ and $E_s = \mathrm{E}\{\|\boldsymbol{s}\|^2\}$.

Defining $\boldsymbol{X}_i = \mathrm{diag}(x[i],\ldots,x[i+N-1])$ and $\boldsymbol{X} = [\boldsymbol{X}_0,\ldots,\boldsymbol{X}_{-N_{\mathsf{delay}}+1}]$, input-output relation (2) can also be written as $\boldsymbol{y} = \sqrt{\rho}\boldsymbol{X}\boldsymbol{h} + \boldsymbol{v}$. Note that, in the sequel, we will use these two input-output representations interchangeably. Due to zero-mean $\boldsymbol{s}$, the pilot and data components of $\boldsymbol{X}$ are $\boldsymbol{P} = \mathrm{E}\{\boldsymbol{X}\}$ and $\boldsymbol{D} = \boldsymbol{X} - \boldsymbol{P}$, respectively. Thus, it follows from (4) that

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{P}\boldsymbol{U}\boldsymbol{\lambda} + \sqrt{\rho}\boldsymbol{D}\boldsymbol{U}\boldsymbol{\lambda} + \boldsymbol{v}. \quad (27)$$

Note that, when the channel statistics $\boldsymbol{U}$ and $\boldsymbol{R}_\lambda$ are known, estimation of $\boldsymbol{h}$ is equivalent to estimation of $\boldsymbol{\lambda}$.

To achieve arbitrarily small probability of decoding error over the block-fading DSC, we construct long codewords that span multiple blocks. Let $\underline{\mathcal{S}}$ denote a codebook in which each codeword $\underline{\boldsymbol{s}}$ spans K blocks. Thus, we can write

$\underline{\boldsymbol{s}} = \left[\boldsymbol{s}^{[0]\top}, \ldots, \boldsymbol{s}^{[K-1]\top}\right]^\top$, where $\boldsymbol{s}^{[k]} \in \mathbb{C}^{N_s \times 1}$ is the "segment" of codeword $\underline{\boldsymbol{s}}$ that corresponds to the kth block. We consider codebooks generated according to a Gaussian distribution, so that each codeword, and its segments, are independently generated with positive-definite segment covariance matrix $\boldsymbol{R}_{\boldsymbol{s}}$. Recall that Gaussian codes are capacity-optimal for coherent Gaussian-noise channels [20].

B. PAT Decoder

We assume that PAT decoding consists of a channel estimation stage followed by a data detection stage. The channel estimator computes the LMMSE estimate of $\boldsymbol{h}$, given the observation $\boldsymbol{y}$, the pilots $\boldsymbol{p}$, and the (joint) second-order statistics of $\boldsymbol{h}$, $\boldsymbol{v}$ and $\boldsymbol{s}$. Specifically, with $\boldsymbol{R}_y = \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}\}$ and $\boldsymbol{R}_{y,h} = \mathrm{E}\{\boldsymbol{y}\boldsymbol{h}^{\mathsf{H}}\}$, the channel estimate is

$$\hat{\boldsymbol{h}} = \boldsymbol{R}_{y,h}^{\mathsf{H}} \boldsymbol{R}_y^{-1} \boldsymbol{y} \tag{28}$$

where, from (27)

$$\boldsymbol{R}_y = \rho \boldsymbol{P}\boldsymbol{U}\boldsymbol{R}_\lambda (\boldsymbol{P}\boldsymbol{U})^{\mathsf{H}} + \rho \mathrm{E}\left\{\boldsymbol{D}\boldsymbol{U}\boldsymbol{R}_\lambda (\boldsymbol{D}\boldsymbol{U})^{\mathsf{H}}\right\} + \boldsymbol{I} \tag{29}$$

$$\boldsymbol{R}_{y,h} = \sqrt{\rho}\boldsymbol{P}\boldsymbol{U}\boldsymbol{R}_\lambda \boldsymbol{U}^{\mathsf{H}}. \tag{30}$$

The channel estimation MSE is given by

$$\sigma_e^2 = \mathrm{E}\left\{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2\right\} = \mathrm{tr}\left\{\boldsymbol{U}\boldsymbol{R}_\lambda \boldsymbol{U}^{\mathsf{H}} - \boldsymbol{R}_{y,h}^{\mathsf{H}} \boldsymbol{R}_y^{-1} \boldsymbol{R}_{y,h}\right\}. \tag{31}$$

We define $\hat{\boldsymbol{H}} \in \mathbb{C}^{N \times (N+N_{\mathsf{delay}}-1)}$ element-wise as $[\hat{\boldsymbol{H}}]_{p,q} = \hat{h}[p; p + N_{\mathsf{delay}} - 1 - q]$ and $\tilde{\boldsymbol{H}} = \boldsymbol{H} - \hat{\boldsymbol{H}}$, which will be used in the sequel.

For data detection, we employ weighted minimum-distance decoding based on the LMMSE channel estimates. Recall that the maximum-likelihood (ML) decoder for coherent Gaussian-noise channels is a weighted minimum-distance decoder [26] and notice that this decoder is relatively simple compared to one that performs joint data detection and channel estimation. Given our multiblock coding scheme, the decoder is specified as

$$\hat{\underline{\boldsymbol{s}}} = \arg\min_{\underline{\boldsymbol{s}} \in \underline{\mathcal{S}}} \sum_{k=0}^{K-1} \left\| \boldsymbol{Q}(\boldsymbol{y}^{[k]} - \sqrt{\rho}\hat{\boldsymbol{H}}^{[k]} \boldsymbol{M}(\boldsymbol{p} + \boldsymbol{B}\boldsymbol{s}^{[k]})) \right\|^2 \tag{32}$$

where $\boldsymbol{y}^{[k]}$ and $\hat{\boldsymbol{H}}^{[k]}$ denote the observation and the estimated channel matrix, respectively, of the kth block. The choice of the weighting matrix $\boldsymbol{Q}$ is, for the moment, arbitrary.

C. Spectral Efficiency of PAT

For PAT, we say that a rate $\mathcal{R}$ is achievable if the probability of decoding error can be made arbitrarily small at that rate. Since our PAT schemes use Gaussian codes, we employ Theorem 1, which bounds the multiplexing gain of noncoherent DSC with continuously distributed inputs, in the following definition.

Definition 1: A PAT scheme is *spectrally efficient* if its achievable rate $\mathcal{R}(\rho)$ over the block-fading CE-BEM DSC satisfies $\lim_{\rho\to\infty} \frac{\mathcal{R}(\rho)}{\log \rho} = \frac{N - N_{\mathsf{Dopp}} N_{\mathsf{delay}}}{N}$.

For the case of flat or frequency-selective channels, MMSE-PAT schemes (i.e., those designed to minimize channel-estimation-error variance) have been shown to be spectrally efficient [1], [2], [7]. In the sequel, we establish that all CP-based affine MMSE-PAT schemes are spectrally inefficient over the CE-BEM strictly DSC and propose a spectrally efficient (nonMMSE) affine PAT scheme.

V. Lossless Linearly Separable PAT

In this section, we focus on affine PAT schemes for which the pilot and data components can be linearly separated without energy loss at the *output* of the CE-BEM DSC channel, i.e., from $\boldsymbol{y}$ in (26) and (27). Practically speaking, these losslessly linearly separable (LLS) PAT schemes are those that enable the receiver to compute channel estimates in the absence of data interference. From (27), it can be seen that the LLS criterion can be stated as

$$(\boldsymbol{P}\boldsymbol{U})^{\mathsf{H}} \boldsymbol{D}\boldsymbol{U} = \boldsymbol{0}, \ \forall\, \boldsymbol{D} \in \mathcal{D} \tag{33}$$

where $\mathcal{D}$ refers to the collection of data matrices constructed from all possible codeword realizations. In the sequel, we use the term MMSE-PAT when referring to any PAT scheme that minimizes the channel-estimation-error variance $\sigma_e^2 = \mathrm{E}\left\{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2\right\}$ subject to a fixed positive pilot energy E_p.

Lemma 3: All MMSE-PAT schemes for the CE-BEM DSC are LLS.

Proof: It has been shown in [9, Theorem 1], [27] that all CP-based affine MMSE-PAT schemes are LLS, and it can be inferred from [8] that ZP-based single-carrier MMSE-PAT schemes are also LLS. ■

A. Achievable Rate

We now analyze the achievable rate of LLS PAT, assuming the encoder/decoder specified in Section IV-B. To do this, we first choose the weighting matrix $\boldsymbol{Q}$ in (32). Let the columns of $\boldsymbol{B}_d$ form an orthonormal basis for the left null space of $\boldsymbol{P}\boldsymbol{U}$. Assuming the LLS condition (33), the projection

$$\boldsymbol{y}_d = \boldsymbol{B}_d^{\mathsf{H}} \boldsymbol{y} = \sqrt{\rho} \underbrace{\boldsymbol{B}_d^{\mathsf{H}} \boldsymbol{H} \boldsymbol{M} \boldsymbol{B}}_{\boldsymbol{H}_d} \boldsymbol{s} + \underbrace{\boldsymbol{B}_d^{\mathsf{H}} \boldsymbol{v}}_{\boldsymbol{v}_d} \tag{34}$$

preserves the data component. Then writing $\boldsymbol{H}_d = \hat{\boldsymbol{H}}_d + \tilde{\boldsymbol{H}}_d$ with estimate $\hat{\boldsymbol{H}}_d = \boldsymbol{B}_d^{\mathsf{H}} \hat{\boldsymbol{H}} \boldsymbol{M} \boldsymbol{B}$ and error $\tilde{\boldsymbol{H}}_d = \boldsymbol{B}_d^{\mathsf{H}} \tilde{\boldsymbol{H}} \boldsymbol{M} \boldsymbol{B}$, we get

$$\boldsymbol{y}_d = \sqrt{\rho}\hat{\boldsymbol{H}}_d \boldsymbol{s} + \underbrace{\sqrt{\rho}\tilde{\boldsymbol{H}}_d \boldsymbol{s} + \boldsymbol{v}_d}_{\boldsymbol{n}}. \tag{35}$$

From [28], we know that the rate-maximizing weighting operator for $\boldsymbol{y}_d$ (under the restricted set of Gaussian codebooks) will be the "whitening operator" $\boldsymbol{R}_n^{-1/2}$, where $\boldsymbol{R}_n = \mathrm{E}\{\boldsymbol{n}\boldsymbol{n}^{\mathsf{H}}\}$. Thus, we use

$$\boldsymbol{Q} = \boldsymbol{R}_n^{-1/2} \boldsymbol{B}_d^{\mathsf{H}} \tag{36}$$

in the decoder (32). [28, Theorem 2] then directly implies[7] the following.

Lemma 4: For an affine PAT scheme that is LLS according to (33) and that uses the weighting factor $\boldsymbol{Q}$ from (36) in decoder (32), the achievable rate is

$$\mathcal{R} = \frac{1}{N}\mathrm{E}\left\{\log\det\left[\boldsymbol{I} + \rho \boldsymbol{R}_n^{-1}\hat{\boldsymbol{H}}_d \boldsymbol{R}_s \hat{\boldsymbol{H}}_d^{\mathsf{H}}\right]\right\}. \tag{37}$$

We note that the rate expression (37) resembles that for the coherent case [20] when $\boldsymbol{n}$ in (35) is considered as "effective" Gaussian noise.

B. Asymptotic Achievable Rate

We now study the achievable rate of LLS PAT in the high-SNR regime. Since the channel estimation error becomes part of the effective noise in (37), the MSE $\sigma_e^2(\rho)$ from (31) directly influences the asymptotic behavior of the achievable rate. The following theorem gives a condition on the MSE $\sigma_e^2(\rho)$ of linearly separable PAT that is sufficient to ensure that the achievable rate's pre-log factor grows in proportion to the data dimension.

Theorem 2: Suppose a $(\boldsymbol{p}, \boldsymbol{B})$ PAT scheme is linearly separable according to (33) and guarantees, for some fixed $\kappa \in \mathbb{R}$, estimation error that satisfies $\sigma_e^2(\rho) = \mathrm{E}\left\|\boldsymbol{h} - \hat{\boldsymbol{h}}\right\|^2 \leq \frac{\kappa}{\rho}$ for all $\rho > 1$. Then its asymptotic achievable rate obeys

$$\lim_{\rho\to\infty}\frac{\mathcal{R}(\rho)}{\log\rho} = \frac{N_s}{N}. \tag{38}$$

Proof: See Appendix D. ∎

When $\sigma_e^2(\rho) < \frac{\kappa}{\rho}$, the effective noise variance $\boldsymbol{R}_n$ remains bounded, enabling the $\log\rho$ growth of achievable rate (37) with pre-log factor equal to the rank of $\hat{\boldsymbol{H}}_d$. The estimation-error condition required for Theorem 2 is quite mild and is satisfied, e.g., by all CP-based affine MMSE-PAT schemes, which is established in Appendix E. In Appendix E, we show that all CP-based affine MMSE-PAT schemes yield $N_s < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ when $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$, i.e., when the channel is *strictly* doubly selective. Putting these two results together, we make the following claim.

Theorem 3 (Spectral Inefficiency): For CE-BEM block-fading DSCs with $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$, all CP-based affine MMSE-PAT schemes are spectrally inefficient.

Proof: See Appendix E. ∎

ZP-based single-carrier MMSE-PAT schemes, as characterized in [8], also yield $N_s < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, and hence are also spectrally inefficient when $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$.

For singly selective channels, however, there do exist spectrally efficient MMSE-PAT schemes, such as those specified for frequency-selective channels (i.e., $N_{\mathsf{Dopp}} = 1$) in [2] and for time-selective channels (i.e., $N_{\mathsf{delay}} = 1$) in [3], [9], [10]. This can be understood by the fact that, in the frequency-(time-) selective case, the effective channel matrix $\boldsymbol{HM}$ has N_{delay} (N_{Dopp}) *deterministic* eigenvectors, known to the transmitter, so that MSE optimal pilot patterns can be designed to estimate the N_{delay} (N_{Dopp}) channel parameters by sacrificing only N_{delay} (N_{Dopp}) signaling dimensions to pilots. In the doubly selective case, however, the eigenvectors of $\boldsymbol{HM}$ are not deterministic and (under our assumptions) unknown to the transmitter, so that pilot patterns that are MSE-optimal for estimation of the $N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ channel parameters consume more than $N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ signaling dimensions. In this sense, Theorem 3 refines the upper bound on the data dimension N_s of CP-based affine MMSE-PAT schemes of strictly doubly selective channels that was presented in [9].

[7]The achievable rate result in [28] is derived assuming MMSE channel estimates. However, when (33) is satisfied, the LMMSE estimates (28) are MMSE because the pilot observations and the channel coefficients are jointly Gaussian.

VI. Spectrally Efficient PAT

As established in Section V, CP-based affine MMSE-PAT schemes, as well as ZP-based single-carrier MMSE-PAT schemes, are spectrally inefficient in strictly doubly selective CE-BEM fading, i.e., when $N_{\mathsf{Dopp}} > 1$ and $N_{\mathsf{delay}} > 1$, because they sacrifice more than $N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ signaling dimensions to pilots. In this section, we design spectrally efficient PAT schemes by side-stepping the MMSE requirement.

Since we have restricted ourselves to nondata-aided channel estimation, we reason that the lossless linear separability criterion (33) is still essential, since, without it, channel estimation would suffer unknown-data interference and, as a result, estimation error would persist even as $\rho \to \infty$. Precise conditions for spectrally efficient PAT are given in the following lemma.

Lemma 5: Suppose that a $(\boldsymbol{p}, \boldsymbol{B})$ PAT scheme satisfies the following conditions:

1) $\boldsymbol{PU}$ is full rank;
2) $\mathrm{rank}(\boldsymbol{B}) = N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$;
3) $\boldsymbol{PU}$ guarantees LLS according to (33).

Then the PAT scheme is spectrally efficient.

Proof: See Appendix F. ∎

In Lemma 5, the first condition avoids an undetermined system of equations during channel estimation, the second enables the transmission of $N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ linearly independent data symbols per block, and the third prevents data-interference during channel estimation. To see how relaxing the MMSE requirements helps in designing spectrally efficient PAT, we recall a necessary requirement for CP-MMSE-PAT [9]: $(\boldsymbol{PU})^{\mathsf{H}}\boldsymbol{PU} = \kappa\boldsymbol{I}$ for some constant κ. The restriction on the pilot $(\boldsymbol{P})$ is less stringent in spectrally efficient PAT design (Lemma 5), allowing us to consider a larger class of PAT schemes. A spectrally efficient PAT (SE-PAT) scheme satisfying these three requirements is now described.

Example 1 (SE-PAT): Assuming N-block transmission over the CE-BEM DSC, consider the pilot index set $\mathcal{P}_s = \{0, N_{\mathsf{delay}}, \ldots, (N_{\mathsf{Dopp}} - 1)N_{\mathsf{delay}}\}$ and the guard index set $\mathcal{G}_s = \{0, \ldots, N_{\mathsf{Dopp}}N_{\mathsf{delay}} - 1\}$. Then construct a ZP-based affine PAT scheme $(\boldsymbol{p}, \boldsymbol{B})$ where

$$p[k] = \begin{cases} \sqrt{\frac{E_p}{N_{\mathsf{Dopp}}}}e^{j\theta[k]} & k \in \mathcal{P}_s \\ 0 & k \notin \mathcal{P}_s \end{cases} \tag{39}$$

for arbitrary $\theta[k] \in \mathbb{R}$ and where $\boldsymbol{B}$ is constructed from the columns of $\boldsymbol{I}_N$ whose indices are *not* in $\mathcal{G}_s$.

For the example scheme, note that the first $N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ time slots are used by pilots while the remaining time slots are used for data transmission, thereby ensuring linear separability. It can be readily verified that $\boldsymbol{B}$ has rank $N-N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ and that $\boldsymbol{PU}$ has full rank, so that all three conditions in Lemma 5 are satisfied. Such SE-PAT schemes are advantageous in that they yield higher achievable rates than spectrally inefficient (e.g., MMSE-PAT) schemes at high SNR.

VII. Conclusion

In this paper, the multiplexing gain (i.e., pre-log factor in the channel capacity expression) of the noncoherent CE-BEM DSC with continuous input distributions was shown to be $\eta = \frac{N-N_{\mathsf{Dopp}}N_{\mathsf{delay}}}{N}$, where N denotes the discrete fading-block interval, N_{Dopp} denotes the channel's discrete Doppler spread, and N_{delay} denotes the channel's discrete delay spread. Furthermore, in the limit of $N \to \infty$, the discrete time CE-BEM DSC model was shown to coincide with a continuous-time WSSUS channel model under pulse-shaped transmission/reception with baud rate $\frac{1}{\mathcal{T}_{\mathsf{s}}} > 2\mathcal{B}_{\mathsf{Dopp}}$ and pulse duration $\mathcal{T}_{\psi} \leq \frac{\mathcal{T}_{\mathsf{s}}}{2}$, where $\mathcal{B}_{\mathsf{Dopp}}$ denotes the single-sided Doppler spread, and $\mathcal{T}_{\mathsf{delay}}$ the single-sided delay spread, of the continuous-time WSSUS channel. When this correspondence holds, the multiplexing gain obeys $\eta \approx 1 - 2\mathcal{B}_{\mathsf{Dopp}}\mathcal{T}_{\mathsf{delay}}$.

In the second part of this paper, CP-based MMSE-PAT schemes (i.e., affine PAT schemes that minimize the channel estimation MSE achieved by a pilot-aided MMSE channel estimator) were shown to be spectrally inefficient (i.e., the pre-log factor in their achievable-rate expressions is less than the channel's multiplexing gain) when the CE-BEM DSC is strictly doubly selective. Sufficient conditions on affine PAT schemes that are spectrally efficient for these channels were then proposed, and an example of a spectrally efficient affine PAT scheme was provided.

Appendix A
Proof of Lemma 1

In this appendix we analyze the statistics of the CE-BEM coefficients $\lambda[k;l]$ by first considering the statistics of the discrete-time impulse response coefficients $h[n;l]$. From (6)–(7) and (10), it can be seen that [see (40)–(44), shown at the bottom of the page]. From (11), we know $\lambda[k;l] = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} h[n;l]e^{-j\frac{2\pi}{N}nk}$ for $k \in \left\{-\frac{N}{2},\ldots,\frac{N}{2}-1\right\}$, so that [see (45)–(47), shown at the bottom of the page], where $u_N[n] = 1$ for $0 \leq n < N$ and $u_N[n] = 0$ otherwise. Combining (44) with (47), we obtain (48)–(49), shown at the bottom of the next page. To analyze the case of large block size N, we make the substitutions $\frac{k}{N} = \zeta$ and $\frac{p}{N} = \xi$ in (49). Then, for the values of $\zeta \in [-\frac{1}{2}, \frac{1}{2})$ and $\xi \in (-1, 1)$ that yield integer-valued

$$
\begin{aligned}
&\mathrm{E}\left\{h[n;l]h^*[n-m;l-q]\right\} \\
&\quad= \int\int\int\int \psi^*(t)\psi(t-\tau)\psi\left(t'\right)\psi^*\left(t'-\tau'\right) \\
&\qquad\times \mathrm{E}\left\{h(t+n\mathcal{T}_{\mathsf{s}};\tau+l\mathcal{T}_{\mathsf{s}})h^*\left(t'+n\mathcal{T}_{\mathsf{s}}-m\mathcal{T}_{\mathsf{s}};\tau'+l\mathcal{T}_{\mathsf{s}}-q\mathcal{T}_{\mathsf{s}}\right)\right\} dtd\tau dt'd\tau' \qquad (40)\\
&\quad= \int\int\int\int \psi^*(t)\psi(t-\tau)\psi(t')\psi^*(t'-\tau') \\
&\qquad\times R_{\mathsf{lag;delay}}(t-t'+m\mathcal{T}_{\mathsf{s}};\tau+l\mathcal{T}_{\mathsf{s}})\delta(\tau-\tau'+q\mathcal{T}_{\mathsf{s}})dtd\tau dt'd\tau' \qquad (41)\\
&\quad= \int\int\int \psi^*(t)\psi(t-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_{\mathsf{s}}) \\
&\qquad\times R_{\mathsf{lag;delay}}(t-t'+m\mathcal{T}_{\mathsf{s}};\tau+l\mathcal{T}_{\mathsf{s}})dtd\tau dt' \qquad (42)\\
&\quad= \int\int\int \psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_{\mathsf{s}})dt' \\
&\qquad\times R_{\mathsf{lag;delay}}(v+m\mathcal{T}_{\mathsf{s}};\tau+l\mathcal{T}_{\mathsf{s}})dvd\tau \qquad (43)\\
&\quad= \int\int\int \psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_{\mathsf{s}})dt' \\
&\qquad\times \int R_{\mathsf{Dopp;delay}}(f;\tau+l\mathcal{T}_{\mathsf{s}})e^{j2\pi f(v+m\mathcal{T}_{\mathsf{s}})}dfdvd\tau \qquad (44)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}\{\lambda[k;l]\lambda^*[k-p;l-q]\} \\
&\quad= \frac{1}{N}\sum_{n'=0}^{N-1}\sum_{n=0}^{N-1}\mathrm{E}\{h[n;l]h^*[n';l-q]\}e^{-j\frac{2\pi}{N}[nk-n'(k-p)]} \qquad (45)\\
&\quad= \frac{1}{N}\sum_{n'=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}u_N[n']u_N[n]\mathrm{E}\{h[n;l]h^*[n';l-q]\}e^{-j\frac{2\pi}{N}[nk-n'(k-p)]} \qquad (46)\\
&\quad= \frac{1}{N}\sum_{n'=-\infty}^{\infty}e^{-j\frac{2\pi}{N}n'p}u_N[n']\sum_{m=-\infty}^{\infty}\mathrm{E}\{h[n'+m;l]h^*[n';l-q]\}e^{-j\frac{2\pi}{N}km}u_N[n'+m] \qquad (47)
\end{aligned}
$$

ζN and ξN, we have (50), shown at the bottom of the page. We now examine the N-dependent term on the RHS of (50): $\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'}$. First, if $|m| > N-1$, then $\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'} = 0$ for all ξ. Next, if $0 \le m \le N-1$, then see (51)–(52), shown at the bottom of the page, and hence, for this range of m

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'} = \begin{cases} 0, & \text{if } \xi \neq 0 \\ 1, & \text{if } \xi = 0. \end{cases} \tag{53}$$

Similarly, if $-N+1 \le m < 0$, then

$$\begin{aligned} &\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'} \\ &= \frac{1}{N}\sum_{n'=-m}^{N-1} e^{-j2\pi\xi n'} \\ &= \frac{1}{N}\sum_{n=0}^{N-1+m} e^{-j2\pi\xi(n-m)} \qquad (54) \\ &= \frac{e^{j2\pi\xi m}}{N}\sum_{n=0}^{N-1+m} e^{-j2\pi\xi n} \qquad (55) \end{aligned}$$

from which an analysis similar to (51)–(52) implies that (53) holds for this range of m as well. Applying these results to (49), we see that, for any $\zeta \in \left[-\frac{1}{2},\frac{1}{2}\right)$, see (56), shown at the bottom of the next page, where we have used the property $\lim_{M\to\infty}\sum_{m=-M}^{M} e^{j2\pi\phi m} = \sum_{i=-\infty}^{\infty}\delta(\phi - i)$. Then, using the fact that $\delta(f\mathcal{T}_s - \zeta - i) = \delta(f - \frac{\zeta}{\mathcal{T}_s} - \frac{i}{\mathcal{T}_s})$, we find that [see (57)–(59), shown at the bottom of the next page, where, for (59), we used the assumptions that $\mathcal{B}_{\text{Dopp}} < \frac{1}{2\mathcal{T}_s}$ and that $R_{\text{Dopp;delay}}(f;\cdot) = 0$ for $f \notin [-\mathcal{B}_{\text{Dopp}}, \mathcal{B}_{\text{Dopp}}]$ in conjunction with the fact that $\zeta \in \left[-\frac{1}{2},\frac{1}{2}\right)$ to write $\sum_{i=-\infty}^{\infty} R_{\text{Dopp;delay}}\left(\frac{\zeta+i}{\mathcal{T}_s};\cdot\right) = R_{\text{Dopp;delay}}\left(\frac{\zeta}{\mathcal{T}_s};\cdot\right)$. Writing (59) in terms of the ambiguity function (12) yields

$$\begin{aligned} &\lim_{N\to\infty}\mathrm{E}\left\{\lambda\left[\lfloor\zeta N\rfloor;l\right]\lambda^*\left[\lfloor\zeta N\rfloor - p;l-q\right]\right\} \\ &= \delta[p]\int A^*\left(\tau,\frac{\zeta}{\mathcal{T}_s}\right)A \\ &\quad\times\left(\tau+q\mathcal{T}_s,\frac{\zeta}{\mathcal{T}_s}\right)R_{\text{Dopp;delay}} \\ &\quad\times\left(\frac{\zeta}{\mathcal{T}_s};\tau+l\mathcal{T}_s\right)d\tau. \end{aligned} \tag{60}$$

$$\begin{aligned} &\mathrm{E}\{\lambda[k;l]\lambda^*[k-p;l-q]\} \\ &= \int\int\int \psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_s)dt' \int R_{\text{Dopp;delay}}(f;\tau+l\mathcal{T}_s)e^{j2\pi fv} \\ &\quad\times\frac{1}{N}\sum_{n'=-\infty}^{\infty} e^{-j\frac{2\pi}{N}n'p}u_N[n']\sum_{m=-\infty}^{\infty} e^{-j\frac{2\pi}{N}km}u_N[n'+m]e^{j2\pi f\mathcal{T}_s m}dfdvd\tau \qquad (48) \\ &= \int\int\int \psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_s)dt' \int R_{\text{Dopp;delay}}(f;\tau+l\mathcal{T}_s)e^{j2\pi fv} \\ &\quad\times\sum_{m=-\infty}^{\infty} e^{j2\pi(f\mathcal{T}_s-\frac{k}{N})m}\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\frac{p}{N}n'}dfdvd\tau \qquad (49) \end{aligned}$$

$$\begin{aligned} &\mathrm{E}\{\lambda[\zeta N;l]\lambda^*[\zeta N-\xi N;l-q]\} \\ &= \int\int\int \psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_s)dt' \int R_{\text{Dopp;delay}}(f;\tau+l\mathcal{T}_s)e^{j2\pi fv} \\ &\quad\times\sum_{m=-\infty}^{\infty} e^{j2\pi(f\mathcal{T}_s-\zeta)m}\frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'}dfdvd\tau \qquad (50) \end{aligned}$$

$$\begin{aligned} \frac{1}{N}\sum_{n'=-\infty}^{\infty} u_N[n']u_N[n'+m]e^{-j2\pi\xi n'} &= \frac{1}{N}\sum_{n'=0}^{N-1-m} e^{-j2\pi\xi n'} \\ &= \begin{cases} \frac{e^{-j2\pi\xi(N-m-1)/2}}{N}\sum_{n=-(N-m-1)/2}^{(N-m-1)/2} e^{-j2\pi\xi n}, & \text{odd } m \\ \frac{e^{-j2\pi\xi(N-m)/2}}{N}\sum_{n=-(N-m)/2}^{(N-m)/2} e^{-j2\pi\xi n} - \frac{1}{N}e^{-j2\pi\xi(N-m)}, & \text{even } m \end{cases} \qquad (51) \\ &= \begin{cases} \frac{\sin(\pi\xi(N-m))}{N\sin(\pi\xi)}e^{-j2\pi\xi(N-m-1)/2}, & \text{odd } m \\ \frac{\sin(\pi\xi(N-m+1))}{N\sin(\pi\xi)}e^{-j2\pi\xi(N-m)/2} - \frac{1}{N}e^{-j2\pi\xi(N-m)}, & \text{even } m \end{cases} \qquad (52) \end{aligned}$$

If the support of $\psi(t)$ is $\left(-\frac{\mathcal{T}_\psi}{2}, \frac{\mathcal{T}_\psi}{2}\right]$, then it can be seen that, when $\mathcal{T}_\psi \leq \frac{\mathcal{T}_\mathsf{s}}{2}$, the functions $A(\tau,\cdot)$ and $A(\tau + q\mathcal{T}_\mathsf{s},\ldots)|_{q\neq 0}$ share no common support, in which case (60) reduces to

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}\left\{\lambda\left[\lfloor\zeta N\rfloor; l\right]\lambda^*\left[\lfloor\zeta N\rfloor - p; l-q\right]\right\} \\ = \delta[p]\delta[q]\int\left|A\left(\tau,\frac{\zeta}{\mathcal{T}_\mathsf{s}}\right)\right|^2 R_{\mathsf{Dopp;delay}} \\ \times\left(\frac{\zeta}{\mathcal{T}_\mathsf{s}};\tau + l\mathcal{T}_\mathsf{s}\right)d\tau. \end{aligned} \tag{61}$$

Appendix B
Proof of Theorem 1

Defining $L = \min(N, N_{\mathsf{delay}}N_{\mathsf{Dopp}})$, we define the vector $\boldsymbol{y}_s = [y[0],\ldots,y[L-1]]^\top$.

Claim: $\lim_{\rho\to\infty}\frac{\mathrm{I}(\boldsymbol{y}_s;\boldsymbol{x}^\rho)}{\log\rho} = 0$.

Proof: Using the chain rule for mutual information [19], we have (62) and (63), shown at the bottom of the page. In the sequel, we analyze each term in (63) separately. In preparation, we define the vectors $\boldsymbol{x}_i^\rho = [x^\rho[i],\ldots,x^\rho[i-N_{\mathsf{delay}}+1]]^\top$ and their "complements" $\bar{\boldsymbol{x}}_i^\rho$, which are composed of elements of $\boldsymbol{x}^\rho$ not in $\boldsymbol{x}_i^\rho$. We also define the channel vectors $\boldsymbol{h}_i = [h[i;0],\ldots,h[i;N_{\mathsf{delay}}-1]]^\top$, where

$$y[i] = \sqrt{\rho}\boldsymbol{h}_i^\top\boldsymbol{x}_i^\rho + v[i]. \tag{64}$$

Next we establish the useful result that $\mathrm{I}(y[i];\boldsymbol{x}_i^\rho) \leq \log\log\rho + \Xi$ for some constant $\Xi\in\mathbb{R}$. Towards this aim, we use a special case of the capacity result from [24, Thm. 4.2], which is stated below.

Lemma 6 (Special Case of Theorem 4.2 From [24]): Consider the following vector input-output relation for CWGN block fading channel model $\boldsymbol{y} = \sqrt{\rho}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}$. The input and

---

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}\left\{\lambda\left[\lfloor\zeta N\rfloor; l\right]\lambda^*\left[\lfloor\zeta N\rfloor - p; l-q\right]\right\} \\ = \delta[p]\int\int\int\psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_\mathsf{s})dt'\int R_{\mathsf{Dopp;delay}}(f;\tau+l\mathcal{T}_\mathsf{s})e^{j2\pi fv} \\ \times\sum_{i=-\infty}^{\infty}\delta(f\mathcal{T}_\mathsf{s}-\zeta-i)df dv d\tau \end{aligned} \tag{56}$$

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}\left\{\lambda\left[\lfloor\zeta N\rfloor; l\right]\lambda^*\left[\lfloor\zeta N\rfloor - p; l-q\right]\right\} \\ = \delta[p]\int\int\int\psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_\mathsf{s})dt' \\ \times\int R_{\mathsf{Dopp;delay}}\left(f'+\frac{\zeta}{\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s}\right)e^{j2\pi(f'+\frac{\zeta}{\mathcal{T}_\mathsf{s}})v}\sum_{i=-\infty}^{\infty}\delta\left(f'-\frac{i}{\mathcal{T}_\mathsf{s}}\right)df' dv d\tau \end{aligned} \tag{57}$$

$$\begin{aligned} = \delta[p]\int\int\int\psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_\mathsf{s})dt' \\ \times\sum_{i=-\infty}^{\infty}R_{\mathsf{Dopp;delay}}(\frac{\zeta+i}{\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s})e^{j2\pi(\zeta+i)v/\mathcal{T}_\mathsf{s}}dv d\tau \end{aligned} \tag{58}$$

$$\begin{aligned} = \delta[p]\int\int\int\psi^*(t'+v)\psi(t'+v-\tau)\psi(t')\psi^*(t'-\tau-q\mathcal{T}_\mathsf{s})dt' \\ \times R_{\mathsf{Dopp;delay}}(\frac{\zeta}{\mathcal{T}_\mathsf{s}};\tau+l\mathcal{T}_\mathsf{s})e^{j2\pi\zeta v/\mathcal{T}_\mathsf{s}}dv d\tau \end{aligned} \tag{59}$$

$$\mathrm{I}(\boldsymbol{y}_s;\boldsymbol{x}^\rho) = \mathrm{I}(y[0];\boldsymbol{x}^\rho) + \sum_{i=1}^{L-1}\mathrm{I}(y[i];\boldsymbol{x}^\rho|y[0],\ldots,y[i-1]) \tag{62}$$

$$\leq \mathrm{I}(y[0];\boldsymbol{x}^\rho) + \sum_{i=1}^{L-1}\mathrm{I}(y[i];\boldsymbol{x}^\rho,y[0],\ldots,y[i-1]) \tag{63}$$

the noise power are constrained as $\mathrm{E}\left\{\|\boldsymbol{x}\|^2\right\} \leq P_1$ and $\mathrm{E}\left\{\|\boldsymbol{v}\|^2\right\} \leq P_2$, respectively, for some positive constants P_1 and P_2. Furthermore assume that the channel fades independently from block to block and only the channel fading statistics are available at both the transmitter and receiver. If the differential entropy (denoted by $\mathrm{h}(\cdot)$) of the channel fading matrix $\boldsymbol{H}$ satisfies $\mathrm{h}(\boldsymbol{H}) > -\infty$, then the high-SNR asymptotic ergodic channel capacity obeys $\limsup_{\rho\to\infty} C(\rho) - \log\log\rho < \infty$.

In our model (64), since the elements of $\boldsymbol{h}_i$ are independent with positive variance, the covariance matrix of $\boldsymbol{h}_i$, denoted by $\boldsymbol{R}_i$, is positive definite, and hence, the differential entropy satisfies $\mathrm{h}(\boldsymbol{h}_i^\top) = \mathrm{h}(\boldsymbol{h}_i) = \log\det\boldsymbol{R}_i > -\infty$. Applying Lemma 6 to (64), it follows that $\mathrm{I}(y[i];\boldsymbol{x}^\rho) \leq \log\log\rho + \Xi$.

The first term in (63) 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, $\mathrm{I}(y[0];\bar{\boldsymbol{x}}_0^\rho|\boldsymbol{x}_0^\rho) = 0$. Since $\mathrm{I}(y[0];\boldsymbol{x}_0^\rho) \leq \log\log\rho + \Xi$, we know $\lim_{\rho\to\infty}\frac{\mathrm{I}(y[0];\boldsymbol{x}^\rho)}{\log\rho} = 0$. Considering the general term inside the summation of (63)

$$\mathrm{I}(y[i];\boldsymbol{x}^\rho, y[0],\ldots,y[i-1]) = \underbrace{\mathrm{I}(y[i];\boldsymbol{x}_i^\rho)}_{\leq\log\log\rho+\Xi} + \underbrace{\mathrm{I}(y[i];\bar{\boldsymbol{x}}_i^\rho|\boldsymbol{x}_i^\rho)}_{=0} + \underbrace{\mathrm{I}(y[i];y[0],\ldots,y[i-1]|\boldsymbol{x}^\rho)}_{T_i}$$

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, $\mathrm{I}(y[i];y[0],\ldots,y[i-1]|\boldsymbol{x}^\rho) = \mathrm{h}(y[i]|\boldsymbol{x}^\rho) - \mathrm{h}(y[i]|\boldsymbol{x}^\rho, y[0],\ldots,y[i-1])$. It follows that

$$\mathrm{h}(y[i]|\boldsymbol{x}^\rho) = \mathrm{E}\{\log(1+\rho\sum_{\ell=0}^{N_{\mathsf{delay}}-1}\mathrm{E}\{|h[i;\ell]|^2\}|x^\rho[i-\ell]|^2)\} \tag{65}$$

where the expectation is with respect to $\boldsymbol{x}^\rho$. Now, given $\{y[0],\ldots,y[i-1]\}$, we split $y[i]$ into MMSE estimate and error as $y[i] = \mathrm{E}\{y[i]|y[0],\ldots,y[i-1],\boldsymbol{x}^\rho\} + \tilde{y}[i]$. Since $\boldsymbol{y}$ is Gaussian given $\boldsymbol{x}^\rho$, we have $\mathrm{h}(y[i]|y[0],\ldots,y[i-1],\boldsymbol{x}^\rho) = \mathrm{E}\log(\mathrm{E}|\tilde{y}[i]|^2)$, where the expectation inside the log is w.r.t. $\boldsymbol{H}$ and $\boldsymbol{v}$ and the expectation outside the log is w.r.t. $\boldsymbol{x}^\rho$. Denoting the covariance of $\boldsymbol{h}_i - \mathrm{E}\{\boldsymbol{h}_i|y[0],\ldots,y[i-1]\}$ by $\tilde{\boldsymbol{R}}_i$, we have $\mathrm{E}|\tilde{y}[i]|^2 = 1 + \rho\boldsymbol{x}_i^{\rho\mathsf{H}}\tilde{\boldsymbol{R}}_i\boldsymbol{x}_i^\rho$. Denote the kth eigenvalue of $\tilde{\boldsymbol{R}}_i$ by $\mu_{i,k}$ and the corresponding unit-norm eigenvector by $\boldsymbol{q}_{i,k}$. Thus, we have $\boldsymbol{x}_i^{\rho\mathsf{H}}\tilde{\boldsymbol{R}}_i\boldsymbol{x}_i^\rho = \sum_{k=0}^{N_{\mathsf{delay}}-1}\mu_{i,k}|\boldsymbol{q}_{i,k}^{\mathsf{H}}\boldsymbol{x}_i^\rho|^2 \geq \mu_{i,\max}|\boldsymbol{q}_{i,\max}^{\mathsf{H}}\boldsymbol{x}_i^\rho|^2$, where $\mu_{i,\max}$ denotes the maximum eigenvalue of $\tilde{\boldsymbol{R}}_i$ and $\boldsymbol{q}_{i,\max}$ denotes the corresponding unit-norm eigenvector. Now define $\kappa_{i,\max} = \inf_{\boldsymbol{x}^\rho\in\mathbb{C}^N}\mu_{i,\max}$. For $i \in \{1,\ldots,L-1\}$, since $\{y[0],\ldots,y[i-1]\}$ correspond to a projection of $\boldsymbol{\lambda}$ onto a subspace of smaller dimension, not *all* the elements of $\boldsymbol{h}_i$ can be estimated perfectly, even in the absence of noise ($\rho=\infty$) (i.e., $\tilde{\boldsymbol{R}}_i \neq \mathbf{0}$), and hence $\kappa_{i,\max} > 0$. Now, $\mathrm{E}|\tilde{y}[i]|^2 \geq 1+\rho\kappa_{i,\max}|\sum_{k=0}^{N_{\mathsf{delay}}-1}q_{i,\max}[k]x^\rho[i-k]|^2$, and hence, see (66), shown at the bottom of the page. Combining (65) and (66), we have $T_i \leq \mathrm{E}\log\frac{1+\rho\sum_{\ell=0}^{N_{\mathsf{delay}}-1}\mathrm{E}\{|h[i;\ell]|^2\}|x^\rho[i-\ell]|^2}{1+\rho\kappa_{i,\max}|\sum_{k=0}^{N_{\mathsf{delay}}-1}q_{i,\max}[k]x^\rho[i-k]|^2}$. Since $\boldsymbol{x}^\rho$ is a sequence of continuous random vectors converging to a continuous random vector, $\lim_{\rho\to\infty}|\sum_{k=0}^{N_{\mathsf{delay}}-1}q_{i,\max}[k]x^\rho[i-k]|^2 > 0$ with probability 1, and $\lim_{\rho\to\infty}\frac{T_i}{\log\rho} = 0$. ∎

Now, if $N \leq N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, the proof is complete since in that case $\boldsymbol{y}_s = \boldsymbol{y}$. For the case $N > N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, we define $\boldsymbol{y}_r = [y[N_{\mathsf{Dopp}}N_{\mathsf{delay}}],\ldots,y[N-1]]^\top$ and, using the chain rule for mutual information, obtain $\mathrm{I}(\boldsymbol{y};\boldsymbol{x}^\rho) = \mathrm{I}(\boldsymbol{y}_s;\boldsymbol{x}^\rho) + \mathrm{I}(\boldsymbol{y}_r;\boldsymbol{x}^\rho|\boldsymbol{y}_s)$. To complete the proof, we need to establish that $\limsup_{\rho\to\infty}\frac{\mathrm{I}(\boldsymbol{y}_r;\boldsymbol{x}^\rho|\boldsymbol{y}_s)}{\log\rho} \leq N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$. For this we have

$$\begin{aligned}\mathrm{I}(\boldsymbol{y}_r;\boldsymbol{x}^\rho|\boldsymbol{y}_s) &= \mathrm{h}(\boldsymbol{y}_r|\boldsymbol{y}_s) - \mathrm{h}(\boldsymbol{y}_r|\boldsymbol{y}_s,\boldsymbol{x}^\rho) &(67)\\ &\leq \mathrm{h}(\boldsymbol{y}_r) - \mathrm{h}(\boldsymbol{y}_r|\boldsymbol{y}_s,\boldsymbol{x}^\rho,\boldsymbol{H}) &(68)\end{aligned}$$

since conditioning reduces entropy. Now, $\mathrm{E}\{|y[n]|^2\} = \mathrm{E}\{|v[n]|^2\} + \rho\sum_{l=0}^{N_{\mathsf{delay}}-1}\mathrm{E}\{|h[n;l]|^2\}\mathrm{E}\{|x[n-l]|^2\} \leq 1+k_1\rho$ for some constant $k_1 \in \mathbb{R}$. Bounding the maximum eigenvalue of the covariance matrix of $\boldsymbol{y}_r$ by the sum of its diagonal elements, we see that $\boldsymbol{R}_{y_r} \preceq (N - N_{\mathsf{delay}}N_{\mathsf{Dopp}})(k_1\rho+1)\boldsymbol{I}_{N-N_{\mathsf{Dopp}}N_{\mathsf{delay}}}$.

Since the Gaussian distribution maximizes the entropy for a given covariance matrix, we have $\mathrm{h}(\boldsymbol{y}_r) \leq \log\det[(N - N_{\mathsf{delay}}N_{\mathsf{Dopp}})(k_1\rho+1)\boldsymbol{I}_{N-N_{\mathsf{Dopp}}N_{\mathsf{delay}}}]$. Finally, $\mathrm{h}(\boldsymbol{y}_r|\boldsymbol{y}_s,\boldsymbol{x}^\rho,\boldsymbol{H})$ is equal to the entropy of the unit variance white noise term in $\boldsymbol{y}_r$, which is bounded and independent of ρ. So, we have $\limsup_{\rho\to\infty}\frac{\mathrm{I}(\boldsymbol{y}_r;\boldsymbol{x}^\rho|\boldsymbol{y}_s)}{\log\rho} \leq N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$.

APPENDIX C
PROOF OF LEMMA 2

Since mutual information is non-negative, it is sufficient to restrict ourselves to the case of $N > N_{\mathsf{Dopp}}N_{\mathsf{delay}}$. We need only to prove that the lower bound on the mutual information with Gaussian inputs satisfies the equality in (22). Using the chain rule for mutual information, we have

$$\begin{aligned}\mathrm{I}(\boldsymbol{y};\boldsymbol{x}) &= \mathrm{I}(\boldsymbol{y};\boldsymbol{x},\boldsymbol{H}) - \mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}) &(69)\\ &\geq \mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H}) - \mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x}). &(70)\end{aligned}$$

$$\mathrm{h}(y[i]|\boldsymbol{x}^\rho, y[0],\ldots,y[i-1]) \geq \mathrm{E}\left\{\log(1+\rho\kappa_{i,\max}\left|\sum_{k=0}^{N_{\mathsf{delay}}-1}q_{i,\max}[k]x^\rho[i-k]\right|^2)\right\} \tag{66}$$

Since $\mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H})$ corresponds to coherent case of perfect receiver CSI and since $\boldsymbol{x}$ is Gaussian with covariance $\boldsymbol{R}_x = \boldsymbol{I}$, we have [20]

$$\mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H}) = \mathrm{E}\{\log\det[\boldsymbol{I}_N + \rho\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}]\}. \tag{71}$$

Since $\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$ is full rank (almost surely), re-using the arguments following (20) yields

$$\lim_{\rho\to\infty}\frac{\mathrm{I}(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{H})}{\log\rho} = N. \tag{72}$$

Now, for matrix $\boldsymbol{X}$ appropriately constructed from the input samples $\{x[i]\}_{i=-N_{\mathsf{delay}}+1}^{N-1}$, (2) can be written as

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

Using the BEM model (4), 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 $\boldsymbol{X}$, the vectors $\boldsymbol{y}$ and $\boldsymbol{\lambda}$ are jointly Gaussian, and hence, using the statistics of $\boldsymbol{\lambda}$ and Jensen's inequality, we have

$$\begin{aligned}\mathrm{I}(\boldsymbol{y};\boldsymbol{\lambda}|\boldsymbol{X}) &= \mathrm{E}\log\det[\boldsymbol{I} + \rho(\boldsymbol{X}\boldsymbol{U})\boldsymbol{R}_\lambda(\boldsymbol{X}\boldsymbol{U})^{\mathsf{H}}] &(73)\\ &\le \log\det\mathrm{E}[\boldsymbol{I} + \rho N(\boldsymbol{X}\boldsymbol{U})^{\mathsf{H}}\boldsymbol{X}\boldsymbol{U}] &(74)\\ &\le N_{\mathsf{Dopp}}N_{\mathsf{delay}}\log\rho + \Xi &(75)\end{aligned}$$

for some constant Ξ, where (75) follows from the fact that $\mathrm{E}(\boldsymbol{X}\boldsymbol{U})^{\mathsf{H}}(\boldsymbol{X}\boldsymbol{U}) \preceq k\boldsymbol{I}_{N_{\mathsf{Dopp}}N_{\mathsf{delay}}}$ for some constant k. So, finally we have

$$\lim_{\rho\to\infty}\frac{\mathrm{I}(\boldsymbol{y};\boldsymbol{H}|\boldsymbol{x})}{\log\rho} \le N_{\mathsf{Dopp}}N_{\mathsf{delay}}. \tag{76}$$

The desired result follows from (70), (72) and (76).

Appendix D
Proof of Theorem 2

According to Lemma 4, a linearly separable PAT scheme with weighting matrix $\boldsymbol{Q}$ in (36) achieves the rate given in (37). To derive a lower bound on the achievable-rate pre-log factor, we first obtain a bound (in the positive semi-definite sense) on $\boldsymbol{R}_n$, the covariance matrix of $\sqrt{\rho}\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{M}\boldsymbol{B}\boldsymbol{s}+\boldsymbol{v}_d$. Because of the orthogonality of pilot and data subspaces of lossless linearly separable PAT, the elements of the (pilot based) channel estimation error matrix $\tilde{\boldsymbol{H}}$ are independent to the noise in the data subspace $\boldsymbol{v}_d$ and also to the data vector $\boldsymbol{s}$. So, we have

$$\begin{aligned}\boldsymbol{R}_n &= \mathrm{E}\{\rho\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{M}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{M}\boldsymbol{B}\boldsymbol{s})^{\mathsf{H}} + \boldsymbol{v}_d\boldsymbol{v}_d^{\mathsf{H}}\}\\ &= \mathrm{E}\{\rho\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{M}\boldsymbol{B}\boldsymbol{R}_s(\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{M}\boldsymbol{B})^{\mathsf{H}}\} + \boldsymbol{I}\\ &\preceq \rho\sigma_e^2E_s\|\boldsymbol{M}\|_F^2\boldsymbol{I} + \boldsymbol{I} &(77)\end{aligned}$$

where the inequality (77) follows from applying the inequalities $\boldsymbol{R}_s \preceq E_s\boldsymbol{I}$, $\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}} \preceq \boldsymbol{I}$, $\mathrm{E}\{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{\mathsf{H}}\} \preceq \sigma_e^2\boldsymbol{I}$, $\boldsymbol{M}\boldsymbol{M}^{\mathsf{H}} \preceq \|\boldsymbol{M}\|_F^2\boldsymbol{I}$ and $\boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{B}_d = \boldsymbol{I}$. Incorporating the condition $\sigma_e^2(\rho) \le \frac{\kappa}{\rho}$, we see that $\boldsymbol{R}_n \preceq C\boldsymbol{I}$ for some constant C, $\forall\rho > 1$. So, we have $\boldsymbol{R}_n^{-1}\hat{\boldsymbol{H}}_d\boldsymbol{R}_s\hat{\boldsymbol{H}}_d^{\mathsf{H}} \succeq \frac{\rho}{C}\hat{\boldsymbol{H}}_d\boldsymbol{R}_s\hat{\boldsymbol{H}}_d^{\mathsf{H}}$ and the achievable rate (37) can be bounded as

$$\begin{aligned}\mathcal{R}(\rho) &\ge \frac{1}{N}\mathrm{E}\{\log\det[\boldsymbol{I} + \frac{\rho}{C}\hat{\boldsymbol{H}}_d\boldsymbol{R}_s\hat{\boldsymbol{H}}_d^{\mathsf{H}}]\} &(78)\\ &\ge \frac{1}{N}\mathrm{E}\{\log\det[\boldsymbol{I} + \frac{\rho\sigma_s^2}{C}\hat{\boldsymbol{H}}_d\hat{\boldsymbol{H}}_d^{\mathsf{H}}]\} &(79)\end{aligned}$$

where σ_s^2 denotes the minimum eigenvalue of $\boldsymbol{R}_s$. Since $\sigma_e^2(\rho) \to 0$ as $\rho\to\infty$, the channel estimates converge almost everywhere to the true channel, i.e., $\lim_{\rho\to\infty}\hat{\boldsymbol{H}} = \boldsymbol{H}$. Also, since $\hat{\boldsymbol{H}}_d\hat{\boldsymbol{H}}_d^{\mathsf{H}}$ has rank equal to $\mathrm{rank}(\boldsymbol{B}) = N_s$, we have $\lim_{\rho\to\infty}\frac{\mathcal{R}(\rho)}{\log\rho} \ge \frac{N_s}{N}$. To derive an upper bound on the achievable-rate's pre-log factor, we use Jensen's inequality to take the expectation inside the $\log\det(\cdot)$ term of (37), thereby obtaining $\lim_{\rho\to\infty}\frac{\mathcal{R}(\rho)}{\log\rho} \le \frac{N_s}{N}$. Together, the upper and lower bounds yield (38).

Appendix E
Proof of Theorem 3

In this proof, we restrict our attention to strictly doubly selective channels, i.e., DSCs for which $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$. Throughout this proof, we consider all indices modulo-N. Let $(\boldsymbol{p},\boldsymbol{B})$ be an arbitrary CP-MMSE PAT scheme for strictly DSC. We establish the desired result in the following two steps:

1) For the CP-MMSE-PAT scheme $(\boldsymbol{p},\boldsymbol{B})$, the achievable rate pre-log factor equals $\mathrm{rank}(\boldsymbol{B})$.
2) For strictly DSCs, any CP-MMSE-PAT scheme $(\boldsymbol{p},\boldsymbol{B})$ obeys $\mathrm{rank}(\boldsymbol{B}) < N - N_{\mathsf{delay}}N_{\mathsf{Dopp}}$.

Step 1) of Proof:

The characterization of CP-based affine MMSE-PAT in [9], [27] establishes that the linear separability condition (33) is satisfied, and furthermore that $\mathrm{E}\{\|\tilde{\boldsymbol{h}}\|^2\} = \mathrm{tr}\{(\boldsymbol{R}_\lambda^{-1} + \frac{\rho E_p}{N}\boldsymbol{I}_{N_{\mathsf{Dopp}}N_{\mathsf{delay}}})^{-1}\}$. Recalling that $\boldsymbol{R}_\lambda$ is diagonal, and defining positive $\alpha_i = [\boldsymbol{R}_\lambda]_{i,i}$, we find $\mathrm{E}\{\|\tilde{\boldsymbol{h}}\|^2\} = \sum_{i=0}^{N_{\mathsf{delay}}N_{\mathsf{Dopp}}-1}(\frac{1}{\alpha_i} + \frac{\rho E_p}{N})^{-1} \le \frac{NN_{\mathsf{delay}}N_{\mathsf{Dopp}}}{\rho E_p}$. Thus, all CP-based affine MMSE-PAT schemes satisfy the hypotheses of Theorem 2, and hence, the pre-log factor of their achievable rates are equal to their corresponding data dimension $\mathrm{rank}(\boldsymbol{B})$.

Step 2) of Proof:

Now, we show that, when $N_{\mathsf{delay}} > 1$ and $N_{\mathsf{Dopp}} > 1$, CP-based affine MMSE-PAT guarantees data dimension $N_s < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$. To establish the condition on N_s, we use the method of contradiction. In particular, we proceed in the following stages.

(i) Assume that there exists a CP-MMSE-PAT scheme for strictly DSC that allows $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$.
(ii) Find the necessary requirements on $\boldsymbol{p}$ and $\boldsymbol{B}$ for such a PAT scheme.
(iii) Establish that the PAT schemes satisfying the requirements obtained in stage (ii) obey $N_s < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, contradicting the initial assumption of stage (i).

Stage (i)—Initial Assumption: Let us assume that there exists a CP-MMSE PAT scheme $(\boldsymbol{p},\boldsymbol{B})$ for strictly DSC that satisfies $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$.

Stage (ii)—Necessary Requirements: To attain the minimal MSE for a given pilot energy, the necessary conditions on CP-based affine MMSE-PAT for the CE-BEM DSC (established in [9], [27]) can be expressed as the pair (80)–(81) using $p[i] = [\boldsymbol{p}]_i, b_q[i] = [\boldsymbol{B}]_{i,q}, \mathcal{N}_{\mathsf{delay}} = \{-N_{\mathsf{delay}}+1,\ldots,N_{\mathsf{delay}}-1\}$, and $\mathcal{N}_{\mathsf{Dopp}} = \{-N_{\mathsf{Dopp}}+1,\ldots,N_{\mathsf{Dopp}}-1\}$

$$\begin{aligned}\sum_{i=0}^{N-1} b_q[i]p^*[i-k]e^{-j\frac{2\pi}{N}mi} &= 0\ \forall\, k\in\mathcal{N}_{\mathsf{delay}}\\ &\forall\, m\in\mathcal{N}_{\mathsf{Dopp}}\\ &\forall\, q\in\{0,\ldots,N_s-1\} &(80)\end{aligned}$$

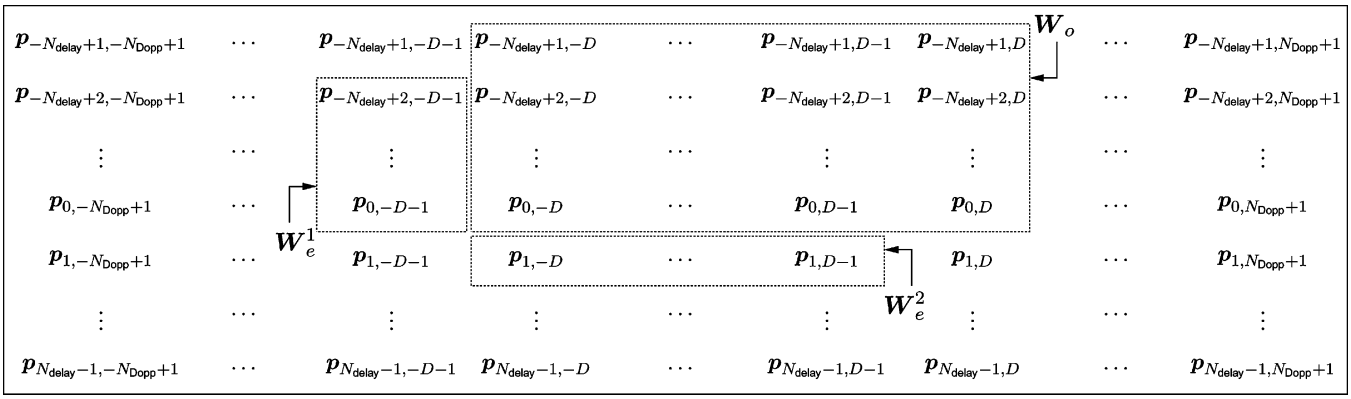

Fig. 1. Elements of the set $\{\boldsymbol{p}_{k,m}, k \in \mathcal{N}_{\mathsf{delay}}, m \in \mathcal{N}_{\mathsf{Dopp}}\}$ arranged in a grid, using $D = (N_{\mathsf{Dopp}} - 1)/2$.

$$\frac{1}{E_p}\sum_{i=0}^{N-1} p[i]p^*[i-k]e^{-j\frac{2\pi}{N}mi} = \delta[k]\delta[m] \quad \forall\, k \in \mathcal{N}_{\mathsf{delay}} \quad \forall\, m \in \mathcal{N}_{\mathsf{Dopp}}. \tag{81}$$

We recall from [9] that condition (80) says that pilots and data should be multiplexed in a way that preserves orthogonality at the channel output, while condition (81) says that pilots should be constructed so that the channel modes are independently excited with equal energy. Also, notice that (80) states the linear separability condition (33) in the case of a CE-BEM DSC. Defining (82), shown at the bottom of the page, as a (normalized) k-time-shifted and m-frequency-shifted version of pilot vector $\boldsymbol{p}$, and constructing matrix $\boldsymbol{W}$ from columns $\{\boldsymbol{p}_{k,m}, k \in \mathcal{N}_{\mathsf{delay}}, m \in \mathcal{N}_{\mathsf{Dopp}}\}$, (80) can be conveniently rewritten as $\boldsymbol{W}^{\mathsf{H}}\boldsymbol{B} = \mathbf{0}$. It will be convenient to visualize the elements of $\{\boldsymbol{p}_{k,m}, k \in \mathcal{N}_{\mathsf{delay}}, m \in \mathcal{N}_{\mathsf{Dopp}}\}$ arranged in a grid, as in Fig. 1. For this, we use the abbreviation $D = (N_{\mathsf{Dopp}} - 1)/2$.

Let $(\boldsymbol{p}, \boldsymbol{B})$ be a CP-based affine MMSE-PAT scheme with data dimension $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$ (i.e., $\mathrm{rank}(\boldsymbol{B}) = N_s$). We now deduce some essential properties of $\boldsymbol{p}$. Defining

$$r_{k,m} \triangleq \langle \boldsymbol{p}_{0,0}, \boldsymbol{p}_{k,m}\rangle = \frac{1}{E_p}\sum_{i=0}^{N-1} p[i]p^*[i+k]e^{-j\frac{2\pi}{N}mi} \tag{83}$$

where $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{y}^H\boldsymbol{x}$ denotes the inner product, the MMSE condition (81) implies that

$$r_{k,m} = \delta[k]\delta[m], \quad \forall\, k \in \mathcal{N}_{\mathsf{delay}}, \quad \forall\, m \in \mathcal{N}_{\mathsf{Dopp}}. \tag{84}$$

Note also that

$$\langle \boldsymbol{p}_{k_1,m_1}, \boldsymbol{p}_{k_2,m_2}\rangle = e^{j\frac{2\pi}{N}(m_2-m_1)k_1} r_{k_2-k_1,m_2-m_1} \tag{85}$$

$$r^*_{k,m} = \langle \boldsymbol{p}_{k,m}, \boldsymbol{p}_{0,0}\rangle = e^{-j\frac{2\pi}{N}mk} r_{-k,-m}. \tag{86}$$

Together, (84) and (85) imply that the elements within any rectangle of height N_{delay} and width N_{Dopp} in Fig. 1 are orthonormal. In addition to being a CP-MMSE-PAT, $(\boldsymbol{p}, \boldsymbol{B})$ satisfies $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, which results in additional restrictions on $\boldsymbol{p}$ and $\boldsymbol{B}$ that are stated in Lemma 7.

Lemma 7: For a CP-MMSE-PAT with $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, either $|r_{0,N_{\mathsf{Dopp}}}| = 1$ or $|r_{N_{\mathsf{delay}},0}| = 1$.

Proof: Let $\boldsymbol{W}_o$ be the matrix constructed from columns $\{\boldsymbol{p}_{k,m}, k \in \{0,-1,\ldots,-N_{\mathsf{delay}}+1\},\ m \in \{-D,\ldots,D\}\}$. Since these columns form a rectangle of height N_{delay} and width N_{Dopp} in Fig. 1, we know they are orthonormal. Furthermore, since these columns form a subset of the columns of $\boldsymbol{W}$, we know that $\mathrm{rank}(\boldsymbol{W}) \geq N_{\mathsf{Dopp}}N_{\mathsf{delay}}$. But, since the MMSE condition (80)$\boldsymbol{W}^{\mathsf{H}}\boldsymbol{B} = \mathbf{0}$ implies that the nullspace of $\boldsymbol{W}^{\mathsf{H}}$ has a dimension of least $N_s = N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, i.e., that $\mathrm{rank}(\boldsymbol{W}) \leq N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, we see that $\mathrm{rank}(\boldsymbol{W}) = N_{\mathsf{Dopp}}N_{\mathsf{delay}}$. Hence, the columns of $\boldsymbol{W}_o$ form an orthonormal basis for the columns of $\boldsymbol{W}$, which implies

$$\boldsymbol{p}_{k,m} = \sum_{i=0}^{N_{\mathsf{delay}}-1}\sum_{j=-D}^{D} \langle \boldsymbol{p}_{k,m}, \boldsymbol{p}_{-i,j}\rangle \boldsymbol{p}_{-i,j} \quad \forall\, k \in \mathcal{N}_{\mathsf{delay}},\ \forall\, m \in \mathcal{N}_{\mathsf{Dopp}}, \tag{87}$$

$$= \sum_{i=0}^{N_{\mathsf{delay}}-1}\sum_{j=-D}^{D} e^{j\frac{2\pi}{N}(j-m)k} r_{-i-k,j-m}\boldsymbol{p}_{-i,j} \quad \forall\, k \in \mathcal{N}_{\mathsf{delay}},\ \forall\, m \in \mathcal{N}_{\mathsf{Dopp}}. \tag{88}$$

Now let $\boldsymbol{W}_e^1 = [\boldsymbol{p}_{0,-D-1}, \ldots, \boldsymbol{p}_{-N_{\mathsf{delay}}+2,-D-1}]$ (see Fig. 1). Considering that we can enclose these elements in a height-N_{delay} and width-N_{Dopp} rectangle in Fig. 1, we can see that the columns of $\boldsymbol{W}_e^1$ form an orthonormal set, and that the columns of $\boldsymbol{W}_e^1$ are orthogonal to most columns in $\boldsymbol{W}_o$. Using (88) to write the columns of $\boldsymbol{W}_e^1$ as a linear combination of those columns in $\boldsymbol{W}_o$ that are not orthogonal to those in $\boldsymbol{W}_e^1$, we have

$$\boldsymbol{W}_e^1 = [\boldsymbol{p}_{0,D}, \boldsymbol{p}_{-1,D}, \ldots, \boldsymbol{p}_{-N_{\mathsf{delay}}+1,D}]\boldsymbol{M}_1 \tag{89}$$

for $\boldsymbol{M}_1 \in \mathbb{C}^{N_{\mathsf{delay}} \times N_{\mathsf{delay}}-1}$ such that (see the first equation shown at the bottom of the next page). Now, letting

$$\boldsymbol{p}_{k,m} = \frac{1}{\sqrt{E_p}}\left[p[k]e^{j\frac{2\pi}{N}m\cdot 0}, p[k+1]e^{j\frac{2\pi}{N}m\cdot 1}, \ldots, p[k+N-1]e^{j\frac{2\pi}{N}m(N-1)}\right]^\top \tag{82}$$

$\boldsymbol{W}_e^2 = [\boldsymbol{p}_{1,-D}, \ldots, \boldsymbol{p}_{1,D-1}]$ and carrying out a similar procedure, we have

$$\boldsymbol{W}_e^2 = [\boldsymbol{p}_{-N_{\mathsf{delay}}+1,-D}, \boldsymbol{p}_{-N_{\mathsf{delay}}+1,-D+1}, \cdots, \boldsymbol{p}_{-N_{\mathsf{delay}}+1,D}]\boldsymbol{M}_2 \tag{90}$$

for $\boldsymbol{M}_2 \in \mathbb{C}^{N_{\mathsf{Dopp}} \times N_{\mathsf{Dopp}}-1}$ such that (see the second equation shown at the bottom of the page).

Notice that the columns of $\boldsymbol{W}_e^1$ must be orthogonal to those in $\boldsymbol{W}_e^2$ since they can all be placed inside a height-N_{delay} and width-N_{Dopp} rectangle in Fig. 1. Since the basis expansions of $\boldsymbol{W}_e^1$ and $\boldsymbol{W}_e^2$ share the common basis vector $\boldsymbol{p}_{-N_{\mathsf{delay}}+1,D}$, the contribution from $\boldsymbol{p}_{-N_{\mathsf{delay}}+1,D}$ to either $\boldsymbol{W}_e^1$ or $\boldsymbol{W}_e^2$ must be zero, i.e., either (91) or (92) must hold

$$\begin{aligned} r_{-1,N_{\mathsf{Dopp}}} &= r_{-2,N_{\mathsf{Dopp}}} = \cdots \\ &= r_{-N_{\mathsf{delay}}+1,N_{\mathsf{Dopp}}} = 0 \end{aligned} \tag{91}$$

$$\begin{aligned} r_{-N_{\mathsf{delay}},1} &= r_{-N_{\mathsf{delay}},2} = \cdots \\ &= r_{-N_{\mathsf{delay}},N_{\mathsf{Dopp}}-1} = 0. \end{aligned} \tag{92}$$

When (91) holds, $\boldsymbol{M}_1$ becomes upper triangular, and (89) implies

$$\boldsymbol{p}_{0,-D-1} = r_{0,N_{\mathsf{Dopp}}}\boldsymbol{p}_{0,D} \tag{93}$$

in which case the unit-norm property of $\boldsymbol{p}_{0,-D-1}$ and $\boldsymbol{p}_{0,D}$ implies that $|r_{0,N_{\mathsf{Dopp}}}| = 1$. When (92) holds, $\boldsymbol{M}_2$ becomes upper triangular, and (90) implies

$$\boldsymbol{p}_{1,-D} = r_{-N_{\mathsf{delay}},0}\boldsymbol{p}_{-N_{\mathsf{delay}}+1,-D} \tag{94}$$

in which case $|r_{-N_{\mathsf{delay}},0}| = 1$. Applying (86), this can be translated to $|r_{N_{\mathsf{delay}},0}| = 1$. ∎

Stage (iii)—Establish Contradiction: Now we examine the implications of either $|r_{N_{\mathsf{delay}},0}| = 1$ or $|r_{0,N_{\mathsf{Dopp}}}| = 1$ on the MMSE pilot vector $\boldsymbol{p}$. In each case, we deduce that $N_s \neq N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$, which contradicts our original assumption, thereby completing the proof.

We start with the first case, where $|r_{0,N_{\mathsf{Dopp}}}| = 1$. Since $r_{0,N_{\mathsf{Dopp}}} = e^{j\theta}$ for some $\theta \in \mathbb{R}$, from (82) and (93), we have

$$p[i](e^{-j\frac{2\pi}{N}N_{\mathsf{Dopp}}i} - e^{j\theta}) = 0, \quad \forall\, i \in \{0, \ldots, N-1\}. \tag{95}$$

Thus, in order to avoid $\boldsymbol{p} = \mathbf{o}$, which would not satisfy the MMSE-PAT requirement (81), we must have $\theta = -\frac{2\pi}{N}N_{\mathsf{Dopp}}q$ for some $q \in \{0, \ldots, N-1\}$. In this case, (95) implies that $p[i]$ will be nonzero only if $i = q + \frac{kN}{N_{\mathsf{Dopp}}}$ for $k \in \mathbb{Z}$ such that $\frac{kN}{N_{\mathsf{Dopp}}} \in \mathbb{Z}$. Now, for $k \in \mathbb{Z}$, we define

$$a_q[k] = \begin{cases} \left|p[q + \frac{kN}{N_{\mathsf{Dopp}}}]\right|^2 & \text{if } \frac{kN}{N_{\mathsf{Dopp}}} \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \tag{96}$$

and use requirement (84) to claim that $\frac{1}{E_p}\sum_{i=0}^{N_{\mathsf{Dopp}}-1} a_q[i]e^{-j\frac{2\pi}{N_{\mathsf{Dopp}}}mi} = \delta[m]$, $\forall m \in \mathcal{N}_{\mathsf{Dopp}}$, which can be met if and only if

$$a_q[i] = \frac{E_p}{N_{\mathsf{Dopp}}}, \quad \forall\, i \in \{0, \ldots, N_{\mathsf{Dopp}} - 1\}. \tag{97}$$

From (96), it follows that (97) can be met if and only if $\frac{N}{N_{\mathsf{Dopp}}} \in \mathbb{Z}$. Now, if $\frac{N}{N_{\mathsf{Dopp}}} \in \mathbb{Z}$, then one can recognize the pilot sequence specified by (96) and (97) as being the time domain Kronecker delta (TDKD) MMSE-PAT scheme from [9], [27], for which $N_s = N - (2N_{\mathsf{delay}} - 1)N_{\mathsf{Dopp}} < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$.

We continue with the second case, where $|r_{N_{\mathsf{delay}},0}| = 1$. Since $r_{N_{\mathsf{delay}},0} = e^{j\bar{\theta}}$ for some $\bar{\theta} \in \mathbb{R}$, from (82), (86), and (94), it follows that

$$p[i] = e^{j\bar{\theta}}p[i + N_{\mathsf{delay}}]. \tag{98}$$

Keeping in mind our modulo-N assumption on time-domain indexing, say that L is the largest integer in $\{1, \ldots, N_{\mathsf{delay}}\}$ for which both $\frac{N}{L} \in \mathbb{Z}$ and $p[i] = e^{j\phi}p[i + L]$ for some $\phi \in \mathbb{R}$. Note that, if $\frac{N}{N_{\mathsf{delay}}} \in \mathbb{Z}$, then $L = N_{\mathsf{delay}}$, else $L < N_{\mathsf{delay}}$. Furthermore, modulo-N indexing implies $\phi = \frac{2\pi}{N}Lq$ for some $q \in \mathbb{Z}$. Let $\breve{\boldsymbol{p}}$ denote the N-point unitary discrete Fourier transform (DFT) of $\boldsymbol{p}$. For a sequence $\boldsymbol{p}$ obeying (98), we have

$$\breve{p}[k] = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} p[i]e^{-j\frac{2\pi}{N}ik} \tag{99}$$

$$= \frac{1}{\sqrt{N}}\sum_{n=0}^{L-1} p[n]e^{-j\frac{2\pi}{N}nk}\sum_{m=0}^{N/L-1} e^{-j\frac{2\pi}{N/L}(k-q)m} \tag{100}$$

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

$$\boldsymbol{M}_2 = \begin{bmatrix} r_{-N_{\mathsf{delay}},0} & e^{-j\frac{2\pi}{N}}r_{-N_{\mathsf{delay}},-1} & \cdots & e^{-j\frac{2\pi}{N}(N_{\mathsf{Dopp}}-2)}r_{-N_{\mathsf{delay}},-N_{\mathsf{Dopp}}+2} \\ e^{j\frac{2\pi}{N}}r_{-N_{\mathsf{delay}},1} & r_{-N_{\mathsf{delay}},0} & \cdots & e^{-j\frac{2\pi}{N}(N_{\mathsf{Dopp}}-3)}r_{-N_{\mathsf{delay}},-N_{\mathsf{Dopp}}+3} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j\frac{2\pi}{N}(N_{\mathsf{Dopp}}-1)}r_{-N_{\mathsf{delay}},N_{\mathsf{Dopp}}-1} & e^{j\frac{2\pi}{N}(N_{\mathsf{Dopp}}-2)}r_{-N_{\mathsf{delay}},N_{\mathsf{Dopp}}-2} & \cdots & e^{j\frac{2\pi}{N}}r_{-N_{\mathsf{delay}},1} \end{bmatrix}$$

and hence $\breve{p}[k] = 0$ for $k \notin \{q, q+\frac{N}{L}, \ldots, q+\frac{N}{L}(L-1)\}$. The MMSE requirement (81) can be written in terms of $\breve{\boldsymbol{p}}$ as [9], [27]

$$\frac{1}{E_p}\sum_{i=0}^{N-1}\breve{p}[i]\breve{p}^*[i-k]e^{-j\frac{2\pi}{N}mi} = \delta[k]\delta[m] \quad \forall\, k \in \mathcal{N}_{\mathsf{Dopp}}, \quad \forall\, m \in \mathcal{N}_{\mathsf{delay}}. \quad (101)$$

Defining $\left|\breve{p}[q+\frac{nN}{L}]\right|^2 = \breve{a}_q[n]$ for $n \in \{0, \ldots, L-1\}$ and using (101) with $k = 0$, we require

$$\frac{1}{E_p}\sum_{n=0}^{L-1}\breve{a}_q[n]e^{-j\frac{2\pi}{N}(n\frac{N}{L}+q)m} = \delta[m], \quad \forall\, m \in \mathcal{N}_{\mathsf{delay}}. \quad (102)$$

Since the magnitude of the left side of (102) is L-periodic, (102) can not be satisfied when $L < N_{\mathsf{delay}}$. Now, if $L = N_{\mathsf{delay}}$, then the only sequence $\{\breve{a}_q[i]\}$ satisfying the requirement (102) is $\breve{a}_q[n] = c$, $\forall n \in \{0, \ldots, N_{\mathsf{delay}} - 1\}$, for constant c. This can be recognized as the frequency domain Kronecker delta (FDKD) MMSE-PAT scheme from [9], [27], for which $N_s = N - (2N_{\mathsf{Dopp}} - 1)N_{\mathsf{delay}} < N - N_{\mathsf{Dopp}}N_{\mathsf{delay}}$.

APPENDIX F
PROOF OF LEMMA 5

First, we establish that, for the PAT schemes satisfying the hypothesis, the total estimation error satisfies $\sigma_e^2 \leq \frac{\kappa}{\rho}, \forall \rho$. Constructing $\boldsymbol{B}_p$ using the orthonormal basis for the column space of $\boldsymbol{PU}$, we consider the projection

$$\boldsymbol{y}_p = \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{y} = \sqrt{\rho}\boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{PU}\boldsymbol{\lambda} + \sqrt{\rho}\boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{DU} + \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{v}. \quad (103)$$

Since the PAT is lossless linearly separable satisfying (33), the projection $\boldsymbol{y}_p$ in (103) captures all the pilot energy and $\boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{DU} = \mathbf{0}$. Denoting $\boldsymbol{G} = \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{PU}$ and $\boldsymbol{v}_p = \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{v}$, we have

$$\boldsymbol{y}_p = \sqrt{\rho}\boldsymbol{G}\boldsymbol{\lambda} + \boldsymbol{v}_p. \quad (104)$$

Since $\boldsymbol{PU}$ is full rank, it follows that the matrix $\boldsymbol{G}$ is full rank. Note that $\sigma_e^2 = \mathrm{E}\{\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2\}$ where $\hat{\boldsymbol{\lambda}}$ denotes the LMMSE estimate of $\boldsymbol{\lambda}$. Using the zero forcing estimate from (104) to upper bound σ_e^2, we have

$$\sigma_e^2 \leq \frac{1}{\rho}\mathrm{tr}\{(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1}\}. \quad (105)$$

Since $\boldsymbol{G}$ is full rank, we have $\mathrm{tr}\{(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1}\} \leq \kappa$ for some $\kappa \in \mathbb{R}$. Now, the desired result follows from the application of Lemma 4.

ACKNOWLEDGMENT

The authors would to thank the two anonymous reviewers for the valuable comments and suggestions.

REFERENCES

[1] L. Zheng and D. Tse, "Communication over the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

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

[3] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[4] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[5] T. Koch and A. Lapidoth, "On multipath fading channels at high SNR," in *Proc. IEEE Internat. Symp. Information Theory*, Toronto, ON, Jul. 2008, pp. 1572–1576.

[6] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Process. Mag.*, vol. 21, no. 11, pp. 12–25, Nov. 2004.

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[8] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[9] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1148–1160, Mar. 2008.

[10] 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*, New York, 2005, pp. 801–805.

[11] R. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.

[12] G. Durisi, H. Bölcskei, and S. Shamai, "Capacity of underspread WSSUS fading channels in the wideband regime," in *Proc. IEEE Int. Symp. Information Theory*, Jul. 2006, pp. 1500–1504.

[13] W. Kozek, "Matched Weyl-Heisenberg Expansions of Nonstationary Environments," Ph.D. dissertation, Vienna Univ. Technology, Vienna, Austria, 1997.

[14] K. Lu, T. Kadous, and A. M. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2583–2603, Nov. 2004.

[15] J. C. Preisig and G. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoust. Soc. Amer.*, vol. 116, pp. 2067–2080, Oct. 2004.

[16] G. Durisi, V. I. Morgenshtern, and H. Bölcskei, "On the sensitivity of noncoherent capacity to the channel model," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, Jun. 2009, pp. 2174–2178.

[17] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adaptive Control Signal Process.*, vol. 10, pp. 159–176, Mar. 1996.

[18] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. New York: Springer, 2001.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley, 1991.

[20] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[21] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[22] R. Kennedy, *Fading Dispersive Communication Channels*. Hoboken, NJ: Wiley, 1969.

[23] T. Kailath, "Measurement on time-variant communication channels," *IRE Trans. Inf. Theory*, vol. 8, pp. S229–S236, Sep. 1962.

[24] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

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

[26] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.

[27] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 789–792.

[28] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.

Arun Pachai Kannu received the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 2004 and 2007, respectively.

From 2007 to 2009, he was a Senior Engineer with Qualcomm, Inc., San Diego, CA. He is currently a visiting faculty in the Department of Electrical Engineering, Indian Institute of Technology, Madras.

Philip Schniter (SM'05) received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1992 and 1993, respectively, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 2000.

From 1993 to 1996, he was with Tektronix, Inc., Beaverton, OR, as a systems engineer. He joined the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, OH, where he is now an Associate Professor and a member of the Information Processing Systems (IPS) Lab. In 2003, he received the National Science Foundation CAREER Award, and in 2008–2009, he was a visiting professor at Eurecom (Sophia Antipolis, France) and Supélec (Gif-sur-Yvette, France). His areas of interest include statistical signal processing, wireless communications and networks, and underwater acoustic communications.

Dr. Schniter currently serves on the IEEE Signal Processing for Communications and Networking (SPCOM) Technical Committee.