Noncoherent Communication over the Doubly Selective Channel via Successive Decoding and Channel Re-Estimation

Sibasish Das and Philip Schniter
Dept. of ECE, The Ohio State University, Columbus, OH 43210
Email: {das,schniter}@ece.osu.edu

***Abstract*—For noncoherent communication over fading channels, pilot-aided transmission is a practical scheme which allows the receiver to compute channel estimates for subsequent use in coherent decoding. We propose an improved scheme whereby the transmitter interleaves several pilot substreams with independently-coded data substreams to facilitate successive decoding and channel re-estimation at the receiver. In particular, an initial pilot-aided channel estimate is used to coherently decode the first data substream, which is then used to refine the channel estimate before coherently decoding the next data substream, and so on. Assuming knowledge of the channel statistics, the pilots and the rates of the data substreams can be chosen to ensure reliable decoding. While similar schemes have been proposed for channels that are either time-selective or frequency-selective, ours is focused on doubly selective channels. We derive a lower bound on the achievable rate of our strategy and further characterize the achievable rate at high SNR. When the channel satisfies a complex-exponential basis expansion model, we show that the pre-log factor of the high-SNR achievable rate expression coincides with that of the ergodic capacity expression. For the same channel, we propose a pilot/data power allocation strategy that maximizes a lower bound on the achievable rate.**[1]

I. Introduction

Practical wireless communication is noncoherent in that the channel state is never known a priori to the transmitter nor receiver. As a result, practical wireless transmissions must be designed with a structure that facilitates reliable reception in the absence of channel state information (CSI). Pilot-aided transmission (PAT) [1], [2] is perhaps the most common means of providing this structure.

With PAT, "one-shot" schemes are common, whereby the receiver computes a pilot-aided channel estimate and subsequently uses it for coherent data decoding. In this case, channel estimation error acts as additional noise which degrades decoding performance and thus the rate of reliable communication [3]. Though channel estimation can be improved by allocating more transmission resources (e.g., rate and power) to pilots, doing so limits the resources that remain for data transmission. Hence, information-theoretic analysis are useful to understand the optimal allocation of resources between pilots and data. Several information-theoretic analyses of PAT with one-shot estimation/decoding have appeared, e.g., in [4]–[10].

[1]This work supported by Motorola Inc., the National Science Foundation under CAREER grant 237037, and the Office of Naval Research under grant N00014-07-1-0209.

As an improvement to one-shot estimation/decoding of PAT, several authors have considered iterative (i.e., "turbo") estimation/decoding strategies, whereby soft decoder outputs are employed to refine channel estimates, which can then be used for improved decoding, and so on [11]–[14]. Such systems are generally suboptimal and difficult to analyze.

More recently, the use of block interleaving with successive decoding has been proposed as a more structured approach to joint estimation/decoding of PAT [15], [16]. There the idea is to split the information stream into independently coded substreams and decode them successively. While a pilot-aided channel estimate is used to decode the first substream, reliably decoded symbols can be employed to refine the channel estimates used by later decoding stages. For long coding blocks and properly chosen substream rates (e.g., assuming known channel statistics), each substream can be reliably decoded, greatly simplifying the design and analysis of such systems. PAT with successive decoding has been used successfully in time-selective and frequency-selective SISO channels [16] as well as time-selective MIMO channels [15].

In this paper, we propose a scheme for noncoherent communication over doubly (i.e, time- and frequency-) selective fading channels that uses successive decoding and channel re-estimation at the receiver. Assuming perfect decoding of each stream, we calculate an achievable-rate lower-bound and use it to infer a set of substream rates which are sufficient to ensure perfect decoding. We also characterize the high-SNR spectral efficiency of the proposed communication strategy. To highlight certain design issues, we consider the special case of doubly selective channel which satisfies a complex-exponential basis expansion model (CE-BEM). For this channel, we design a suitable pilot pattern and through it verify that the pre-log factor of the high-SNR achievable rate expression coincides with that of the ergodic capacity [10], [17]. We also propose a pilot/data power allocation strategy that maximizes a lower bound on the achievable rate.

The paper is organized as follows: A description of the system model appears in Section II, the reception strategy is described in Section III, and achievable-rate expressions are derived in Section III-B. The specific case of the doubly selective CE-BEM channel is considered in Section IV, and conclusions are drawn in Section V.

Notation: In the manuscript, $(\cdot)^T$ denotes transpose, $(\cdot)^*$ denotes conjugate, and $(\cdot)^H$ denotes conjugate-transpose. $[\boldsymbol{B}]_{m,n}$ denotes the element in the m^{th} row and n^{th} column of $\boldsymbol{B}$, where row and column indices begin with zero. $\mathbf{0}_{m\times n}$ denotes the $m \times n$ zero matrix, $\boldsymbol{I}_K$ denotes the $K \times K$ identity matrix, and $\boldsymbol{e}_K^{(q)}$ denotes the q^{th} column of $\boldsymbol{I}_K$. For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \geq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semi-definite. The trace of a matrix is denoted by $\operatorname{tr}(\cdot)$, and the Kronecker product of two matrices is denoted by $\otimes$. Also, δ_l denotes the Kronecker delta sequence, $\langle \cdot \rangle$ denotes the modulo operator, and $\mathbb{C}$ the set of all complex numbers. Expectation is denoted by $\mathrm{E}(\cdot)$ and auto-covariance by $\boldsymbol{\Sigma}_{\boldsymbol{b}} := \mathrm{E}(\boldsymbol{b}\boldsymbol{b}^H) - \mathrm{E}(\boldsymbol{b})\,\mathrm{E}(\boldsymbol{b}^H)$.

II. System Model

A. Transmission Model

Consider a scheme in which information is transmitted through N_s substreams, each of which uses the channel N_b times. In particular, say that $s_k(i)$ denotes the i^{th} sample of the k^{th} substream. The first N_p substreams (i.e., $\{s_k(i)\}_{i=0}^{N_b-1}$ for $k = 0 \dots N_p - 1$) are dedicated to pilots while the remaining $N_s - N_p$ substreams (i.e., $\{s_k(i)\}_{i=0}^{N_b-1}$ for $k = N_p \dots N_s - 1$) are dedicated to data. The data substreams are independently encoded at rates that ensure reliable decoding, as will be discussed later. For this, we assume that the transmitter knows the channel statistics, but not the channel state.

The average transmission power is constrained to E_{tot} Joules per-channel-use, E_p of which is allocated to pilots and the remainder of which is divided equally among the data substreams. Thus, each data substream has power

$$\sigma_s^2 \;=\; \frac{E_{\text{tot}} - E_p}{N_s - N_p}. \tag{1}$$

For analytical tractability, we assume the use of i.i.d. Gaussian codebooks. With this assumption, the power constraints can be expressed as

$$\sum_{k=0}^{N_p-1} |s_k(i)|^2 \;=\; E_p \qquad \forall i \tag{2}$$

$$\mathrm{E}\{\underline{\boldsymbol{s}}_{N_p}(i_1)\underline{\boldsymbol{s}}_{N_p}(i_2)^H\} \;=\; \sigma_s^2 \boldsymbol{I}_{N_s-N_p}\delta_{i_1-i_2} \tag{3}$$

where

$$\underline{\boldsymbol{s}}_k(i) \;:=\; [s_k(i), \dots, s_{N_s-1}(i)]^T. \tag{4}$$

In the sequel, we will also make frequent use of the notation

$$\boldsymbol{s}_k(i) \;:=\; [s_0(i), \dots, s_{k-1}(i)]^T \tag{5}$$

$$\boldsymbol{s}(i) \;:=\; [s_0(i), \dots, s_{N_s-1}(i)]^T \;=\; [\boldsymbol{s}_k(i)^T, \underline{\boldsymbol{s}}_k(i)^T]^T. \tag{6}$$

B. Channel Model

We assume that the receiver observes the noisy inter-substream-interference (ISSI) corrupted samples $\{y_k(i)\}_{i=0}^{N_b-1}$, for $k \in \{0, \dots, N-1\}$, where

$$y_k(i) \;=\; \sum_{l=0}^{N_h-1} h_{k,l}(i) s_{k-l}(i) + w_k(i). \tag{7}$$

Here, $N := N_s + N_h - 1$, where N_h denotes the ISSI length, and $s_k(i) := 0$ for $k \notin \{0, \dots, N_s - 1\}$. The observations can be written in vector form as $\boldsymbol{y}(i) = [y_0(i), \dots, y_{N-1}(i)]^T$, where

$$\boldsymbol{y}(i) \;=\; \boldsymbol{H}(i)\boldsymbol{s}(i) + \boldsymbol{w}(i) \tag{8}$$

for $\boldsymbol{w}(i) = [w_0(i), \dots, w_{N-1}(i)]^T$ and

$$\boldsymbol{H}(i) \;=\; \begin{bmatrix} h_{0,0}(i) & & \\ \vdots & \ddots & \\ h_{N_h-1,N_h-1}(i) & & h_{N_s-1,0}(i) \\ & \ddots & \vdots \\ & & h_{N-1,N_h-1}(i) \end{bmatrix}. \tag{9}$$

Notice that the channel model suffices to describe either single-carrier or multi-carrier transmission. In the single-carrier case, (7) corresponds to N_b blocks of $(N_h - 1)$-zero-padded N_s-block transmission through a doubly selective fading channel with time-varying inter-symbol interference (ISI) of length N_h. In the multi-carrier case, (7) corresponds to N_b symbols of an N-subcarrier system with N_s active subcarriers and an inter-carrier interference (ICI) response of length N_h.

It will sometimes be convenient to write the system model as

$$\boldsymbol{y}(i) \;=\; \boldsymbol{S}(i)\boldsymbol{h}(i) + \boldsymbol{w}(i) \tag{10}$$

with $\boldsymbol{h}(i) \in \mathbb{C}^{N_s N_h}$ such that

$$\begin{aligned} \boldsymbol{h}(i) \;:=\; & [h_{0,0}(i), \dots, h_{N_s-1,0}(i), h_{1,1}(i), \dots, h_{N_s,1}(i), \cdots \\ & h_{N_h-1,N_h-1}(i), \dots, h_{N-1,N_h-1}(i)]^T, \end{aligned} \tag{11}$$

and with

$$\boldsymbol{S}(i) := \begin{bmatrix} s_0(i) & & & 0 & \cdots & 0 & & 0 & \cdots & 0 \\ & \ddots & & s_0(i) & & & \cdots & \vdots & & \vdots \\ & & s_{N_s-1}(i) & & \ddots & & & 0 & \cdots & 0 \\ 0 & \cdots & 0 & & & s_{N_s-1}(i) & & s_0(i) & & \\ \vdots & & \vdots & 0 & \cdots & 0 & \cdots & & \ddots & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & & & & s_{N_s-1}(i) \end{bmatrix}.$$

In the sequel, we make use of $\boldsymbol{y}_k(i) := [y_0(i), \dots, y_{k-1}(i)]^T$, $\boldsymbol{w}_k(i) := [w_0(i), \dots, w_{k-1}(i)]^T$, and $\boldsymbol{S}_k(i)$, the latter defined as the matrix formed by the first k rows of $\boldsymbol{S}(i)$. Note that

$$\boldsymbol{y}_k(i) \;=\; \boldsymbol{S}_k(i)\boldsymbol{h}(i) + \boldsymbol{w}_k(i) \tag{12}$$

where the entries in $\boldsymbol{S}_k(i)$ come from $\boldsymbol{s}_k(i)$ but not from $\underline{\boldsymbol{s}}_k(i)$.

We assume that the channel coefficients $\boldsymbol{h}(i)$ are zero-mean circular Gaussian with $\mathrm{E}\{\boldsymbol{h}(i)\boldsymbol{h}(i)^H\} = \boldsymbol{\Sigma}_{\boldsymbol{h}}$, where $\operatorname{rank}(\boldsymbol{\Sigma}_{\boldsymbol{h}}) = N_m < N_s$ and where $\operatorname{tr}(\boldsymbol{\Sigma}_{\boldsymbol{h}}) = N_s$ (i.e., the channel is energy-preserving). Similarly, we assume that the noise coefficients are zero-mean circular Gaussian with $\mathrm{E}\{\boldsymbol{w}(i)\boldsymbol{w}(i)^H\} = \sigma^2 \boldsymbol{I}_N$ and independent across i.

III. Noncoherent Pilot-and-Data-Aided Communication

A. Description of Scheme

We now summarize the noncoherent communication scheme, elaborating on the details after the summary.

0. The MMSE estimate of $\boldsymbol{h}(i)$ from $\boldsymbol{y}_{N_p}(i)$ is computed for each $i \in \{0, \ldots, N_b - 1\}$, leveraging the fact that $\boldsymbol{y}_{N_p}(i)$ is a function of the pilots $\boldsymbol{s}_{N_p}(i)$ but not the unknown data $\underline{\boldsymbol{s}}_{N_p}(i)$. Denoting this pilot-aided channel estimate by $\hat{\boldsymbol{h}}^{(N_p)}(i)$, the first data substream $\{s_{N_p}(i)\}_{i=0}^{N_b-1}$ is then coherently decoded using the pilot-aided channel estimate $\hat{\boldsymbol{h}}^{(N_p)}(i)$. With large enough N_b and suitable choice of code rate, this data substream can be reliably decoded.
1. Using the decoded substream in conjunction with pilots, a refined MMSE channel estimate $\hat{\boldsymbol{h}}^{(N_p+1)}(i)$ is computed from $\boldsymbol{y}_{N_p+1}(i)$ for each $i \in \{0, \ldots, N_b - 1\}$, leveraging the fact that $\boldsymbol{y}_{N_p+1}(i)$ is not a function of the not-yet-decoded data. The next data substream $\{s_{N_p+1}(i)\}_{i=0}^{N_b-1}$ is then coherently decoded using the refined channel estimate $\hat{\boldsymbol{h}}^{(N_p+1)}(i)$. With a suitable choice of code rate, this second data substream can also be reliably decoded.
2. Using the two decoded substreams in conjunction with pilots, the refined MMSE channel estimate $\hat{\boldsymbol{h}}^{(N_p+2)}(i)$ is computed from $\boldsymbol{y}_{N_p+2}(i)$ for each $i \in \{0, \ldots, N_b - 1\}$. The next data substream $\{s_{N_p+2}(i)\}_{i=0}^{N_b-1}$ is then coherently decoded using the most recent channel estimate $\hat{\boldsymbol{h}}^{(N_p+2)}(i)$, where decoding can be made reliable via proper rate selection.

*. The procedure continues this way until all N_s data substreams have been decoded.

Next we elaborate on the channel estimation and data decoding procedures. Rate allocation will be detailed in Section III-B.

The MMSE channel estimate $\hat{\boldsymbol{h}}^{(k)}(i)$, i.e., the estimate of $\boldsymbol{h}(i)$ from $\boldsymbol{y}_k(i)$ given perfect knowledge of $\boldsymbol{S}_k(i)$, can be written as [18]

$$\hat{\boldsymbol{h}}^{(k)}(i) = \boldsymbol{\Sigma}_{\boldsymbol{h}} \boldsymbol{S}_k(i)^H \times \big(\boldsymbol{S}_k(i)\boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{S}_k(i)^H + \sigma^2 \boldsymbol{I}_k\big)^{-1} \boldsymbol{y}_k(i). \quad (13)$$

Conditioned on $\boldsymbol{s}_k(i)$, the estimation error $\tilde{\boldsymbol{h}}^{(k)}(i) := \boldsymbol{h}(i) - \hat{\boldsymbol{h}}^{(k)}(i)$ has covariance [18]

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} = \boldsymbol{\Sigma}_{\boldsymbol{h}} - \boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{S}_k(i)^H \times \big(\boldsymbol{S}_k(i)\boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{S}_k(i)^H + \sigma^2 \boldsymbol{I}_k\big)^{-1} \boldsymbol{S}_k(i)\boldsymbol{\Sigma}_{\boldsymbol{h}}. \quad (14)$$

Now we describe the decoding of data substream $\{s_k(i)\}_{i=0}^{N_b-1}$ for $k \in \{N_p, \ldots, N_s - 1\}$. In doing so, we make use of the partition $\boldsymbol{H}(i) = [\boldsymbol{H}_k(i), \boldsymbol{h}_k(i), \underline{\boldsymbol{H}}_{k+1}(i)]$, where $\boldsymbol{H}_k(i) \in \mathbb{C}^{N\times k}$, $\boldsymbol{h}_k(i) \in \mathbb{C}^{N\times 1}$, and $\underline{\boldsymbol{H}}_{k+1}(i) \in \mathbb{C}^{N\times(N_s-k-1)}$, so that (8) becomes

$$\boldsymbol{y}(i) = \boldsymbol{H}_k(i)\boldsymbol{s}_k(i) + \boldsymbol{h}_k(i)s_k(i) + \underline{\boldsymbol{H}}_{k+1}(i)\underline{\boldsymbol{s}}_{k+1}(i) + \boldsymbol{w}(i). \quad (15)$$

In addition, we construct $\hat{\boldsymbol{H}}^{(k)}(i)$ from $\hat{\boldsymbol{h}}^{(k)}(i)$, and $\tilde{\boldsymbol{H}}^{(k)}(i)$ from $\tilde{\boldsymbol{h}}^{(k)}(i)$, in the same way that we constructed $\boldsymbol{H}^{(k)}(i)$ from $\boldsymbol{h}^{(k)}(i)$, and we make the corresponding partition $\hat{\boldsymbol{H}}^{(k)}(i) = [\hat{\boldsymbol{H}}_k^{(k)}(i), \hat{\boldsymbol{h}}_k^{(k)}(i), \hat{\underline{\boldsymbol{H}}}_{k+1}^{(k)}(i)]$. The first stage of decoding involves interference cancellation and linear combining:

$$\boldsymbol{r}^{(k)}(i) = \boldsymbol{y}(i) - \hat{\boldsymbol{H}}_k^{(k)}(i)\boldsymbol{s}_k(i) \quad (16)$$

$$z_k(i) = \boldsymbol{c}^{(k)}(i)^H \boldsymbol{r}^{(k)}(i). \quad (17)$$

Recall that we have assumed that, at the time of decoding $\{s_k(i)\}_{i=0}^{N_b-1}$, the substreams $\{\boldsymbol{s}_k(i)\}_{i=0}^{N_b-1}$ are known through reliable decoding or as pilots. Using (15) and (16), we see that

$$\boldsymbol{r}^{(k)}(i) = \hat{\boldsymbol{h}}_k^{(k)}(i)s_k(i) + \boldsymbol{v}^{(k)}(i), \quad (18)$$

$$\boldsymbol{v}^{(k)}(i) = \hat{\underline{\boldsymbol{H}}}_{k+1}^{(k)}(i)\underline{\boldsymbol{s}}_{k+1}(i) + \tilde{\boldsymbol{H}}^{(k)}(i)\boldsymbol{s}(i) + \boldsymbol{w}(i) \quad (19)$$

from which the post-combining SINR $\gamma^{(k)}(i)$ becomes

$$\gamma^{(k)}(i) = \frac{|\boldsymbol{c}^{(k)}(i)^H \hat{\boldsymbol{h}}_k^{(k)}(i)|^2 \sigma_s^2}{\boldsymbol{c}^{(k)}(i)^H \boldsymbol{\Sigma}_{\boldsymbol{v}^{(k)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} \boldsymbol{c}^{(k)}(i)} \quad (20)$$

for $\boldsymbol{\Sigma}_{\boldsymbol{v}^{(k)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} := \mathrm{E}\{\boldsymbol{v}^{(k)}(i)\boldsymbol{v}^{(k)}(i)^H | \boldsymbol{s}_k(i), \hat{\boldsymbol{h}}^{(k)}(i)\}$. The combiner output can then be written as

$$z_k(i) = g_k(i)s_k(i) + n_k(i), \quad (21)$$

for $g_k(i) := \boldsymbol{c}^{(k)}(i)^H \hat{\boldsymbol{h}}_k^{(k)}(i)$ and $n_k(i) := \boldsymbol{c}^{(k)}(i)^H \boldsymbol{v}^{(k)}(i)$. After $\{z_k(i)\}_{i=0}^{N_b-1}$ and $\{g_k(i)\}_{i=0}^{N_b-1}$ have been computed, coherent decoding based on (21) can be applied.

B. Achievable-Rate Analysis

Notice that, for $k \in \{N_p, \ldots, N_s - 1\}$, the effective noise $n_k(i)$ is non-Gaussian:

$$n_k(i) = \boldsymbol{c}^{(k)}(i)^H \big(\hat{\underline{\boldsymbol{H}}}_{k+1}^{(k)}(i)\underline{\boldsymbol{s}}_{k+1}(i) + \underbrace{\tilde{\boldsymbol{H}}^{(k)}(i)\boldsymbol{s}(i)}_{\text{non-Gaussian}} + \boldsymbol{w}(i)\big). \quad (22)$$

Medard [3] and Hassibi [4] showed that the worst-case noise distribution with respect to mutual information is the Gaussian one. Additionally, the achievable rate for Gaussian signaling in the presence of Gaussian distributed noise can be expressed in terms of the post-combining SINR [19]. Therefore, assuming adequately large N_b, the achievable rate across data substream $k \in \{N_p, \ldots, N_s - 1\}$ can be bounded from below, in units of nats-per-(scalar)-channel-use, as

$$R_k \geq \mathrm{E}\Big\{\log\big(1 + \gamma^{(k)}(i)\big)\Big\}, \quad (23)$$

where the SINR $\gamma^{(k)}(i)$ was given in (20) and the expectation in (23) is taken over the joint distribution of $\hat{\boldsymbol{h}}^{(k)}(i)$ and $\boldsymbol{s}_k(i)$.

Note that the bound in (23) holds for general linear combiners $\boldsymbol{c}^{(k)}(i)$. The tightest bound can be obtained by choosing the max-SINR combiner

$$\boldsymbol{c}^{(k)}(i) \;=\; \boldsymbol{\Sigma}^{-1}_{\boldsymbol{v}^{(k)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} \hat{\boldsymbol{h}}_k^{(k)}(i), \qquad (24)$$

in which case the bound (23) becomes

$$R_k \;\geq\; \mathrm{E}\Big\{\log\Big(1+\sigma_s^2\hat{\boldsymbol{h}}_k^{(k)}(i)^H \times \boldsymbol{\Sigma}^{-1}_{\boldsymbol{v}^{(k)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} \hat{\boldsymbol{h}}_k^{(k)}(i)\Big)\Big\} \qquad (25)$$

leading to the following bound on the overall achievable rate, given in nats-per-channel-use.

$$R_{\text{tot}} \;\geq\; \frac{1}{N}\sum_{k=N_p}^{N_s-1} \mathrm{E}\Big\{\log\Big(1+\sigma_s^2\hat{\boldsymbol{h}}_k^{(k)}(i)^H \times \boldsymbol{\Sigma}^{-1}_{\boldsymbol{v}^{(k)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} \hat{\boldsymbol{h}}_k^{(k)}(i)\Big)\Big\}. \qquad (26)$$

To facilitate reliable decoding, the data substream rates should be chosen in accordance with (25).

C. Asymptotic Achievable-Rate Analysis

In this section, we analyze the achievable rate R_{tot} at high SNR. For this, we define the SNR $\rho := \frac{E_{\text{tot}}}{N\sigma^2}$ and examine $R_{\text{tot}}(\rho)$ as $\rho \to \infty$.

To provide some intuition on the high-SNR behavior, consider for the moment choosing a zero-forcing (ZF) combiner $\boldsymbol{c}^{(k)}(i)$, $||\boldsymbol{c}^{(k)}(i)|| = 1$, such that

$$\boldsymbol{c}^{(k)}(i)^H \underline{\hat{\boldsymbol{H}}}^{(k)}_{k+1}(i) \;=\; \boldsymbol{0} \qquad (27)$$

which implies that $g_k(i) = \boldsymbol{c}^{(k)}(i)^H \hat{\boldsymbol{h}}_k^{(k)}(i)$ and (via (22))

$$n_k(i) \;=\; \boldsymbol{c}^{(k)}(i)^H\big(\tilde{\boldsymbol{H}}^{(k)}(i)\boldsymbol{s}(i) + \boldsymbol{w}^{(k)}(i)\big). \qquad (28)$$

Note that this ZF combiner exists w.p.1. We reason that, for a "well-designed" channel estimation procedure, the covariance of the channel estimation error $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}$ (defined in (14)) should vanish as $\sigma^2 \to 0$. In particular, there should exist a pilot pattern $\boldsymbol{s}_{N_p}$ that, for fixed E_{tot} and $N_p < N_s$, guarantees the existence of σ-invariant $\boldsymbol{A}$ such that $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} \leq \sigma^2\boldsymbol{A}$ for all $\sigma > 0$ and for all $k \in \{N_p,\dots,N_s-1\}$. When this is the case, the use of a zero-forcing combiner ensures that $\sigma^2_{n_k(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} := \mathrm{E}\{|n_k(i)|^2 \,|\, \boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)\}$ will also vanish as $\sigma^2 \to 0$ for each information substream. In particular, there will exist σ-invariant α such that $\sigma^2_{n_k(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} \leq \sigma^2\alpha$ for all $\sigma > 0$ and for all $k \in \{N_p,\dots,N_s-1\}$. Then we see that $\gamma^{(k)}(i) \geq \frac{|g_k(i)|^2\sigma_s^2}{\alpha\sigma^2} = \frac{|g_k(i)|^2 N}{\alpha(N_s-N_p)}(1-\frac{E_p}{E_{\text{tot}}})\rho$ for each $k \in \{N_p,\dots,N_s-1\}$. When this is the case, (23) implies

$$\lim_{\rho\to\infty}\frac{R_{\text{tot}}(\rho)}{\log\rho} \;\geq\; \frac{N_s-N_p}{N}. \qquad (29)$$

We now make these statements more precise.

The condition under which the covariance of the estimation error $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}$ vanishes with increasing SNR is given by the following lemma.

Lemma 1. *Let the columns of $\boldsymbol{B} \in \mathbb{C}^{N_hN_s\times N_m}$ be the eigenvectors corresponding to non-zero eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{h}}$. Then there exists σ-invariant $\boldsymbol{A}$ such that, for every $k \in \{N_p,\dots,N_s-1\}$,*

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} \;\leq\; \sigma^2\boldsymbol{A} \qquad \forall\sigma > 0 \qquad (30)$$

if and only if

$$\operatorname{rank}\big(\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\big) = N_m. \qquad (31)$$

Proof: See Appendix A ■

In the sequel, we refer to condition (31) as the "rank condition." Lemma 1 says that the N_p pilot substreams must excite all N_m channel modes in order to obtain channel estimates whose error vanishes with increasing SNR.

Theorem 1. *For the class of channels that enable pilots $\boldsymbol{S}_{N_m}(i)$ to yield $\operatorname{rank}(\boldsymbol{S}_{N_m}(i)\boldsymbol{B}) = N_m$, the achievable rate of our scheme obeys*

$$\lim_{\rho\to\infty}\frac{R_{\text{tot}}(\rho)}{\log\rho} \;=\; \frac{N_s-N_m}{N}. \qquad (32)$$

Proof: See Appendix B ■

IV. Illustrative Example

We now consider the specific case of a doubly selective fading channel which obeys a complex-exponential basis expansion model (CE-BEM) [20], [21]. In particular, the channel coefficients in $\boldsymbol{h}(i)$ are parameterized by $N_m = (2D+1)N_h$ uncorrelated Gaussian random variables $\{\phi_{m,l}(i) : m \in \{-D,\dots,D\}, l \in \{0,\dots,N_h-1\}\}$ via

$$h_{k,l}(i) \;=\; \frac{1}{\sqrt{N_s}}\sum_{m=-D}^{D}\phi_{m,l}(i)\, e^{j\frac{2\pi}{N_s}m(k-l)}. \qquad (33)$$

For simplicity, we assume that the random variables $\{\phi_{m,l}(i)\}_{m,l}$ have equal variance. Hence $\mathrm{E}\big[\phi_{m_1,l_1}(i_1)\phi^*_{m_2,l_2}(i_2)\big] = \frac{N_s}{(2D+1)N_h}\delta_{m_1-m_2}\delta_{l_1-l_2}\delta_{i_1-i_2}$. In (33), $D \approx \lceil f_{\mathsf{D}}T_sN_s\rceil$ where $f_{\mathsf{D}}T_s$ is the single-sided normalized Doppler spread. For this CE-BEM channel, the eigenvector matrix $\boldsymbol{B}$ (defined in Lemma 1) has the form $\boldsymbol{B} = \boldsymbol{I}_{N_h}\otimes\boldsymbol{F}$, where the $N_s\times(2D+1)$ matrix $\boldsymbol{F}$ is defined element-wise as $[\boldsymbol{F}]_{m_1,m_2} = \frac{1}{\sqrt{N_s}}e^{j\frac{2\pi}{N_s}m_1(m_2-D)}$.

We first address the issue of choosing a suitable pilot pattern for this example. Estimating the channel coefficients $\{h_{k,l}(i)\}_{k=l}^{N_s+l-1}$ for each l is equivalent to estimating the $(2D+1)$ random variables $\{\phi_{m,l}\}_{m=-D}^{D}$. This can be accomplished by exciting the channel with a set of $(2D+1)$ impulse sequences of length N_h-1. This motivates the use of the $N_p = N_m = (2D+1)N_h$-length pilot pattern

$$s_k(i) \;=\; \sqrt{\frac{N_hE_p}{N_m}}\,\delta_{\langle k\rangle_{N_h}}, \qquad 0 \leq k < N_m. \qquad (34)$$

The proposed scheme, in conjunction with this pilot pattern, leads to the following achievable-rate characterization.

Proposition 1. *For the CE-BEM doubly selective fading channel, the achievable rate of our scheme obeys*

$$\lim_{\rho\to\infty} \frac{R_{\mathsf{tot}}(\rho)}{\log\rho} = \frac{N_s - N_m}{N}. \tag{35}$$

Proof: See Appendix C. ∎

Interestingly, [10], [17] has shown that, under continuously distributed inputs, the maximum spectral efficiency than can be achieved on the CE-BEM doubly selective block-fading channel is $\frac{N_s-N_m}{N}$. Thus, using the pilot pattern (34), the proposed scheme becomes "spectrally-efficient".

Next, we tackle the issue of power allocation between pilot and data substreams. Let $E_p = \alpha_p E_{\mathsf{tot}}$ for some $\alpha_p \in (0,1)$. Then $\sigma_s^2 = (1-\alpha_p)E_{\mathsf{tot}}/(N_s - N_p)$. We propose a "minimax" approach whereby we choose α_p to maximize a lower-bound on the achievable rate of the weakest data substream. Though this power allocation strategy can be used with an arbitrarily chosen pilot pattern, we restrict ourselves to the pilot pattern (34) for simplicity. Recall that the channel estimate is refined after decoding each data substream, thereby increasing the effective SINR. Thus, the first data substream $\{s_{N_p}(i)\}_{i=0}^{N_b-1}$ must be the weakest. Recalling that $\boldsymbol{e}_N^{(N_p)}$ denotes the N_p^{th} column of $\boldsymbol{I}_N$, we have the following result.

Proposition 2. *For the CE-BEM doubly selective fading channel, the pilot power allocation*

$$\alpha_{p,*} = \arg\max_{\alpha_p\in(0,1)} \frac{\hat{\sigma}_{N_p}^2 \sigma_s^2}{\tilde{\sigma}_{N_p}^2 \sigma_s^2 + \sigma^2}, \tag{36}$$

where

$$\hat{\sigma}_{N_p}^2 := \left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)}\right]_{N_p,N_p} \tag{37}$$

$$\tilde{\sigma}_{N_p}^2 := \left[\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)}\right]_{N_p,N_p}, \tag{38}$$

maximizes a lower-bound on the achievable-rate of the weakest data substream, in particular, the lower bound that follows from the use of the (sub-optimal) combiner $\boldsymbol{c}^{(N_p)}(i) = \boldsymbol{e}_N^{(N_p)}$ in (23).

Proof: See Appendix D ∎

Figure 1 plots the power allocation parameter $\alpha_{p,*}$ versus SNR for single-carrier transmission with $N_s = 128$ substreams across the CE-BEM doubly selective fading channel with $N_h = 8$ taps of ISSI and $D \in \{1,2\}$. These parameters correspond to, e.g., a channel with bandwidth 1.5 MHz, carrier frequency 60 GHz, delay spread 5.4 μs, and mobile and reflector velocities of $\{69, 138\}$ km/hr, in a "triple Doppler" scenario [22]. Figure 1 suggests that, at low values of SNR, additive noise level dictates performance and more power is allocated to the data substreams. However, as the SNR grows, the effect of channel estimation error on performance becomes more pronounced, and the pilots are given more power to keep the estimation error in check.

V. Conclusion

In this paper, we designed and analyzed a communication scheme for the doubly dispersive channel based on pilot-aided transmission and successive decoding with channel re-estimation. We derived a lower bound on the achievable rate and characterized the pre-log factor of the high-SNR achievable rate expression. For the special case of the CE-BEM doubly selective fading channel, we found the proposed communication system to be spectrally efficient. Finally, we designed a pilot/data power allocation strategy based on the maximization of an achievable-rate lower-bound.

Appendix A
Proof for Lemma 1

We know that $\operatorname{rank}(\boldsymbol{\Sigma}_{\boldsymbol{h}}) = N_m$, so that $\boldsymbol{\Sigma}_{\boldsymbol{h}} = \boldsymbol{B}\boldsymbol{\Lambda}_h\boldsymbol{B}^H$ for some positive definite diagonal matrix $\boldsymbol{\Lambda}_h \in \mathbb{C}^{N_m\times N_m}$. We can then express $\boldsymbol{h}(i)$ using the Karhunen-Loeve transform as

$$\boldsymbol{h}(i) = \boldsymbol{B}\boldsymbol{\lambda}(i), \tag{39}$$

where $\boldsymbol{\lambda}(i) \in \mathbb{C}^{N_m}$ is a zero-mean complex Gaussian random vector with covariance $\boldsymbol{\Lambda}_h$. Furthermore, there exists a unitary matrix $\boldsymbol{U}(i)$ and positive semi-definite diagonal matrix $\boldsymbol{\Lambda}_{N_p}(i)$ such that $\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H\boldsymbol{S}_{N_p}(i)\boldsymbol{B} = \boldsymbol{U}(i)\boldsymbol{\Lambda}_{N_p}(i)\boldsymbol{U}(i)^H$. Using $\boldsymbol{\lambda}'(i) := \boldsymbol{U}(i)^H\boldsymbol{\lambda}(i)$, the observations $\boldsymbol{y}_{N_p}(i)$ can be expressed as

$$\begin{aligned} \boldsymbol{y}_{N_p}(i) &= \boldsymbol{S}_{N_p}(i)\boldsymbol{h}(i) + \boldsymbol{w}_{N_p}(i) && (40)\\ &= \boldsymbol{S}_{N_p}(i)\boldsymbol{B}\boldsymbol{\lambda}(i) + \boldsymbol{w}_{N_p}(i) && (41)\\ &= \boldsymbol{\Lambda}_{N_p}^{\frac{1}{2}}(i)\boldsymbol{\lambda}'(i) + \boldsymbol{w}_{N_p}(i). && (42) \end{aligned}$$

We first show that the rank condition is a necessary condition. Realize from (40)-(42) that estimating $\boldsymbol{h}(i)$ is equivalent to estimating $\boldsymbol{\lambda}'(i)$. Let $\operatorname{rank}(\boldsymbol{S}_{N_p}\boldsymbol{B}) = N_m' < N_m$. Then w.l.o.g. the first N_m' entries along the diagonal of $\boldsymbol{\Lambda}_{N_p}(i)$ are positive and the rest are zero. Consequently, the MMSE estimates of the last $N_m - N_m'$ components of $\boldsymbol{\lambda}'(i)$ are identically zero and the estimation error for these $N_m - N_m'$ components of $\boldsymbol{\lambda}'(i)$ does not depend on the noise variance σ^2. Then there is no hope of finding a σ-invariant $\boldsymbol{A}$ satisfying $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} \le \sigma^2\boldsymbol{A}$ when $\operatorname{rank}(\boldsymbol{S}_{N_p}\boldsymbol{B}) = N_m' < N_m$. This establishes that the rank condition is a necessary condition.

We now show that the rank condition is a sufficient condition. We first write the estimation error from (14) as

$$\begin{aligned} \boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} &= \boldsymbol{\Sigma}_{\boldsymbol{h}} - \boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{S}_{N_p}(i)^H \\ &\quad\times \left(\boldsymbol{S}_{N_p}(i)\boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{S}_{N_p}(i)^H + \sigma^2\boldsymbol{I}_{N_p}\right)^{-1}\boldsymbol{S}_{N_p}(i)\boldsymbol{\Sigma}_{\boldsymbol{h}} \\ &= \boldsymbol{B}\Big[\boldsymbol{\Lambda}_h - \boldsymbol{\Lambda}_h\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H\big(\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\boldsymbol{\Lambda}_h\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H \\ &\quad + \sigma^2\boldsymbol{I}_{N_p}\big)^{-1}\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\boldsymbol{\Lambda}_h\Big]\boldsymbol{B}^H \\ &= \boldsymbol{B}\left(\boldsymbol{\Lambda}_h^{-1} + \sigma^{-2}\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\right)^{-1}\boldsymbol{B}^H. \end{aligned} \tag{43}$$

The last step above is an application of the matrix inversion lemma [23]. Realize that $\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H\boldsymbol{S}_{N_p}(i)\boldsymbol{B}$ is σ-invariant, positive definite and invertible if $\mathrm{rank}(\boldsymbol{S}_{N_p}\boldsymbol{B}) = N_m$. We choose $\boldsymbol{A}' = \left(\boldsymbol{B}^H\boldsymbol{S}_{N_p}(i)^H\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\right)^{-1}$, and apply the matrix inversion lemma on (43) to obtain

$$\begin{aligned}\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} &= \\ &\boldsymbol{B}\left(\sigma^2\boldsymbol{A}' - \sigma^4\boldsymbol{A}'(\Lambda_h + \sigma^2\boldsymbol{A}')^{-1}\boldsymbol{A}'\right)\boldsymbol{B}^H. \end{aligned} \tag{44}$$

Then for the choice $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{A}'\boldsymbol{B}^H$, (44) shows that

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} - \sigma^2\boldsymbol{A} \;\le\; \boldsymbol{0}. \tag{45}$$

This shows that the rank condition is sufficient for $k = N_p$. It remains to be shown that the rank condition is sufficient for each $k > N_p$. Let $\breve{\boldsymbol{s}}_k(i)^H$ be the k^{th} row of $\boldsymbol{S}(i)$, so that $\boldsymbol{S}_{k+1}(i) = [\boldsymbol{S}_k(i)^H \;\; \breve{\boldsymbol{s}}_{k+1}(i)]^H$. Then

$$\begin{aligned}&\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k+1)}(i)|\boldsymbol{s}_{k+1}(i)} \\ &= \boldsymbol{B}\left(\boldsymbol{\Lambda}_h^{-1} + \sigma^{-2}\boldsymbol{B}^H\boldsymbol{S}_{k+1}(i)^H\boldsymbol{S}_{k+1}(i)\boldsymbol{B}\right)^{-1}\boldsymbol{B}^H \\ &= \boldsymbol{B}\Big[\boldsymbol{\Lambda}_h^{-1} + \sigma^{-2}\boldsymbol{B}^H\boldsymbol{S}_k(i)^H\boldsymbol{S}_k(i)\boldsymbol{B} \\ &\quad + \sigma^{-2}\boldsymbol{B}^H\breve{\boldsymbol{s}}_{k+1}(i)\breve{\boldsymbol{s}}_{k+1}(i)^H\boldsymbol{B}\Big]^{-1}\boldsymbol{B}^H \\ &= \boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} \\ &\quad - \sigma^2\frac{\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}\breve{\boldsymbol{s}}_{k+1}(i)\breve{\boldsymbol{s}}_{k+1}(i)^H\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}}{1+\sigma^2\breve{\boldsymbol{s}}_{k+1}(i)^H\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}\breve{\boldsymbol{s}}_{k+1}(i)}, \end{aligned} \tag{46}$$

where (46) results from applying the matrix inversion lemma to the penultimate expression and then substituting (43) (with indices k instead of N_p). This clearly implies that

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k+1)}(i)|\boldsymbol{s}_{k+1}(i)} - \boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} \;\le\; \boldsymbol{0}. \tag{47}$$

We conclude from (45) and (47) that

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)} \;\le\; \sigma^2\boldsymbol{A},\; N_p \le k \le N_s - 1, \tag{48}$$

which establishes that the rank criterion is also a sufficient condition and completes the proof.

Appendix B
Proof for Theorem 1

We first show that the spectral efficiency of the proposed communication strategy is *at least* $\frac{N_s - N_m}{N}$ when the rank criterion holds. Recall that

$$R_k(\rho) \;\ge\; \mathrm{E}\left\{\log\left(1+\gamma^{(k)}(i)\right)\right\}. \tag{49}$$

Our approach will be to show that the required spectral efficiency can be achieved by using a sub-optimal zero forcing combiner $\boldsymbol{c}^{(k)}(i) = \boldsymbol{e}_N^{(k)}$, the k^{th} column of $\boldsymbol{I}_N$. Recall that we also used this combiner to split transmit power between pilot and data substreams. This choice of combiner implies that only observations $\{y_k(i)\}_{i=0}^{N_b-1}$ are used in decoding the k^{th} data substream. Recall from the system model that $y_k(i)$ is influenced by symbols $\{s_m(i)\}_{m=k-N_h+1}^{k}$, of which the symbols except $s_k(i)$ are known (from previously decoded substreams or as pilots). There is no interference from yet-to-be-decoded substreams. It can be shown that the combiner output is

$$z_k(i) = [\hat{\boldsymbol{h}}_k^{(k)}(i)]_k s_k(i) + n_k(i) \tag{50}$$

$$n_k(i) = \breve{\boldsymbol{s}}_k(i)^H\tilde{\boldsymbol{h}}^{(k)}(i) + w_k(i), \tag{51}$$

where $\breve{\boldsymbol{s}}_k(i)^H$ is the k^{th}-*row* of $\boldsymbol{S}(i)$. First, we bound the variance of the noise term in (51). In doing so, we can make use of Lemma 1 since we have assumed that a suitable pilot pattern that satisfies the rank criterion is used. Then we can write $\sigma^2_{n_k(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} := \mathrm{E}\left[|n_k(i)|^2|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}_k^{(k)}(i)\right]$ as

$$\begin{aligned}\sigma^2_{n_k(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)} &= \mathrm{E}\left[\breve{\boldsymbol{s}}_k(i)^H\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(k)}(i)|\boldsymbol{s}_k(i)}\breve{\boldsymbol{s}}_k(i)\right] + \sigma^2 \\ &\le \mathrm{E}\left[\sigma^2\breve{\boldsymbol{s}}_k(i)^H\boldsymbol{A}\breve{\boldsymbol{s}}_k(i)\right] + \sigma^2 \\ &\le \sigma^2\alpha_k, \end{aligned} \tag{52}$$

for some positive semi-definite σ-invariant matrix $\boldsymbol{A}$ and σ-invariant $\alpha_k > 1$. Then (52) can be used to express the SINR as

$$\gamma^{(k)}(i) \;\ge\; \frac{\left|[\hat{\boldsymbol{h}}_k^{(k)}(i)]_k\right|^2\sigma_s^2}{\sigma^2\alpha_k} \tag{53}$$

$$= \frac{(1-\frac{E_p}{E_{\mathrm{tot}}})}{(N_s-N_p)\alpha_k}\left|[\hat{\boldsymbol{h}}_k^{(k)}(i)]_k\right|^2\rho. \tag{54}$$

Then for the class of channels that enable pilots $\boldsymbol{S}_{N_m}(i)$ to yield $\mathrm{rank}(\boldsymbol{S}_{N_m}(i)\boldsymbol{B}) = N_m$, we first show that

$$\liminf_{\rho\to\infty}\frac{R_k(\rho)}{\log\rho} \;\ge\; 1 \quad \forall k \in \{N_p, \cdots, N_s-1\}. \tag{55}$$

In this direction, we define $\psi_k := [\hat{\boldsymbol{h}}_k^{(k)}(i)]_k$ and $q_k := \frac{(1-E_p/E_{\mathrm{tot}})}{(N_s-N_p)\alpha_k}$ to simplify the notation. With these definitions, we can say that

$$\frac{R_k(\rho)}{\log\rho} \;\ge\; \frac{\mathrm{E}\log(1+q_k|\psi_k|^2\rho)}{\log\rho} \tag{56}$$

$$\ge\; 1 + \frac{\mathrm{E}\log(\rho^{-1}+q_k|\psi_k|^2)}{\log\rho}. \tag{57}$$

Notice that the estimate $\psi_k = [\hat{\boldsymbol{h}}_k^{(k)}(i)]_k$ is zero-mean Gaussian distributed. Then

$$\lim_{\rho\to\infty}\frac{\log(\rho^{-1}+q_k|\psi_k|^2)}{\log\rho} = \begin{cases}0, & \psi_k \ne 0\\ -1, & \psi_k = 0\end{cases} \tag{58}$$

$$= 0 \quad \text{w.p.1}. \tag{59}$$

Taking limit infimum on both sides of (57), we see that

$$\begin{aligned}\liminf_{\rho\to\infty}\frac{R_k(\rho)}{\log\rho} &\ge 1 + \liminf_{\rho\to\infty}\frac{\mathrm{E}\log\left(\rho^{-1}+q_k|\psi_k|^2\right)}{\log\rho} \\ &\ge 1 + \mathrm{E}\left(\liminf_{\rho\to\infty}\frac{\log(\rho^{-1}+q_k|\psi_k|^2)}{\log\rho}\right) \\ &\ge 1. \end{aligned} \tag{60}$$

In the above, the penultimate step is an application of Fatou's Lemma [24] and the last step applies (59). Using (60) and the fact that

$$R_{\mathsf{tot}}(\rho) \;=\; \frac{1}{N}\sum_{k=N_m}^{N_s-1} R_k(\rho), \tag{61}$$

we obtain

$$\lim_{\rho\to\infty} \frac{R_{\mathsf{tot}}(\rho)}{\log\rho} \;\geq\; \frac{N_s-N_m}{N}. \tag{62}$$

Thus the proposed communication strategy attains a spectral efficiency of at least (62).

On the other hand, consider that perfect CSI is available at the receiver through a genie. In this situation, well known results in [19], [25] dictate that the spectral efficiency of a communication strategy that transmits $(N_s - N_m)$ data substreams over N channel-uses cannot exceed $\frac{N_s-N_m}{N}$ even with optimal joint decoding. The proposed communication strategy has poorer performance than the genie aided strategy since it uses imperfect CSI, and can only have a poorer spectral efficiency. This observation leads us to conclude that

$$\lim_{\rho\to\infty} \frac{R_{\mathsf{tot}}(\rho)}{\log\rho} \;=\; \frac{N_s-N_m}{N}. \tag{63}$$

This completes the proof.

Appendix C
Proof for Proposition 1

We need to demonstrate that $\operatorname{rank}\big(\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\big) = N_m = N_h(2D+1)$. Recall that the pilot pattern used is

$$s_k(i) \;=\; \sqrt{\frac{N_h E_p}{N_m}}\delta_{\langle k\rangle_{N_h}}, \qquad 0 \leq k < N_m. \tag{64}$$

Recalling the structure of $\boldsymbol{S}(i)$ and the fact that, for the CE-BEM channel, $\boldsymbol{B} = \boldsymbol{I}_{N_h} \otimes \boldsymbol{F}$, where the $N_s \times (2D+1)$ matrix $\boldsymbol{F}$ is defined element-wise as $[\boldsymbol{F}]_{m_1,m_2} = \frac{1}{\sqrt{N_s}} e^{j\frac{2\pi}{N_s}m_1(m_2-D)}$. Under these conditions, it is straightforward to show that

$$\boldsymbol{S}_{N_p}(i)\boldsymbol{B} \;=\; \sqrt{\frac{N_h E_p}{N_m}}\boldsymbol{P}\left(\boldsymbol{I}_{N_h}\otimes\boldsymbol{M}\right), \tag{65}$$

In (65), the $(2D+1)N_h \times (2D+1)N_h$ row-permutation matrix $\boldsymbol{P}$ is defined element-wise by $[\boldsymbol{P}]_{m_1,m_2(2D+1)+m_3} = [\boldsymbol{I}_{(2D+1)N_h}]_{m_1,m_3N_h+m_2}$ where $0 \leq m_1 < (2D+1)N_h$, $0 \leq m_2 < N_h$ and $0 \leq m_3 < (2D+1)$. Furthermore, the $(2D+1)\times(2D+1)$ complex matrix $\boldsymbol{M}$ is defined element-wise as $[\boldsymbol{M}]_{m_1,m_2} = \frac{1}{\sqrt{N_s}}\exp(\frac{j2\pi N_h m_1 m_2}{N_s})$. Then,

$$\begin{aligned}
\operatorname{rank}\big(\boldsymbol{S}_{N_p}(i)\boldsymbol{B}\big) \;&=\; \operatorname{rank}(\boldsymbol{P}(\boldsymbol{I}_{N_h}\otimes\boldsymbol{M})) && (66)\\
&=\; \operatorname{rank}(\boldsymbol{I}_{N_h}\otimes\boldsymbol{M}) && (67)\\
&=\; N_h \operatorname{rank}(\boldsymbol{M}) && (68)\\
&=\; N_h(2D+1) = N_m. && (69)
\end{aligned}$$

In the above, (66) is obtained using (65), (67) is a result of $\boldsymbol{P}$ being, by definition, a permutation of the columns of $\boldsymbol{I}_{(2D+1)N_h}$, (68) is a standard result for block-diagonal matrices and the final step is a result of $\boldsymbol{M}$ being a full rank Vandermonde matrix.

Then applying Lemma 1 and Theorem 1 we see that when the proposed communication strategy is used for single carrier transmission over doubly selective fading channels,

$$\lim_{\rho\to\infty} \frac{R_{\mathsf{tot}}(\rho)}{\log\rho} \;=\; \frac{N_s-N_m}{N}. \tag{70}$$

This concludes the proof.

Appendix D
Proof for Proposition 2

Recall that the post-combining SINR for the N_p^{th} substream is

$$\gamma^{(N_p)}(i) \;=\; \frac{|\boldsymbol{c}^{(N_p)}(i)^H\hat{\boldsymbol{h}}^{(N_p)}_{N_p}(i)|^2\sigma_s^2}{\boldsymbol{c}^{(N_p)}(i)^H\boldsymbol{\Sigma}_{\boldsymbol{v}^{(N_p)}(i)|\boldsymbol{s}_k(i),\hat{\boldsymbol{h}}^{(k)}(i)}\boldsymbol{c}^{(N_p)}(i)}. \tag{71}$$

For the sub-optimal combiner $\boldsymbol{c}^{(N_p)}(i) = \boldsymbol{e}_N^{(N_p)}$, the N_p^{th} column of $\boldsymbol{I}_N$, only observations $\{y_{N_p}(i)\}_{i=0}^{N_b-1}$ are used to decode the N_p^{th} substream. Recall from the system model that $y_{N_p}(i)$ is influenced by symbols $\{s_m(i)\}_{m=N_p-N_h+1}^{N_p}$, of which, all the symbols except $s_{N_p}(i)$ are known as pilots. For the pilot pattern used (34), $s_k(i) = 0$ for $N_p - N_h + 1 \leq k \leq N_p - 1$. As a result, only the estimation error from the estimate of $h_{N_p,0}(i)$ affects $y_{N_p}(i)$. Under these circumstances, it is straightforward to show that

$$\gamma^{(N_p)}(i) \;\geq\; \frac{\hat{\sigma}^2_{N_p}\sigma_s^2\zeta}{\tilde{\sigma}^2_{N_p}\sigma_s^2 + \sigma^2}, \tag{72}$$

In (72), ζ is a zero-mean complex Gaussian random variable with unit variance, $\hat{\sigma}^2_{N_p}$ is the variance of the estimate of $h_{N_p,0}(i)$, and $\tilde{\sigma}^2_{N_p}$ the variance of the corresponding estimation error. The variances $\hat{\sigma}^2_{N_p}$ and $\tilde{\sigma}^2_{N_p}$ can be calculated as per their definitions in (37) and (38). In doing so, the covariance matrices of the estimates used to decode the N_p^{th} substream and the corresponding estimation error for the pilot pattern in (34) and the CE-BEM channel is given by

$$\begin{aligned}
\boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} \;&=\; \boldsymbol{\Sigma}_{\boldsymbol{h}(i)} - \boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} && (73)\\
\boldsymbol{\Sigma}_{\tilde{\boldsymbol{h}}^{(N_p)}(i)|\boldsymbol{s}_{N_p}(i)} \;&=\; \boldsymbol{B}\bigg[\frac{(2D+1)N_h}{N_s}\boldsymbol{I}_{N_m} + \frac{\alpha_p E_{\mathsf{tot}}}{2D+1}\\
&\qquad \times \boldsymbol{I}_{N_h}\otimes(\boldsymbol{M}^H\boldsymbol{M})\bigg]^{-1}\boldsymbol{B}^H, && (74)
\end{aligned}$$

where, the $(2D+1)\times(2D+1)$ matrix $\boldsymbol{M}$ is defined element-wise as $[\boldsymbol{M}]_{m_1,m_2} = \frac{1}{\sqrt{N_s}}\exp(\frac{j2\pi N_h m_1 m_2}{N_s})$. (See Appendix C for details.) The choice $\alpha_{p,*}$ that maximizes this lower-bound on the post-combining SINR and consequently the lower-bound on the achievable rate is

$$\alpha_{p,*} \;=\; \arg\max_{\alpha_p\in(0,1)} \frac{\hat{\sigma}^2_{N_p}\sigma_s^2}{\tilde{\sigma}^2_{N_p}\sigma_s^2+\sigma^2}. \tag{75}$$

This completes the proof.

References

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

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

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

[4] 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.

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

[6] 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.

[7] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 453–466, Mar. 2005.

[8] 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 Processing Adv. in Wireless Commun.*, pp. 388–392, June 2003.

[9] A. P. Kannu and P. Schniter, "Capacity analysis of MMSE pilot patterns for doubly selective channels," in *Proc. IEEE Workshop Signal Processing Adv. in Wireless Commun.*, 2005.

[10] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective channels," in *Proc. Allerton Conf. Commun., Control, and Computing*, 2006.

[11] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Select. Areas In Commun.*, vol. 19, pp. 1706–1717, Sept. 2001.

[12] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas In Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.

[13] K. Fu and A. Anastasopoulos, "Analysis and design of LDPC codes for time-selective complex-fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1175–1185, May 2005.

[14] A. Krishnamoorthy and A. Anastasopoulos, "Code and receiver design for the noncoherent fast-fading channel," *IEEE J. Select. Areas In Commun.*, vol. 23, pp. 1769–1778, Sept. 2005.

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

[16] T. Li and O. M. Collins, "A successive decoding strategy for channels with memory," *IEEE Trans. Inform. Theory*, vol. 53, pp. 628–646, Feb. 2007.

[17] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective block-fading channels," *IEEE Trans. Inform. Theory*. Submitted.

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

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

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

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

[22] T. A. Thomas, T. P. Krauss, and F. W. Vook, "CHAMPS: A near-ML joint Doppler frequency/TOA search for channel characterization," in *Proc. IEEE Veh. Tech. Conf.*, Fall 2003.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.

[24] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 3rd ed., 1976.

[25] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," tech. rep., AT&T Bell Labs, June 1995.

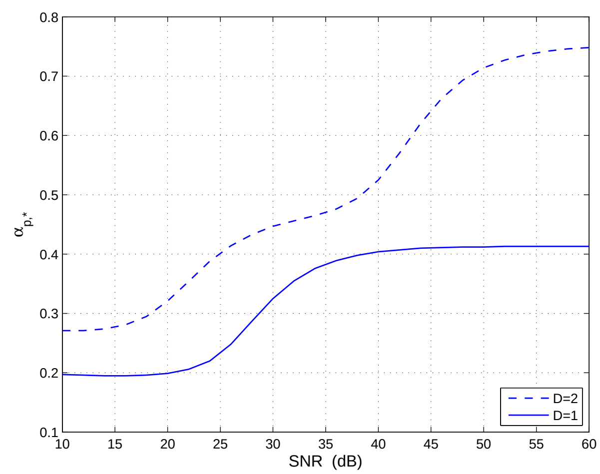

Fig. 1. Power allocation parameter $\alpha_{p,*}$ at various SNRs for a $N_s = 128$-substream spectrally efficient transmission over a CE-BEM doubly selective fading channel with $N_h = 8$ taps of ISSI and $D \in \{1, 2\}$.