Design and Analysis of MMSE Pilot-Aided Cyclic-Prefixed Block Transmissions for Doubly Selective Channels

Arun P. Kannu, *Member, IEEE*, and Philip Schniter, *Senior Member, IEEE*

***Abstract*—This paper considers affine cyclic-prefixed block-based pilot-aided transmission (PAT) over the single-antenna doubly selective channel, where the channel is assumed to obey a complex-exponential basis expansion model. First, a tight lower bound on the mean-squared error (MSE) of pilot-aided channel estimates is derived, along with necessary and sufficient conditions on the pilot/data pattern that achieves this bound. From these conditions, novel minimum-MSE (MMSE) PAT schemes are proposed and upper/lower bounds on their ergodic achievable rates are derived. A pilot/data power allocation technique is also developed. A high-SNR asymptotic analysis of the ergodic achievable rate of affine MMSE-PAT is then performed which suggests that the channel's spreading parameters should be taken into account when choosing among affine MMSE-PAT schemes. Specifically, we establish that multicarrier MMSE-PAT achieves higher rates than single-carrier MMSE-PAT when the channel's delay-spread dominates its Doppler-spread, and vice versa.**

***Index Terms*—Achievable rates, channel estimation, doubly dispersive channels, doubly selective channels, minimum mean-squared error (MMSE), noncoherent channels, optimal training, pilot-assisted wireless transmissions, pilot symbols.**

I. Introduction

WIRELESS channels that exhibit multipath fading are typically modeled as a linear transformation parameterized by random fading coefficients. Several coherent coding and decoding techniques have been developed to combat fading for the case that the receiver and/or transmitter knows these fading coefficients. In many practical scenarios, however, neither the transmitter nor the receiver has this channel state information (CSI). In the latter case, it is common to employ pilot-aided transmission (PAT), whereby the transmitter embeds known pilot (i.e., training) signals that the receiver can use to estimate the channel. The estimated channel coefficients can then be used by coherent decoding techniques. Cavers [1] authored one of the first analytical studies of PAT. Since then, there has been a growing interest in PAT design. See [2] for a recent comprehensive overview.

A popular PAT design criterion is the minimization of channel estimate mean-square error (MSE). For frequency selective channels, minimum-MSE (MMSE) PAT schemes were obtained in [3]–[5] for orthogonal frequency division multiplexing (OFDM) systems. For time selective channels, MMSE-PAT schemes were obtained in [6] for single-carrier modulation (SCM) systems. For slowly time-varying frequency-selective channels, MMSE-PAT design for OFDM was discussed in [7], [8], and [9] for single-antenna and multiple-antenna systems, respectively.

Since the bandwidth and power consumed by pilot symbols compromise the bandwidth and power available for information-bearing data symbols, PAT schemes may suffer a loss of information rate relative to noncoherent communication schemes that do not explicitly transmit pilots. To address this issue, several authors have studied the PAT design problem using information-theoretic tools (e.g., [10]–[15]). For example, [10]–[14] optimized the location, power, and number of pilot symbols by maximizing a lower bound on ergodic channel capacity, and, in doing so, uncovered connections between MSE minimization and achievable-rate maximization. In fact, for some channel models, MMSE-PAT and capacity-bound-maximizing PAT schemes coincide [10]–[14]. Furthermore, in the high signal-to-noise ratio (SNR) regime, some of these MMSE-PAT schemes are first-order capacity-optimal, i.e., their achievable rates exhibit the same growth-rate (versus SNR) as the noncoherent channel's capacity [10], [11], [16], [17].

In this paper, we consider the PAT design problem for doubly selective channels (DSC), i.e., channels which disperse the transmitted signal in both the time and frequency domains. Such channels are expected for future wireless systems communicating broadband data in highly mobile environments. In particular, we consider single-antenna DSCs whose time-evolution is well approximated by a complex-exponential (CE) basis expansion model (BEM) [18, p. 65], [19]. For these channels, time- and frequency-domain dispersion is characterized by delay and Doppler-spread, respectively. Motivated by the information-theoretic optimality of MMSE-PAT schemes for certain flat and frequency-selective channels [10], [11], [16], we focus on the design and analysis of MMSE-PAT schemes for the DSC. While most studies on MMSE-PAT [3], [5], [6], [9] and achievable-rate-maximizing PAT [10]–[14], [20] assume a specific modulation scheme (e.g., OFDM or SCM),

Manuscript received February 6, 2007; revised July 5, 2007. This work was supported in part by an NSF CAREER Grant 237037 and by the Office of Naval Research under Grant N00014-07-1-0209. Portions of this work were presented at the 2005 IEEE Conference on Acoustics, Speech, and Signal Processing, and at the 2005 IEEE Workshop on Signal Processing Advances in Wireless Communications. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. William A. Sethares.

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

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

Digital Object Identifier 10.1109/TSP.2007.908969

our study applies to the general class of cyclic-prefixed (CP) affine precoding [21] schemes. Affine precoding subsumes all schemes which transmit the sum of an energy-constrained pilot vector and a linearly precoded data vector, where the choice of the precoding matrix is arbitrary. For this setting, we derive necessary and sufficient conditions on the combination of pilot vector and precoding matrix that yield MMSE channel estimates. These conditions reveal the fundamental structure of all PAT schemes (i.e., irrespective of modulation format) that are MMSE for this DSC, paving the way for the design of novel MMSE-PAT schemes. Using these conditions, we then analyze the achievable rate of affine MMSE-PAT transmission and propose pilot/data power-allocation guidelines for it. The ergodic achievable rate is further analyzed in the high-SNR regime, and the spectral efficiency of systems which transmit a stream of affine MMSE-PAT blocks are characterized. This latter analysis provides insight into the roles of delay- and Doppler-spread on MMSE-PAT performance. Finally, numerical examples are presented to illustrate the theoretical results.

We now comment on related work. In [12] and [14], the authors investigated DSC PAT design for SCM with the goal of maximizing achievable rate. Due to connections between achievable-rate maximization and MSE-minimization, the PAT schemes proposed in [12] and [14] turn out to be MMSE. However, as a consequence of our more general affine-precoding framework, we are able to show that the MMSE-PAT scheme obtained in [12] is but one of many affine MMSE-PAT schemes possible for the DSC. Furthermore, our achievable-rate analysis provides a means of comparing among different affine MMSE-PAT schemes and establishes that a particular multicarrier MMSE-PAT scheme achieves higher rates than the single-carrier MMSE-PAT scheme from [12] when the DSC's discrete delay-spread dominates its discrete Doppler-spread.

The paper[1] is organized as follows. In Section II, we derive MMSE conditions and design MMSE-PAT schemes for single-block transmission. In Section III, we perform an information-theoretic analysis of MMSE-PAT, also for single-block transmission. In Section IV, we study the spectral efficiency of schemes which transmit a stream of MMSE-PAT blocks. In Section V, we present numerical results, and, in Section VI, we conclude.

II. MMSE-PAT Design

In this section, we consider affine precoding for CP block transmission over the single-antenna DSC, where a CE-BEM is used to characterize the channel variation over the block duration. We outline a procedure for designing MMSE-PAT schemes for this DSC and provide several novel MMSE-PAT examples.

A. Cyclic-Prefix Block Transmission Model

The sampled complex-baseband DSC output $\{y(i)\}$ is related to the transmitted signal $\{x(i)\}$ via

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

where $\{v(i)\}$ is zero-mean σ_v^2-variance circular white Gaussian noise (CWGN) and $h(i,\ell)$ is the time-i channel response to an impulse applied at time $i-\ell$. Here, N_t denotes the channel's delay-spread normalized to the sampling interval T_s. To avoid interference from the preceding block, each length-N transmission block[2] $\{x(i)\}_{i=0}^{N-1}$ is preceded by a CP of length N_t-1. The received vector is formed as $\boldsymbol{y} = [y(0),\ldots,y(N-1)]^\top$, from which it can be seen that the CP portion is discarded. The effect of CP overhead will be discussed at length in Section IV. Throughout this section, we assume modulo-N indexing, i.e., $z(i) = z(\langle i\rangle_N)$. With the definitions $\boldsymbol{X} = [\boldsymbol{X}_0\cdots\boldsymbol{X}_{-N_t+1}]$, $\boldsymbol{X}_i = \mathrm{diag}(x(i),\ldots,x(i+N-1))$, $\boldsymbol{h} = [\boldsymbol{h}_0^\top\cdots\boldsymbol{h}_{N_t-1}^\top]^\top$, $\boldsymbol{h}_\ell = [h(0,\ell),\ldots,h(N-1,\ell)]^\top$, $\boldsymbol{v} = [v(0),\ldots,v(N-1)]^\top$, the DSC block transmission model (1) can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{v} \tag{2}$$

where $\boldsymbol{X} \in \mathbb{C}^{N\times NN_t}$ and $\boldsymbol{h} \in \mathbb{C}^{NN_t\times 1}$.

The transmitted signal is constructed as $x(i) = p(i) + d(i)$, where $\{p(i)\}$ is the pilot sequence and $\{d(i)\}$ is the zero-mean data sequence. Using $p(i)$ and $d(i)$ to construct $\boldsymbol{P}$ and $\boldsymbol{D}$, respectively, in the manner of $\boldsymbol{X}$, we find

$$\boldsymbol{X} = \boldsymbol{P} + \boldsymbol{D} \tag{3}$$

with $\boldsymbol{P} = E\{\boldsymbol{X}\}$. The data vector $\boldsymbol{d} = [d(0),\ldots,d(N-1)]^\top$ is constructed via linear precoding as $\boldsymbol{d} = \bar{\boldsymbol{B}}\boldsymbol{s}$, where $\boldsymbol{s} = [s(0),\ldots,s(N_s-1)]^\top$ is the information-bearing symbol vector and $\bar{\boldsymbol{B}} \in \mathbb{C}^{N\times N_s}$ is an arbitrary full-rank precoding matrix. Thus, our transmission strategy coincides with the general case of affine precoding [21], i.e.,

$$\boldsymbol{x} = \boldsymbol{p} + \boldsymbol{B}\boldsymbol{s} \tag{4}$$

with $\boldsymbol{x} = [x(0),\ldots,x(N-1)]^\top$ and $\boldsymbol{p} = [p(0),\ldots,p(N-1)]^\top$. As we will see, the channel estimate MSE is a direct function of the pilot energy

$$\|\boldsymbol{p}\|^2 = E_p. \tag{5}$$

Our generic transmission model allows us to analyze PAT for many modulation schemes under a common framework. Notice that, for SCM and OFDM systems, the columns of $\bar{\boldsymbol{B}}$ are chosen from those of $\boldsymbol{I}_N$ and $\boldsymbol{F}_N$, respectively.

[1]*Notation*: Matrices (column vectors) are denoted by upper-case (lower-case) bold-face letters. Hermitian is denoted by $(\cdot)^{\mathsf{H}}$, transpose by $(\cdot)^\top$, and conjugate by $(\cdot)^*$. The determinant and Frobenius norm are denoted by $\det(\cdot)$ and $\|\cdot\|_F$, respectively. The expectation, trace, Kronecker delta, Kronecker product, modulo-N and integer ceiling operations are denoted by $E\{\cdot\}$, $\mathrm{tr}\{\cdot\}$, $\delta(\cdot)$, $\otimes$, $\langle\cdot\rangle_N$ and $\lceil\cdot\rceil$, respectively. The null space and column space of a matrix are denoted by $\mathrm{null}(\cdot)$ and $\mathrm{col}(\cdot)$, respectively, and the dimension of a vector space is denoted by $\dim(\cdot)$. The operation $[\cdot]_{n,m}$ extracts the $(n,m)^{th}$ element of a matrix, where the row/column indices n,m begin with 0, while $\mathrm{diag}(\cdot)$ denotes a diagonal matrix constructed from the vector-valued argument. The $N\times N$ unitary DFT and identity matrices are denoted by $\boldsymbol{F}_N$ and $\boldsymbol{I}_N$, respectively, and appropriately dimensioned identity and all-zero matrices are denoted by $\boldsymbol{I}$ and $\boldsymbol{0}$, respectively. The union, intersection, and set-minus operators are denoted by $\cup$, $\cap$, and $\setminus$, respectively, while the empty set is denoted by $\emptyset$. Finally, the sets of integers, reals, positive reals, and complex numbers are denoted by $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}^+$, and $\mathbb{C}$, respectively.

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

B. Doubly Selective Channel Model

The following CE-BEM [18, p. 65], [19] will be used to describe the channel response over the N-length block duration. For $i \in \{0,\dots,N-1\}$ and $\ell \in \{0,\dots,N_t-1\}$, we assume that

$$h(i,\ell) = \frac{1}{\sqrt{N}} \sum_{k=-(N_f-1)/2}^{(N_f-1)/2} \lambda(k,\ell) e^{j\frac{2\pi}{N}ki} \tag{6}$$

where the CE-BEM coefficients $\{\lambda(k,\ell)\}$ are zero-mean and uncorrelated with positive variance. The CE-BEM (6) has been widely used to model time-varying communication channels (e.g., [12], [17]–[19]) and can be interpreted as an N_f-term truncated Fourier-series approximation of each of the N_t coefficient trajectories $\{h(0,\ell),\dots,h(N-1,\ell)\}_{\ell=0}^{N_t-1}$. The application of truncated Fourier series can be motivated by the bandlimited nature of coefficient trajectories that results from finite mobile velocities. Specifically, path lengths which vary by at most $v_{\max}$ meters per second imply a maximum single-sided Doppler-spread of $f_{\mathsf{d}} = 2v_{\max}/\lambda_c$ Hz, where λ_c denotes the carrier wavelength [18]. Since the use of $N_f = 2\lfloor f_{\mathsf{d}} T_s N \rfloor + 1$ terms in the Fourier series yields a reasonably accurate approximation to each trajectory, we assume this value of N_f throughout. We allow CE-BEM coefficients with possibly unequal variances in order to model arbitrary delay profiles and Doppler spectra.

We emphasize that, like all models, this CE-BEM *approximates* the DSC; it does not yield a "perfect" description. (See [17] for a thorough discussion on the validity of the CE-BEM.) While alternative BEMs have been proposed that, in some cases, yield "better" approximations (e.g., the polynomial [22] and Slepian [23] BEMs), they make MMSE-PAT analysis difficult, if not impossible. We adopt the CE-BEM (6) because it is reasonably accurate and yields a tractable MMSE-PAT analysis with insightful results.

With $\bar{\boldsymbol{F}} \in \mathbb{C}^{N\times N_f}$ defined element-wise as $[\bar{\boldsymbol{F}}]_{n,m} = (1/\sqrt{N}) e^{j(2\pi/N)n(m-(N_f-1)/2)}$, $\boldsymbol{U} = \boldsymbol{I}_{N_t} \otimes \bar{\boldsymbol{F}} \in \mathbb{C}^{NN_t\times N_fN_t}$, $\boldsymbol{\lambda}_\ell = [\lambda(-((N_f-1)/2),\ell),\dots,\lambda(((N_f-1)/2),\ell)]^\top$, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_0^\top \cdots \boldsymbol{\lambda}_{N_t-1}^\top]^\top \in \mathbb{C}^{N_fN_t}$, (6) can be written compactly as

$$\boldsymbol{h} = \boldsymbol{U}\boldsymbol{\lambda}. \tag{7}$$

Note that, because $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U} = \boldsymbol{I}_{N_fN_t}$ and because $\boldsymbol{R}_\lambda = E\{\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathsf{H}}\}$ is diagonal and positive definite, (7) gives the Karhunen-Lòeve (KL) expansion of $\boldsymbol{h}$. We assume that the DSC has an average energy gain of unity, so that $(1/N)E\{\|\boldsymbol{h}\|^2\} = (1/N)\mathrm{tr}\{\boldsymbol{R}_\lambda\} = 1$.

In the sequel, we refer to N_t as the "discrete delay-spread" and to N_f as the "discrete Doppler-spread." In addition, we refer to $f_{\mathsf{d}}T_s$ as the "normalized Doppler-spread" and to $\gamma = N_tN_f/N \approx 2f_{\mathsf{d}}T_sN_t$ as the channel's "spreading index." We restrict our focus to channels for which $\gamma < 1$, known as "underspread" channels [24], so that N_tN_f, the number of independent channel parameters per block, is less than the block length N. The underspread assumption is standard for radio-frequency channels [25].

C. MSE Lower Bound

The linear-MMSE (LMMSE) estimate of $\boldsymbol{h}$ given the knowledge of $\{\boldsymbol{y},\boldsymbol{P}\}$ and the knowledge of the second-order statistics of $\{\boldsymbol{h},\boldsymbol{D},\boldsymbol{v}\}$ is [26]

$$\hat{\boldsymbol{h}} = \boldsymbol{R}_{y,h}^{\mathsf{H}}\boldsymbol{R}_y^{-1}\boldsymbol{y} \tag{8}$$

where $\boldsymbol{R}_{y,h} = E\{\boldsymbol{y}\boldsymbol{h}^{\mathsf{H}}\}$ and $\boldsymbol{R}_y = E\{\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}\}$. With our transmission and channel models, we have $\boldsymbol{y} = \boldsymbol{PU}\boldsymbol{\lambda} + \boldsymbol{DU}\boldsymbol{\lambda} + \boldsymbol{v}$, and hence $\boldsymbol{R}_{y,h} = \boldsymbol{PUR}_\lambda\boldsymbol{U}^{\mathsf{H}}$ and $\boldsymbol{R}_y = \boldsymbol{PUR}_\lambda\boldsymbol{U}^{\mathsf{H}}\boldsymbol{P}^{\mathsf{H}} + E\{\boldsymbol{DUR}_\lambda\boldsymbol{U}^{\mathsf{H}}\boldsymbol{D}^{\mathsf{H}}\} + \sigma_v^2\boldsymbol{I}_N$. Our channel estimator is pilot-aided, noniterative and nondecision-aided. The channel estimation error $\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}}$ has variance $\sigma_e^2 = E\{\|\tilde{\boldsymbol{h}}\|^2\}$ given by [26]

$$\sigma_e^2 = \mathrm{tr}\left\{\boldsymbol{U}\boldsymbol{R}_\lambda\boldsymbol{U}^{\mathsf{H}} - \boldsymbol{R}_{y,h}^{\mathsf{H}}\boldsymbol{R}_y^{-1}\boldsymbol{R}_{y,h}\right\}. \tag{9}$$

In this paper, we are interested PAT schemes which minimize the MSE (9) of the LMMSE channel estimate (8) subject to a pilot power constraint E_p. Specifically, we are interested in finding the combination of E_p-constrained pilot vector $\boldsymbol{p}$ and orthonormal data modulation matrix $\boldsymbol{B}$ which minimize the estimation MSE σ_e^2. We refer to such combinations of $(\boldsymbol{p},\boldsymbol{B})$ as affine MMSE-PAT schemes. We shall see later that there are many $(\boldsymbol{p},\boldsymbol{B})$ combinations which lead to minimal MSE. We now proceed to the characterization of these affine MMSE-PAT schemes.

Theorem 1 (MSE Lower Bound): For N-block CP affine PAT over the CE-BEM DSC with the pilot-aided channel estimator (8), the channel estimate MSE (9) obeys

$$\sigma_e^2 \geq \mathrm{tr}\left\{\left(\boldsymbol{R}_\lambda^{-1} + \frac{E_p}{N\sigma_v^2}\boldsymbol{I}_{N_fN_t}\right)^{-1}\right\} \tag{10}$$

where equality in (10) occurs if and only if the following two conditions hold:

$$\text{Pilot-Data Orthogonality}: (\boldsymbol{PU})^{\mathsf{H}}\boldsymbol{DU} = \boldsymbol{0},\ \forall \boldsymbol{D} \tag{11}$$

$$\text{Optimal Excitation}: (\boldsymbol{PU})^{\mathsf{H}}\boldsymbol{PU} = \frac{E_p}{N}\boldsymbol{I}_{N_fN_t}. \tag{12}$$

When (11) and (12) are met, $\boldsymbol{R}_{\tilde{h}} = E\{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathsf{H}}\}$ and $\boldsymbol{R}_{\hat{h}} = E\{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^{\mathsf{H}}\}$ are given by

$$\boldsymbol{R}_{\tilde{h}} = \boldsymbol{U}\left(\boldsymbol{R}_\lambda^{-1} + \frac{E_p}{N\sigma_v^2}\boldsymbol{I}_{N_fN_t}\right)^{-1}\boldsymbol{U}^{\mathsf{H}} \tag{13}$$

$$\boldsymbol{R}_{\hat{h}} = \boldsymbol{U}\left[\boldsymbol{R}_\lambda - \left(\boldsymbol{R}_\lambda^{-1} + \frac{E_p}{N\sigma_v^2}\boldsymbol{I}_{N_fN_t}\right)^{-1}\right]\boldsymbol{U}^{\mathsf{H}}. \tag{14}$$

Proof: See Appendix I. ∎

Condition (11) says that pilots and data should be multiplexed in a way that preserves orthogonality at the channel output, while condition (12) says that pilots should be constructed so that the channel modes are independently excited with equal energy. Defining $b_q(i) = [\boldsymbol{B}]_{i,q}$, we rephrase the MSE optimality requirements (11), (12) in terms of the pilot sequence $\{p(i)\}$ and modulation sequences $\{b_q(i)\}$, using the index sets $\mathcal{N}_t = \{-N_t+1,\dots,N_t-1\}$ and $\mathcal{N}_f = \{-N_f+1,\dots,N_f-1\}$.

Lemma 1 (Time-Domain Conditions): For N-block CP affine PAT over the CE-BEM DSC, the pair (15), (16) form necessary and sufficient conditions for equality in (10)

$$\sum_{i=0}^{N-1} p(i)p^*(i-k)e^{-j\frac{2\pi}{N}mi} = E_p\delta(k)\delta(m) \quad \forall k \in \mathcal{N}_t,\ \forall m \in \mathcal{N}_f \tag{15}$$

$$\sum_{i=0}^{N-1} b_q(i)p^*(i-k)e^{-j\frac{2\pi}{N}mi} = 0 \quad \forall k \in \mathcal{N}_t,\ \forall m \in \mathcal{N}_f,\ \forall q \in \{0,\ldots,N_s-1\}. \tag{16}$$

Proof: See Appendix II. ∎

We are not aware of previous results which specify the conditions on affine transmission parameters that are necessary and sufficient to minimize the MSE of LMMSE DSC estimates (8). Previous work on DSC PAT design [12], [14] was based on the maximization of a channel-capacity lower bound for the specific case of SCM. Requirements on pilot-data orthogonality to minimize the least squares (LS) estimation error variance were discussed, for frequency selective channels, in [4]. Our pilot-data orthogonality requirement (16) establishes that, for DSC affine MMSE-PAT, the data modulation basis must be orthogonal to certain time- and frequency-shifts of the pilot vector.

D. Bounds on "Data Dimension" N_s

In this section, we obtain bounds on the data dimension N_s of affine MMSE-PAT schemes, i.e., schemes for which $(\boldsymbol{p},\boldsymbol{B})$ satisfies (15) and (16). Defining $\breve{\boldsymbol{F}} \in \mathbb{C}^{2N_f-1\times N}$ element-wise as $[\breve{\boldsymbol{F}}]_{n,m} = (1/\sqrt{N})e^{-j(2\pi/N)(n-N_f+1)m}$, and defining $\boldsymbol{P}_i = \mathrm{diag}(p(i),\ldots,p(i+N-1))$, $\boldsymbol{W}_i = \breve{\boldsymbol{F}}\boldsymbol{P}_i^{\mathsf{H}}$, as well as $\boldsymbol{W} = [\boldsymbol{W}_{-N_t+1}^{\top}\cdots\boldsymbol{W}_{N_t-1}^{\top}]^{\top}$, we see that (16) can be written as $\boldsymbol{W}\boldsymbol{b}_q = \boldsymbol{0}$, implying that each $\boldsymbol{b}_q$ must lie in the null space of $\boldsymbol{W}$. Clearly, N_s, the number of columns in $\boldsymbol{B}$, must obey

$$N_s \le \dim\left(\mathrm{null}(\boldsymbol{W})\right). \tag{17}$$

We restrict[3] our attention to the case when equality is attained in (17) by choosing $\boldsymbol{B}$ such that its columns form a basis for $\mathrm{null}(\boldsymbol{W})$. Based on the structure of $\boldsymbol{W}$, we can bound N_s as follows. From the construction of $\boldsymbol{W}$, it follows that the N_fN_t rows of $(\boldsymbol{PU})^{\mathsf{H}}$ are contained within the $(2N_f-1)(2N_t-1)$ rows of $\boldsymbol{W}$. To satisfy (12), the rows of $(\boldsymbol{PU})^{\mathsf{H}}$ must be orthogonal. Thus, $N_fN_t \le \mathrm{rank}(\boldsymbol{W}) \le (2N_f-1)(2N_t-1)$, and so (15), (16) imply

$$N-(2N_f-1)(2N_t-1) \le N_s \le N-N_fN_t. \tag{18}$$

From (18), we see that MMSE-PAT dedicates at least N_tN_f, but no more than $(2N_t-1)(2N_f-1)$, signaling dimensions to pilots. The implication is that the PAT schemes which achieve equality in (10) obey $N_s \le N-N_fN_t$. This result is intuitive, because, to minimize MSE, the pilot symbols must estimate all N_fN_t independent BEM coefficients, thereby consuming at least N_fN_t signaling dimensions. Our analysis also shows that PAT schemes can support a data dimension of at least $N-(2N_f-1)(2N_t-1)$ without sacrificing MSE.

[3]In Section III, we establish that higher N_s leads to higher data rates in the high-SNR regime.

E. Time-Frequency Duality of MMSE-PAT

Time-frequency duality is a well-known and fundamental concept in communication theory (see, e.g., [24]). While it has been recently applied to the design of space-time codes (e.g., [27]), we now demonstrate that it can be applied to the design of affine MMSE-PAT schemes for the DSC. We refer to $(\breve{\boldsymbol{p}},\breve{\boldsymbol{B}})$, for $\breve{\boldsymbol{p}} = \boldsymbol{F}_N\boldsymbol{p}$ and $\breve{\boldsymbol{B}} = \boldsymbol{F}_N\boldsymbol{B}$, as the "frequency-domain counterpart" of $(\boldsymbol{p},\boldsymbol{B})$. Using this notation, Lemma 2 gives the frequency-domain analogy of Lemma 1.

Lemma 2 (Frequency-Domain Conditions): For N-block CP affine PAT over the CE-BEM DSC, the pair (19), (20) form necessary and sufficient conditions for equality in (10)

$$\sum_{i=0}^{N-1} \breve{p}(i)\breve{p}^*(i-k)e^{-j\frac{2\pi}{N}mi} = E_p\delta(k)\delta(m) \quad \forall k \in \mathcal{N}_f,\ \forall m \in \mathcal{N}_t \tag{19}$$

$$\sum_{i=0}^{N-1} \breve{b}_q(i)\breve{p}^*(i-k)e^{-j\frac{2\pi}{N}mi} = 0 \quad \forall k \in \mathcal{N}_f,\ \forall m \in \mathcal{N}_t,\ \forall q \in \{0,\ldots,N_s-1\}. \tag{20}$$

Proof: Plugging $p(i) = (1/\sqrt{N})\sum_{q=0}^{N-1}\breve{p}(q)e^{j(2\pi/N)qi}$ into (15), it is straightforward to show that (15) and (19) are equivalent. A similar technique proves the equivalence between (16) and (20). ∎

Note that the conditions in Lemma 2 mirror those in Lemma 1, except that the discrete delay-spread N_t and discrete Doppler-spread N_f have interchanged their roles. Our duality result can be stated concisely as follows.

Lemma 1 (Duality): If $(\boldsymbol{p},\boldsymbol{B})$ parameterizes N-block CP MMSE-PAT over the CE-BEM DSC with discrete delay-spread N_1 and discrete Doppler-spread N_2, then $(\boldsymbol{F}_N\boldsymbol{p},\boldsymbol{F}_N\boldsymbol{B})$ parameterizes N-block CP MMSE-PAT for the CE-BEM DSC with discrete delay-spread N_2 and discrete Doppler-spread N_1.

To our knowledge, the application of time-frequency duality to the design of DSC MMSE-PAT schemes is novel. Similarities between the structure of MMSE-PAT schemes for SCM over time-selective channels [28] and OFDM over frequency selective channels [4] have been previously noted in [28]. However, our result on the duality of DSC MMSE-PAT is more general, and applies to any affine modulation scheme (including, but not limited to, SCM and OFDM), as illustrated by the MMSE-PAT examples given in Section II-F.

F. Examples of Affine MMSE-PAT

Here we give several examples of N-block CP affine MMSE-PAT schemes for the CE-BEM DSC with discrete delay-spread N_t and discrete Doppler-spread N_f, using the $(\boldsymbol{p},\boldsymbol{B})$ parameterization. It is straightforward to verify that the following examples satisfy the MMSE-PAT conditions (15), (16) and (19), (20).

Example 1 (TDKD): Assuming $(N/N_f) \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_t^{(l)} = \{l, l+(N/N_f),\ldots,l+((N_f-1)N/N_f)\}$ and the guard index set $\mathcal{G}_t^{(l)} = \bigcup_{k\in\mathcal{P}_t^{(l)}}\{-N_t+1+k,\ldots,N_t-1+k\}$. An N-block CP MMSE-PAT scheme for the CE-BEM DSC is given by the pilot vector $p(k) = \sqrt{E_p/N_f}e^{j\theta(k)}$ if $k \in \mathcal{P}_t^{(l)}$, $p(k) = 0$ if $k \notin \mathcal{P}_t^{(l)}$ and by $\boldsymbol{B}$ constructed from

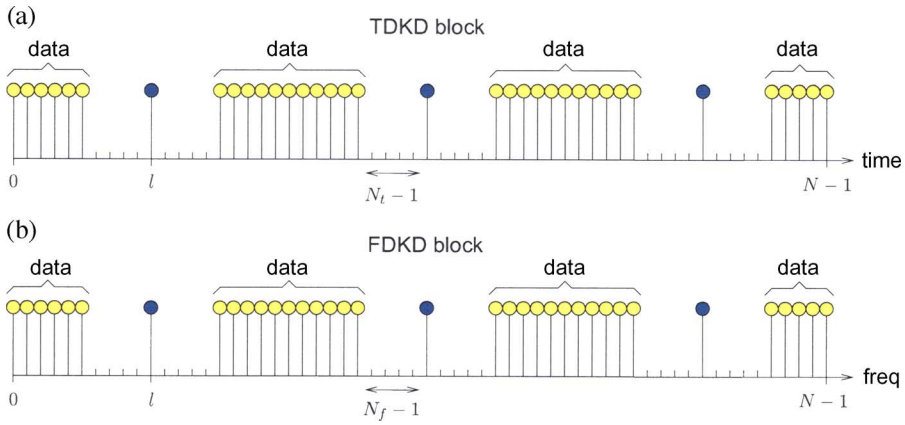


Fig. 1. (a) Structure of TDKD with $N_t = 5$ and $N_f = 3$. (b) Structure of FDKD with $N_t = 3$ and $N_f = 5$.

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

Example 1 specifies a PAT scheme in which the data and pilot sequences are nonoverlapping in the time domain, where the pilot pattern consists of periodic time-domain bursts, and where each burst has a truncated Kronecker-delta structure, with l controlling the time-offset of the first burst. (See Fig. 1). Note that the burst period $N/N_f \approx 1/2f_{\mathrm{d}}T_s$ satisfies a Nyquist-sampling criterion. This "time-domain Kronecker delta" (TDKD) scheme bears similarity to the PAT scheme proposed in [29] (heuristically) and later in [12], with the difference that [12] focused on zero-padded (ZP) block transmission, which allows for efficient concatenation of blocks. We make a thorough comparison of CP and ZP schemes in Section IV.

Example 2 (FDKD): Assuming $(N/N_t) \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_f^{(l)} = \{l, l + (N/N_t), \ldots, l + ((N_t - 1)N/N_t)\}$ and the guard index set $\mathcal{G}_f^{(l)} = \bigcup_{k \in \mathcal{P}_f^{(l)}} \{-N_f + 1 + k, \ldots, N_f - 1 + k\}$. An N-block CP MMSE-PAT scheme for the CE-BEM DSC is given by the pilot vector $\boldsymbol{p} = \boldsymbol{F}_N^{\mathsf{H}} \breve{\boldsymbol{p}}$, with $\breve{p}(k) = \sqrt{E_p/N_t} e^{j\theta(k)}$ if $k \in \mathcal{P}_f^{(l)}$, $\breve{p}(k) = 0$ if $k \notin \mathcal{P}_f^{(l)}$ and by $\boldsymbol{B}$ constructed from the columns of the IDFT matrix $\boldsymbol{F}_N^{\mathsf{H}}$ with indices *not* in the set $\mathcal{G}_f^{(l)}$. Both $l \in \{0, \ldots, (N/N_t) - 1\}$ and $\theta(k) \in \mathbb{R}$, are arbitrary. The corresponding data dimension is $N_s = N - N_t(2N_f - 1)$.

Example 2 specifies a PAT scheme in which the data and pilot sequences are nonoverlapping in the frequency domain, where the pilot pattern consists of periodic subcarrier clusters, and where each cluster has a truncated Kronecker-delta structure, with l controlling the offset of the first cluster (see Fig. 1). Note that the cluster spacing N/N_t satisfies a frequency-domain Nyquist-sampling criterion. This "frequency-domain Kronecker delta" (FDKD) scheme is the time-frequency dual of TDKD and bears similarity to the heuristic PAT schemes proposed in [29] and [30]. For the special case of frequency-selective channels (i.e., $N_f = 1$), FDKD coincides with the MSE-optimal OFDM system identified in [3] and [4], where the pilot clusters reduce to pilot tones.

Example 3 (Time-Domain Chirps): Assuming even N, an N-block CP MMSE-PAT scheme for the CE-BEM DSC is given by $p(k) = \sqrt{E_p/N} e^{j(2\pi/N)(N_f/2)k^2}$ and $[\boldsymbol{B}]_{k,q} = (1/\sqrt{N}) e^{j(2\pi/N)(q+N_fN_t)k} e^{j(2\pi/N)(N_f/2)k^2}$, for $k \in \{0, \ldots, N-1\}$ and $q \in \{0, \ldots, N_s - 1\}$, where the data dimension $N_s = N - 2N_fN_t + 1$.

Example 4 (Frequency-Domain Chirps): Assuming even N, an N-block CP MMSE-PAT scheme for the CE-BEM DSC is given by $\boldsymbol{p} = \boldsymbol{F}_N^{\mathsf{H}} \breve{\boldsymbol{p}}$ and $\boldsymbol{B} = \boldsymbol{F}_N^{\mathsf{H}} \breve{\boldsymbol{B}}$, with $\breve{p}(k) = \sqrt{E_p/N} e^{j(2\pi/N)(N_t/2)k^2}$ and $[\breve{\boldsymbol{B}}]_{k,q} = (1/\sqrt{N}) e^{j(2\pi/N)(q+N_fN_t)k} e^{j(2\pi/N)(N_t/2)k^2}$, for $k \in \{0, \ldots, N-1\}$ and $q \in \{0, \ldots, N_s - 1\}$, where the data dimension $N_s = N - 2N_fN_t + 1$.

To our knowledge, the use of "chirp" signaling for DSC MMSE PAT is novel; though a chirp-based PAT scheme was suggested in [31], it was not MMSE. We refer to examples 3 and 4 as "chirp" schemes because the pilot and data waveforms are linear chirps, as evidenced by the k^2 term in the complex exponentials. With our chirp schemes, the pilots and data are superimposed in both the time and frequency domains, but still *linearly separable* at the channel output, as required by (11). Note that the PATs in Example 3 and Example 4 are time-frequency duals.

III. Achievable-Rate Analysis of MMSE-PAT

We now calculate bounds on the ergodic achievable rate of affine precoded MMSE-PAT schemes for the CE-BEM DSC, paying special attention to the high-SNR regime. Lemma 1, which distills the fundamental properties of these MMSE-PAT schemes, allows this achievable-rate analysis to be conducted in a unified manner. In particular, we derive bounds on the mutual information between the input and output of the DSC when an affine MMSE-PAT scheme is employed. In addition, we propose a pilot/data power allocation procedure, recalling that, until now, we have assumed fixed pilot energy E_p and arbitrary data energy. Finally, we deepen our analysis in the high-SNR regime and compare the MMSE-PAT examples from Section II-F.

A. Definitions

As before, we consider the transmitter and channel models discussed in Sections II-A and II-B, respectively. Suppose that the MMSE-PAT scheme $(\boldsymbol{p}, \boldsymbol{B})$ has pilot energy E_p [recall (5)]

and yields data dimension N_s. It will be convenient to write the input-output relation (2) as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{p} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{s} + \boldsymbol{v}. \tag{21}$$

where $\boldsymbol{H} \in \mathbb{C}^{N\times N}$ is defined element-wise as

$$[\boldsymbol{H}]_{n,m} = h\left(n, \langle n-m\rangle_N\right) \tag{22}$$

with $h(n,\ell) = 0$ for $\ell \neq \{0,\ldots,N_t-1\}$. We also define $\hat{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}$ in the same way as $\boldsymbol{H}$, but from $\hat{\boldsymbol{h}}$ and $\tilde{\boldsymbol{h}}$, respectively. It can be verified that $E\{\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\} = \boldsymbol{I}$, $E\{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}}\} = \boldsymbol{I}\mathrm{tr}\{\boldsymbol{R}_{\hat{h}}\}/N$ and $E\{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{\mathsf{H}}\} = \boldsymbol{I}\mathrm{tr}\{\boldsymbol{R}_{\tilde{h}}\}/N$, where $\boldsymbol{R}_{\tilde{h}}$ and $\boldsymbol{R}_{\hat{h}}$ were given in (13) and (14), respectively.

For Section III, we make two additional assumptions on our model. First, we assume that the channel coefficient vector $\boldsymbol{h}$ is Gaussian and varies independently from block to block. In practice, independent fading across blocks can be achieved through block-level interleaving. Second, we assume that the columns of precoding matrix $\boldsymbol{B}$ are orthonormal (whereas earlier $\boldsymbol{B}$ was specified as full rank). Note that, for the purpose of ergodic achievable-rate analysis, this assumption can be made w.l.o.g., since the mutual information between $\boldsymbol{s}$ and $\boldsymbol{y}$ remains unaffected by invertible transformations of $\boldsymbol{s}$. We then define the data energy $E_s = E\{\|\boldsymbol{d}\|^2\} = E\{\|\boldsymbol{s}\|^2\}$, the total energy $E_x = E_p + E_s$, and the transmission power $\sigma_x^2 = (1/N)\sum_{i=0}^{N-1} E\{|x(i)|^2\} = E_x/N$. In addition, we define $\mathsf{SNR} = \sigma_x^2/\sigma_v^2$, as motivated by our energy-preserving channel model. Finally, we define the normalized data power $\sigma_s^2 = E_s/N_s$. Note that these energy and power definitions do not take the CP into account; the effects of the CP will be discussed in Section IV.

B. Achievable-Rate Bounds

In this section, we obtain achievable-rate bounds assuming a zero-mean i.i.d. Gaussian distribution for $\boldsymbol{s}$. The general distribution case will be treated in Section III-D. To present the bounds, we use the normalized channel estimate $\bar{\boldsymbol{H}} = \hat{\boldsymbol{H}}\sqrt{N/\mathrm{tr}\{\boldsymbol{R}_{\hat{h}}\}}$, where $E\{\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}\} = \boldsymbol{I}$.

Theorem 2 (Achievable-Rate Bounds): For the N-block CP affine MMSE-PAT scheme $(\boldsymbol{p},\boldsymbol{B})$ with zero-mean i.i.d. Gaussian $\boldsymbol{s} \in \mathbb{C}^{N_s}$ over the CE-BEM DSC, the per-block ergodic achievable rate $\mathcal{C}_{\mathsf{mmse-gau-blk}}$ in bits/block obeys $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} \le \mathcal{C}_{\mathsf{mmse-gau-blk}} \le \mathcal{C}_{\mathsf{mmse-gau-blk-ub}}$, where

$$\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} = E\left\{\log\det\left[\boldsymbol{I}_{N_s} + \rho_{\mathsf{lb}}\boldsymbol{B}^{\mathsf{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}\bar{\boldsymbol{H}}\boldsymbol{B}\right]\right\} \tag{23}$$

$$\mathcal{C}_{\mathsf{mmse-gau-blk-ub}} = E\left\{\log\det\left[\boldsymbol{I}_{N_s} + \rho_{\mathsf{ub}}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B}\right]\right\} \tag{24}$$

$$\rho_{\mathsf{lb}} = \frac{\sigma_s^2\mathrm{tr}\{\boldsymbol{R}_{\hat{h}}\}}{\sigma_s^2\mathrm{tr}\{\bar{\boldsymbol{R}}_{\tilde{h}}\} + N\sigma_v^2} \tag{25}$$

$$\rho_{\mathsf{ub}} = \frac{\sigma_s^2}{\sigma_v^2}. \tag{26}$$

Proof: See Appendix III. ∎

The lower bound (23) describes the "worst case" scenario of channel estimation error acting as additive Gaussian noise. This concept was previously used in, e.g., [10]. The upper bound (24) describes the "best case" scenario of perfect channel estimates. The achievable rate $\mathcal{C}_{\mathsf{mmse-gau-blk}}$ specifies a rate below which codes can be designed so that the probability of the decoding error is arbitrarily small. Though we have not considered any particular decoding technique in Theorem 2, it follows from [32] that weighted minimum distance decoding can facilitate a rate of $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$. In Section IV, we consider the effects of the CP and write the achievable rate in units of bits/s/Hz.

It is insightful to compare $\mathcal{C}_{\mathsf{mmse-gau-blk}}$, the achievable rate of affine MMSE-PAT, to $\mathcal{C}_{\mathsf{csi-gau-blk}}$, the achievable rate of a system with perfect receiver-CSI, under equal transmission power σ_x^2. (With perfect CSI, there is, of course, no need for pilots.) With i.i.d. Gaussian data, the achievable rates are known to be [33] $\mathcal{C}_{\mathsf{csi-gau-blk}} = E\{\log\det[\boldsymbol{I}_N + \rho_{\mathsf{csi}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}]\}$ bits/block with $\rho_{\mathsf{csi}} = \sigma_x^2/\sigma_v^2$. Two principle factors separate $\mathcal{C}_{\mathsf{csi-gau-blk}}$ from $\mathcal{C}_{\mathsf{mmse-gau-blk}}$. First, MMSE-PAT suffers from channel estimation error, which degrades $\mathcal{C}_{\mathsf{mmse-gau-blk}}$ by affecting the "effective SNR" (25). Second, MMSE-PAT uses only N_s out of N total signaling dimensions for data transmission. Note, from (21), (23), and (24), that MMSE-PAT communicates the data $\boldsymbol{s}$ through the "effective channel" $\boldsymbol{H}\boldsymbol{B} \in \mathbb{C}^{N\times N_s}$, which offers only N_s degrees of freedom. The perfect-receiver-CSI system, on the other hand, communicates data through the effective channel $\boldsymbol{H} \in \mathbb{C}^{N\times N}$, which offers N degrees of freedom. Our asymptotic analysis in Section III-D provides further insight on these issues.

C. Pilot/Data Power Allocation

Until now, our MMSE-PAT schemes were designed using fixed pilot energy E_p. Now we consider the problem of judiciously allocating a fixed energy E_x between pilots and data. Notice the inherent tradeoff: increasing the pilot power decreases the channel estimation error but also decreases the data power, which in turn increases the sensitivity to noise and channel estimation error. Intuitively, power should be allocated to maximize an "effective SNR" which takes into account both the noise and channel estimation errors. The approach we take is to maximize ρ_{lb}.

Let $\alpha \in (0,1)$ denote the fraction of energy allocated to the data symbols, i.e., $E_s = \alpha E_x$ and $E_p = (1-\alpha)E_x$. We are interested in finding $\alpha_\star = \arg\max_\alpha \rho_{\mathsf{lb}}(\alpha)$. Recall that $\alpha_\star$ must satisfy $(\partial\rho_{\mathsf{lb}}(\alpha)/\partial\alpha)|_{\alpha=\alpha_\star} = 0$, which is equivalent to satisfying

$$(\psi(\alpha_\star)+1)\left(\frac{E_x}{N_s\sigma_v^2} - \psi'(\alpha_\star)\right) = \left(\frac{\alpha_\star E_x}{N_s\sigma_v^2} - \psi(\alpha_\star)\right)\psi'(\alpha_\star) \tag{27}$$

with

$$\psi(\alpha) = \sum_{i=0}^{N_f N_t - 1} \frac{\alpha E_x}{NN_s\sigma_v^2[\boldsymbol{R}_\lambda]_{i,i}^{-1} + (1-\alpha)E_x N_s} \tag{28}$$

and $\psi'(\alpha) = \partial\psi/\partial\alpha$. Numerical techniques can be used to find the roots, within the interval (0,1), of the polynomial (27). Among these roots, $\alpha_\star$ is the one which maximizes ρ_{lb}.

In the case of identically distributed BEM coefficients, i.e., $\boldsymbol{R}_\lambda = (N/N_f N_t)\boldsymbol{I}_{N_f N_t}$, it can be shown that the maximizer of $\rho_{\mathsf{lb}}(\alpha)$ is

$$\alpha_{\star,\mathsf{iid}} = \begin{cases} \beta - \sqrt{\beta^2 - \beta}, & \text{if } N_s > N_t N_f \\ \beta + \sqrt{\beta^2 - \beta}, & \text{if } N_s < N_t N_f \\ \frac{1}{2}, & \text{if } N_s = N_t N_f \end{cases} \tag{29}$$

with $\beta = [1 + (N_f N_t \sigma_v^2/E_x)]/[1 - (N_f N_t/N_s)]$. Furthermore, it can be shown that $\alpha_{\star,\mathsf{iid}}$ maximizes the achievable-rate lower bound given in (23). To see this, note that the channel $\boldsymbol{h}$ and the normalized channel estimate $\bar{\boldsymbol{h}} = \hat{\boldsymbol{h}}\sqrt{N/\mathrm{tr}\{\boldsymbol{R}_{\hat{h}}\}}$ have the same covariance, and hence $\boldsymbol{H}$ and $\bar{\boldsymbol{H}}$ have the same distribution, so that the power allocation fraction affects the $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$ only through ρ_{lb}. Since $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$ is an increasing function of ρ_{lb}, maximizing ρ_{lb} is equivalent to maximizing $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$.

For the case of general $\boldsymbol{R}_\lambda$, closed-form solutions for $\alpha_{\max}$ are possible in the high-SNR and low-SNR asymptotic cases. It can be shown that $\arg\max_\alpha \lim_{\mathsf{SNR}\to 0} \rho_{\mathsf{lb}} = 1/2$ and $\arg\max_\alpha \lim_{\mathsf{SNR}\to\infty} \rho_{\mathsf{lb}} = \lim_{\mathsf{SNR}\to\infty} \alpha_{\star,\mathsf{iid}}$, where $\alpha_{\star,\mathsf{iid}}$ is calculated from (29) using $\beta = (1 - N_f N_t/N_s)^{-1}$. Note that all affine MMSE-PAT schemes with data dimension N_s have ρ_{lb} maximized by the same pilot/data power allocation.

D. Asymptotic Achievable Rates

In Section III-B, we derived ergodic achievable-rate bounds on affine MMSE-PAT assuming that the distribution of $\boldsymbol{s}$ is i.i.d. Gaussian. In this section, by focusing on the high-SNR regime, we find the maximal achievable rate optimized over all input distributions.

Theorem 3 (High-SNR Achievable Rate): For the N-block CP MMSE-PAT scheme $(\boldsymbol{p}, \boldsymbol{B})$ with data dimension N_s over the CE-BEM DSC, the per-block maximal ergodic rate is given by

$$\mathcal{C}_{\mathsf{mmse-blk}} = N_s \log \mathrm{SNR} + O(1) \text{ bits/block} \tag{30}$$

as $\mathrm{SNR} \to \infty$. In addition, an i.i.d. Gaussian input distribution achieves the first-order term in (30).

Proof: See Appendix IV. ∎

Note that, in the high-SNR regime, the difference in achievable rate between affine MMSE-PAT with an i.i.d. Gaussian input and with the optimal input is only a finite constant; this lends weight to the Gaussian analysis in Section III-B.

In the case of N-block transmission under perfect receiver-CSI, it is well known that the asymptotic capacity obeys $\mathcal{C}_{\mathsf{csi-blk}} = N \log \mathrm{SNR} + O(1)$ bits/block as $\mathrm{SNR} \to \infty$. Notice that, in the high-SNR regime, $\mathcal{C}_{\mathsf{mmse-blk}}$ and $\mathcal{C}_{\mathsf{csi-blk}}$ increase linearly in $\log \mathrm{SNR}$ with slopes N_s and N, respectively. Recalling the range of N_s from (18), it is evident that affine MMSE-PAT suffers a pre-log factor penalty relative to the perfect-receiver-CSI system. In fact, at high SNR, the loss in pre-log factor becomes the dominant cost of imperfect receiver-CSI, since there the MMSE-PAT channel estimates will be accurate and losses due to channel estimation error will be insignificant. For channels with low spreading index (i.e., $\gamma \ll 1$), N_s will be close to N (recall $\gamma = N_f N_t/N$), and, hence, the pre-log factors of $\mathcal{C}_{\mathsf{mmse-blk}}$ and $\mathcal{C}_{\mathsf{csi-blk}}$ will be similar. Thus, for low spreading indices and moderately high SNR, the rates achieved by affine MMSE-PAT should not be far from those of a perfect-receiver-CSI system. This is intuitively satisfying, since smaller γ implies relatively few unknown channel parameters and thus relatively small pilot overhead. On the other hand, for channels with high spreading index (γ close to 1), N_s will be significantly less than N [recall (18)]. In this case, $\mathcal{C}_{\mathsf{mmse-blk}}$ can deviate significantly from $\mathcal{C}_{\mathsf{csi-blk}}$, even at moderately high SNR. These trends are confirmed by the numerical results in Section V.

For specific modulation schemes, PAT achievable rates have been previously investigated for flat [10], [16], frequency-selective [11] and time-selective [17] fading channels. For SCM over the DSC, PAT parameters have been optimized with the goal of achievable-rate maximization in [12], yielding a scheme which resembles the TDKD example in Section II-F. In contrast, our achievable-rate analyses and pilot/data power allocation procedure apply to *any* affine CP MMSE-PAT scheme over the CE-BEM DSC, including the MMSE-PAT examples given in Section II-F. Furthermore, our unified rate analysis allows a *comparison* of different affine MMSE-PAT schemes according to their achievable rates. To this end, Theorem 3 implies that, among affine MMSE-PAT schemes, the one with the highest data-dimension N_s will yield the highest information rate in the high-SNR regime. Comparing the examples in Section II-F for channel parameters (N_t, N_f), TDKD maximizes N_s when $N_f > N_t$ and FDKD maximizes N_s when $N_f < N_t$. Apart from the trivial case $N_f = N_t = 1$, the chirp-based MMSE-PAT schemes yield smaller N_s than both TDKD and FDKD.

IV. Spectral Efficiency of Streaming MMSE-PAT

In this section, we analyze the spectral efficiency of systems which transmit a stream of blocks, where each block is constructed according to the affine MMSE-PAT principles discussed earlier and separated from its neighbors by time-domain guard intervals. We consider guards based on CP (as assumed in Section II) as well as ZP (as assumed, e.g., in [12]). In particular, we examine the time-bandwidth resources consumed by these systems and analyze their achievable rates in units of bits/s/Hz. To quantify time-bandwidth consumption, we consider the use of an arbitrary continuous-time baseband-equivalent pulse. The analysis in this section facilitates a direct comparison between the CP-based MMSE-PAT schemes in Section II-F and the ZP-based SCM scheme from [12].

A. Streaming Transmission Models

Let $\{x^{(m)}(i)\}_{i=0}^{N-1}$ denote the discrete-time transmitted sequence within the m^{th} block, and let $c(t)$ denote a continuous-time (complex baseband) pulse shape. We assume that $\{x^{(m)}(i)\}_{i=0}^{N-1}$ is constructed in the manner of $\{x(i)\}_{i=0}^{N-1}$ from (4), but with pilot and data that satisfy the affine MMSE-PAT conditions in Lemma 1. We assume that $c(t)$ is negligible for $t \notin [-(T_s/2), (T_s/2)]$ and that $C(f) = \int_{-\infty}^{\infty} c(t) e^{-j2\pi f t} dt$

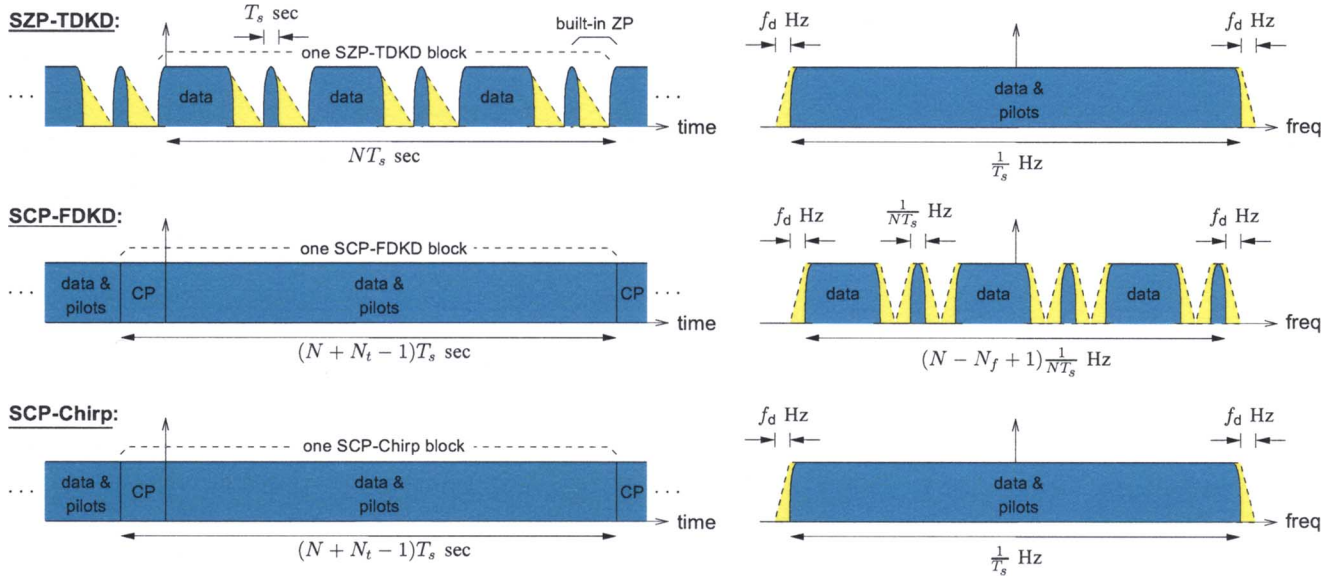


Fig. 2. Time and bandwidth occupation for several streaming MMSE-PAT systems designed for $N_t = 3$ and $N_f = 3$. Darkly shaded shows channel input and lightly shaded shows channel output. Note the use of a time-frequency concentrated modulation pulse $c(t)$.

is negligible for $f \notin [-(1/2T_s), (1/2T_s)]$. Using $x_c(t)$ to denote the continuous-time transmitted waveform, the streaming cyclic-prefixed (SCP) modulator generates

$$x_c(t) = \sum_{m\in\mathbb{Z}} \sum_{i=-N_t+1}^{N-1} x^{(m)}(i)\, c\,(t - iT_s - m(N+N_t-1)T_s) \tag{31}$$

where modulo-N indexing is assumed for $x^{(m)}(i)$ in (31). We refer to SCP with $\{x^{(m)}(i)\}_{i=0}^{N-1}$ constructed according to the FDKD example from Example 2, with $l = N/N_t - N_f$ and arbitrary $\theta(k)$, as "SCP-FDKD." Similarly, we use "SCP-Chirp" to refer to the corresponding schemes constructed from either the time- or frequency-domain Chirp examples from Section II-F. (See Fig. 2.) Because of the CP, the SCP block period equals $(N+N_t-1)T_s$ seconds. For the streaming zero-padded (SZP) PAT scheme from [12], the modulator generates

$$x_c(t) = \sum_{m\in\mathbb{Z}} \sum_{i=0}^{N-1} x^{(m)}(i) c(t - iT_s - mNT_s). \tag{32}$$

We refer to SZP with $\{x^{(m)}(i)\}_{i=0}^{N-1}$ constructed according to the TDKD example from Example 1, with $l = N/N_f - N_t$ and arbitrary $\theta(k)$, as "SZP-TDKD." (See Fig. 2.) Noting that the TDKD samples $\{x^{(m)}(i)\}_{i=N-N_t+1}^{N-1}$ are zero-valued for this choice of l, it can be seen that SZP-TDKD takes advantage of a "built in" ZP, which permits efficient concatenation of blocks (Fig. 2) at a block period of NT_s seconds.

The continuous-time channel output is $y_c(t) = \int h_c(t,\tau) x_c(t-\tau) + v_c(t)$, where $h_c(t,\tau)$ denotes the (time-varying) continuous-time DSC impulse response and $v_c(t)$ denotes the noise waveform. Assuming that the pulse $c(t)$ is also used at the receiver, $h_c(t,\tau)$ can be related to the discrete-time impulse response $h(i,\ell)$ from (1) via $h(i,\ell) = \int\int c(t) h_c(t+iT_s; \tau+\ell T_s) c(t-\tau) dt d\tau$ [25] and $v_c(t)$ can be related to the sampled noise $v(i)$ from (1) via $v(i) = \int c(t) v_c(t+iT_s) dt$.

B. Time-Bandwidth Consumption

We now compute the spectral efficiencies of SZP-TDKD, SCP-FDKD and SCP-Chirp schemes in bits/s/Hz by quantifying the per-block time-bandwidth resources consumed by these systems. To do this, we examine the continuous-time *channel output* signal $y_c(t)$, since we reason that channel output properties are more relevant than channel input properties when analyzing egress onto adjacent frequency bands. In the sequel, we make the approximation $2f_\mathsf{d} \approx (N_f - 1)/NT_s$, which is accurate for large N.[4]

As illustrated in Fig. 2, each block of SCP-FDKD and SCP-Chirp schemes consumes $(N+N_t-1)T_s$ seconds while that of SZP-TDKD consumes NT_s seconds. To find the bandwidth consumption of these systems, we assume uncorrelated information symbols $\{s^{(m)}(k)\}$ (e.g., from an i.i.d. Gaussian codebook). Observe that SZP-TDKD is an SCM system whose transmit signal occupies a bandwidth of $1/T_s$ Hz. Because of the frequency dispersion of DSC, the channel output has bandwidth of $(1/T_s) + 2f_\mathsf{d} = (N+N_f-1)/NT_s$ Hz, so that SZP-TDKD consumes $N+N_f-1$ Hz s per block. In contrast, SCP-FDKD is a multicarrier system transmitting pilots and data in the frequency domain, where each subcarrier consumes a bandwidth of $1/NT_s$ Hz. Because of zero-valued pilot subcarriers at the band edges (with the choice of $l = (N/N_t) - N_f$), the SCP-FDKD output signal consumes a bandwidth of $((N-N_f+1)/NT_s) + 2f_\mathsf{d} = 1/T_s$ Hz. Thus, SCP-FDKD consumes $N - N_t + 1$ Hz s per block. Notice that SCP-FDKD consumes more time resources, but less frequency resources, (per block) than SZP-TDKD. This follows from the fact that

[4]By writing $N_f = 2\lfloor f_\mathsf{d} T_s N \rfloor + 1 = 2 f_\mathsf{d} T_s N - \epsilon + 1$ where $\epsilon \in [0,1)$, we see that $2f_\mathsf{d} = ((N_f-1)/NT_s) + (\epsilon/NT_s)$. We desire that the approximation error satisfies $\epsilon/NT_s \ll 1/T_s$, or equivalently $N \gg \epsilon$, and this is guaranteed when $N \gg 1$.

SCP-FDKD contains "built-in" frequency-domain guard intervals while SZP-TDKD contains "built-in" time-domain guard intervals. Both the SCP-Chirp schemes consume a bandwidth of $(1/T_s) + 2f_{\mathsf{d}} = (N + N_f - 1)/NT_s$ Hz, so that their time-bandwidth consumption is $(N+N_t-1)(N+N_f-1)/N$ Hz s per block.

C. On the Definition of SNR

For CP-based systems, we have assumed throughout that the receiver discards the samples corresponding to the CP. (Recall Section II-A.) In contrast, the SZP-TDKD receiver does not discard samples. Thus, when comparing these systems, we must be careful when defining SNR.

We define SNR as *the ratio of signal power to noise power observed at the output of the receiver's pulse-shaping filter*. Because we assume an energy-preserving channel $\{h(i,\ell)\}$, this SNR can be equivalently described as the ratio of transmitted signal power to received noise power. Since SZP-TDKD has a built-in ZP, the transmitted signal power is $\sigma^2_{\mathsf{zp}} = (1/N)\sum_{i=0}^{N-1}\mathrm{E}\{|x^{(m)}(i)|^2\} = E_x/N$. For SCP schemes, the transmitted signal power is $\sigma^2_{\mathsf{cp}} = (1/(N + N_t - 1))\sum_{i=-N_t+1}^{N-1}\mathrm{E}\{|x^{(m)}(i)|^2\}$, taking the cyclic prefix into account. However, since it is easily verified that SCP-FDKD and SCP-Chirp both guarantee that $(1/(N + N_t - 1))\sum_{i=-N_t+1}^{N-1}\mathrm{E}\{|x^{(m)}(i)|^2\} = 1/N\sum_{i=0}^{N-1}\mathrm{E}\{|x^{(m)}(i)|^2\}$, we find that $\sigma^2_{\mathsf{zp}} = \sigma^2_{\mathsf{cp}}$ for these schemes. In other words, SZP-TDKD, SCP-FDKD, and SCP-Chirp all have the same average transmitted power, and thus can be fairly compared under the definition SNR $= E_x/N\sigma_v^2$ that was originally introduced in Section III-A.

D. Spectral Efficiency of Streaming MMSE-PAT

In this section, we combine the per-block achievable rates (from Section III) with the per-block time-bandwidth consumption (from Section IV-B) in order to write the achievable rates in units of bits/s/Hz. As shown in Section IV-C, the definition of SNR from Section III-A can be fairly applied to SZP-TDKD, SCP-FDKD, and SCP-Chirp. In addition, the channel matrix definition (22) holds for SZP-TDKD as well as SCP schemes because, as evident from Fig. 2, the ZP portion of each SZP-TDKD block acts as if it were a CP for the next SZP-TDKD block. Thus, with i.i.d. Gaussian input, upper and lower bounds on the achievable rate of the streaming schemes under consideration are

$$\mathcal{C}_{\mathsf{mmse-gau-ub}} = \kappa\mathcal{C}_{\mathsf{mmse-gau-blk-ub}} \quad \text{bits/s/Hz} \tag{33}$$

$$\mathcal{C}_{\mathsf{mmse-gau-lb}} = \kappa\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} \quad \text{bits/s/Hz} \tag{34}$$

where $\mathcal{C}_{\mathsf{mmse-gau-blk-ub}}$ and $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$ were defined in Theorem 2 and where κ equals $1/[N+N_f-1]$, $1/[N+N_t-1]$ and $N/[(N + N_t - 1)(N + N_f - 1)]$ for SZP-TDKD, SCP-FDKD, and SCP-Chirp schemes, respectively. Using similar arguments, the achievable rate of an SCP system with i.i.d. Gaussian inputs and perfect receiver-CSI (and no pilots) is

$$\mathcal{C}_{\mathsf{csi-gau}} = \frac{N}{(N+N_t-1)(N+N_f-1)}\mathcal{C}_{\mathsf{csi-gau-blk}} \ \text{bits/s/Hz}. \tag{35}$$

The asymptotic achievable rates in bits/s/Hz can then be written as

$$\mathcal{C}_{\text{SZP-TDKD}} = \underbrace{\left(\frac{N-(2N_t-1)N_f}{N+N_f-1}\right)}_{\eta\ \text{SZP-TDKD}}\log(\text{SNR}) + O(1) \tag{36}$$

$$\mathcal{C}_{\text{SCP-FDKD}} = \underbrace{\left(\frac{N-(2N_f-1)N_t}{N+N_t-1}\right)}_{\eta\text{SCP-FDKD}}\log(\text{SNR}) + O(1) \tag{37}$$

$$\mathcal{C}_{\text{SCP-Chirp}} = \underbrace{\left(\frac{(N-2N_fN_t+1)N}{(N+N_f-1)(N+N_t-1)}\right)}_{\eta_{\mathsf{SCP-Chirp}}}\log(\text{SNR}) + O(1) \tag{38}$$

for SNR $\to \infty$. For integer N_t and N_f, it easily follows that $N_t \geq N_f \Leftrightarrow \eta_{\mathsf{SCP-FDKD}} \geq \eta_{\mathsf{SZP-TDKD}}$ and vice versa. Thus, in the high-SNR regime, SCP-FDKD dominates SZP-TDKD (from [12]) when $N_t > N_f$, while SZP-TDKD dominates SCP-FDKD when $N_t < N_f$.

V. Numerical Examples

In this section, we present numerical examples of the spectral efficiencies of several streaming MMSE-PAT schemes. For this purpose, we evaluate the achievable rate bounds derived in Section IV-D (in units of bits/s/Hz) for the SZP-TDKD, SCP-FDKD, and SCP-Chirp schemes, using the power allocation procedure described in Section III-C. In all cases, we consider block size $N = 120$ and plot the bounds over the SNR range of practical interest. As discussed in Section IV-C, the definition of SNR allows a fair comparison of the three schemes under consideration. We used a Monte Carlo approach to evaluate the expectation in the bounding expressions (23) and (24).

In Fig. 3, we show results for a channel with i.i.d. BEM coefficients (i.e., $\boldsymbol{R}_\lambda = (N/N_fN_t)\boldsymbol{I}$) such that $N_t = N_f = 3$. These discrete spreading parameters correspond to a spreading index of $\gamma \approx 0.07$, as results from, e.g., carrier frequency $f_c = 80$ GHz, sampling interval $T_s = 1$ μs, mobile velocity $v = 150\,\mathrm{km/hr}$, and a channel delay-spread of 3 μs. These physical channel parameters are related to f_{d} via $f_{\mathsf{d}} = f_cv/c$, where c denotes the speed of light. Both the SCP-Chirp schemes yield identical bounds since they have the same spectral efficiency. Also, as can be seen from Fig. 3, SZP-TDKD and SCP-FDKD yield identical bounds on achievable rate, which is expected since $N_t = N_f$. The bounds for SCP-Chirp schemes are uniformly lower than those for SZP-TDKD and SCP-FDKD, which is also expected since $N_f \neq 1 \neq N_t$. For reference, Fig. 3 also plots the performance of an SCP system with perfect receiver CSI (and no pilots) via (35).

In Fig. 4, we show results for a channel with i.i.d. BEM coefficients such that $N_t = 15$ and $N_f = 3$. This channel is primarily time-spreading with spreading index $\gamma \approx 0.37$ and results from, e.g., the same physical channel parameters as before, but with a channel delay-spread of 15 μs. Note that, since $N_t > N_f$, the SCP-FDKD bounds dominate SZP-TDKD bounds, which in turn dominate the SCP-Chirp bounds. Compared to Fig. 3, there is a much larger gap between the MMSE-PAT bounds and

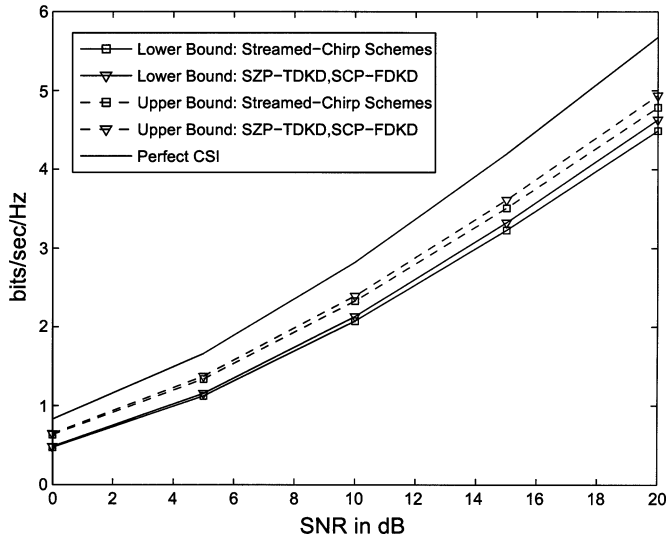


Fig. 3. Bounds on ergodic achievable rates for $N_t = N_f = 3$.

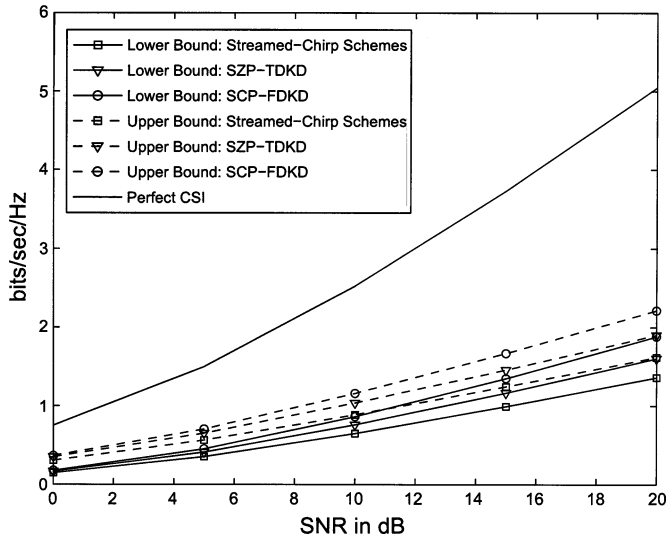


Fig. 4. Bounds on ergodic achievable rates for $N_t = 15$ and $N_f = 3$.

the perfect-receiver-CSI bounds, as a consequence of the higher spreading index.

In Fig. 5, we compare SZP-TDKD to SCP-FDKD on two channels with $N_t = 15$ and $N_f = 3$. One has i.i.d. BEM coefficients (i.e., $\mathrm{E}|\lambda(k,\ell)|^2 = N/N_fN_t$ for $k \in \{-((N_f-1)/2),\ldots,((N_f-1)/2)\}$ and $\ell \in \{0,\ldots,N_t-1\}$), and the other has independent but nonidentically distributed BEM coefficients, as would result from a nonuniform delay profile and a nonuniform Doppler spectrum. In particular, we consider a channel with a "Jakes" Doppler spectrum and an exponential delay profile, for which $\mathrm{E}\{|\lambda(k,\ell)|^2\} = \xi e^{-0.1l}(f_{\mathrm{d}}^2 - k^2(NT_s)^{-2})^{-0.5}$ for $k \in \{-((N_f-1)/2),\ldots,((N_f-1)/2)\}$ and $\ell \in \{0,\ldots,N_t-1\}$, and where ξ is chosen such that $\mathrm{tr}\{\boldsymbol{R}_\lambda\} = N$. For the channel with i.i.d. BEM coefficients, we allocate pilot/data power according to the procedure in Section III-C, while, for the non-i.i.d. channel, we allocate equal power between pilots and data. In both cases, we see

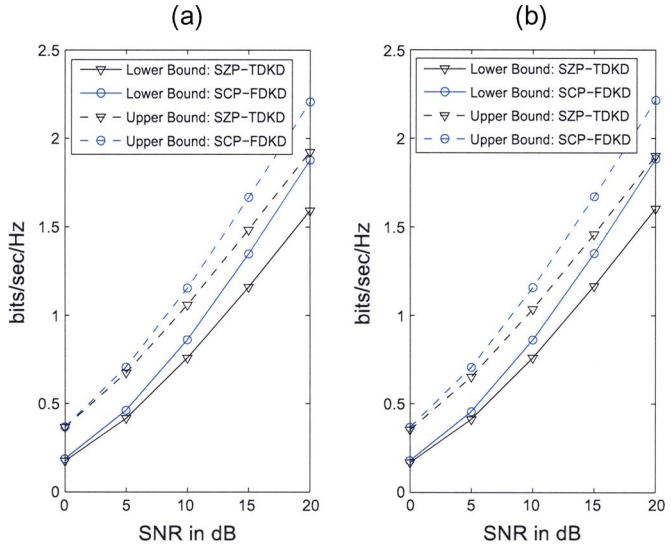


Fig. 5. Bounds on ergodic achievable rates for $N_t = 15$ and $N_f = 3$ for (a) nonuniform and (b) uniform decay and Doppler profiles.

that the achievable rate bounds grow (asymptotically) linearly in $\log \mathrm{SNR}$ with slopes proportional to the pre-log factors in the asymptotic rate expressions (36) and (37). Since, for this channel, $N_t > N_f$, SCP-FDKD's higher pre-log factor translates into significant rate gains over SZP-TDKD at high SNR.

VI. Conclusion

This paper presented necessary and sufficient conditions for cyclic-prefixed block-based affine PAT to yield MMSE channel estimates for the single-antenna CE-BEM DSC. Using these conditions, novel MMSE-PAT schemes were proposed, and certain heuristic PAT schemes were shown to be MSE-optimal. In addition, bounds on the ergodic achievable rates of affine MMSE-PAT systems were derived, and a pilot/data power allocation scheme was developed. A high-SNR analysis was then presented which suggested that, to maximize high-SNR achievable rate, the channel's spreading parameters should be taken into account when choosing among affine MMSE-PAT schemes.

Appendix I
Proof of Theorem 1

We begin with the singular value decompositions $\boldsymbol{PU} = \boldsymbol{V}_{\mathrm{p}}\boldsymbol{\Sigma}_{\mathrm{p}}\boldsymbol{Q}_{\mathrm{p}}^{\mathsf{H}}$ and $\boldsymbol{DU} = \boldsymbol{V}_{\mathrm{d}}\boldsymbol{\Sigma}_{\mathrm{d}}\boldsymbol{Q}_{\mathrm{d}}^{\mathsf{H}}$, where $\boldsymbol{\Sigma}_{\mathrm{p}}$ and $\boldsymbol{\Sigma}_{\mathrm{d}}$ are diagonal and full-rank. Let $K \le N_fN_t$ denote the rank of $\boldsymbol{\Sigma}_{\mathrm{p}}$. Defining $\boldsymbol{z} = \boldsymbol{V}_{\mathrm{p}}^{\mathsf{H}}\boldsymbol{y}$ and using (2), (3), and (7)

$$\boldsymbol{z} = \underbrace{\boldsymbol{\Sigma}_{\mathrm{p}}\boldsymbol{Q}_{\mathrm{p}}^{\mathsf{H}}}_{\boldsymbol{A}_{\mathrm{p}}}\boldsymbol{\lambda} + \underbrace{\boldsymbol{V}_{\mathrm{p}}^{\mathsf{H}}\boldsymbol{V}_{\mathrm{d}}\boldsymbol{\Sigma}_{\mathrm{d}}\boldsymbol{Q}_{\mathrm{d}}^{\mathsf{H}}}_{\boldsymbol{A}_{\mathrm{d}}}\boldsymbol{\lambda} + \underbrace{\boldsymbol{V}_{\mathrm{p}}^{\mathsf{H}}\boldsymbol{v}}_{\boldsymbol{n}}.$$

Since projection onto $\mathrm{col}(\boldsymbol{V}_{\mathrm{p}})$ does not attenuate the pilot component of $\boldsymbol{y}$, the pilot-aided MMSE channel estimate given $\{\boldsymbol{y},\boldsymbol{P}\}$ is equal to that given $\{\boldsymbol{z},\boldsymbol{P}\}$. With $\boldsymbol{R}_{z,\lambda} = E\{\boldsymbol{z}\boldsymbol{\lambda}^{\mathsf{H}}\}$ and $\boldsymbol{R}_z = E\{\boldsymbol{z}\boldsymbol{z}^{\mathsf{H}}\}$, the MMSE estimate of $\boldsymbol{\lambda}$ given $\{\boldsymbol{z},\boldsymbol{P}\}$ is $\hat{\boldsymbol{\lambda}} = \boldsymbol{R}_{z,\lambda}^{\mathsf{H}}\boldsymbol{R}_z^{-1}\boldsymbol{z}$, where $\boldsymbol{R}_{z,\lambda} = \boldsymbol{A}_{\mathrm{p}}\boldsymbol{R}_\lambda$ and

$$\boldsymbol{R}_z = \underbrace{\boldsymbol{A}_{\mathrm{p}}\boldsymbol{R}_\lambda\boldsymbol{A}_{\mathrm{p}}^{\mathsf{H}} + \sigma_v^2\boldsymbol{I}_K}_{\boldsymbol{\Delta}} + \underbrace{E\{\boldsymbol{A}_{\mathrm{d}}\boldsymbol{R}_\lambda\boldsymbol{A}_{\mathrm{d}}^{\mathsf{H}}\}}_{\boldsymbol{U}_{\mathrm{d}}\boldsymbol{\Lambda}_{\mathrm{d}}\boldsymbol{U}_{\mathrm{d}}^{\mathsf{H}}}$$

with diagonal $\boldsymbol{\Lambda}_{\mathsf{d}} \geq 0$ and $\boldsymbol{U}_{\mathsf{d}}^{\mathsf{H}}\boldsymbol{U}_{\mathsf{d}} = \boldsymbol{I}$. Note that the MMSE estimate of $\hbar$ is $\hat{\hbar} = \boldsymbol{U}\hat{\boldsymbol{\lambda}}$ and that $\sigma_e^2 = E\{\|\hbar - \hat{\hbar}\|^2\} = E\{\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2\}$. We have $\boldsymbol{R}_{\tilde{\lambda}} = E\{(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})^{\mathsf{H}}\} = \boldsymbol{R}_{\lambda} - \boldsymbol{R}_{z,\lambda}^{\mathsf{H}}\boldsymbol{R}_z^{-1}\boldsymbol{R}_{z,\lambda}$ and

$$\begin{aligned}\sigma_e^2 &= \mathrm{tr}\left\{\boldsymbol{R}_{\lambda} - \boldsymbol{R}_{z,\lambda}^{\mathsf{H}}\boldsymbol{R}_z^{-1}\boldsymbol{R}_{z,\lambda}\right\} \\ &= \mathrm{tr}\left\{\boldsymbol{R}_{\lambda} - \boldsymbol{R}_{\lambda}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\left(\boldsymbol{\Delta} + \boldsymbol{U}_{\mathsf{d}}\boldsymbol{\Lambda}_{\mathsf{d}}\boldsymbol{U}_{\mathsf{d}}^{\mathsf{H}}\right)^{-1}\boldsymbol{A}_{\mathsf{p}}\boldsymbol{R}_{\lambda}\right\} \\ &= \mathrm{tr}\Big\{\boldsymbol{R}_{\lambda} - \boldsymbol{R}_{\lambda}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\Big[\boldsymbol{\Delta}^{-1} - \boldsymbol{\Delta}^{-1}\boldsymbol{U}_{\mathsf{d}}\left(\boldsymbol{\Lambda}_{\mathsf{d}}^{-1} + \boldsymbol{U}_{\mathsf{d}}^{\mathsf{H}}\boldsymbol{\Delta}^{-1}\boldsymbol{U}_{\mathsf{d}}\right)^{-1} \\ &\qquad \times \boldsymbol{U}_{\mathsf{d}}^{\mathsf{H}}\boldsymbol{\Delta}^{-1}\Big]\boldsymbol{A}_{\mathsf{p}}\boldsymbol{R}_{\lambda}\Big\} \qquad (39) \\ &\geq \mathrm{tr}\left\{\boldsymbol{R}_{\lambda} - \boldsymbol{R}_{\lambda}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\boldsymbol{\Delta}^{-1}\boldsymbol{A}_{\mathsf{p}}\boldsymbol{R}_{\lambda}\right\} \qquad (40)\end{aligned}$$

where we used the matrix inversion lemma in (39). Inequality (40) follows since $\boldsymbol{\Delta} > 0$ and $\boldsymbol{\Lambda}_{\mathsf{d}} \geq 0$. Since $\boldsymbol{\Sigma}_{\mathsf{p}}$ is full rank, $\boldsymbol{A}_{\mathsf{p}}$ has full row rank, and so equality in (40) is achieved if and only if

$$\boldsymbol{U}_{\mathsf{d}}\boldsymbol{\Lambda}_{\mathsf{d}}\boldsymbol{U}_{\mathsf{d}}^{\mathsf{H}} = \boldsymbol{0} \Leftrightarrow E\left\{\boldsymbol{A}_{\mathsf{d}}\boldsymbol{R}_{\lambda}\boldsymbol{A}_{\mathsf{d}}^{\mathsf{H}}\right\} = \boldsymbol{0}. \qquad (41)$$

Since $\boldsymbol{R}_{\lambda} > 0$, (41) is satisfied if and only if $\boldsymbol{A}_{\mathsf{d}} = \boldsymbol{0}$, which is equivalent to (11), since $\boldsymbol{\Sigma}_{\mathsf{p}}$ and $\boldsymbol{\Sigma}_{\mathsf{d}}$ are full rank square matrices. We proceed further assuming that (11) is satisfied. With

$$\begin{aligned}\boldsymbol{A}_{\mathsf{d}} = \boldsymbol{0}, \sigma_e^2 &= \mathrm{tr}\{\boldsymbol{R}_{\lambda} - \boldsymbol{R}_{\lambda}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}(\boldsymbol{A}_{\mathsf{p}}\boldsymbol{R}_{\lambda}\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}} + \sigma_v^2\boldsymbol{I}_K)^{-1}\boldsymbol{A}_{\mathsf{p}}\boldsymbol{R}_{\lambda}\} \\ &= \mathrm{tr}\{(\boldsymbol{R}_{\lambda}^{-1} + (1/\sigma_v^2)\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}})^{-1}\}\end{aligned}$$

using the matrix inversion lemma. With $\sigma_{\lambda_m}^2 = [\boldsymbol{R}_{\lambda}]_{m,m}$, diagonal $\boldsymbol{R}_{\lambda}$ implies

$$\sigma_e^2 \geq \sum_{m=0}^{N_f N_t - 1}\left(\frac{1}{\sigma_{\lambda_m}^2} + \frac{\alpha_m}{\sigma_v^2}\right)^{-1} \qquad (42)$$

where $\alpha_m = [\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}}]_{m,m}$ and equality in (42) is achieved if and only if $\boldsymbol{A}_{\mathsf{p}}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{p}} = (\boldsymbol{P}\boldsymbol{U})^{\mathsf{H}}\boldsymbol{P}\boldsymbol{U}$ is diagonal. (See [34] for additional details.) Since all the columns of $\boldsymbol{P}\boldsymbol{U}$ have norm E_p/N, the bound (42) is written as (10). When (11) and (12) are met, it follows straightforwardly that $\boldsymbol{R}_{\tilde{h}}$ is given by (13).

Appendix II
Proof of Lemma 1

Notice that $(\boldsymbol{P}\boldsymbol{U})^{\mathsf{H}}\boldsymbol{P}\boldsymbol{U}$ is composed of $N_f \times N_f$ blocks $\bar{\boldsymbol{P}}_{k_2,k_1} = \bar{\boldsymbol{F}}^{\mathsf{H}}\boldsymbol{P}_{-k_2}^{\mathsf{H}}\boldsymbol{P}_{-k_1}\bar{\boldsymbol{F}}$ for $k_1, k_2 \in \{0, \ldots, N_t - 1\}$. For these k_1, k_2 and for $m_1, m_2 \in \{0, \ldots, N_f - 1\}$, (12) becomes

$$\left[\bar{\boldsymbol{P}}_{k_2,k_1}\right]_{m_1,m_2} = \frac{E_p}{N}\delta(k_1 - k_2)\delta(m_1 - m_2). \qquad (43)$$

The definitions of $\bar{\boldsymbol{F}}$ and $\bar{\boldsymbol{P}}_{-i}$ imply

$$[\bar{\boldsymbol{P}}_{k_2,k_1}]_{m_1,m_2} = (1/N)\sum_{i=0}^{N-1} p(i - k_1)p^*(i - k_2)$$

$e^{-j(2\pi/N)(m_1 - m_2)i}$. Setting $k = k_2 - k_1$ and $m = m_1 - m_2$, so that $k \in N_t$ and $m \in N_f$, we have

$$\begin{aligned}\bar{\boldsymbol{P}}_{k_2,k_1}\big]_{m_1,m_2} &= \frac{1}{N}\sum_{q=-k_1}^{N-1-k_1} p(q)p^*(q-k)e^{-j\frac{2\pi}{N}m(q+k_1)} \\ &= \frac{e^{-j\frac{2\pi}{N}mk_1}}{N}\sum_{q=0}^{N-1} p(q)p^*(q-k)e^{-j\frac{2\pi}{N}mq} \qquad (44)\end{aligned}$$

where in (44) we exploited the fact that $p(-q) = p(N - q)$ for $1 \leq q < N_t$. Combining (43) and (44), we obtain (15). Similarly, we can show that (11) is equivalent to

$$\sum_{i=0}^{N-1} d(i)p^*(i - k)e^{-j\frac{2\pi}{N}mi} = 0 \quad \forall k \in \mathcal{N}_t,\ \forall m \in \mathcal{N}_f. \qquad (45)$$

Since *any* data vector $\boldsymbol{d}$ should satisfy (45), and since the information symbols $\boldsymbol{s}$ are arbitrary, the requirements (16) and (45) are equivalent.

Appendix III
Proof of Theorem 2

Let the columns of $\boldsymbol{B}_p \in \mathbb{C}^{N \times N_t N_f}$ form an orthonormal basis for $\mathrm{col}(\boldsymbol{P}\boldsymbol{U})$ and the columns of $\bar{\boldsymbol{B}}_d \in \mathbb{C}^{N \times (N - N_t N_f)}$ form an orthonormal basis for the left null space of $\boldsymbol{P}\boldsymbol{U}$. Notice that, with $\bar{\boldsymbol{B}} = [\boldsymbol{B}_p\boldsymbol{B}_d]$, we have $\bar{\boldsymbol{B}}^{\mathsf{H}}\bar{\boldsymbol{B}} = \bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^{\mathsf{H}} = \boldsymbol{I}_N$. The pilot-data orthogonality of MMSE-PAT [recall (11)] implies that $\boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B} = \boldsymbol{0}$ and $\boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{H}\boldsymbol{p} = \boldsymbol{0}$. Projecting the observed vector $\boldsymbol{y}$ onto the pilot and data subspaces, we obtain $\boldsymbol{y}_p = \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{y}$ and $\boldsymbol{y}_d = \boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{y}$, respectively, where

$$\boldsymbol{y}_p = \boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{H}\boldsymbol{p} + \underbrace{\boldsymbol{B}_p^{\mathsf{H}}\boldsymbol{v}}_{\boldsymbol{v}_p} \qquad (46)$$

$$\boldsymbol{y}_d = \boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B}\boldsymbol{s} + \underbrace{\boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{v}}_{\boldsymbol{v}_d}. \qquad (47)$$

Clearly, $\boldsymbol{v}_p$ and $\boldsymbol{v}_d$ are CWGN with variance σ_v^2. Since the projection (46) does not compromise pilot energy, the LMMSE estimate of $\hbar$ given $\boldsymbol{y}$ equals the LMMSE estimate given $\boldsymbol{y}_p$. Similarly, since the projection (47) does not compromise data energy, there is no loss in mutual information, i.e., $I(\boldsymbol{y}; \boldsymbol{s}) = I(\boldsymbol{y}_d; \boldsymbol{s})$.

Splitting the channel matrix $\boldsymbol{H}$ into the estimate and error component as $\boldsymbol{H} = \hat{\boldsymbol{H}} + \tilde{\boldsymbol{H}}$, we have

$$\boldsymbol{y}_d = \boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\boldsymbol{s} + \underbrace{\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{B}\boldsymbol{s} + \boldsymbol{v}_d}_{\boldsymbol{v}_e}. \qquad (48)$$

The orthogonality principle guarantees that $\boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\boldsymbol{s}$ will be uncorrelated with $\boldsymbol{v}_e$. In this case, it has been shown [10] that the worst-case distribution for $\boldsymbol{v}_e$ (from a mutual information perspective) is Gaussian. With an i.i.d. Gaussian distribution on $\boldsymbol{s}$, we have the following mutual information bound [33]:

$$I(\boldsymbol{y}_d; \boldsymbol{s}) \geq E\left\{\log\det\left[\boldsymbol{I} + \sigma_s^2\boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\boldsymbol{R}_{v_e}^{-1}\left(\boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\right)^{\mathsf{H}}\right]\right\} \qquad (49)$$

where $\boldsymbol{R}_{v_e} = E\{\boldsymbol{v}_e\boldsymbol{v}_e^{\mathsf{H}}\}$. Furthermore, we have $\boldsymbol{R}_{v_e} = \sigma_v^2\boldsymbol{I} + \sigma_s^2 E\{\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{B}(\boldsymbol{B}_d^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{B})^{\mathsf{H}}\} \leq \sigma_v^2\boldsymbol{I} + \sigma_s^2\boldsymbol{B}_d^{\mathsf{H}}E\{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^{\mathsf{H}}\}\boldsymbol{B}_d$, in the

positive definite sense, since the columns of $\boldsymbol{B}$ are orthonormal. Since $E\{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathsf{H}}\} = (\mathrm{tr}\{\boldsymbol{R}_{\tilde{h}}\}/N)\boldsymbol{I}$ and $\boldsymbol{B}_d^{\mathsf{H}}\boldsymbol{B}_d = \boldsymbol{I}$, we have

$$\boldsymbol{R}_{v_e} \leq \left(\sigma_v^2 + \sigma_s^2 \frac{\mathrm{tr}\{\boldsymbol{R}_{\tilde{h}}\}}{N}\right)\boldsymbol{I}. \tag{50}$$

Substituting (50) into (49), we have

$$I(\boldsymbol{y}_d;\boldsymbol{s}) \geq E\left\{\log\det\left[\boldsymbol{I} + \frac{N\sigma_s^2}{\sigma_s^2\mathrm{tr}\{\boldsymbol{R}_{\tilde{h}}\} + N\sigma_v^2} \times \boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\left(\boldsymbol{B}_d^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B}\right)^{\mathsf{H}}\right]\right\}. \tag{51}$$

It follows that $\boldsymbol{B}_p^{\mathsf{H}}\hat{\boldsymbol{H}}\boldsymbol{B} = \boldsymbol{0}$ due to the similarity in the structure of $\boldsymbol{H}$ and $\hat{\boldsymbol{H}}$. Using the determinant identity $\det(\boldsymbol{I} + \boldsymbol{G}_1\boldsymbol{G}_2) = \det(\boldsymbol{I} + \boldsymbol{G}_2\boldsymbol{G}_1)$ and the fact that $\bar{\boldsymbol{B}}_d\bar{\boldsymbol{B}}_d^{\mathsf{H}} = \boldsymbol{I}_N - \bar{\boldsymbol{B}}_p\bar{\boldsymbol{B}}_p^{\mathsf{H}}$, the lower bound on the mutual information (51) can be rewritten as (23) after the normalization of $\hat{\boldsymbol{H}}$.

For the upper bound on the mutual information, we consider the "best case" scenario of perfect channel estimates, i.e., $\hat{\boldsymbol{H}} = \boldsymbol{H}$ and $\tilde{\boldsymbol{H}} = 0$. For this scenario, with i.i.d. Gaussian input distribution, it is straightforward [33] to obtain the bound (24).

Appendix IV
Proof of Theorem 3

We bound the asymptotic achievable rates from both below and above and get the complete characterization. To start, $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} \leq \mathcal{C}_{\mathsf{mmse-blk}}$. First, let us characterize the asymptotic behavior of $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$. With any *fixed* power allocation fraction $\alpha \in (0,1)$, $E_s = \alpha E_x$ and $E_p = (1-\alpha)E_x$, as $\mathsf{SNR} \to \infty$, estimation error variance goes to zero, i.e., $\lim_{\mathsf{SNR}\to\infty}\mathrm{tr}\{\boldsymbol{R}_{\tilde{h}}\} = 0$ and, hence, it follows that $\lim_{\mathsf{SNR}\to\infty}\mathrm{tr}\{\boldsymbol{R}_{\hat{h}}\} = N$ and $\lim_{\mathsf{SNR}\to\infty}\rho_{\mathsf{lb}} = \sigma_s^2/\sigma_v^2$. Also, as $\mathsf{SNR} \to \infty$, the normalized channel estimates are equal to the actual channel almost surely (a.s.), $\lim_{\mathsf{SNR}\to\infty}\bar{\boldsymbol{H}} = \boldsymbol{H}$. Using Fatou's lemma [35] we find the lower bound on the asymptotics of $\mathcal{C}_{\mathsf{mmse-gau-blk-lb}}$, by interchanging the limit and the expectation. We have

$$\begin{aligned}\lim_{\mathsf{SNR}\to\infty} &\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} \\ &\geq E\left\{\lim_{\mathsf{SNR}\to\infty}\log\det[\boldsymbol{I} + \rho_{\mathsf{lb}}\boldsymbol{B}^{\mathsf{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}\bar{\boldsymbol{H}}\boldsymbol{B}]\right\} &(52)\\ &= E\left\{\log\det\lim_{\mathsf{SNR}\to\infty}[\boldsymbol{I} + \rho_{\mathsf{lb}}\boldsymbol{B}^{\mathsf{H}}\bar{\boldsymbol{H}}^{\mathsf{H}}\bar{\boldsymbol{H}}\boldsymbol{B}]\right\} &(53)\end{aligned}$$

which follows from the continuity of $\log\det(\cdot)$. Since $\exists k$ such that $\sigma_s^2/\sigma_v^2 \geq k\mathsf{SNR}$, $\forall \mathsf{SNR} \geq 1$, we have $\lim_{\mathsf{SNR}\to\infty}\mathcal{C}_{\mathsf{mmse-gau-blk-lb}} \geq E\{\log\det\lim_{\mathsf{SNR}\to\infty}[\boldsymbol{I} + k\mathsf{SNR}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B}]\} = E\{\log\prod_{i=0}^{N_s-1}(1 + k\mathsf{SNR}\mu_i)\}$, where μ_i are the eigen values of $\boldsymbol{B}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B}$. Since $\boldsymbol{H}\boldsymbol{B}$ is full rank (N_s) with probability 1, we have the following lower bound on $\mathcal{C}_{\mathsf{mmse-blk}}$

$$\begin{aligned}\mathcal{C}_{\mathsf{mmse-blk}}(\mathsf{SNR}) &\geq \mathcal{C}_{\mathsf{mmse-gau-blk-lb}}(\mathsf{SNR}) \\ &\geq N_s\log(\mathsf{SNR}) + O(1) &(54)\end{aligned}$$

as $\mathsf{SNR} \to \infty$.

Now, to bound the asymptotic achievable rates from above, we consider the "coherent" case of zero channel estimation error. Since the pilot component $\boldsymbol{H}\boldsymbol{p}$ can be subtracted from $\boldsymbol{y}$ (21) without affecting the mutual information, we start with

$$\boldsymbol{y}_c = \boldsymbol{H}\boldsymbol{B}\boldsymbol{s} + \boldsymbol{v}. \tag{55}$$

Using $\mathcal{C}_{\mathsf{coh-blk}}$ to denote the capacity of (55) with power constraint σ_x^2, it is evident that $\mathcal{C}_{\mathsf{mmse-blk}} \leq \mathcal{C}_{\mathsf{coh-blk}}$. In fact, for (55), the capacity maximizing input distribution is zero-mean Gaussian [33], so that

$$\begin{aligned}\mathcal{C}_{\mathsf{mmse-blk}} &\leq \mathcal{C}_{\mathsf{coh-blk}} \\ &= \sup_{\mathrm{tr}\{\boldsymbol{R}_s\}\leq N\sigma_x^2} E\left\{\log\det\left[\boldsymbol{I}_{N_s} + \frac{1}{\sigma_v^2}(\boldsymbol{H}\boldsymbol{B})^{\mathsf{H}}\boldsymbol{R}_s\boldsymbol{H}\boldsymbol{B}\right]\right\} &(56)\end{aligned}$$

where $\boldsymbol{R}_s$ denotes the covariance of $\boldsymbol{s}$. Note that, for any $\boldsymbol{R}_s$ satisfying the power constraint in (56), we have $\boldsymbol{R}_s \leq N\sigma_x^2\boldsymbol{I}_{N_s}$, in the positive semidefinite sense. Using this in (56), we find

$$\begin{aligned}\mathcal{C}_{\mathsf{mmse-blk}} &\leq E\left\{\log\det\left[\boldsymbol{I}_{N_s} + \frac{N\sigma_x^2}{\sigma_v^2}(\boldsymbol{H}\boldsymbol{B})^{\mathsf{H}}\boldsymbol{H}\boldsymbol{B}\right]\right\} &(57)\\ &= N_s\log\mathsf{SNR} + O(1) &(58)\end{aligned}$$

as $\mathsf{SNR} \to \infty$. Theorem 3 follows from (54) and (58).

References

[1] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[2] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

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

[4] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with applications to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.

[5] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[6] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmission over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[7] M. Dong, L. Tong, and B. M. Sadler, "Optimal pilot placement for channel tracking in OFDM," in *Proc. IEEE Military Commun. Conf.*, Oct. 2002.

[8] J.-W. Choi and Y.-H. Lee, "Optimal pilot pattern for channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2083–2088, Sep. 2005.

[9] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, June 2003.

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

[11] 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.*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.

[12] 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.

[13] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[14] L. Yang, X. Ma, and G. B. Giannakis, Eds., "Optimal training for MIMO fading channels with time- and frequency-selectivity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 3, pp. 821–824.

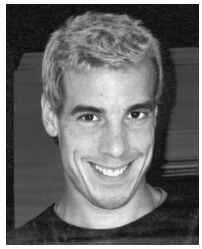
[15] S. Misra, A. Swami, and L. Tong, "Cutoff rate analysis of the Gauss-Markov fading channel with adaptive energy allocation," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2003, pp. 388–392.

[16] L. Zheng and D. Tse, "Communication on 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.

[17] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.
[18] W. C. Jakes, *Microwave Mobile Communications*. : Wiley, 1974.
[19] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adapt. Contr. Signal Process.*, vol. 10, pp. 159–176, Mar. 1996.
[20] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2138–2145, Sep. 2004.
[21] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Jun. 2000, vol. 5, pp. 2749–2752.
[22] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.
[23] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.
[24] P. A. Bello, "Time-frequency duality," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 1, pp. 18–33, Jan. 1964.
[25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.
[26] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
[27] X. Ma, G. Leus, and G. B. Giannakis, "Space-time-Doppler block coding for correlated time-selective fading channels," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2167–2182, June 2005.
[28] S. Ohno and G. B. Giannakis, "Average rate optimal PSAM transmissions over time selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.
[29] B. D. Hart and D. P. Taylor, "Extended MLSE diversity receiver for the time and frequency selective channel," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 322–333, Mar. 1997.
[30] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Estimation of fast fading channels in OFDM," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2002, vol. 1, pp. 465–470.
[31] S. Barbarossa and A. Scaglione, "Theoretical bounds on the estimation and prediction of multipath time-varying channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Jun. 2000, vol. 5, pp. 2545–2548.
[32] H. Weingarten, Y. Steinberg, and S. Shamai, "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.
[33] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
[34] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 514–557, Feb. 2004.
[35] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: Wiley, 1999.

Arun P. Kannu (M'07) received the M.S. and Ph.D. degrees in electrical and computer engineering from The Ohio State University, Columbus, in 2004 and 2007, respectively.

Currently, he is a Senior Systems Engineer with the Corporate R&D division of Qualcomm Inc., San Diego, CA.



Philip Schniter (S'03–M'93–SM'05) received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively. In 2000, he received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY.

From 1993 to 1996, he was with Tektronix Inc., Beaverton, OR, as a Systems Engineer. Subsequently, he joined the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, where he is now an Associate Professor. His research interests include signal processing, communication theory, and wireless networks.

Dr. Schniter received the National Science Foundation CAREER Award in 2003, and he currently serves on the IEEE Signal Processing for Communications Technical Committee.