MSE-OPTIMAL TRAINING FOR LINEAR TIME-VARYING 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

We consider pilot-aided transmission (PAT) for a general class of systems encompassing linear modulation and a linear time-varying channel. For these systems, and given a pilot energy constraint, we derive a tight lower bound on the mean squared error (MSE) of pilot-aided channel estimates as well as necessary and sufficient conditions on PAT to attain this bound. We then apply these results to the design of single-antenna PAT for doubly selective channels and arrive at novel MSE-optimal PAT schemes. In this application, we assume a block-based cyclic-prefix PAT and a basis expansion model for the channel.

1. INTRODUCTION

The wireless communication channel is typically modeled as linear transformation and parameterized by a set of time-varying coefficients. Often, the receiver estimates these coefficients for subsequent use in data detection, so that high-quality channel estimates are desired. In the pilot-aided approach to channel estimation, a known pilot (or "training") sequence is embedded in the otherwise unknown transmitted sequence.

Tong, Sadler, and Dong published a recent overview of pilot-aided transmission (PAT) [1]. They argued that the PAT design problem can be separated into two sub-problems: pilot pattern design and pilot/data power allocation. In this work we target the first subproblem, i.e., pilot pattern design given a fixed pilot power allocation. Previous work on pilot pattern design (see, e.g., [1]) assumed a specific modulation type and either non-overlapping pilot/data or persistent data with superimposed pilots.

We follow a different approach. First, we consider a general linear modulation scheme (e.g., single-carrier, multicarrier, code-multiplexed) with data and pilot patterns that may or may not overlap. Second, we consider a general linear time-varying channel based on zero-mean random parameters with arbitrary correlation structure. For this class of systems, and for a constraint on the pilot power, we derive an expression for the minimum mean-squared error (MSE) of pilot-aided channel estimates as well as necessary and sufficient conditions on the pilot/data pattern to attain this minimum MSE. Applying these conditions to the single-antenna doubly-selective channel (DSC) using a basis expansion model (BEM), we outline a procedure for MSE-optimal PAT design that yields novel pilot/data patterns. We also uncover an inherent duality between time- and frequency-domain PAT systems.

It should be noted that several authors (e.g., [2–4]) have established close connections between the capacity and MSE criteria for pilot pattern design. Though these connections apply to our work as well, these issues (as well as pilot/data power allocation) are treated elsewhere for reasons of space. The paper is organized as follows. In Section II, we derive the MSE lower bound and achievability conditions. In Section III, we apply these results to the DSC. In Section IV, we conclude.

Supported in part by NSF CAREER Grant CCR-0237037.

2. MSE-OPTIMAL PILOT-AIDED TRANSMISSION

In this section, we derive a lower bound on the MSE of pilot-aided channel estimates assuming linear modulation, a linear time-varying channel, and constrained pilot power. We also establish necessary and sufficient conditions to achieve this bound.

2.1. System Model

For linear modulation and a linear time-varying channel, the received complex-baseband vector $\boldsymbol{y} \in \mathbb{C}^N$ can be written

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{h} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{T} \in \mathbb{C}^{N \times G}$ contains transmitted symbols, $\boldsymbol{h} \in \mathbb{C}^G$ channel coefficients, and $\boldsymbol{v} \in \mathbb{C}^N$ zero-mean circular white Gaussian noise (CWGN) with variance σ_v^2. $\boldsymbol{T}$ is formed by superimposing pilots $\boldsymbol{S}$ and unknown data $\boldsymbol{X}$:

$$\boldsymbol{T} = \boldsymbol{S} + \boldsymbol{X}. \tag{2}$$

We assume zero-mean data, so that $\boldsymbol{S} = E\{\boldsymbol{T}\}$, and

$$\boldsymbol{h} = \boldsymbol{U}\boldsymbol{\lambda}, \tag{3}$$

where $\boldsymbol{\lambda} \in \mathbb{C}^M$ is zero-mean Gaussian with $\boldsymbol{R}_\lambda = E\{\boldsymbol{\lambda}\boldsymbol{\lambda}^H\} = \operatorname{diag}(\sigma_{\lambda_0}^2 \cdots \sigma_{\lambda_{M-1}}^2) > 0$ and $\boldsymbol{U}$ is fixed with $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_M$. Finally, we assume that $\boldsymbol{v}$, $\boldsymbol{X}$, and $\boldsymbol{\lambda}$ are uncorrelated.

2.2. MSE lower bound

Here we derive an MSE lower bound for estimation of $\boldsymbol{h}$ given knowledge of $\{\boldsymbol{y}, \boldsymbol{S}\}$, statistical knowledge of $\{\boldsymbol{h}, \boldsymbol{X}, \boldsymbol{v}\}$, and pilot energy constraint $\|\boldsymbol{S}\|_F^2 \leq P'$. We begin by taking SVDs, $\boldsymbol{S}\boldsymbol{U} = \boldsymbol{V}_{\mathsf{s}}\boldsymbol{\Sigma}_{\mathsf{s}}\boldsymbol{Q}_{\mathsf{s}}^H$ and $\boldsymbol{X}\boldsymbol{U} = \boldsymbol{V}_{\mathsf{x}}\boldsymbol{\Sigma}_{\mathsf{x}}\boldsymbol{Q}_{\mathsf{x}}^H$, where $\boldsymbol{\Sigma}_{\mathsf{s}}$ and $\boldsymbol{\Sigma}_{\mathsf{x}}$ are diagonal and full-rank. Let $K \leq M$ denote the rank of $\boldsymbol{\Sigma}_{\mathsf{s}}$. Defining $\boldsymbol{z} := \boldsymbol{V}_{\mathsf{s}}^H\boldsymbol{y}$ and using (3), we have

$$\boldsymbol{z} = \underbrace{\boldsymbol{\Sigma}_{\mathsf{s}}\boldsymbol{Q}_{\mathsf{s}}^H}_{\boldsymbol{A}_{\mathsf{s}}}\boldsymbol{\lambda} + \underbrace{\boldsymbol{V}_{\mathsf{s}}^H\boldsymbol{V}_{\mathsf{x}}\boldsymbol{\Sigma}_{\mathsf{x}}\boldsymbol{Q}_{\mathsf{x}}^H}_{\boldsymbol{A}_{\mathsf{x}}}\boldsymbol{\lambda} + \underbrace{\boldsymbol{V}_{\mathsf{s}}^H\boldsymbol{v}}_{\boldsymbol{n}}. \tag{4}$$

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

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{R}_{z,\lambda}^H \boldsymbol{R}_z^{-1} \boldsymbol{z}, \tag{5}$$

$$\boldsymbol{R}_{z,\lambda} = \boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda + \underbrace{E\{\boldsymbol{A}_{\mathsf{x}}\}}_{\mathbf{0}} \boldsymbol{R}_\lambda + \underbrace{E\{\boldsymbol{n}\boldsymbol{\lambda}^H\}}_{\mathbf{0}}, \tag{6}$$

$$\boldsymbol{R}_z = \underbrace{\boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda \boldsymbol{A}_{\mathsf{s}}^H + \sigma_v^2 \boldsymbol{I}_K}_{\boldsymbol{\Delta}} + \underbrace{E\{\boldsymbol{A}_{\mathsf{x}} \boldsymbol{R}_\lambda \boldsymbol{A}_{\mathsf{x}}^H\}}_{\boldsymbol{U}_{\mathsf{x}} \boldsymbol{\Lambda}_{\mathsf{x}} \boldsymbol{U}_{\mathsf{x}}^H}, \tag{7}$$

with diagonal $\boldsymbol{\Lambda}_{\mathsf{x}} \geq 0$ and $\boldsymbol{U}_{\mathsf{x}}^H \boldsymbol{U}_{\mathsf{x}} = \boldsymbol{I}$. Note that the MMSE estimate of $\boldsymbol{h}$ is $\hat{\boldsymbol{h}} = \boldsymbol{U}\hat{\boldsymbol{\lambda}}$ and that $\sigma_e^2 := E\{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2\} = E\{\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2\}$. The energy constraint on $\boldsymbol{S}$ implies

$$\mathrm{tr}\{(\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{S}\boldsymbol{U}\} = \mathrm{tr}(\boldsymbol{A}_{\mathsf{s}}^H \boldsymbol{A}_{\mathsf{s}}) \leq P, \tag{8}$$

for some P, where the relationship of P to P' depends on the structure of $\boldsymbol{S}$ and $\boldsymbol{U}$. Given constraint (8), a tight lower bound on σ_e^2, as well as necessary and sufficient conditions to achieve this bound, are stated in Theorem 1.

Theorem 1 (MSE Lower Bound).

$$\sigma_e^2 \geq \sum_{m=0}^{M-1} \left(\frac{1}{\sigma_{\lambda_m}^2} + \frac{\alpha_m^{opt}}{\sigma_v^2} \right)^{-1}, \tag{9}$$

$$\alpha_m^{opt} = \left[\gamma - \frac{\sigma_v^2}{\sigma_{\lambda_m}^2} \right]^+, \tag{10}$$

where $[x]^+ := \max(0, x)$ *and* $\gamma \in \mathbb{R}$ *satisfies*

$$\sum_{m=0}^{M-1} \left[\gamma - \frac{\sigma_v^2}{\sigma_{\lambda_m}^2} \right]^+ = P. \tag{11}$$

Equality in (9) occurs if and only if (12)-(13) hold:

$$\forall \boldsymbol{X},\ (\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{X}\boldsymbol{U} = \mathbf{0}. \tag{12}$$

$$(\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{S}\boldsymbol{U} = \mathrm{diag}(\alpha_0^{opt}, ..., \alpha_{M-1}^{opt}). \tag{13}$$

Proof. For the estimator (5) we have

$$\begin{aligned}
\sigma_e^2 &= \mathrm{tr}\{\boldsymbol{R}_\lambda - \boldsymbol{R}_{z,\lambda}^H \boldsymbol{R}_z^{-1} \boldsymbol{R}_{z,\lambda}\} \\
&= \mathrm{tr}\{\boldsymbol{R}_\lambda - \boldsymbol{R}_\lambda^H \boldsymbol{A}_{\mathsf{s}}^H (\boldsymbol{\Delta} + \boldsymbol{U}_{\mathsf{x}} \boldsymbol{\Lambda}_{\mathsf{x}} \boldsymbol{U}_{\mathsf{x}}^H)^{-1} \boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda\} \\
&= \mathrm{tr}\{\boldsymbol{R}_\lambda - \boldsymbol{R}_\lambda^H \boldsymbol{A}_{\mathsf{s}}^H [\boldsymbol{\Delta}^{-1} - \boldsymbol{\Delta}^{-1} \boldsymbol{U}_{\mathsf{x}} (\boldsymbol{\Lambda}_{\mathsf{x}}^{-1} \\
&\quad + \boldsymbol{U}_{\mathsf{x}}^H \boldsymbol{\Delta}^{-1} \boldsymbol{U}_{\mathsf{x}})^{-1} \boldsymbol{U}_{\mathsf{x}}^H \boldsymbol{\Delta}^{-1}] \boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda\}, \qquad (14) \\
&\geq \mathrm{tr}\{\boldsymbol{R}_\lambda - \boldsymbol{R}_\lambda^H \boldsymbol{A}_{\mathsf{s}}^H \boldsymbol{\Delta}^{-1} \boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda\}, \qquad (15)
\end{aligned}$$

where we used the matrix inversion lemma in (14). The inequality (15) follows since $\boldsymbol{\Delta} > 0$ and $\boldsymbol{\Lambda}_{\mathsf{x}} \geq 0$. Since $\boldsymbol{\Sigma}_{\mathsf{s}}$ is full rank, $\boldsymbol{A}_{\mathsf{s}}$ has full row rank, and so equality in (15) is achieved if and only if

$$\boldsymbol{U}_{\mathsf{x}} \boldsymbol{\Lambda}_{\mathsf{x}} \boldsymbol{U}_{\mathsf{x}}^H = \mathbf{0} \Leftrightarrow E\{\boldsymbol{A}_{\mathsf{x}} \boldsymbol{R}_\lambda \boldsymbol{A}_{\mathsf{x}}^H\} = \mathbf{0}. \tag{16}$$

Since $\boldsymbol{R}_\lambda > 0$, (16) is satisfied if and only if $\boldsymbol{A}_{\mathsf{x}} = \mathbf{0}$, which is equivalent to (12), since $\boldsymbol{\Sigma}_{\mathsf{s}}$ and $\boldsymbol{\Sigma}_{\mathsf{x}}$ are full rank square matrices. We proceed further assuming that (12) is satisfied. With $\boldsymbol{A}_{\mathsf{x}} = \mathbf{0}$,

$$\begin{aligned}
\sigma_e^2 &= \mathrm{tr}\{\boldsymbol{R}_\lambda - \boldsymbol{R}_\lambda^H \boldsymbol{A}_{\mathsf{s}}^H (\boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda \boldsymbol{A}_{\mathsf{s}} + \sigma_v^2 \boldsymbol{I}_K)^{-1} \boldsymbol{A}_{\mathsf{s}} \boldsymbol{R}_\lambda\}, \\
&= \mathrm{tr}\{(\boldsymbol{R}_\lambda^{-1} + \frac{1}{\sigma_v^2} \boldsymbol{A}_{\mathsf{s}}^H \boldsymbol{A}_{\mathsf{s}})^{-1}\} \qquad (17)
\end{aligned}$$

using the matrix inversion lemma. Diagonal $\boldsymbol{R}_\lambda$ implies

$$\sigma_e^2 \geq \sum_{m=0}^{M-1} \left(\frac{1}{\sigma_{\lambda_m}^2} + \frac{\alpha_m}{\sigma_v^2} \right)^{-1}, \tag{18}$$

where $\alpha_m = [\boldsymbol{A}_{\mathsf{s}}^H \boldsymbol{A}_{\mathsf{s}}]_{m,m}$. Equality in (18) is achieved if and only if $\boldsymbol{A}_{\mathsf{s}}^H \boldsymbol{A}_{\mathsf{s}} = (\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{S}\boldsymbol{U}$ is diagonal. To find the lower bound on MSE, we minimize the right hand side of (18) with respect to $\{\alpha_m\}$ given the constraints (8) and $\alpha_m > 0\ \forall m$. The method of Lagrange multipliers yields the optimal $\{\alpha_m\}$ given by (10)-(11). □

The MSE-optimality condition (12) says that pilots and data should be multiplexed in a way that preserves orthogonality at the channel output. Condition (13) says that pilot signal should be constructed so that the channel modes are independently excited with energies specified by the water-filling expression (10).

Corollary 1. *If* $\boldsymbol{R}_\lambda = \sigma_\lambda^2 \boldsymbol{I}_M$, *then* $\alpha_m^{opt} = \frac{P}{M}\ \forall m$ *and*

$$\sigma_e^2 \geq M \left(\frac{1}{\sigma_\lambda^2} + \frac{P}{\sigma_v^2 M} \right)^{-1}, \tag{19}$$

with equality if and only if $(\boldsymbol{S}\boldsymbol{U})^H \boldsymbol{S}\boldsymbol{U} = \frac{P}{M} \boldsymbol{I}_M$ *and (12) holds.*

3. PAT FOR THE DOUBLY SELECTIVE CHANNEL

Using results of Section 2.2, we now outline a procedure for designing MSE-optimal pilot and data patterns for block transmission over single-antenna doubly-selective channels (DSCs) that fit a basis expansion model (BEM). For these channels, we show an inherent duality between time- and frequency-domain PAT.

3.1. Cyclic-Prefix Block Transmission Model

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

$$y(n) = \sum_{\ell=0}^{N_t-1} h(n, \ell) t(n - \ell) + v(n), \tag{20}$$

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 $n - \ell$. The time spread of the channel is N_t. Furthermore, we assume a length-N block transmission $\{t(n)\}_{n=0}^{N-1}$ prepended with a length $N_t - 1$ cyclic prefix (CP). By considering arbitrarily large N, the CP overhead becomes insignificant. We form the received vector $\boldsymbol{y} := [y(0), \ldots, y(N-1)]^t$ by discarding the CP contribution. To fit the model (1), we set $\boldsymbol{T} := [\boldsymbol{T}_0 \cdots \boldsymbol{T}_{-N_t+1}]$ with $\boldsymbol{T}_{-i} := \mathrm{diag}(t(-i), ..., t(-i + N - 1))$, $\boldsymbol{h} := [\boldsymbol{h}_0^t \cdots \boldsymbol{h}_{N_t-1}^t]^t$ with $\boldsymbol{h}_i := [h(0, i), \ldots, h(N-1, i)]^t$, and $\boldsymbol{v} := [v(0), \ldots, v(N-1)]^t$.

The transmit signal $\{t(i)\}$ is composed of pilot portion $s(i) = E\{t(i)\}$ and zero-mean data portion $x(i) = t(i) - s(i)$. We employ a pilot power constraint of the form

$$\frac{1}{N} \sum_{n=0}^{N-1} |s(n)|^2 \leq \sigma_s^2. \tag{21}$$

With $\boldsymbol{S}_i := E\{\boldsymbol{T}_i\}$, $\boldsymbol{X}_i := \boldsymbol{T}_i - \boldsymbol{S}_i$, $\boldsymbol{S} := [\boldsymbol{S}_0 \cdots \boldsymbol{S}_{-N_t+1}]$, and $\boldsymbol{X} := [\boldsymbol{X}_0 \cdots \boldsymbol{X}_{-N_t+1}]$, we fit the model (2). In the sequel we use $\boldsymbol{s} := [s(0), \ldots, s(N-1)]^t$ and $\boldsymbol{x} := [x(0), \ldots, x(N-1)]^t$.

3.2. Doubly-Selective Channel Model

The following BEM (see, e.g., [4]) characterizes the DSC response over the block duration:

$$h(n,\ell) = N^{-\frac{1}{2}} \sum_{k=-(N_f-1)/2}^{(N_f-1)/2} \lambda(k,\ell) e^{j\frac{2\pi}{N}kn}, \tag{22}$$

where $n \in \{0,\ldots,N-1\}$ and $\ell \in \{0,\ldots,N_t-1\}$. In (22), $\lambda(k,\ell)$'s are zero-mean uncorrelated Gaussian with variance $\frac{N}{N_f N_t}$. This model approximates wide-sense stationary uncorrelated scattering (WSSUS) with uniform PSD

$$S_{hh}(f) = \begin{cases} \frac{1}{2N_t f_{\mathsf{d}} T_s}, & |f| < f_{\mathsf{d}} T_s, \\ 0, & |f| \geq f_{\mathsf{d}} T_s, \end{cases} \tag{23}$$

where $f_{\mathsf{d}} T_s$ is the one-sided Doppler spread normalized to the symbol rate and where $N_f := \lfloor 2 f_{\mathsf{d}} T_s N \rfloor$. We refer to N_f as the frequency spread of the channel, and assumed it to be an odd integer. Also, we assume $2 f_{\mathsf{d}} T_s N_t < 1$, so that the channel is underspread.

Denoting the N-point unitary DFT matrix by $\boldsymbol{F}_N$, we rewrite (22) as $\boldsymbol{h}_\ell = \bar{\boldsymbol{F}} \boldsymbol{\lambda}_\ell$ with $\bar{\boldsymbol{F}} := \boldsymbol{F}_N^*(:, -\frac{N_f-1}{2} : \frac{N_f-1}{2})$, $\boldsymbol{h}_\ell := [h(0,\ell),\ldots,h(N-1,\ell)]^t$, and $\boldsymbol{\lambda}_\ell := [\lambda(-\frac{N_f-1}{2},\ell),\ldots,\lambda(\frac{N_f-1}{2},\ell)]^t$. Notice that $\bar{\boldsymbol{F}}^H \bar{\boldsymbol{F}} = \boldsymbol{I}_{N_f}$. If we define

$$\left.\begin{aligned} \boldsymbol{U} &:= \boldsymbol{I}_{N_t} \otimes \bar{\boldsymbol{F}} \\ \boldsymbol{h} &:= [\boldsymbol{h}_0^t \ \cdots \ \boldsymbol{h}_{N_t-1}^t]^t \\ \boldsymbol{\lambda} &:= [\boldsymbol{\lambda}_0^t \ \cdots \ \boldsymbol{\lambda}_{N_t-1}^t]^t \end{aligned}\right\} \tag{24}$$

then $\boldsymbol{h} = \boldsymbol{U}\boldsymbol{\lambda}$ with $\boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I}_{N_f N_t}$ and $\boldsymbol{R}_\lambda = \frac{N}{N_f N_t} \boldsymbol{I}_{N_f N_t}$, which is compatible with the channel model in Section 2.1.

The transmitted pilot-power constraint (21) yields limits on the received pilot-power as in (8). Since $\boldsymbol{SU} = [\boldsymbol{S}_0 \bar{\boldsymbol{F}} \cdots \boldsymbol{S}_{-N_t+1} \bar{\boldsymbol{F}}]$, $\boldsymbol{S}_{-i}^H \boldsymbol{S}_{-i}$ is diagonal, and all diagonal elements of $\bar{\boldsymbol{F}} \bar{\boldsymbol{F}}^H$ equal $\frac{N_f}{N}$, we find

$$\begin{aligned} \operatorname{tr}\{(\boldsymbol{SU})^H \boldsymbol{SU}\} &= \sum_{i=0}^{N_t-1} \operatorname{tr}\{\bar{\boldsymbol{F}} \bar{\boldsymbol{F}}^H \boldsymbol{S}_{-i}^H \boldsymbol{S}_{-i}\} \\ &= \frac{N_f}{N} \sum_{i=0}^{N_t-1} \operatorname{tr}\{\boldsymbol{S}_{-i}^H \boldsymbol{S}_{-i}\} = N_f N_t \sigma_s^2. \end{aligned}$$

Thus, to match (8), we set $P = N_f N_t \sigma_s^2$.

3.3. MSE-Optimal Cyclic-prefix PAT for the DSC

We now state the MSE-optimality requirements on pilot/data pattern for the block-transmission model in Section 3.1 and the DSC model in Section 3.2. We will use the index sets $\mathcal{N}_t := \{-N_t+1, ..., N_t-1\}$ and $\mathcal{N}_f := \{-N_f+1, ..., N_f-1\}$.

Lemma 1. *For N-block CP transmission over the doubly selective channel (22), the necessary and sufficient conditions for MSE-optimal PAT can be written as follows.* $\forall k \in \mathcal{N}_t,\ \forall m \in \mathcal{N}_f$,

$$\frac{1}{N} \sum_{i=0}^{N-1} s(i) s^*(i-k) e^{-j\frac{2\pi}{N}mi} = \sigma_s^2 \delta(k)\delta(m) \tag{25}$$

$$\sum_{i=0}^{N-1} x(i) s^*(i-k) e^{-j\frac{2\pi}{N}mi} = 0. \tag{26}$$

Proof. According to Corollary 1, we require

$$(\boldsymbol{SU})^H \boldsymbol{SU} = \sigma_s^2 \boldsymbol{I}_{N_f N_t} \tag{27}$$

and (12). Notice that $(\boldsymbol{SU})^H \boldsymbol{SU}$ is composed of $N_f \times N_f$ blocks $\bar{\boldsymbol{S}}_{k_2,k_1} := \bar{\boldsymbol{F}}^H \boldsymbol{S}_{-k_2}^H \boldsymbol{S}_{-k_1} \bar{\boldsymbol{F}}$ for $k_1, k_2 \in \{0,\ldots,N_t-1\}$. For these k_1, k_2 and for $m_1, m_2 \in \{0,\ldots,N_f-1\}$, (27) becomes

$$[\bar{\boldsymbol{S}}_{k_2,k_1}]_{m_1,m_2} = \sigma_s^2 \delta(k_1-k_2)\delta(m_1-m_2). \tag{28}$$

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

$$[\bar{\boldsymbol{S}}_{k_2,k_1}]_{m_1,m_2} = \frac{1}{N} \sum_{i=0}^{N-1} s(i-k_1) s^*(i-k_2) e^{-j\frac{2\pi}{N}(m_1-m_2)i} \tag{29}$$

Setting $k := k_2 - k_1$ and $m := m_1 - m_2$, so that $k \in \mathcal{N}_t$ and $m \in \mathcal{N}_f$, (29) becomes

$$\begin{aligned} [\bar{\boldsymbol{S}}_{k_2,k_1}]_{m_1,m_2} &= \frac{1}{N} \sum_{q=-k_1}^{N-1-k_1} s(q) s^*(q-k) e^{-j\frac{2\pi}{N}m(q+k_1)}, \\ &= \frac{e^{-j\frac{2\pi}{N}mk_1}}{N} \sum_{q=0}^{N-1} s(q) s^*(q-k) e^{-j\frac{2\pi}{N}mq} \end{aligned} \tag{30}$$

where in (30) we exploited the fact that $s(-q) = s(N-q)$ for $1 \leq q < N_t$. Combining (28) and (30), we obtain (25). The equivalence of (12) and (26) can be shown similarly. □

Using the quantities $s_f(i) := \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} s(q) e^{j\frac{2\pi}{N}qi}$ and $x_f(i) := \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} x(q) e^{j\frac{2\pi}{N}qi}$, Lemma 1 can be easily translated to the frequency domain.

Corollary 2. *For N-block CP transmission over doubly-selective channel (22), the necessary and sufficient conditions for MSE-optimal PAT can be written as follows.* $\forall k \in \mathcal{N}_t,\ \forall m \in \mathcal{N}_f$,

$$\frac{1}{N} \sum_{i=0}^{N-1} s_f(i) s_f^*(i-m) e^{-j\frac{2\pi}{N}ki} = \sigma_s^2 \delta(k)\delta(m) \tag{31}$$

$$\sum_{i=0}^{N-1} x_f(i) s_f^*(i-m) e^{-j\frac{2\pi}{N}ki} = 0. \tag{32}$$

3.4. Examples of MSE-Optimal PAT for the DSC

The pilot/data patterns specified by Lemma 1 are not unique. A PAT design procedure is described in brief below, followed by several examples including single-carrier cyclic prefix (SCCP) and cyclic-prefix OFDM. For more details, see [5].

The "time-domain Kronecker delta" (TDKD) family of pilot patterns follows from the choice $\boldsymbol{s} = \boldsymbol{b} \otimes [1\,0\,\cdots\,0]^t$, for $\boldsymbol{b} \in \mathbb{C}^L$ such that $L := \frac{N}{N_t} \in \mathbb{Z}$ and

$$N\sigma_s^2 \delta(m) = \sum_{i=0}^{L-1} |b(i)|^2 e^{-j\frac{2\pi}{L}mi} \ \ \forall m \in \mathcal{N}_f. \tag{33}$$

If $L < N_f$, no solution to (33) exists. If $N_f \leq L < 2N_f$, the elements in $\boldsymbol{b}$ must have equal magnitude. When $L \geq N_f$, however, the design of $\boldsymbol{b}$ is less constrained. (See [5].) Another family of pilot patterns—the "frequency-domain Kronecker delta" (FDKD) family—results from setting $\boldsymbol{s}_f = \boldsymbol{b}_f \otimes [1\,0\,\cdots\,0]^t$ with

$\boldsymbol{b}_f \in \mathbb{C}^{L'}$ and $L' := \frac{N}{N_f} \in \mathbb{Z}$. A third family of MSE-optimal pilot patterns can be constructed from linear chirp sequences.

Given a pilot pattern, (26) imposes requirements on the MSE-optimal data pattern. These can be rewritten as $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{0}$ via

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

In other words, data must be transmitted in the nullspace of $\boldsymbol{W}$. To do this, we construct $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{d}$, where $\boldsymbol{d}$ contains $N_d := \dim(\text{null}(\boldsymbol{W}))$ data symbols and where the columns of $\boldsymbol{B} \in \mathbb{C}^{N \times N_d}$ form an orthonormal basis for $\text{null}(\boldsymbol{W})$. SCCP follows naturally from TDKD, whereas CP-OFDM follows naturally from FDKD.

It is possible to bound N_d for the DSC (22). Note 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 (27), those rows must be orthogonal. Thus, $N_f N_t \leq \text{rank}(\boldsymbol{W}) \leq (2N_f - 1)(2N_t - 1)$, which implies $N - (2N_f - 1)(2N_t - 1) \leq N_d \leq N - N_f N_t$.

The examples below specify various MSE-optimal PAT schemes using their $(\boldsymbol{s}, \boldsymbol{B})$ parameterization.

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

$$\mathcal{P}_t^{(i)} := \{i, i + \tfrac{N}{N_f}, ..., i + \tfrac{(N_f-1)N}{N_f}\} \tag{34}$$
$$\mathcal{G}_t^{(i)} := \bigcup_{k \in \mathcal{P}_t^{(i)}} \{-N_t + 1 + k, ..., N_t - 1 + k\}. \tag{35}$$

An MSE-optimal PAT scheme is given by

$$s(q) = \begin{cases} \sigma_s \sqrt{\frac{N}{N_f}} e^{j\theta(q)} & q \in \mathcal{P}_t^{(i)} \\ 0 & q \notin \mathcal{P}_t^{(i)} \end{cases} \tag{36}$$

and by $\boldsymbol{B}$ constructed from the columns of $\boldsymbol{I}_N$ with indices not *in the set $\mathcal{G}_t^{(i)}$. 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 with FDKD). *Assuming $\frac{N}{N_t} \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_f^{(i)}$ and the guard index set $\mathcal{G}_f^{(i)}$:*

$$\mathcal{P}_f^{(i)} := \{i, i + \tfrac{N}{N_t}, ..., i + \tfrac{(N_t-1)N}{N_t}\} \tag{37}$$
$$\mathcal{G}_f^{(i)} := \bigcup_{k \in \mathcal{P}_f^{(i)}} \{-N_f + 1 + k, ..., N_f - 1 + k\} \tag{38}$$

An MSE-optimal PAT scheme is given by

$$s_f(q) = \begin{cases} \sigma_s \sqrt{\frac{N}{N_t}} e^{j\theta(q)} & q \in \mathcal{P}_f^{(i)} \\ 0 & q \notin \mathcal{P}_f^{(i)} \end{cases} \tag{39}$$

and by $\boldsymbol{B}$ constructed from the columns of $\boldsymbol{F}_N$ with indices not *in the set $\mathcal{G}_f^{(i)}$. 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 (Superimposed Chirps). *Assuming even N, an MSE-optimal PAT scheme is given by*

$$s(q) = \sigma_s e^{j\frac{2\pi}{N}\frac{N_f}{2}q^2} \tag{40}$$
$$[\boldsymbol{B}]_{q,p} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(p+N_f N_t)q} e^{j\frac{2\pi}{N}\frac{N_f}{2}q^2}, \tag{41}$$

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

3.5. Discussion

A few comments are in order. The scheme in Example 1 was shown to be MSE-optimal in [4]. To our knowledge, the scheme in Example 2 is novel; the suggestion to cluster pilots in the frequency domain was given by Stamoulis et al. [6], though details were lacking. To our knowledge, the scheme in Example 3 is also novel; a chirp-based training scheme was suggested in [7], but data and pilots were transmitted in different frames. Note that, relative to TDKD or FDKD, chirp systems may have advantages in peak-to-average power ratio.

Though all three PAT examples above yield MSE-optimal channel estimates, they differ in the dimension of their data subspace N_d. While a capacity analysis is outside the scope of this manuscript (see [5] instead), it should be intuitively clear that larger N_d lead to higher capacity. Notice that, among the three examples above, FDKD yields the largest N_d when $N_t > N_f > 1$, while TDKD yields the largest N_d when $N_f > N_t > 1$. At the moment it is not clear, though, whether there exists an MSE-optimal PAT scheme for the DSC with even higher N_d.

4. CONCLUSION

For a general class of systems encompassing linear modulation and a linear time-varying channel, we derived a lower bound on the MSE of pilot-based channel estimates assuming a pilot energy constraint. In addition, we derived necessary and sufficient conditions for PAT schemes to achieve this lower bound. Applying these results to the case of single-antenna block-transmission over a DSC, we gave three examples of MSE-optimal PAT schemes, two of them novel. Our future work will strive to tighten the link between PAT design based on MSE and capacity criteria using the framework developed here.

5. REFERENCES

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

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

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

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

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

[6] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Estimation of fast fading channels in OFDM," in *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 1, pp. 465–470, Mar. 2002.

[7] S. Barbarossa and A. Scaglione, "Theoretical bounds on the estimation and prediction of multipath time-varying channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2545–2548, 2000.