CAPACITY ANALYSIS OF MMSE PILOT-AIDED TRANSMISSION FOR DOUBLY SELECTIVE CHANNELS

Arun P. Kannu and Philip Schniter

Dept. ECE, The Ohio State University, 2015 Neil Ave, Columbus, OH 43210
pachaik@ece.osu.edu, schniter@ece.osu.edu

ABSTRACT

In previous work, we established necessary and sufficient conditions on pilot/data patterns that minimize the mean-squared error of pilot-aided channel estimates, given a pilot power constraint and a general linear time-varying channel model. Here we derive upper and lower bounds on the ergodic channel capacity of these MSE-optimal pilot-aided transmission (PAT) schemes. These bounds may be used to allocate power between pilot and data, as well as to choose among MSE-optimal pilot/data patterns. For example, the bounds show that frequency-domain PAT achieves higher capacity than time-domain PAT, or vice-versa, depending on the relative time- and frequency-dispersion of the channel.

1. INTRODUCTION

The wireless communication channel is typically modeled as linear transformation and parameterized by a set of time-varying coefficients. Coherent receivers estimate these channel coefficients for subsequent use in data detection. In pilot aided transmission (PAT), a known signal is embedded in the transmitted stream to facilitate channel estimation.

As noted in the recent overview [1], capacity-optimal PAT designs have thus far proven elusive. A considerable amount of work has been devoted to the simpler problem of MSE-optimal pilot pattern design, however (e.g., [2–7]). Furthermore, several authors have observed connections between the MSE-optimal and capacity-optimal criteria. For example, an information-theoretic analysis of imperfect channel knowledge appeared in [8]. More recently, PAT designs that maximize a lower bound on ergodic capacity were studied for MIMO flat-fading [9], frequency-selective [10–12], and doubly-selective [4] channels.

In [7], necessary and sufficient conditions on MSE-optimal pilot/data patterns were specified for a general class of linear time-varying channels. Here we present a capacity analysis of these MSE-optimal PAT schemes for the doubly-selective channel case. The work [4], while similar in its intent, considers only non-superimposed pilots, thereby omitting important instances of MSE-optimal pilot/data patterns. Our more general analysis provides insight into the advantages and disadvantages of different MSE-optimal PAT schemes. Specifically, we find that the channel's spreading characteristics imply relative advantages for time-domain versus frequency-domain PAT schemes. Numerical results confirm the theoretical analysis.

This paper is organized as follows. We present the system model in Section 2 and review MSE-optimal PAT in Section 3. We present a capacity analysis in Section 4 and conclude in Section 5.

Supported in part by NSF CAREER Grant CCR-0237037.

2. SYSTEM MODEL

We consider block linear modulation over doubly selective channels, where a basis expansion model (BEM) is used to characterize channel variation over the block.

2.1. Cyclic-Prefix Block Transmission Model

The output signal $\{y(n)\}$ is related to the transmit signal $\{t(n)\}$ via

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

where $\{v(n)\}$ is σ_v^2-variance CWGN and $\{h(n,\ell)\}$ is the time-n channel response to an impulse applied at time $n-\ell$. Here, N_t denotes the channel's time spread normalized to the symbol duration. The length-N block transmission $\{t(n)\}_{n=0}^{N-1}$ includes a cyclic prefix (CP) of length N_t-1. With large N, the CP overhead becomes insignificant. The received vector $\boldsymbol{y} := [y(0),\dots,y(N-1)]^t$ is formed after discarding the CP contribution. Defining

$$\begin{aligned}
\boldsymbol{T} &:= [\boldsymbol{T}_0 \cdots \boldsymbol{T}_{-N_t+1}] \\
\boldsymbol{T}_{-i} &:= \mathrm{diag}(t(-i),...,t(-i+N-1)) \\
\boldsymbol{h} &:= [\boldsymbol{h}_0^t \cdots \boldsymbol{h}_{N_t-1}^t]^t \\
\boldsymbol{h}_i &:= [h(0,i),\dots,h(N-1,i)]^t \\
\boldsymbol{v} &:= [v(0),\dots,v(N-1)]^t,
\end{aligned}$$

we obtain

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

The transmit signal is constructed as $t(i) = s(i)+x(i)$, where $\{s(i)\}$ is the pilot sequence and $\{x(i)\}$ is the zero-mean data sequence. Note the superposition of pilots and data. The pilot power is constrained as follows.

$$\frac{1}{N}\sum_{n=0}^{N-1} |s(n)|^2 \le \sigma_s^2. \tag{3}$$

Constructing $\boldsymbol{S}$ and $\boldsymbol{X}$ in the manner of $\boldsymbol{T}$, we find $\boldsymbol{T} = \boldsymbol{S} + \boldsymbol{X}$.

Defining $\boldsymbol{s} := [s(0),\dots,s(N-1)]^t$, $\boldsymbol{x} := [x(0),\dots,x(N-1)]^t$ and an $N \times N$ matrix $[\boldsymbol{H}]_{n,m} = h(n,\langle n-m\rangle_n)$, the input-output relation (2) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}. \tag{4}$$

Equation (2) will be convenient for MSE-optimal pattern design and (4) for capacity analysis.

2.2. Doubly-Selective Channel Model

The following BEM [4] describes the channel response over the block duration. For $n \in \{0,\dots,N-1\}$ and $\ell \in \{0,\dots,N_t-1\}$,

$$h(n,\ell) = N^{-\frac{1}{2}} \sum_{k=-(N_f-1)/2}^{(N_f-1)/2} \lambda(k,\ell) e^{j\frac{2\pi}{N}kn}, \tag{5}$$

where $\{\lambda(k,\ell)\}$ are zero-mean i.i.d. Gaussian with variance $\frac{N}{N_f N_t}$. The model (5) approximates wide-sense stationary uncorrelated scattering (WSSUS) with uniform PSD

$$S_{hh}(f) = \begin{cases} \frac{1}{2N_t f_{\mathrm{d}} T_s}, & |f| < f_{\mathrm{d}} T_s, \\ 0, & |f| \geq f_{\mathrm{d}} T_s, \end{cases} \tag{6}$$

where $f_{\mathrm{d}} T_s$ denotes the one-sided Doppler spread normalized to the symbol rate. The quantity $N_f := \lceil 2 f_{\mathrm{d}} T_s N \rceil + 1$ will be referred to as the channel's "frequency spread" and, for simplicity, assumed to be an odd integer. The quantity $\gamma := 2 f_{\mathrm{d}} T_s N_t$ will be referred to as the channel's "spreading index." Note that $\gamma \approx N_t N_f / N$. We assume an underspread channel, for which $\gamma < 1$.

Denoting the N-point unitary DFT matrix as $\boldsymbol{F}_N$, we define

$$\begin{aligned} \bar{\boldsymbol{F}} &:= \boldsymbol{F}_N^*(:, -\tfrac{N_f-1}{2} : \tfrac{N_f-1}{2}) \\ \boldsymbol{h}_\ell &:= [h(0,\ell),\dots,h(N-1,\ell)]^t \\ \boldsymbol{\lambda}_\ell &:= [\lambda(-\tfrac{N_f-1}{2},\ell),\dots,\lambda(\tfrac{N_f-1}{2},\ell)]^t, \end{aligned}$$

where modulo-N indexing is assumed. Notice that $\bar{\boldsymbol{F}}^H \bar{\boldsymbol{F}} = \boldsymbol{I}_{N_f}$. With these definitions, (5) becomes $\boldsymbol{h}_\ell = \bar{\boldsymbol{F}} \boldsymbol{\lambda}_\ell$. If we define

$$\begin{aligned} \boldsymbol{U} &:= \boldsymbol{I}_{N_t} \otimes \bar{\boldsymbol{F}} \\ \boldsymbol{h} &:= [\boldsymbol{h}_0^t \ \cdots \ \boldsymbol{h}_{N_t-1}^t]^t \\ \boldsymbol{\lambda} &:= [\boldsymbol{\lambda}_0^t \ \cdots \ \boldsymbol{\lambda}_{N_t-1}^t]^t, \end{aligned}$$

then

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

with $\boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I}_{N_f N_t}$ and $\boldsymbol{R}_\lambda := \mathrm{E}\{\boldsymbol{\lambda}\boldsymbol{\lambda}^H\} = \frac{N}{N_f N_t} \boldsymbol{I}_{N_f N_t}$.

3. MSE OPTIMAL PILOT/DATA PATTERNS

In this section, we review results from [7] on MSE-optimal pilot/data pattern design, including necessary/sufficient conditions and design examples.

3.1. MSE-Optimality Conditions

For the model of Section 2, (8)-(9) provide necessary and sufficient conditions on the pilot/data pattern such that the MSE incurred in Wiener estimation of $\boldsymbol{h}$, given $\boldsymbol{y}$ and $\boldsymbol{s}$, is minimal [7].

$$\forall \boldsymbol{X},\ (\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{X} \boldsymbol{U} = \boldsymbol{0}. \tag{8}$$

$$(\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{S}\boldsymbol{U} = \sigma_s^2 \boldsymbol{I}_{N_f N_t}. \tag{9}$$

Condition (8) says that the transmit signal should be constructed so that the pilot and data subspaces remain orthogonal at the channel output. Condition (9) says that pilot signal should be constructed so that the channel modes are independently excited with equal power. Let $\hat{h}(n,\ell)$ denote the Wiener estimate of $h(n,\ell)$ and $\tilde{h}(n,\ell) := h(n,\ell) - \hat{h}(n,\ell)$ denote the estimation error. When (8) and (9) are satisfied, the variances $\sigma_{\tilde{h}}^2 := E\{|\tilde{h}(n,\ell)|^2\}$ and $\sigma_{\hat{h}}^2 := E\{|\hat{h}(n,\ell)|^2\}$ are given by

$$\sigma_{\tilde{h}}^2 = \frac{N_f}{N} \left(\frac{N_f N_t}{N} + \frac{\sigma_s^2}{\sigma_v^2}\right)^{-1}, \tag{10}$$

$$\sigma_{\hat{h}}^2 = \frac{1}{N_t} - \frac{N_f}{N} \left(\frac{N_f N_t}{N} + \frac{\sigma_s^2}{\sigma_v^2}\right)^{-1} \tag{11}$$

for $n \in \{0,\cdots,N-1\},\ \ell \in \{0,\cdots,N_t-1\}$.

The conditions (8)-(9) can be rewritten in terms of $\{s(i)\}$ and $\{x(i)\}$ using the index sets $\mathcal{N}_t := \{-N_t+1, ..., N_t-1\}$ and $\mathcal{N}_f := \{-N_f+1, ..., N_f-1\}$.

Lemma 1 ([7]) *For N-block CP transmission over the doubly selective channel (5), the necessary and sufficient conditions for MSE-optimal PAT can be written as follows. $\forall k \in \mathcal{N}_t$, $\forall m \in \mathcal{N}_f$,*

$$\frac{1}{N} \sum_{i=0}^{N-1} s(i) s^*(i-k) e^{-j\frac{2\pi}{N}mi} = \sigma_s^2 \delta(k)\delta(m) \tag{12}$$

$$\sum_{i=0}^{N-1} x(i) s^*(i-k) e^{-j\frac{2\pi}{N}mi} = 0. \tag{13}$$

Given a pilot pattern that satisfies (12), data patterns that satisfy (13) can be constructed as follows. Defining

$$\begin{aligned} \boldsymbol{W}_k &:= \boldsymbol{F}_N(-N_f+1 : N_f-1, :) \boldsymbol{S}_k^H \\ \boldsymbol{W} &:= [\boldsymbol{W}_{-N_t+1}^t \cdots \boldsymbol{W}_{N_t-1}^t]^t. \end{aligned}$$

condition (13) becomes $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{0}$, implying that data must be transmitted in the nullspace of $\boldsymbol{W}$. Thus, we structure the data as

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{d}, \tag{14}$$

where the columns of $\boldsymbol{B} \in \mathbb{C}^{N \times N_d}$ form an orthonormal basis for $\mathrm{null}(\boldsymbol{W})$ and where $\boldsymbol{d}$ contains $N_d := \dim(\mathrm{null}(\boldsymbol{W}))$ data symbols.

The data dimension N_d can be bounded as follows. Note from (8) that the $N_f N_t$ rows of $(\boldsymbol{S}\boldsymbol{U})^H$ are contained within the $(2N_f-1)(2N_t-1)$ rows of $\boldsymbol{W}$. In order to satisfy (9), those rows must be orthogonal. Thus, $N_f N_t \leq \mathrm{rank}(\boldsymbol{W}) \leq (2N_f-1)(2N_t-1)$, which means that MSE-optimality implies

$$N - (2N_f-1)(2N_t-1) \ \leq N_d \leq \ N - N_f N_t.$$

3.2. Examples of MSE-Optimal Pilot/Data Patterns

Example MSE-optimal pilot/data patterns from [7] are given below in terms of their $(\boldsymbol{s}, \boldsymbol{B})$ parameterization, where $\boldsymbol{s}$ specifies the pilot sequence and $\boldsymbol{B}$ the data basis.

Example 1 (SCCP) *Assuming $\frac{N}{N_f} \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_t^{(i)}$ and the guard index set $\mathcal{G}_t^{(i)}$:*

$$\mathcal{P}_t^{(i)} := \{i, i+\tfrac{N}{N_f}, ..., i+\tfrac{(N_f-1)N}{N_f}\} \tag{15}$$

$$\mathcal{G}_t^{(i)} := \bigcup_{k \in \mathcal{P}_t^{(i)}} \{-N_t+1+k, ..., N_t-1+k\}. \tag{16}$$

An MSE-optimal PAT scheme is given by

$$s(q) = \begin{cases} \sigma_s \sqrt{\frac{N}{N_f}} e^{j\theta(q)} & q \in \mathcal{P}_t^{(i)} \\ 0 & q \notin \mathcal{P}_t^{(i)} \end{cases} \quad (17)$$

and by $\boldsymbol{B}$ constructed from the columns of $\boldsymbol{I}_N$ with indices in the set $\{0,\ldots,N-1\}\backslash\mathcal{G}_t^{(i)}$. Both $i \in \{0,\ldots,\frac{N}{N_f}-1\}$ and $\theta(q) \in \mathbb{R}$ are arbitrary. Here, $N_d = N - N_f(2N_t-1)$.

Example 2 (CP-OFDM) *Assuming $\frac{N}{N_t} \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_f^{(i)}$ and the guard index set $\mathcal{G}_f^{(i)}$:*

$$\mathcal{P}_f^{(i)} := \{i, i+\tfrac{N}{N_t}, ..., i+\tfrac{(N_t-1)N}{N_t}\} \quad (18)$$

$$\mathcal{G}_f^{(i)} := \bigcup_{k\in\mathcal{P}_f^{(i)}} \{-N_f+1+k, ..., N_f-1+k\} \quad (19)$$

An MSE-optimal PAT scheme is given by $\boldsymbol{s} = \boldsymbol{F}_N^H \boldsymbol{s}_f$ with

$$s_f(q) = \begin{cases} \sigma_s \sqrt{\frac{N}{N_t}} e^{j\theta(q)} & q \in \mathcal{P}_f^{(i)} \\ 0 & q \notin \mathcal{P}_f^{(i)} \end{cases} \quad (20)$$

and $\boldsymbol{B}$ constructed from the columns of the IDFT matrix $\boldsymbol{F}_N^H$ with indices in the set $\{0,\ldots,N-1\}\setminus\mathcal{G}_f^{(i)}$. Both $i \in \{0,\ldots,\frac{N}{N_t}-1\}$ and $\theta(q) \in \mathbb{R}$ are arbitrary. Here, $N_d = N - N_t(2N_f-1)$.

Example 3 (Chirps) *Assuming even N, an MSE- optimal PAT scheme is given by*

$$s(q) = \sigma_s e^{j\frac{2\pi}{N}\frac{N_f}{2}q^2} \quad (21)$$

$$[\boldsymbol{B}]_{q,p} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(p+N_fN_t)q} e^{j\frac{2\pi}{N}\frac{N_f}{2}q^2}, \quad (22)$$

for $q \in \{0,\ldots,N-1\}$ and $p \in \{0,\ldots,N_d-1\}$. Here, $N_d = N - 2N_fN_t + 1$.

Example 1 corresponds to a single carrier cyclic prefix (SCCP) system with uniformly spaced Kronecker-delta (KD) pilot bursts. This system was analyzed in [4]. Example 2 corresponds to a CP-OFDM system with uniformly spaced KD pilot tones. Example 3 corresponds to a cylic prefixed "chirp signaling" scheme.

Note that the three MSE-optimal examples above differ in their data subspace dimension N_d. Intuitively, a larger N_d should lead to a higher data rate. Notice that, among the three examples above, CP-OFDM yields the largest N_d when $N_t > N_f$, while SCCP yields the largest N_d when $N_f > N_t$. The capacity analysis in the sequel explores these issues further.

4. CAPACITY ANALYSIS

Here we present bounds on the ergodic capacity of MSE-optimal PAT for the system of Section 2. A pilot/data power allocation scheme based on maximization of the capacity lower bound is also presented. The three example schemes in Section 3.2 are then compared using these bounds. We find that SCCP dominates when the channel is primarily frequency-spreading, while CP-OFDM dominates when the channel is primarily time-spreading.

4.1. Capacity Bounds

Let $(\boldsymbol{s},\boldsymbol{B})$ denote an MSE-optimal system with pilot power σ_s^2 as in (3). Combining (4) with (14), the observation becomes

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{d} + \boldsymbol{v}. \quad (23)$$

In the sequel, we impose the data power constraint

$$\frac{1}{N}\sum_{n=0}^{N-1} E\{|x(n)|^2\} = \frac{1}{N}\sum_{n=0}^{N_d-1} E\{|d(n)|^2\} \le \sigma_d^2. \quad (24)$$

In addition, we assume that the channel fading coefficients $\boldsymbol{\lambda}$ are Gaussian and fade independently from block to block.

Lemma 2 *For the MMSE PAT system $(\boldsymbol{s},\boldsymbol{B})$ with i.i.d. Gaussian symbols $\boldsymbol{d} \in \mathbb{C}^{N_d}$, the ergodic capacity C_{mse} obeys $C_{\text{mse-lb}} \le C_{\text{mse}} \le C_{\text{mse-ub}}$, where*

$$C_{\text{mse-lb}} := \frac{1}{N} E\{\log\det[\boldsymbol{I} + \rho_l \boldsymbol{B}^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{B}]\} \quad (25)$$

$$C_{\text{mse-ub}} := \frac{1}{N} E\{\log\det[\boldsymbol{I} + \rho_u \boldsymbol{B}^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{B}]\} \quad (26)$$

$$\rho_l := \frac{N\sigma_d^2}{N_d\sigma_v^2}\left(\frac{\frac{N\sigma_s^2}{N_fN_t}}{\frac{N\sigma_d^2}{N_d}+\frac{N\sigma_s^2}{N_fN_t}+\sigma_v^2}\right) \quad (27)$$

$$\rho_u := \frac{N\sigma_d^2}{N_d\sigma_v^2} \quad (28)$$

A proof is given in [13]. In brief, the lower bound (25) describes the "worst case" scenario, where channel estimation error acts as AWGN. This concept was previously used in [8] and [9]. The upper bound (26) describes the "best case" scenario of perfect channel estimates. Recalling (3) and (24), the quantities $N\sigma_s^2/(N_fN_t)$ and $N\sigma_d^2/N_d$ can be interpreted as the average pilot and data energy per *essential* pilot and data subspace dimension, respectively. In (25) and (26), we ignore the overhead due to the prefix portion, in consistent with [12].

While a lower bound on the channel capacity for a subcase of Example 1 was presented in [4], the bounds here apply more generally to any MSE-optimal PAT scheme with pilot power σ_s^2 and data power σ_d^2 over the doubly selective channel of Section 2.2.

For reference, we consider the "ideal system" with perfect receiver channel state information (CSI) and transmitted power constraint σ_t^2. In this ideal setting, there is no need to transmit pilots. For i.i.d. Gaussian data, the ergodic capacity is [14]

$$C_{\text{ideal}} = \frac{1}{N} E\{\log\det[\boldsymbol{I} + \rho_i \boldsymbol{H}^H\boldsymbol{H}]\} \quad (29)$$

$$\rho_i := \frac{\sigma_t^2}{\sigma_v^2}. \quad (30)$$

Several factors contribute to the difference between C_{ideal} and C_{mse}. First, the PAT system suffers from channel estimation error, which contributes a form of interference that degrades C_{mse}. In comparing C_{ideal} to $C_{\text{mse-ub}}$, we remove the effect of channel estimation error. Note, from (28) and (30), that ρ_u and ρ_i describe the effective SNR *within the data subspaces* of PAT and the ideal system, respectively. In this sense, ρ_u and ρ_i are equivalent. The primary difference between C_{ideal} and $C_{\text{mse-ub}}$, then, is a result of the fact that MMSE PAT uses only N_d out of N total dimensions for data transmission. As N_fN_t increases, so does the difference between N_d and N, widening the gap between C_{ideal} and $C_{\text{mse-ub}}$, and hence between C_{ideal} and C_{mse}.

4.2. Pilot/Data Power Allocation

Section 3 described MSE-optimal pilot and data patterns that assume a fixed allocation of pilot power σ_s^2. Here we consider the problem of allocating a fixed transmit power $\sigma_t^2 = \sigma_s^2 + \sigma_d^2$ between pilots and data. Notice the inherent tradeoff. Allocating more power to pilots results in less channel estimation error but also results in less data power, which in turn increases the sensitivity to noise and estimation error. Roughly speaking, power should be allocated to maximize the *effective* SNR. We approach this problem through the maximization of $C_{\text{mse-lb}}$.

Let $\alpha \in [0,1]$ denote the fraction of power allocated to the data symbols, i.e., $\sigma_d^2 = \alpha\sigma_t^2$ and $\sigma_s^2 = (1-\alpha)\sigma_t^2$. We are interested in finding $\alpha_\star := \arg\max_\alpha C_{\text{mse-lb}}(\alpha)$. Because α affects $C_{\text{mse-lb}}$ only through the term ρ_l, and because $C_{\text{mse-lb}}$ is strictly increasing in ρ_l, it suffices to maximize ρ_l w.r.t. α. The value of $\alpha_\star$ is readily obtained by finding the value of α which sets $\partial\rho_l/\partial\alpha = 0$:

$$\alpha_\star \;=\; \begin{cases} \frac{b}{a} - \sqrt{(\frac{b}{a})^2 - \frac{b}{a}} & \text{if } a \neq 0 \\ \frac{1}{2} & \text{if } a = 0 \end{cases} \tag{31}$$

$$a := \frac{N\sigma_t^2}{N_f N_t} - \frac{N\sigma_t^2}{N_d},\; b \;:=\; \frac{N\sigma_t^2}{N_f N_t} + \sigma_v^2 \tag{32}$$

It can be verified that $\alpha_\star \in [0,1]$.

According to (31) and (32), when $N_f N_t = N_d$, power should be allocated equally to pilots and data. It is interesting to note that the authors in [9] came to a similar conclusion: when the number of time slots used for training is equal to that used for data transmission, then the training and data components should be given equal power.

Note that $\alpha_\star$ is invariant to $(\boldsymbol{s}, \boldsymbol{B})$. Thus, all MSE-optimal PAT schemes with the same data dimension N_d have $C_{\text{mse-lb}}$ maximized by the same pilot/data power allocation. While the $C_{\text{mse-lb}}$-maximizing power allocation for the SCCP scheme was derived previously in [4], the results in this paper hold more generally.

4.3. Comparison of MSE-optimal PAT Schemes

Although all MSE-optimal PAT schemes yield equivalent channel estimates, they differ in data subspace dimension N_d and hence in achievable rate. To understand the relative merits of these schemes, we would like closed-form expressions for C_{mse}. Unfortunately, the expectations in (25), (26) and (29) are difficult to evaluate. Instead, we compare approximations of the capacity bounds. Specifically, we use Jensen's inequality to upper bound $C_{\text{mse-lb}}$, $C_{\text{mse-ub}}$ and C_{ideal} by $\bar{C}_{\text{mse-lb}}$, $\bar{C}_{\text{mse-ub}}$ and $\bar{C}_{\text{ideal}}$, respectively. Using $E\{\log\det \boldsymbol{A}\} \leq \log\det E\{\boldsymbol{A}\}$ and the facts that $E\{\boldsymbol{H}^H\boldsymbol{H}\} = \boldsymbol{I}_N$ and $\boldsymbol{B}^H\boldsymbol{B} = \boldsymbol{I}_{N_d}$, we find

$$\bar{C}_{\text{mse-lb}} \;=\; \frac{N_d}{N}\log(1+\rho_l) \tag{33}$$

$$\bar{C}_{\text{mse-ub}} \;=\; \frac{N_d}{N}\log(1+\rho_u), \tag{34}$$

$$\bar{C}_{\text{ideal}} \;=\; \log(1+\rho_i). \tag{35}$$

We now show that $\bar{C}_{\text{mse-ub}}$ is increasing in N_d for any power allocation $\alpha \in [0,1]$. Treating N_d as a continuous parameter,

$$\frac{\partial \bar{C}_{\text{mse-ub}}}{\partial N_d} \;=\; \frac{1}{N}\left(\log(1+\rho_u) + \frac{N_d}{1+\rho_u}\frac{(-1)N\sigma_d^2}{N_d^2\sigma_d^2}\right) \tag{36}$$

$$= \frac{1}{N}\left(\log(1+\rho_u) - \frac{\rho_u}{1+\rho_u}\right) \;\geq\; 0, \tag{37}$$

since $\log(1+x) - \frac{x}{1+x}$ is increasing in x for $x \geq 0$. Using a similar approach, it can be shown that $\bar{C}_{\text{mse-lb}}$ is also increasing in N_d. These results provide evidence that, among MSE-optimal PAT schemes, those with larger N_d achieve higher rate. Similar conclusions were reached for MIMO flat-fading channels in [9].

Taking a closer look at these upper bounds, we see that, in the high-SNR regime, both $\bar{C}_{\text{mse-lb}}$ and $\bar{C}_{\text{mse-ub}}$ increase linearly with $\log(\text{SNR})$ at slope N_d/N. In contrast, $\bar{C}_{\text{ideal}}$ increases linearly with $\log(\text{SNR})$ at slope 1. For channels with low spreading index (i.e., $\gamma \ll 1$), N_d/N will be close to 1, and hence $\bar{C}_{\text{mse-lb}}$ and $\bar{C}_{\text{mse-ub}}$ will both have slopes close to that of $\bar{C}_{\text{ideal}}$. Thus, for small spreading index, the capacity achieved by MSE-optimal PAT is comparable to that of the ideal perfect-receiver-CSI system. But, for channels with significant spreading, $\gamma \approx N_f N_t/N$ will be significantly greater than 0 and hence N_d/N will be significantly less than 1. In this case, C_{mse} will deviate significantly from C_{ideal} at high SNR. These trends are confirmed by the numerical results in Section 4.5.

4.4. SCCP versus CP-OFDM

The previous section showed that, among MSE-optimal PAT schemes, those with higher N_d yield higher rate. For a given pair (N_t, N_f), each of the examples in Section 3.2 supported a different value of N_d. Specifically, SCCP maximizes N_d when $N_f > N_t$ and CP-OFDM maximizes N_d when $N_f < N_t$. Apart from the trivial case $N_f = N_t = 1$, the chirp signaling in Example 3 is dominated by both SCCP and CP-OFDM. When $N_t = N_f$, it can be proven that the capacity lower bounds for SCCP and CP-OFDM coincide, as do the capacity upper bounds [13].

4.5. Numerical Results

Here we numerically investigate the capacity bounds from Section 4.1 for the SCCP and CP-OFDM schemes in Examples 1 and 2 using the power allocation (31). In all cases, we consider block size $N = 128$ and plot the bounds as a function of SNR $:= \sigma_t^2/\sigma_v^2$.

Figure 1 plots $C_{\text{mse-lb}}$ and $C_{\text{mse-ub}}$ for the MSE-optimal PAT assuming a channel with $N_t = 2 = N_f = 2$. These channel parameters correspond to a spreading index of $\gamma \approx 0.03$. Recall that, when $N_f = N_t$, both the SCCP and CP-OFDM schemes support the same value of N_d. From the bounds and C_{ideal}, we conclude that the capacity of C_{mse} is close to that of the "ideal" data-only system with perfect receiver CSI. We also note that the capacity curves are approximately linear in $\log(\text{SNR})$, as suggested in Section 4.3.

Figure 2 repeats the investigation in Fig. 1 but with $N_t = N_f = 4$, corresponding to $\gamma \approx 0.12$. Compared to Fig. 1, there is a more significant difference between the ideal system and the MSE-optimal PAT systems at high SNR. Note that the "slope" arguments, given in Section 4.3 for the approximations $\bar{C}_{\text{mse-lb}}$, $\bar{C}_{\text{mse-ub}}$ and $\bar{C}_{\text{ideal}}$, fit the actual curves $C_{\text{mse-lb}}$, $C_{\text{mse-ub}}$ and C_{ideal} quite well.

Figure 3 investigates the case where $N_t = 16$ and $N_f = 2$, i.e., where the channel is primarily time-spreading. In this case, the CP-OFDM system shows SNR gains of several dB over the SCCP system. This corresponds to a channel for the scenario with the carrier frequency $f_c = 20$GHz, system bandwidth 500kHz, mobile speed 100 km/hour and channel delay spread 32μs.

5. CONCLUSIONS

This paper presented a capacity analysis of MSE-optimal PAT schemes. The transmission model covered a wide range of linear modulation schemes and supported superimposed pilot/data patterns. The doubly dispersive channel was modeled via standard basis expansion. Lower and upper bounds on the ergodic capacity were derived for MSE-optimal PAT, and the lower bound was maximized as a means of allocating power between pilots and data. The capacity bounds were used to compare two MSE-optimal PAT schemes based on SCCP and CP-OFDM. The CP-OFDM scheme was found to acheive higher rate when the channel is primarily time-dispersive, while the SCCP acheives higher rate when the channel is primarily frequency-dispersive. A numerical evaluation of the capacity bounds confirmed the theoretical analysis.

6. REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov. 2004.

[2] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," in *IEEE Trans. on Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.

[3] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.

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

[5] M. Biguesh and A. B. Gershman, "MIMO channel estimation: optimal training and tradeoffs between estimation techniques," in *Proc. IEEE Intern. Conf. on Communication*, vol. 5, pp. 2658–2662, June 2004.

[6] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmission over time-varying flat fading channels," *IEEE Trans. on Signal Processing*, May 2004.

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

[8] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. on Information Theory*, vol. 46, pp. 933–946, May 2000.

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

[10] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Trans. on Information Theory*, vol. 48, no. 8, pp. 2338–2353, 2002.

[11] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 2138–2145, Sept. 2004.

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

[13] A. P. Kannu and P. Schniter, "On pilot aided transmission in doubly dispersive channels," manuscript in preparation.

[14] I. E. Teletar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.

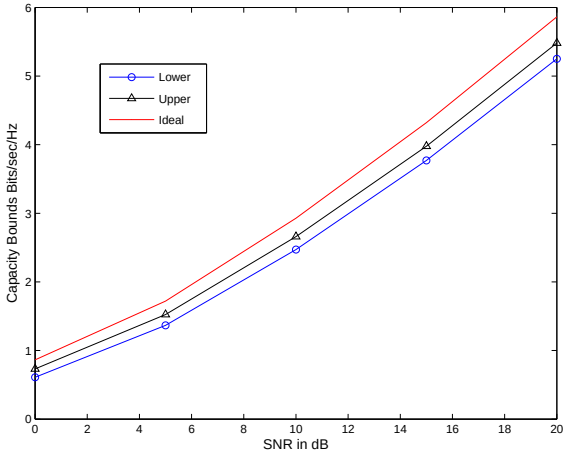

Fig. 1. Bounds on Ergodic Channel Capacity $N_t = 2$, $N_f = 2$.

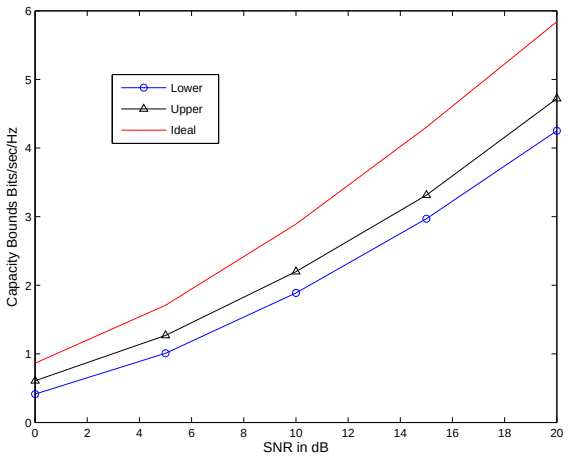

Fig. 2. Bounds on Ergodic Channel Capacity $N_t = 4$, $N_f = 4$.

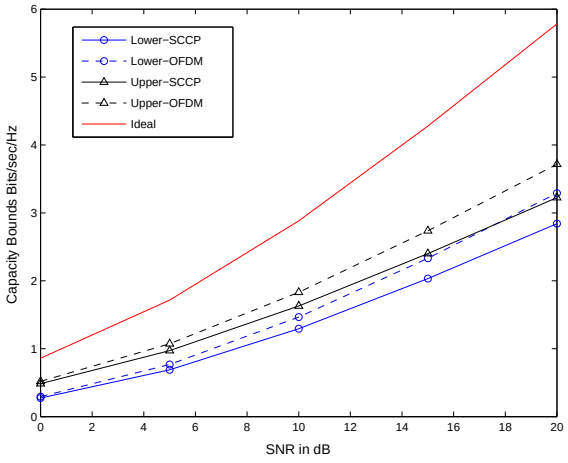

Fig. 3. Bounds on Ergodic Channel Capacity $N_t = 16$, $N_f = 2$.