

How Much Coherent Interval Should be Dedicated to Non-Redundant Diagonal Precoding for Blind Channel Estimation in Single-Carrier Block Transmission?

Jwo-Yuh Wu, *Member, IEEE*

Abstract—Transmit precoding is a key technique for facilitating blind channel estimation at the receiver but the impact due to precoding on the channel capacity is scarcely addressed in the literature. In this paper we consider the single-carrier block transmission with cyclic prefix, in which a recently proposed diagonal-precoding assisted blind channel estimation scheme via covariance matching is adopted to acquire the channel information. It is shown that, when perfect channel knowledge is available at the receiver, the optimal noise resistant precoder proposed in the literature incurs the worst-case capacity penalty. When the coherent interval is finite, channel mismatch occurs due to finite-sample covariance matrix estimation. Thus, we aim to determine how much of the coherent interval should be dedicated to precoding in order to trade channel estimation accuracy for the maximal capacity. Toward this end, we leverage the matrix perturbation theory to derive a closed-form capacity measure which explicitly takes account of the channel uncertainty in the considered blind estimation setup. Such a capacity metric is seen to be a complicated function of the precoding interval. To facilitate analysis, an approximate formula for the derived capacity measure is further given. This allows us to find a closed-form estimate of the capacity-maximizing precoding time fraction, and can also provide insights into the optimal tradeoff between channel estimation accuracy and achievable capacity. Numerical simulations are used for evidencing the proposed analytic study.

Index Terms—Blind channel estimation; channel capacity; precoding; single-carrier block transmission; cyclic prefix; matrix perturbation analysis; sample covariance matrix; circulant matrix.

I. INTRODUCTION

A. Motivation and Paper Contributions

BLIND channel estimation is widely known as a bandwidth-efficient alternative as opposed to the training technique for acquiring the channel information at the receiver [11]. Among the existing blind estimation methods, the transmit-precoding assisted solutions attracted considerable

attention in recent years [11]. Various channel estimation algorithms associated with different precoding schemes have been proposed, either in the serial transmission case [7], [17], [25], [30], or in the block transmission counterpart, e.g., [4], [18], [23], [26], [31], [33]. Unlike the multi-channel receive-diversity estimation algorithms, e.g., [20], [28], the transmit-precoding assisted approach is quite robust against channel order mismatch and the channel zero locations [17], [25], [30]. Nevertheless, most of the precoding based methods are developed under the assumption that certain second-order statistics of the received signal can be perfectly obtained. The resultant channel estimation performance, therefore, is dominated by the finite-sample estimation errors in the computed data statistics. Moreover, symbol precoding at the transmitter may have significant impact on the channel capacity [6]. Under the perfect channel knowledge assumption the capacity performances attained by various redundant and non-redundant precoding schemes were analyzed in [6]. More in-depth study of the achievable system capacity that explicitly takes into account the channel mismatch effect in the context of precoding-based blind estimation has not been seen in the literature yet.

In this paper we consider the single-carrier block transmission with cyclic prefix (CP)¹ [10], in which the transmitter implements a non-redundant diagonal precoder and, at the receiver the channel information is acquired through the precoding-assisted blind estimation scheme [31]. The main purpose of this paper is to investigate the optimal noise-resistant precoder [31] from a capacity perspective, and to further characterize an inherent tradeoff between channel estimation quality and the achievable system capacity. Specifically, it is shown that when the received signal covariance matrix is perfectly obtained, thus the channel estimate is exact², the optimal noise-resistant precoder results in the minimal capacity in the high SNR regime. Hence, if we adopt the considered precoder to improve the channel estimation accuracy, there is a potential loss in the achievable information

Manuscript received July 8, 2009; revised November 21, 2009; accepted April 29, 2010. The associate editor coordinating the review of this paper and approving it for publication was C. Tellambura.

This work is sponsored by the National Science Council of Taiwan under grants NSC 97-2221-E-009-101-MY3; by the Ministry of Education of Taiwan under the MoE ATU Program; by the Academia-Research Cooperation Program of MoEA under grant 98-EC-17-A-02-S2-0050; by the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd. under grant TL-99-G107; and by the MediaTek Research Center at National Chiao Tung University, Taiwan.

The author is with the Department of Electrical Engineering, and the Institute of Communications Engineering, National Chiao Tung University, Taiwan (e-mail: jywu@cc.nctu.edu.tw).

Digital Object Identifier 10.1109/TWC.2010.061010.091019

¹CP-based single carrier systems have been considered as one next-generation wireless standard, e.g., SC-OFDMA in LTE uplink [19].

²It is well-known that all blind estimation schemes can identify the channel only up to a scalar ambiguity, which has to be removed by further inserting some pilot symbols [11]. As in many previous works regarding performance analysis of blind algorithms [7], [32], we assume for analytical simplicity that the ambiguity is removed. In this sense, the channel estimate is considered to be exact if the received signal covariance matrix can be obtained without errors.

rate. We then turn to consider the more realistic block fading environment: (1) the channel remains constant over a finite time duration and can change independently across different time frames; (2) during each coherent interval the blind algorithm [31] is implemented for obtaining the channel estimate. Clearly, if all the symbol blocks within a coherent interval are precoded, and the receiver uses all the available data blocks to form a sample covariance matrix, one can come up with the utmost channel estimation accuracy since the covariance matrix estimation error is kept as small as possible. Such an advantage, however, comes at the expense of capacity loss due to precoding. On the contrary, if only a certain fraction of the coherent time is spent for precoding, the quality of channel estimation could be relatively poor though the precoding-induced capacity loss can be reduced. This naturally motivates the following question: How much the coherent interval should be dedicated to precoding so that one can achieve the optimal tradeoff between channel estimation accuracy and the maximal system capacity ?

To pin down such a tradeoff, one needs a capacity measure that can explicitly reflect the channel mismatch effect in the considered blind estimation setup. By leveraging the matrix perturbation analysis and following the technique in [12], we derive one such capacity measure, which is seen to be a complicated function of the precoding interval. To facilitate analysis, we further derive an approximate expression for the obtained capacity metric that has the following threefold advantages. Firstly, it allows us to specify the aggregate capacity cost incurred by the considered blind channel estimation scheme as a sum of two terms, one due to precoding whereas the other caused by channel estimation errors (or finite-sample covariance matrix estimation). Secondly, based on this decomposition there is a simple procedure for obtaining an analytic estimate of the capacity-maximizing time fraction spent for precoding. Thirdly, there are very informative interpretations regarding the two penalty terms that will provide further insights into the optimal tradeoff. The capacity results predicted by our analysis are further corroborated via numerical simulations.

B. Paper Organization and Notation List

The rest of this paper is organized as follows. Section II introduces the signal model, briefly reviews the blind channel estimation scheme in [31], and highlights the essentials regarding the optimal noise-resistant precoder. Section III discusses the capacity performance of the considered precoder, assuming that perfect channel knowledge is available at the receiver. Section IV considers the block fading channel environment and derives the capacity measure for specifying the design tradeoff. Section V addresses the optimal precoding interval selection problem. Section VI provides several numerical experiments for evidencing the proposed analytical results. Finally, Section VII concludes this paper.

Notation: Let $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ be the sets of real and complex matrices. Denote by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$, respectively the transpose, complex conjugate, and Hermitian operations. The symbols \mathbf{I}_m and \mathbf{O}_m denote the $m \times m$ identity and zero matrices; $\mathbf{O}_{m \times n}$ is the $m \times n$ zero matrix. The notation \otimes stands for the Kronecker product [14, p-242], and $\text{vec}(\mathbf{X})$

is the vectorized operation of the matrix \mathbf{X} [14]. For $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n] \in \mathbb{C}^{m \times n}$ and $\mathbf{Y} = [\mathbf{y}_1 \ \cdots \ \mathbf{y}_n] \in \mathbb{C}^{k \times n}$, $\mathbf{X} \square \mathbf{Y} := [\mathbf{x}_1 \otimes \mathbf{y}_1 \ \cdots \ \mathbf{x}_n \otimes \mathbf{y}_n]$ denotes the column-wise Kronecker product [32]. For $\mathbf{x} \in \mathbb{C}^m$, let $\text{diag}\{\mathbf{x}\}$ be the $m \times m$ diagonal matrix with the elements of \mathbf{x} on the main diagonal. The notation $E\{y\}$ stands for the expected value of the random variable y . Denote by $\text{Tr}[\mathbf{M}]$ the trace of the square matrix \mathbf{M} .

II. DIAGONAL PRECODING BASED BLIND CHANNEL ESTIMATION

A. Signal Model

We consider a precoded single-carrier CP-based system over an L -order frequency-selective fading channel described as [31]

$$\mathbf{y}_k = \mathbf{G}\mathbf{P}\mathbf{s}_k + \mathbf{v}_k, \quad k \geq 0, \quad (2.1)$$

where $\mathbf{s}_k \in \mathbb{C}^N$ and $\mathbf{y}_k \in \mathbb{C}^N$ are, respectively, the source symbol and the received signal blocks (with N denoting the dimension of the source symbol vector), $\mathbf{v}_k \in \mathbb{C}^N$ is the noise vector,

$$\mathbf{P} := \text{diag}\{[p(0) \ \cdots \ p(N-1)]\} \in \mathbb{R}^N, \quad p(n) \in \mathbb{R}, \quad (2.2)$$

is a diagonal precoding matrix, and $\mathbf{G} \in \mathbb{C}^{N \times N}$ is the circulant channel matrix whose first column is given by

$$\mathbf{g} := [h(0) \ \cdots \ h(L) \ 0 \ \cdots \ 0]^T \in \mathbb{C}^N, \quad (2.3)$$

with $h(n)$ being the n th channel tap, $0 \leq n \leq L$. The purpose of diagonal precoding is to deliberately induce certain transmit power variation so as to facilitate blind channel estimation at the receiver. The following assumptions are made throughout the paper.

- The source \mathbf{s}_k is a white vector sequence with zero mean and unit variance.
- The noise \mathbf{v}_k is white circularly complex Gaussian with zero mean, covariance $\sigma_v^2 \mathbf{I}$, and is independent of the source signal \mathbf{s}_k .

B. Blind Channel Estimation Algorithm

The approach in [31] exploits the circulant structure of the channel matrix \mathbf{G} as well as a resultant covariance-matching channel estimation setup. More specifically, since \mathbf{G} is circulant, it can be expressed in terms of its first column as

$$\mathbf{G} = [\mathbf{g} \ \mathbf{J}\mathbf{g} \ \cdots \ \mathbf{J}^{N-2}\mathbf{g} \ \mathbf{J}^{N-1}\mathbf{g}], \quad (2.4)$$

where

$$\mathbf{J} := \begin{bmatrix} \mathbf{0}_{1 \times (N-1)} & 1 \\ \mathbf{I}_{N-1} & \mathbf{0}_{(N-1) \times 1} \end{bmatrix} \in \mathbb{R}^{N \times N} \quad (2.5)$$

is the permutation matrix. We assume for the moment that the channel noise is absent, hence $\mathbf{v}_k = \mathbf{0}$. With (2.1), (2.2), and (2.4), the covariance matrix of the received signal is easily shown to be

$$\begin{aligned} \mathbf{R}_y &:= E\{\mathbf{y}_k \mathbf{y}_k^H\} \\ &= \mathbf{G}\mathbf{P}^2\mathbf{G}^H = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g} \mathbf{g}^H (\mathbf{J}^T)^n. \end{aligned} \quad (2.6)$$

We observe that, for a given \mathbf{R}_y , (2.6) defines a set of *linear* equations with the *product channel coefficients* $h(k)h(l)^*$ (i.e., entries in $\mathbf{g}\mathbf{g}^H$) as unknowns. To exploit such an inherent linear signal structure for channel estimation, we shall first rearrange (2.6) into a standard linear equation form. By resorting to the $\text{vec}(\cdot)$ operation, (2.6) can be rearranged into (2.7), shown at the bottom of this page.

To solve for the product unknowns $h(k)h(l)^*$ via (2.7), we have to first reduce the dimension of the equations by removing the null variables in $\text{vec}(\mathbf{g}\mathbf{g}^H)$. Specifically, it can be shown that (2.7) is equivalent to [31]

$$\tilde{\mathbf{Q}} \text{vec}(\mathbf{h}\mathbf{h}^H) = \text{vec}(\mathbf{R}_y), \quad (2.8)$$

where $\mathbf{h} := [h(0) \ \dots \ h(L)]^T$ is the desired channel impulse response vector, and

$$\tilde{\mathbf{Q}} := \mathbf{Q}\mathbf{J}_1 (\mathbf{I}_{L+1} \otimes \mathbf{J}_2) \quad (2.9)$$

with

$$\mathbf{J}_1 := \begin{bmatrix} \mathbf{I}_{N(L+1)} \\ \mathbf{0}_{N(N-L-1) \times N(L+1)} \end{bmatrix} \in \mathbb{R}^{N^2 \times N(L+1)}$$

$$\mathbf{J}_2 := \begin{bmatrix} \mathbf{I}_{L+1} \\ \mathbf{0}_{(N-L-1) \times (L+1)} \end{bmatrix} \in \mathbb{R}^{N \times (L+1)}.$$

Assume that the matrix $\tilde{\mathbf{Q}}$ is of full column rank; an associated sufficient condition characterized in terms of the precoder $p(n)$ is established in [31, Proposition 4.1]. Then the product channel coefficient vector can be computed as

$$\text{vec}(\mathbf{h}\mathbf{h}^H) = (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \text{vec}(\mathbf{R}_y). \quad (2.10)$$

Once $\text{vec}(\mathbf{h}\mathbf{h}^H)$ is obtained, let us form the rank-one matrix

$$\mathbf{H} := \mathbf{h}\mathbf{h}^H = [h(k)h(l)^*]_{0 \leq k, l \leq L}. \quad (2.11)$$

The channel impulse response vector \mathbf{h} can then be estimated, up to a scalar ambiguity, by computing the dominant eigenvector associated with the matrix \mathbf{H} . We note that a similar matrix outer-product approach has also been adopted in the previous works [9], [16], [17], [26], [32]. To obtain the full channel estimate, some pilot symbols should be inserted in the symbol block for removing the scalar ambiguity; the detail is referred to [31, p-1119].

C. Optimal Noise-Resistant Precoder

With noise corruption the autocorrelation matrix in (2.6) instead reads

$$\mathbf{R}_y = E \{ \mathbf{y}_k \mathbf{y}_k^H \} = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g}\mathbf{g}^H (\mathbf{J}^T)^n + \sigma_v^2 \mathbf{I}. \quad (2.12)$$

The identification equation (2.8) is then accordingly modified as

$$\text{vec}(\mathbf{R}_y) = \tilde{\mathbf{Q}} \text{vec}(\mathbf{h}\mathbf{h}^H) + \sigma_v^2 \text{vec}(\mathbf{I}). \quad (2.13)$$

In (2.13) we can think of the product channel coefficients $\text{vec}(\mathbf{h}\mathbf{h}^H)$ as the signal of interest, hence the range space of $\tilde{\mathbf{Q}}$ defining the *signal subspace*, and treat the white-noise perturbation $\text{vec}(\mathbf{I})$ as spanning the *noise subspace*. It is noted that the matrix $\tilde{\mathbf{Q}}$ depends entirely on the precoder coefficients $p(n)$'s. To mitigate the noise effect on the estimated channel, one natural approach is thus to design $p(n)$ so that the signal and noise subspaces are rendered as close to being orthogonal as possible. In [31] the precoder design problem is specifically formulated as minimizing the largest correlation index among the noise signature $\text{vec}(\mathbf{I})$ and the columns of $\tilde{\mathbf{Q}}$, subject to the power normalization constraint

$$\sum_{n=0}^{N-1} p(n)^2 = N \quad (2.14)$$

and the threshold requirement

$$p(n)^2 \geq \delta > 0, \quad \forall 0 \leq n \leq N-1, \quad (2.15)$$

where δ is the power threshold, which is constrained to be strictly positive to avoid symbol nulling. The optimal noise-resistant precoder admits the following two-level form (see [31, p-1121]): for a fixed but arbitrary $0 \leq m \leq N-1$,

$$p(m)^2 = N(1-\delta) + \delta \quad \text{and} \quad p(n)^2 = \delta \quad \text{for} \quad n \neq m. \quad (2.16)$$

The rest of this paper aims to study the optimal noise-resistant precoder (2.16) from a capacity perspective, and to address the optimal tradeoff between channel estimation accuracy and achievable capacity when the coherent interval is finite and \mathbf{R}_y is estimated via a finite amount of data.

Remark: With the optimal precoder (2.16), it is shown in [31] that a small threshold δ results in small noise corruption on the product channel coefficients, thereby improving the estimation accuracy. However, an unlimitedly small δ should be avoided since it not only renders the symbol decision process quite prone to noise (see [7], [31]), but will also incur a high capacity penalty as will be shown next. For $N = 32$ and $L = 8$ (with the CP interval set equal to the channel order), our simulation study in Section VI (see Figures 4 and 5) indicates that $\delta \approx 0.9$ is the compromising choice regarding the tradeoff between channel estimation accuracy and the achievable capacity.

$$\underbrace{\begin{bmatrix} p(0)^2 \mathbf{I}_N & p(N-1)^2 \mathbf{J}^{N-1} & \dots & p(2)^2 \mathbf{J}^2 & p(1)^2 \mathbf{J} \\ p(1)^2 \mathbf{J} & p(0)^2 \mathbf{I}_N & \dots & p(3)^2 \mathbf{J}^3 & p(2)^2 \mathbf{J}^2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ p(N-2)^2 \mathbf{J}_{N-2} & p(N-3)^2 \mathbf{J}^{N-3} & \dots & p(0)^2 \mathbf{I}_N & p(N-1)^2 \mathbf{J}^{N-1} \\ p(N-1)^2 \mathbf{J}_{N-1} & p(N-2)^2 \mathbf{J}^{N-2} & \dots & p(1)^2 \mathbf{J} & p(0)^2 \mathbf{I}_N \end{bmatrix}}_{:=\tilde{\mathbf{Q}}} \text{vec}(\mathbf{g}\mathbf{g}^H) = \text{vec}(\mathbf{R}_y). \quad (2.7)$$

III. CAPACITY PERFORMANCE WITH EXACT CHANNEL KNOWLEDGE

This section investigates the impact of the precoder (2.16) on capacity, assuming that the channel is perfectly known to the receiver. In such an idealized case, the ergodic capacity (in bits per block transmission, neglecting the CP overhead) of the system (2.1) is well-known to be [27]

$$I = E \left\{ \log \det(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H) \right\}. \quad (3.1)$$

Since \mathbf{G} is circulant, we have $\mathbf{G} = \mathbf{F}^H \mathbf{D} \mathbf{F}$, where \mathbf{F} is the FFT matrix and \mathbf{D} is a diagonal matrix containing the channel frequency responses. This implies

$$\mathbf{G} \mathbf{P}^2 \mathbf{G}^H = \mathbf{F}^H \mathbf{D} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \mathbf{D}^H \mathbf{F}, \quad (3.2)$$

and (3.1) becomes

$$\begin{aligned} I &= E \left\{ \log \det(\mathbf{I} + \sigma_v^{-2} \mathbf{F}^H \mathbf{D} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \mathbf{D}^H \mathbf{F}) \right\} \\ &\stackrel{(a)}{=} E \left\{ \log \det(\mathbf{I} + \sigma_v^{-2} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \mathbf{D}^H \mathbf{F} \mathbf{F}^H \mathbf{D}) \right\} \\ &\stackrel{(b)}{=} E \left\{ \log \det(\mathbf{I} + \sigma_v^{-2} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \tilde{\mathbf{D}}) \right\}, \end{aligned} \quad (3.3)$$

where (a) holds due to $\det(\mathbf{I} + \mathbf{A} \mathbf{B}) = \det(\mathbf{I} + \mathbf{B} \mathbf{A})$ for \mathbf{A} and \mathbf{B} with compatible dimensions, and (b) follows since $\mathbf{F} \mathbf{F}^H = \mathbf{I}$ and by defining $\tilde{\mathbf{D}} := \mathbf{D}^H \mathbf{D}$, which is a positive definite diagonal matrix. According to the Hadamard's inequality [13, p-477], the term $\log \det(\mathbf{I} + \sigma_v^{-2} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \tilde{\mathbf{D}})$ in the capacity expression (3.3) is maximized if $p(n)$ is chosen so that $\mathbf{F} \mathbf{P}^2 \mathbf{F}^H \tilde{\mathbf{D}}$ is diagonal. Since $\tilde{\mathbf{D}}$ is diagonal, the matrix $\mathbf{F} \mathbf{P}^2 \mathbf{F}^H \tilde{\mathbf{D}}$ can be diagonalized only when $\mathbf{F} \mathbf{P}^2 \mathbf{F}^H$ is also diagonal. Subject to the fact that $\mathbf{F} \mathbf{P}^2 \mathbf{F}^H$ is *circulant* [8], the only such diagonal $\mathbf{F} \mathbf{P}^2 \mathbf{F}^H$ is a scalar multiple of the identity matrix. As a result, the capacity-maximizing $p(n)$ should be chosen so that

$$\mathbf{F} \mathbf{P}^2 \mathbf{F}^H = \alpha \mathbf{I} \quad \text{for some } \alpha > 0. \quad (3.4)$$

The unique $p(n)$ which satisfies (3.4) as well as the two constraints (2.14) and (2.15) is

$$p(n) = 1, \quad \text{for } 0 \leq n \leq N-1. \quad (3.5)$$

Hence, for i.i.d. sources, the equal-power scheme (3.5) is capacity-optimal³. The modulated symbol power induced by diagonal precoding, therefore, will inevitably incur a capacity loss (i.e., precoding is harmful from the capacity perspective). To characterize the capacity performance of the two-level precoder (2.16) we further note that, in the high SNR region, (3.3) is well approximated by [22]

$$\begin{aligned} I &\approx E \left\{ \log \det \left(\sigma_v^{-2} \mathbf{F} \mathbf{P}^2 \mathbf{F}^H \tilde{\mathbf{D}} \right) \right\} \\ &= \log \left\{ \sigma_v^{-2N} \prod_{n=0}^{N-1} p(n)^2 \right\} + E \left\{ \log \det \tilde{\mathbf{D}} \right\}. \end{aligned} \quad (3.6)$$

From (3.6) we can see that, when SNR is high, the impact on the capacity due to precoding is entirely characterized by the product term $\prod_{n=0}^{N-1} p(n)^2$. The smaller such a product

is, the larger the capacity penalty will be. The optimal noise-resistant precoder (2.16), however, turns out to be the worst-case choice regarding channel capacity, as established in the next theorem (see Appendix A for a proof).

Theorem 3.1: Among the $p(n)$ satisfying the constraints (2.14) and (2.15), the precoder (2.16) minimizes the quantity $\prod_{n=0}^{N-1} p(n)^2$, yielding

$$\min \prod_{n=0}^{N-1} p(n)^2 = \delta^{N-1} [N - (N-1)\delta]. \quad (3.7)$$

Discussions:

1. It is easy to check that the minimal product in (3.7) increases as the threshold δ is increased toward unity; when $\delta = 1$, (2.16) reduces to the capacity-optimal equal-power scheme (3.5). This implies that a large δ , though resulting in poor channel estimation accuracy [31, Sec. V], will limit the capacity penalty incurred by the precoding scheme (2.16). When channel estimation error occurs, the compromising choice of δ between the channel estimation accuracy and achievable capacity is investigated in the simulation section.
2. The capacity (3.1) obtained under the perfect channel knowledge assumption can be regarded as a yardstick upper bound for the realistic situation when channel error occurs. In this sense, the optimal noise-resistant precoder (2.16) leads to the worst-case benchmark capacity when SNR is high. Hence, if one adopts the precoder (2.16) for improving the channel estimation accuracy, there could be a substantial loss in the achievable system capacity.
3. The acquisition of exact channel knowledge via the blind technique shown in Section III calls for an infinitely long coherent interval, during which the received covariance matrix \mathbf{R}_y in (2.6) can be estimated, e.g., via the time average, without errors. In the finite coherent interval case, we can only obtain a sample covariance matrix by using a finite number of data blocks, and the resultant channel estimate will be no longer exact. If a large portion of the source symbol blocks within the coherent interval are precoded by (2.16), the finite sample error in estimating \mathbf{R}_y is relatively small, and a more accurate channel estimate can be obtained. However, this would come at the expense of reduced capacity since (2.16) is the worst-case choice regarding the benchmark capacity performance. On the other hand, if a small fraction of the coherent interval is spent for precoding, the quality of the channel estimate could be quite poor, though the capacity penalty due to precoding can be limited. The optimal tradeoff between channel estimation accuracy and the achievable system capacity regarding the precoding interval selection is elaborated on next.

IV. CAPACITY MEASURE OVER FINITE COHERENT INTERVALS

In the sequel we focus on the block fading environment, in which the channel remains constant over some interval of T symbol block periods, after which it changes independently to another value that it holds for the next interval T , and so on. In

³If the uniform scheme (3.5) is used, the matrix $\tilde{\mathbf{Q}}$ in (2.9) will however be rank deficient and the channel is then rendered unidentifiable.

the training based counterpart, the coherent interval is typically divided into two phases, one for placing training pilots and the other for carrying data symbols, and the capacity-optimal training period has been addressed in [1], [12], [28] for various scenarios. Motivated by these works and to study the optimal tradeoff problem in the considered precoding based blind estimation setup, we thus assume that, during each coherent interval, only the initial $1 < T_p \leq T$ source symbol blocks are precoded by (2.16) to facilitate channel estimation. The signal model within a coherent interval of T blocks can then be described in a two-phase form as⁴ (4.1), shown at the bottom of this page. The T_p received blocks in the precoding phase are used to form the sample covariance matrix

$$\widehat{\mathbf{R}}_y = \frac{1}{T_p} \sum_{k=1}^{T_p} \mathbf{y}_k \mathbf{y}_k^H \quad (4.2)$$

for channel estimation via the blind technique shown in Section III; the remaining $T - T_p$ time slots are left for direct data transmission to boost capacity. To seek for the optimal T_p which attains the optimal tradeoff between blind channel estimation accuracy and the achievable capacity, we need a capacity metric that can explicitly take the channel mismatch effect (or imperfect estimation of $\widehat{\mathbf{R}}_y$) into account. This is the main focus of this section.

A. Capacity Measure

For a given channel estimate $\widehat{\mathbf{h}}$, and hence the associated channel matrix $\widehat{\mathbf{G}}$, let us follow the idea of [1], [12], [28] to rewrite the system (2.1) by treating the estimated $\widehat{\mathbf{G}}$ as the known channel matrix, and relegating the channel mismatch into the noise component so that

$$\begin{aligned} \mathbf{y}_k &= \mathbf{G}\mathbf{P}\mathbf{s}_k + \mathbf{v}_k \\ &= \widehat{\mathbf{G}}\mathbf{P}\mathbf{s}_k + \underbrace{\widetilde{\mathbf{G}}\mathbf{P}\mathbf{s}_k + \mathbf{v}_k}_{:=\widetilde{\mathbf{v}}_k}, \quad \text{where } \widetilde{\mathbf{G}} = \mathbf{G} - \widehat{\mathbf{G}}. \end{aligned} \quad (4.3)$$

We note that, while in (2.1) the channel remains unknown, the channel matrix $\widehat{\mathbf{G}}$ in (4.3) is otherwise known to the receiver. Also, although the additive noise \mathbf{v}_k in (2.1) is white Gaussian, the effective noise $\widetilde{\mathbf{v}}_k$ in (4.3), which depends also on the channel estimation error $\widetilde{\mathbf{G}}$, could be neither white nor Gaussian. To characterize the capacity performance of the channel (4.3), one needs to specify the covariance of the effective noise $\widetilde{\mathbf{v}}_k$. For this we first observe that, since $\widehat{\mathbf{R}}_y$ in (4.2) is unbiased, the computed outer-product $\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H$ as a linear function of $\widehat{\mathbf{R}}_y$ (cf. (2.10)) is also an unbiased estimator of $\mathbf{h}\mathbf{h}^H$. However, the resultant channel estimate $\widehat{\mathbf{h}}$, obtained as the dominant eigenvector of $\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H$, is an unbiased estimator of the true \mathbf{h} only when $T_p \rightarrow \infty$ [2], [15], [24]. For a finite T_p , the statistical property of $\widehat{\mathbf{h}}$ is quite difficult to characterize [24]; exact expressions for the bias term $E\{\widehat{\mathbf{h}} - \mathbf{h}\}$, and

⁴Without loss of generality we consider the initial coherent interval to simplify notation.

consequently the covariance of $\widetilde{\mathbf{v}}_k$, are thus intractable. To facilitate analysis in the finite-sample case, we propose to adopt the approach similar to [21] and [32], in which T_p is assumed to be large so that the unbiased-ness condition $E\{\widehat{\mathbf{h}} - \mathbf{h}\} = \mathbf{0}$ (or $E\{\widetilde{\mathbf{G}}\} = \mathbf{0}$) is deemed to hold from the first-order perturbation perspective⁵. Under this assumption and by using a similar technique as in [1], [12], [28], a capacity lower bound for the channel (4.3) is derived in the next lemma (see Appendix B for a proof).

Lemma 4.1: Let $\bar{I}(\mathbf{y}_k; \mathbf{s}_k | \widehat{\mathbf{G}}) = \max I(\mathbf{y}_k; \mathbf{s}_k | \widehat{\mathbf{G}})$, namely, the maximal mutual information (over the source distributions) between the source and received signals for a fixed channel estimate $\widehat{\mathbf{G}}$. Then the following inequality holds for a large T_p :

$$\begin{aligned} \bar{I}(\mathbf{y}_k; \mathbf{s}_k | \widehat{\mathbf{G}}) &\geq \\ &\log \det \left\{ \mathbf{I} + \left[E \left\{ \widetilde{\mathbf{G}}\mathbf{P}^2\widetilde{\mathbf{G}}^H \right\} + \sigma_v^2\mathbf{I} \right]^{-1} \widehat{\mathbf{G}}\mathbf{P}^2\widehat{\mathbf{G}}^H \right\}. \end{aligned} \quad (4.4)$$

The lower bound in (4.4) is particularly appealing, as it is a function of the channel mismatch $\widetilde{\mathbf{G}} = \mathbf{G} - \widehat{\mathbf{G}}$ and can therefore serve as a capacity measure when channel estimation error occurs. To address the optimal tradeoff based on the capacity lower bound in (4.4), the first task is to find an explicit expression of $E\{\widetilde{\mathbf{G}}\mathbf{P}^2\widetilde{\mathbf{G}}^H\}$ in terms of the precoding interval T_p . This is done in the next subsection.

B. Covariance of the Channel Estimation Error: A Perturbation Analysis

To proceed, we note from (2.4) that the circulant nature of $\widetilde{\mathbf{G}}$ again yields

$$\widetilde{\mathbf{G}} = [\widetilde{\mathbf{g}} \quad \mathbf{J}\widetilde{\mathbf{g}} \quad \cdots \quad \mathbf{J}^{N-2}\widetilde{\mathbf{g}} \quad \mathbf{J}^{N-1}\widetilde{\mathbf{g}}], \quad (4.5)$$

in which \mathbf{J} is defined in (2.5) and $\widetilde{\mathbf{g}} = [\widetilde{\mathbf{h}}^H \quad 0 \quad \cdots \quad 0]^T$ is the zero-padded channel estimation error vector with $\widetilde{\mathbf{h}} = \widehat{\mathbf{h}} - \mathbf{h}$. Since $\widetilde{\mathbf{g}} = \mathbf{J}_2\widetilde{\mathbf{h}}$, where \mathbf{J}_2 is defined in (2.9), based on (4.5) it is straightforward to verify

$$E\{\widetilde{\mathbf{G}}\mathbf{P}^2\widetilde{\mathbf{G}}^H\} = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{J}_2 E\{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}^H\} \mathbf{J}_2^T (\mathbf{J}^T)^n. \quad (4.6)$$

Equation (4.6) shows that $E\{\widetilde{\mathbf{G}}\mathbf{P}^2\widetilde{\mathbf{G}}^H\}$ is completely determined by $E\{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}^H\}$, namely, the covariance of the channel estimation error. Hence it remains to find a closed-form expression of $E\{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}^H\}$ in terms of T_p . For this let us recall from Section III-A that, if the perfect covariance matrix \mathbf{R}_y can be obtained, the exact channel \mathbf{h} is identified via a "two-step" approach: first compute the rank-one outer-product matrix (2.11) followed by an eigen-decomposition for finding

⁵Through our simulation study the normalized average bias per channel tap, namely, $E\{\|\widehat{\mathbf{h}} - \mathbf{h}\|^2\} / [E\{\|\mathbf{h}\|^2\}(L+1)]$, is below -75 dB for a wide SNR range even if T_p is as small as $T_p = 100$. Hence, in the finite-sample case, $E\{\widehat{\mathbf{h}} - \mathbf{h}\} = \mathbf{0}$ is a plausible assumption, and the lower bound (4.4) can be a valid capacity measure.

$$\begin{cases} \mathbf{y}_k = \mathbf{G}\mathbf{P}\mathbf{s}_k + \mathbf{v}_k, & 1 \leq k \leq T_p, \\ \mathbf{y}_k = \mathbf{G}\mathbf{s}_k + \mathbf{v}_k, & T_p + 1 \leq k \leq T, \end{cases} \quad \begin{array}{l} \text{(precoding \& channel acquisition phase)} \\ \text{(direct data transmission phase).} \end{array} \quad (4.1)$$

the associated dominant eigenvector. When only a sample covariance matrix $\widehat{\mathbf{R}}_{\mathbf{y}}$ as in (4.1) is available, the channel mismatch $\tilde{\mathbf{h}}$ is entirely caused by the deviation of $\widehat{\mathbf{R}}_{\mathbf{y}}$ from the true $\mathbf{R}_{\mathbf{y}}$. Although the outer-product $\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H$ is linear in the entries of $\widehat{\mathbf{R}}_{\mathbf{y}}$ (see (2.10)), the eigen-decomposition procedure on $\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H$, however, will render the exact expression for $\tilde{\mathbf{h}}$ in terms of $\widehat{\mathbf{R}}_{\mathbf{y}} = \widetilde{\mathbf{R}}_{\mathbf{y}} - \mathbf{R}_{\mathbf{y}}$ intractable. By assuming T_p to be sufficiently large so that the deviation $\widetilde{\mathbf{R}}_{\mathbf{y}}$ is small, we can nonetheless leverage the matrix perturbation technique [3], [21], [32] to find one approximate, yet tractable, such expression; the result will further enable us to derive a formula of $E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\}$, and hence $E\{\widehat{\mathbf{G}}\mathbf{P}^2\widehat{\mathbf{G}}^H\}$, in terms of T_p . More specifically, through first-order perturbation analysis we have the following linear relation between $\tilde{\mathbf{h}}$ and $\widetilde{\mathbf{R}}_{\mathbf{y}}$ (see Appendix C for a derivation)

$$\tilde{\mathbf{h}} = \mathbf{A} \text{vec}(\widetilde{\mathbf{R}}_{\mathbf{y}}), \quad (4.7)$$

where

$$\mathbf{A} := \frac{1}{\|\tilde{\mathbf{h}}\|^2} \boldsymbol{\Sigma}_{\mathbf{h}} \boldsymbol{\Sigma}_{\mathbf{h}}^H \sum_{i=0}^L h(i) \mathbf{e}_i^T \otimes \mathbf{I}_{L+1} (\widetilde{\mathbf{Q}}^T \widetilde{\mathbf{Q}})^{-1} \widetilde{\mathbf{Q}}^T,$$

in which \mathbf{e}_i denotes the i th column of \mathbf{I}_{L+1} and $\boldsymbol{\Sigma}_{\mathbf{h}} \in \mathbb{C}^{(L+1) \times L}$ is a matrix whose columns are orthonormal and span the L -dimensional subspace orthogonal to \mathbf{h} . Based on (4.7), we have the following lemma (see also Appendix C for a proof).

Lemma 4.2: Assume that the source symbols are drawn from a complex constellation with a finite fourth-order cumulant κ_{4s} . Then we have

$$E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\} = \frac{1}{T_p} \mathbf{A} \{ \kappa_{4s} (\mathbf{G}^* \mathbf{P}^* \square \mathbf{G} \mathbf{P}) (\mathbf{G}^* \mathbf{P}^* \square \mathbf{G} \mathbf{P})^H + \mathbf{R}_{\mathbf{y}}^* \otimes \mathbf{R}_{\mathbf{y}} \} \mathbf{A}^H, \quad (4.8)$$

where the matrix \mathbf{A} is defined in (4.7).

Combining (4.6) and (4.8) leads to the crucial relation (4.9), shown at the bottom of this page. Equation (4.9) specifies the weighted channel error covariance matrix in terms of the precoding interval T_p . In particular, as T_p increases, the error covariance will decay at the rate $1/T_p$.

C. Capacity Lower Bound

With (4.4) and (4.9), the ergodic capacity lower bound (per block transmission) during the precoding phase is thus

$$\underline{I}_p(T_p) = E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/T_p + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \mathbf{P}^2 \widehat{\mathbf{G}}^H \right] \right\}; \quad (4.10)$$

similarly one such bound for the direct data transmission phase can be accordingly obtained from (4.10) by setting $\mathbf{P} = \mathbf{I}$:

$$\underline{I}_d(T_p) = E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/T_p + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \widehat{\mathbf{G}}^H \right] \right\}. \quad (4.11)$$

Based on (4.10) and (4.11), the average ergodic capacity lower bound over the entire T symbol periods is given by

$$\begin{aligned} \underline{I}(T_p) &= \frac{T_p}{T} \underline{I}_p(T_p) + \frac{(T - T_p)}{T} \underline{I}_d(T_p) \\ &= \frac{T_p}{T} E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/T_p + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \mathbf{P}^2 \widehat{\mathbf{G}}^H \right] \right\} \\ &\quad + \frac{(T - T_p)}{T} E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/T_p + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \widehat{\mathbf{G}}^H \right] \right\}. \end{aligned} \quad (4.12)$$

The problem of selecting T_p toward maximizing $\underline{I}(T_p)$ is addressed in the next section.

We note that the capacity measure (4.12) is a function of not only the precoding interval T_p but also the precoder coefficients $p(n)$ (in particular, the power threshold δ if the two-level precoder (2.16) is used). Hence, true capacity maximization should be done based on joint optimization over both T_p and δ . However, as one can see from (4.9), the channel error covariance \mathbf{R}_e involves the multiplication of column-wise Kronecker products of the precoding matrix and is therefore a function of $p(n)^2 p(i) p(j) p(k) p(m)$. This shows that the capacity metric (4.12) is highly nonlinear in terms of δ , and joint optimization $\underline{I}(T_p)$ of over both T_p and δ appears completely intractable. To overcome this difficulty, a reasonable suboptimal approach as we shall adopt in the sequel is to determine the best T_p under a fixed δ (as we will see later, even for such a suboptimal scheme the analysis turns out to be totally nontrivial).

V. SELECTION OF PRECODING INTERVAL

To facilitate subsequent analysis and discussions, let us define

$$\tau_p = T_p/T, \quad 0 < \tau_p \leq 1, \quad (5.1)$$

to be the relative time fraction of the precoding phase normalized with respect to the coherent interval T . We can thus alternatively express the capacity lower bound (4.12) in terms of τ_p as

$$\begin{aligned} \underline{I}(\tau_p) &= \tau_p E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \mathbf{P}^2 \widehat{\mathbf{G}}^H \right] \right\} \\ &\quad + (1 - \tau_p) E \left\{ \log \det \left[\mathbf{I} + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \widehat{\mathbf{G}} \widehat{\mathbf{G}}^H \right] \right\}. \end{aligned} \quad (5.2)$$

For a given T , the optimal tradeoff problem can be formulated as maximizing $\underline{I}(\tau_p)$ in (5.2) with respect to all $0 < \tau_p \leq 1$. We shall note that, since both T_p and T are positive integers, $0 < \tau_p \leq 1$ is a rational number. To ease analysis τ_p is relaxed to be a positive real number; once the best such τ_p is found, the corresponding T_p (though suboptimal) can be determined as the lower integer floor of $\tau_p T$. An exact closed-form solution to the considered optimization problem,

$$E\{\widehat{\mathbf{G}}\mathbf{P}^2\widehat{\mathbf{G}}^H\} = \frac{1}{T_p} \underbrace{\sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{J}_2 \mathbf{A} \{ \kappa_{4s} (\mathbf{G}^* \mathbf{P}^* \square \mathbf{G} \mathbf{P}) (\mathbf{G}^* \mathbf{P}^* \square \mathbf{G} \mathbf{P})^H + \mathbf{R}_{\mathbf{y}}^* \otimes \mathbf{R}_{\mathbf{y}} \} \mathbf{A}^H \mathbf{J}_2^T (\mathbf{J}^T)^n}_{:= \mathbf{R}_e}. \quad (4.9)$$

however, is formidable to derive since the objective function (5.2) is highly nonlinear in τ_p . Even though one can instead resort to numerical simulation to search for a solution, in what follows we shall aim to analytically characterize the optimal τ_p . Specifically, we will first derive an approximate, but more tractable, expression of the capacity lower bound in (5.2). The result will lead to very simple, and insightful, procedures for determining a closed-form estimate of the optimal τ_p .

A. Approximate Capacity Lower Bound Expression

The proposed approach is based on the key result shown in the next lemma (see Appendix D for a proof).

Lemma 5.1: Let $\underline{I}(\tau_p)$ be defined in (5.2). Assuming that SNR is high, so that σ_v^2 is small, and⁶ $\mathbf{I}/T \ll \sigma_v^2 \tau_p \mathbf{R}_e^{-1}$, we have the following approximation

$$\begin{aligned} \underline{I}(\tau_p) \approx & E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \right\} \\ & + \frac{1}{T \sigma_v^4} E \left\{ \text{Tr} \left[\left(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H \right)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{G}^H \right. \right. \\ & \left. \left. - \left(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \right\} - f(\tau_p), \end{aligned} \quad (5.3)$$

where

$$f(\tau_p) := \alpha \tau_p + \frac{\beta}{\tau_p} \quad (5.4)$$

with

$$\begin{aligned} \alpha := & E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \right\} \\ & - E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \right\} \end{aligned} \quad (5.5)$$

and

$$\beta := \frac{1}{T \sigma_v^4} E \left\{ \text{Tr} \left[\left(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H \right)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{G}^H \right] \right\}. \quad (5.6)$$

Several important comments are in order.

1. Lemma 5.1 is quite appealing in that the dependency of $\underline{I}(\tau_p)$ on the design parameter τ_p is completely characterized by $f(\tau_p)$ defined in (5.4). Compared with $\underline{I}(\tau_p)$ in (5.2), $f(\tau_p)$ is a simple rational function in τ_p ; it is such an attractive feature that can facilitate analytical study of the performance tradeoff, as will be shown later. We also note that, even though the approximation (5.3) is derived based on the high-SNR assumption, the proposed analytic estimate based on (5.3) can well predict the true optimal solution, as well as the resultant capacity performance, over a wide SNR region (this will be seen in the simulation section).
2. We observe that α in (5.5) represents the capacity gap of two Gaussian channels characterized by, respectively, the channel matrices \mathbf{G} and $\mathbf{G} \mathbf{P}$. Based on the discussions in Section III we have $\alpha \geq 0$, with equality attained when $\mathbf{P} = \mathbf{I}$. Roughly speaking, we can think of α

⁶By $\mathbf{A} \ll \mathbf{B}$ (\mathbf{A} and \mathbf{B} both Hermitian and positive definite) we mean (i) $\mathbf{B} - \mathbf{A}$ is positive definite and, (ii) if the eigenvalues of \mathbf{A} and \mathbf{B} are arranged in the same (either increasing or decreasing) order, then $\lambda_k(\mathbf{A}) \ll \lambda_k(\mathbf{B})$ (we note that the positive-definiteness of $\mathbf{B} - \mathbf{A}$ does guarantee $\lambda_k(\mathbf{A}) < \lambda_k(\mathbf{B})$ [13, p-471]). Hence, if $\mathbf{A} \ll \mathbf{B}$, we have $\|\mathbf{A}\|_F^2 = \sum_{k=1}^n \lambda_k^2(\mathbf{A}) \ll \sum_{k=1}^n \lambda_k^2(\mathbf{B}) = \|\mathbf{B}\|_F^2$, and $\|\mathbf{A}\|_2^2 = \max_k \lambda_k^2(\mathbf{A}) \ll \max_k \lambda_k^2(\mathbf{B}) = \|\mathbf{B}\|_2^2$, which are two commonly used conditions for “ \mathbf{A} is small when compared with \mathbf{B} ” in the context of matrix perturbation analysis.

as the *worst-case average capacity penalty induced by precoding with $T_p = T$* . In light of this point and since $0 < \tau_p \leq 1$, the first term $\alpha \tau_p$ in $f(\tau_p)$ thus reflects the proportional reduction in the penalty when symbol precoding is performed only over a τ_p fraction of the coherent time.

3. On the other hand, the quantity β in (5.6) accounts for the channel mismatch effect, and can be treated as the *minimal capacity loss incurred by channel estimation errors* (also attained when $T_p = T$). The second term $\beta \tau_p^{-1}$ in (5.4), therefore, specifies the enlargement of the capacity penalty beyond this minimum if a mere τ_p portion of the coherent interval is spent for symbol precoding to aid channel estimation.
4. With the above facts in mind, $\alpha \tau_p$ and $\beta \tau_p^{-1}$ thus specify the capacity loss due to, respectively, precoding over a τ_p fraction of the coherent interval and the resultant channel estimation errors. The term $f(\tau_p)$ in (5.4), therefore, can be deemed as the *aggregate capacity penalty* incurred by the precoding-based blind channel estimation scheme [31]. With the aid of Lemma 5.1 and the informative interpretations of $f(\tau_p)$, there is a simple yet insightful way of finding an analytic estimate of the optimal τ_p , as shown below.

B. Optimal Performance Tradeoff: An Analytic Characterization

Based on (5.3) and (5.4), the first-order derivative of $\underline{I}(\tau_p)$ with respect to τ_p can be approximately obtained as

$$\underline{I}'(\tau_p) \approx -f'(\tau_p) = (\beta - \alpha \tau_p^2) \tau_p^{-2}. \quad (5.7)$$

With (5.7) it is straightforward to show that the maximum of $\underline{I}(\tau_p)$ occurs nearby

$$\bar{\tau}_p := \sqrt{\beta/\alpha}. \quad (5.8)$$

When $\tau_p < \sqrt{\beta/\alpha}$, it is expected from (5.7) that $\underline{I}'(\tau_p) > 0$ and the capacity lower bound increases with τ_p . As τ_p is enlarged beyond $\sqrt{\beta/\alpha}$, we have $\underline{I}'(\tau_p) < 0$ and $\underline{I}(\tau_p)$ will instead be a decreasing function of τ_p . Since we are only concerned about $0 < \tau_p \leq 1$, the selection of τ_p toward the maximal capacity lower bound depends on whether $\sqrt{\beta/\alpha}$ exceeds unity or not.

Case 1: If $1 \leq \sqrt{\beta/\alpha}$ and hence $\alpha \leq \beta$, i.e., the worst-case average capacity loss due to precoding is less severe than the minimal capacity penalty caused by channel mismatch, $\underline{I}(\tau_p)$ is an increasing function in $0 < \tau_p \leq 1$. This implies that we shall simply set $\tau_p = 1$, i.e., to precode the symbols throughout the entire coherent interval, to maximize $\underline{I}(\tau_p)$. This is intuitively reasonable since, as long as the channel mismatch effect is more deleterious, a long precoding interval is needed to reduce the channel estimation errors, and in turn enlarge the total capacity.

Case 2: If $1 > \sqrt{\beta/\alpha}$ and hence $\alpha > \beta$, meaning that the precoding-induced capacity loss is more harmful, in this case placing τ_p at $\sqrt{\beta/\alpha}$ can attain the maximal capacity lower bound. This reflects the fact that, when the precoding effect is more detrimental, symbol

precoding throughout the entire coherent interval should be avoided. Rather, one should limit the fraction of the precoding phase to $\sqrt{\beta/\alpha}$ in order to realize the largest capacity gain.

Based on the above discussions, the proposed closed-form estimate of the capacity-maximizing time fraction for precoding is thus

$$\tilde{\tau}_p = \min \left\{ 1, \sqrt{\beta/\alpha} \right\}, \quad (5.9)$$

where α and β defined, respectively, in (5.5) and (5.6). We note that both α and β involve the average over the true channel realizations, and can thus be computed off-line once the channel statistics (e.g., Gaussian) are known. The estimated optimal precoding interval is then given by the lower integer floor of $T\tilde{\tau}_p$. The accuracy of the proposed analytic estimate (5.9) is assessed through numerical simulations as shown in the next section.

Remark: Lemma 5.1 together with the analytic estimate of the optimal precoding fraction (5.9) allow us to investigate the achievable capacity performance in the asymptotic regime $T \rightarrow \infty$. Assume that the optimal precoding fraction $\tilde{\tau}_p$ is adopted. Then, from (5.3), the difference between the idealized capacity $E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \}$ and the capacity lower bound $\underline{I}(\tau_p)$ when $T \rightarrow \infty$ can be immediately obtained as (5.10), shown at the bottom of this page. Also, from (5.9) and by definition of α and β in (5.5) and (5.6), it is easy to see $\lim_{T \rightarrow \infty} \tilde{\tau}_p = 0$. This result together with (5.10) assert

$$E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \} - \underline{I}(\tilde{\tau}_p) \approx 0, \quad \text{as } T \rightarrow \infty. \quad (5.11)$$

i.e., the achievable capacity of the considered blind scheme [31] converges to the idealized performance. An intuitive reason for (5.11) is that, when the length of the coherent interval grows without bound ($T \rightarrow \infty$), an arbitrarily small precoding fraction $\tilde{\tau}_p$ can provide a sufficient amount of precoded data blocks for obtaining a quite accurate channel estimate. As a result, the capacity penalty caused by both precoding and channel mismatch can be kept arbitrarily small if $T \rightarrow \infty$. Our simulation results (see Section VI-C) confirm this tendency.

C. On Selection of the Threshold δ

As mentioned in the last paragraph of Section IV, the capacity lower bound (4.12) is a highly nonlinear function of the power threshold δ ; the effect of different δ on the capacity performance is thus extremely difficult to characterize. To provide some guidelines about the selection of δ and to also

keep the analysis tractable, a plausible approach is to focus on some special case in which the dependence of the lower bound (4.12) on τ_p (or T_p) is suppressed. Specifically, we will consider the situation that $\tau_p = 1$ (or $T_p = T$), i.e., the entire coherent interval is used for precoding; based on our simulation such a scenario typically occurs when the coherent interval is not large, and thus accounts for the more realistic mobile environment. Based on (5.3) and with $\tau_p = 1$, the gap between the idealized capacity $E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \}$ and the lower bound $\underline{I}(1)$ can be approximated as (5.12) shown at the bottom of this page (see Appendix E for a derivation). A very rough interpretation of (5.12) is that, when the entire coherent interval⁷ is dedicated to precoding, the channel estimation error can be mitigated, and the performance gap is mainly caused by the capacity loss due to precoding. In this case, the achievable capacity can be enhanced if the negative effect due to precoding can be reduced. In Section III-A it has been shown that $E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \} \leq E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \}$, with equality attained by the uniform precoding scheme $p(n) = 1, 0 \leq n \leq N-1$. Hence, to limit the capacity loss when the two-level precoder (2.16) is used, we shall enlarge the power threshold δ so that the resultant power pattern is close to being uniform; however, the selection $\delta = 1$, which results in uniform power allocation, should be precluded since this will render the channel unidentifiable (cf. footnote 3). Hence, small δ should be disregarded in the algorithm implementation. We shall note that the exact δ that can yield the maximal capacity remains difficult to characterize even in the considered special case; the best δ is a highly complicated function of several system parameters, e.g., length of symbol block N , coherent interval T , and SNR. Through simulation a rule of thumb is $0.6 \leq \delta \leq 0.9$ ($\delta \geq 0.9$ usually leads to poor channel estimation error, and, thus, is not plausible from the equalization perspective, cf., [31]).

VI. SIMULATION RESULTS

This section provides numerical results for corroborating the proposed analytic guidelines for τ_p selection. In each coherent interval the channel taps are drawn from i.i.d. complex Gaussian random variables with zero mean and unit-variance. As in [31] the system parameters are likewise set as $N = 32, L = 8$ (the length of CP is the same as the channel order L), the symbol constellation is QPSK, and the optimal noise-resistant precoder (2.16) is used for channel estimation. To remove the scalar ambiguity we perform a least-squares fit between the computed dominant eigenvector and the true channel; such

⁷Since (5.12) is derived on the basis of Lemma 5.1, the coherent interval is implicitly assumed to be not too small, say, $T > 100$, so that the channel estimate is reasonably accurate and perturbation analysis is valid (cf. the discussions before Lemma 4.1 and footnote 5).

$$E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \} - \underline{I}(\tilde{\tau}_p) \approx \tilde{\tau}_p [E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \} - E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \}] \quad (5.10)$$

$$E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \} - \underline{I}(1) \approx E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \} - E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \} \quad (5.12)$$

TABLE I
THE ESTIMATED AND OPTIMAL PRECODING INTERVALS (SNR = 25 dB).

	$T = 200$	$T = 2000$	$T = 8000$	$T = 20000$
$\sqrt{\beta/\alpha}$	12.23	3.54	2.84	1.22
Estimated τ_p by (5.9)	1	1	1	1
Optimal τ_p by simulation	1	1	1	1

TABLE II
THE ESTIMATED AND OPTIMAL PRECODING INTERVALS (SNR = 0 dB).

	$T = 200$	$T = 2000$	$T = 8000$	$T = 20000$
$\sqrt{\beta/\alpha}$	1.22	0.55	0.32	0.21
Estimated τ_p by (5.9)	1	0.55	0.32	0.21
Optimal τ_p by simulation	1	0.60	0.35	0.25

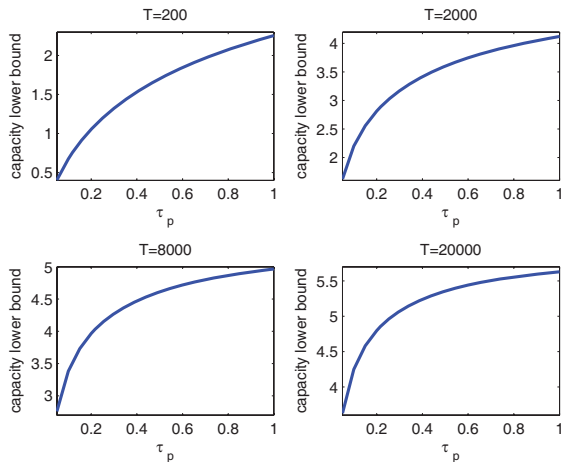


Fig. 1. Capacity lower bounds for different coherent intervals (SNR = 25 dB).

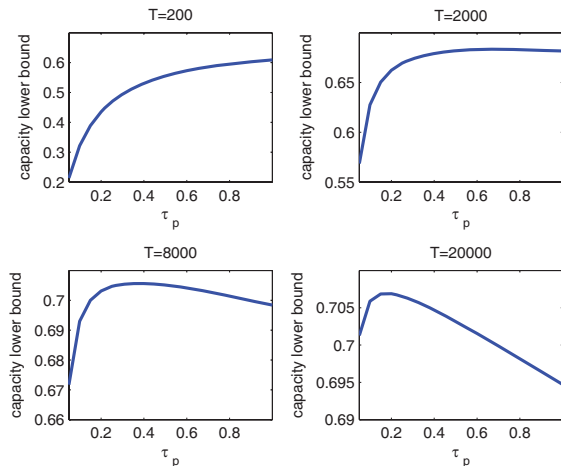


Fig. 2. Capacity lower bounds for different coherent intervals (SNR = 0 dB).

a technique has been adopted in, e.g., [9], [16], [17], for fixing the channel estimate. The capacity lower bounds in all figures are plotted in the unit of bits per channel use, that is, $\underline{I}(\tau_p)/(N + L)$. In Simulations A~C the threshold of the precoder (2.16) is set to be $\delta = 0.9$.

A. Precoding Interval Selection in the High SNR Regime

We set SNR = 25 dB, and consider four different cases of coherent intervals: $T = 200, 2000, 8000, 20000$. Associated

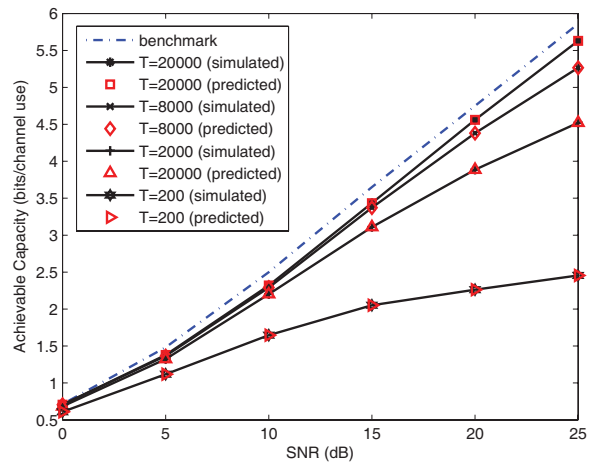


Fig. 3. Achievable capacity lower bounds versus SNR for different coherent intervals.

with each T the value of $\sqrt{\beta/\alpha}$ are computed and listed in Table I. As we can see from the table, $\sqrt{\beta/\alpha}$ exceeds unity for all T . Hence, based on (5.9) $\tau_p = 1$ is expected to maximize the capacity lower bound. Figure 1 plots the experimental $\underline{I}(\tau_p)$ (computed based on (5.2) via averaging the results of Monte-Carlo trials) with respect to the four choices of T . All the curves of $\underline{I}(\tau_p)$ are seen to increase with τ_p , and attain the maximum at $\tau_p = 1$: This thus confirms the analytical study. A rough interpretation of the observed tendency of $\underline{I}(\tau_p)$ is that, when SNR is high, the effective background noise in the system (4.3) is dominated by channel estimation errors, and the achievable capacity advantage can be realized by mitigating channel uncertainty through a long precoding period.

B. Precoding Interval Selection in the Low SNR Regime

We repeat the above experiment by instead setting SNR = 0 dB. The computed $\sqrt{\beta/\alpha}$ with respect to the four different T are listed in Table II. The results show that $\sqrt{\beta/\alpha} > 1$ when $T = 200$; as T increases, the value of $\sqrt{\beta/\alpha}$ gradually falls below unity. Toward the maximal capacity bound, (5.9) implies that we shall thus set $\tau_p = 1$ for $T = 200$, and place τ_p at $\sqrt{\beta/\alpha}$ for the other three T . Figure 2 plots the experimental $\underline{I}(\tau_p)$, based on which the true optimal τ_p for each T is determined and also included in Table II. The result

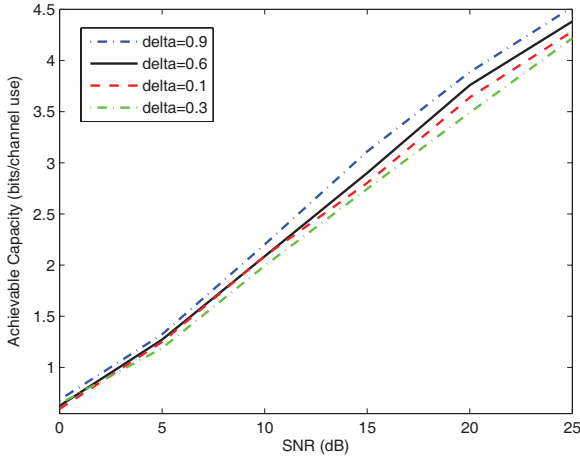


Fig. 4. Achievable capacity lower bounds versus SNR for precoding thresholds $0.1 \leq \delta \leq 0.9$.

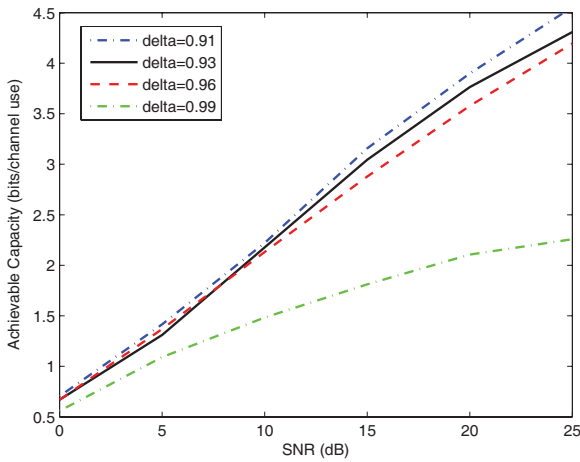


Fig. 5. Achievable capacity lower bounds versus SNR for different precoding thresholds $0.91 \leq \delta \leq 0.99$.

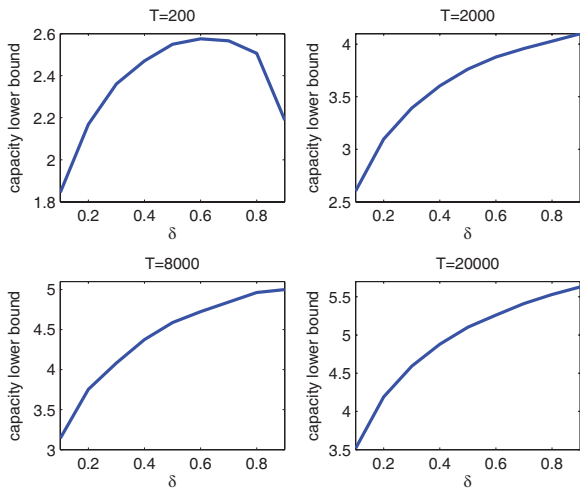


Fig. 6. Capacity lower bounds for $0.1 \leq \delta \leq 0.9$ with $T_p = T$ (SNR = 25 dB).

shows that there is a good agreement between the estimated solutions via (5.9) and the optimal τ_p . Even though a slight discrepancy remains, as we will see in the next simulation the difference between the resultant capacity performances turns

out to be negligible. A plausible rationale for the capacity tendency seen in Figure 2 is that, when SNR is low and T is small, the quality of channel estimation is likely to be quite poor and will be the dominant factor for capacity loss. The entire coherent interval should then be dedicated to precoding for improving channel estimation accuracy, and $\tau_p = 1$ thus maximizes the capacity bound. However, as T gets larger, spending too much coherent interval for precoding cannot largely reduce the channel estimation errors (since the error covariance decays only at the rate $1/T_p = 1/\tau_p T$, cf. (4.8)), but, rather, will enlarge the precoding induced penalty. As a result, τ_p should be kept below unity so that the maximal capacity bound can be attained.

C. Achievable System Capacity

For each considered T , the peak capacity lower bounds at different SNR levels are further determined based on, respectively, the simulated $\underline{I}(\tau_p)$ and the proposed analytical solution (5.9). The results are depicted in Figure 3; the idealized performance measure

$$I_0 := \frac{1}{(N+L)} E \{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \}, \quad (6.1)$$

where \mathbf{G} is the exact channel matrix, is also included as the benchmark. The figure shows that the analytic solutions via (5.9) do accurately predict the experimental counterparts. Also we can see that, when T is small, there is a large capacity gap between the achievable lower bound and the benchmark performance (6.1). The reason is that, for small T , the channel estimation quality is likely to be poor, even when the entire coherent interval ($\tau_p = 1$) is used for achieving the optimal tradeoff. As a result, there tends to be a large capacity loss due to potentially severe channel estimation errors as well as the use of a large precoding time fraction. As T increases, the figure shows that the lower bounds then improve. This is because, as T is large, a relatively small τ_p will suffice to yield a good channel estimate and achieve the maximal capacity bound (recalling from (5.5) and (5.6) that $\sqrt{\beta/\alpha}$ is inversely proportional to \sqrt{T} and, eventually, $\tilde{\tau}_p = \sqrt{\beta/\alpha} < 1$ as T continuously increases). Hence the degradation due to both channel mismatch and precoding can be limited, resulting in a large average capacity gain. We thus conclude that, when the coherent interval is large, the capacity performance attained by the blind estimation scheme [31] approaches the idealized bound (6.1).

D. Impact of the Precoding Threshold

In the last experiment we first test the achievable capacity lower bounds when different thresholds δ of the precoder (2.16) are used. For the coherent interval $T = 2000$, Figures 4 and 5 show the achievable capacity results for two threshold sets $\{0.1, 0.3, 0.6, 0.9\}$ and $\{0.91, 0.93, 0.96, 0.99\}$. In Figure 4 the best capacity performance is seen to be attained with $\delta = 0.9$. This reflects the fact that, although a large δ results in a less accurate channel (see [31, Sec. V]), it can otherwise limit the capacity loss due to precoding (cf. Discussion 1 in Section III). However, if δ increases beyond 0.9, severe channel estimation error occurs: this will then dominate the performance and degrades the capacity, as can be seen from

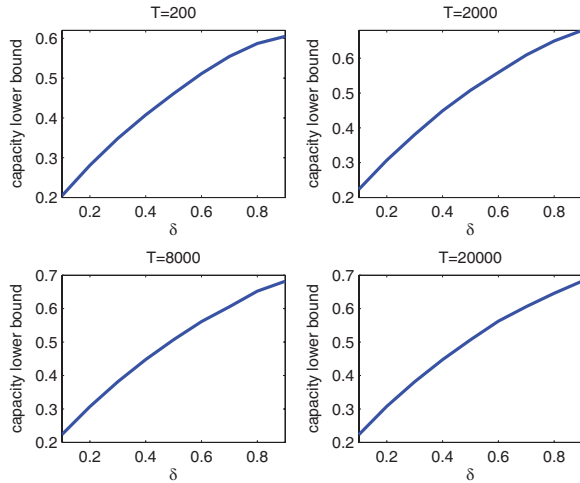


Fig. 7. Capacity lower bounds for $0.1 \leq \delta \leq 0.9$ with $T_p = T$ (SNR = 0 dB).

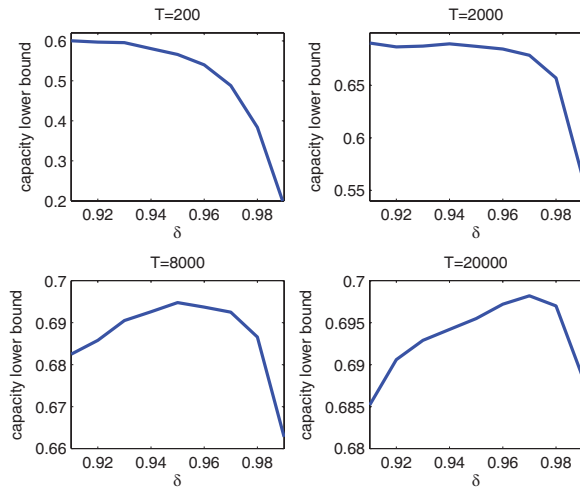


Fig. 8. Capacity lower bounds for $0.91 \leq \delta \leq 0.99$ with $T_p = T$ (SNR = 0 dB).

Figure 5. We thus conclude that $\delta \approx 0.9$ is the compromising choice for $T = 2000$ under the current system setup. Recall that, in the proposed scheme, the threshold δ is fixed and optimization is conducted only with respect to τ_p . We go on to investigate the capacity performance by fixing $T_p = T$ (i.e., the whole coherent interval is used for precoding) and varying the parameter δ . For SNR = 25 dB, Figure 6 plots the proposed capacity lower bound as a function of δ for $T = 200, 2000, 8000, 20000$. Notably it is seen that, for $T = 200$, $\delta \approx 0.6$ (but not $\delta = 0.9$ as used in the previous simulation) attains the maximal capacity. Our further simulation confirms that, for $T = 200$, the optimal precoding fraction is $\tau_p = 1$ for all considered δ . Hence, when SNR is high and T is small, it is plausible to just set $T_p = T$ and choose δ to maximize the capacity. However, even though the formula of the proposed capacity lower bound (4.12) is somewhat simplified with $T_p = T$ (or $\tau_p = 1$), it is still a quite involved function of δ and exact characterization of the optimal δ in this scenario remains difficult. For SNR= 0 dB, the capacity results corresponding to the two threshold sets $0.1 \leq \delta \leq 0.9$ and $0.91 \leq \delta \leq 0.99$ are, respectively, shown

in Figures 7 and 8. As we can see, for $T = 200$, $\delta = 0.9$ attains the maximal capacity. When T is large ($T = 8000$ and 20000), the peak capacity through varying δ is below 0.7 bits/channel use, and is less than the achievable capacity via optimal precoding fraction design with fixed $\delta = 0.9$ (above 0.7 bits/channel use, from Fig. 2). Hence, in this case, optimization with respect to τ_p should be explicitly taken into account in order to realize the maximal capacity advantage.

VII. CONCLUSION

This paper, to the best of our knowledge, is the first contribution in the literature that investigates the capacity performance for wireless communication systems in the blind channel estimation setup. We focus on the CP-based single-carrier system scenario which employs the diagonal-precoding assisted blind channel estimation scheme [31]. When the channel is perfectly known, we show that the optimal noise-resistant two-level precoder proposed in [31] tends to incur the largest capacity penalty. In case that the coherent time is finite and the sample covariance matrix estimation is subject to errors, the optimal tradeoff between channel estimation accuracy and the achievable capacity through precoding interval selection is addressed. By leveraging the matrix perturbation techniques, we derived a closed-form capacity metric in the presence of channel mismatch that is a complicated function of the precoding interval. To simplify analysis an associated tractable approximation of the considered capacity metric is also given. The established results facilitate an analytic approach for finding an estimate of the capacity-maximizing precoding interval, and also allow for informative interpretations regarding the optimal tradeoff. Computer simulations show that the proposed analytic solution can very well predict the experimental results. Also, it is seen that, for small coherent intervals (hence a high mobility environment), channel estimation error is the dominant factor and a large precoding fraction is needed, irrespective of the SNR.

APPENDIX A

PROOF OF THEOREM 3.1

The optimization problem considered is

$$\text{Minimize } \prod_{n=0}^{N-1} p(n)^2 \quad (\text{A.1})$$

subject to the constraints (2.14) and (2.15),

or equivalently,

$$\text{Minimize } \ln \left(\prod_{n=0}^{N-1} p(n)^2 \right) = \sum_{n=0}^{N-1} \ln [p(n)^2] \quad (\text{A.2})$$

subject to the constraints (2.14) and (2.15),

since $\ln(\cdot)$ is a monotone increasing function. Let us define $q_n := p(n)^2 - \delta$, for $0 \leq n \leq N-1$. Then the optimization problem (A.2) becomes

$$\begin{aligned} & \text{Minimize } \sum_{n=0}^{N-1} \ln [q_n + \delta] \\ & \text{subject to } \sum_{n=0}^{N-1} q_n = N(1 - \delta) \\ & q_n \geq 0, \quad \text{for all } 0 \leq n \leq N-1. \end{aligned} \quad (\text{A.3})$$

Since the cost function in (A.3) is concave, the minimizer has to either lie on the boundary or make the partial derivative of the Lagrangian to be zero. Let us assume that there are only k components of the minimizer on the boundary, say, $\bar{q}_0 = \bar{q}_1 = \dots = \bar{q}_{k-1} = 0$ without loss of generality. Then we must have

$$\frac{\partial}{\partial q_n} \left\{ \sum_{n=0}^{N-1} \ln[q_n + \delta] - \lambda \sum_{n=0}^{N-1} q_n \right\}_{q_n = \bar{q}_n} = 0, \quad \text{for } k \leq n \leq N-1. \quad (\text{A.4})$$

By invoking the equality constraint in (A.3) to solve (A.4), it is easy to check $\bar{q}_n = \frac{N(1-\delta)}{N-k}$, for $k \leq n \leq N-1$. Hence the minimal cost function (as a function of k) is

$$\begin{aligned} f(k) &= \sum_{n=0}^{k-1} \ln(\delta) + \sum_{n=k}^{N-1} \ln \left[\frac{N(1-\delta)}{N-k} + \delta \right] \\ &= k \ln(\delta) + \sum_{n=k}^{N-1} \ln \left[\frac{N-k\delta}{N-k} \right]. \end{aligned} \quad (\text{A.5})$$

It can be checked that $f(k)$ is a decreasing function in k , because

$$\begin{aligned} f(k+1) - f(k) &= \ln(\delta) - \ln \left(\frac{N-k\delta}{N-k} \right) \\ &= \ln \left(\frac{\delta(N-k)}{N-k\delta} \right) = \ln \left(\frac{\delta N - k\delta}{N - k\delta} \right) < 0. \end{aligned} \quad (\text{A.6})$$

This implies that the minimal value occurs when $k = N-1$. As a result,

$$\bar{q}_0 = \bar{q}_1 = \dots = \bar{q}_{N-2} = 0 \text{ and } \bar{q}_{N-1} = N(1-\delta) \quad (\text{A.7})$$

solves (A.3), and the optimal solution to (A.1) is

$$\begin{cases} p(0)^2 = p(1)^2 = \dots = p(N-2)^2 = \delta, \\ p(N-1)^2 = N(1-\delta) + \delta. \end{cases} \quad (\text{A.8})$$

Due to the symmetric nature of the problem, the index at which the peak value $N(1-\delta) + \delta$ occurs can be arbitrary, and the two-level precoder (2.16) is thus the minimizing solution.

APPENDIX B PROOF OF LEMMA 4.1

To derive (4.4), we first observe from (4.3) that the mutual information between \mathbf{s}_k and \mathbf{y}_k , for a given channel estimate $\hat{\mathbf{G}}$, is by definition given by

$$I(\mathbf{y}_k; \mathbf{s}_k | \hat{\mathbf{G}}) = h(\mathbf{s}_k | \hat{\mathbf{G}}) - h(\mathbf{s}_k | \hat{\mathbf{G}}, \mathbf{y}_k). \quad (\text{B.1})$$

For a fixed $\hat{\mathbf{G}}$, let us choose \mathbf{s}_k to be Gaussian (which is not necessarily the capacity achieving distribution with imperfect channel information). Then we have

$$\bar{I}(\mathbf{y}_k; \mathbf{s}_k | \hat{\mathbf{G}}) \geq \log \det(\pi e \mathbf{I}) - h(\mathbf{s}_k | \hat{\mathbf{G}}, \mathbf{y}_k). \quad (\text{B.2})$$

It is known that the term $h(\mathbf{s}_k | \hat{\mathbf{G}}, \mathbf{y}_k)$ can be upper bounded as [12, p-963]

$$h(\mathbf{s}_k | \hat{\mathbf{G}}, \mathbf{y}_k) \leq \log \det(\pi e \mathbf{R}_{LMMSE}), \quad (\text{B.3})$$

where

$$\begin{aligned} \mathbf{R}_{LMMSE} &:= E \left\{ [\mathbf{s}_k - (E\{\mathbf{s}_k \mathbf{y}_k^H\})(E\{\mathbf{y}_k \mathbf{y}_k^H\})^{-1} \mathbf{y}_k] \right. \\ &\quad \left. [\mathbf{s}_k - (E\{\mathbf{s}_k \mathbf{y}_k^H\})(E\{\mathbf{y}_k \mathbf{y}_k^H\})^{-1} \mathbf{y}_k]^H \right\} \\ &= \mathbf{I} - (E\{\mathbf{s}_k \mathbf{y}_k^H\})(E\{\mathbf{y}_k \mathbf{y}_k^H\})^{-1} (E\{\mathbf{s}_k \mathbf{y}_k^H\})^H \end{aligned} \quad (\text{B.4})$$

is the error covariance of the linear minimum-mean-square error estimate of \mathbf{s}_k . With the assumption $E\{\mathbf{G}\} = \mathbf{0}$ and by computations, we have

$$E\{\mathbf{s}_k \mathbf{y}_k^H\} = \mathbf{P}^H \hat{\mathbf{G}}^H + \mathbf{P}^H (E\{\tilde{\mathbf{G}}\})^H = \mathbf{P}^H \hat{\mathbf{G}}^H, \quad (\text{B.5})$$

and

$$\begin{aligned} E\{\mathbf{y}_k \mathbf{y}_k^H\} &= \hat{\mathbf{G}} \mathbf{P}^2 \hat{\mathbf{G}}^H + \hat{\mathbf{G}} \mathbf{P}^2 (E\{\tilde{\mathbf{G}}^H\}) \\ &\quad + (E\{\tilde{\mathbf{G}}\}) \mathbf{P}^2 \hat{\mathbf{G}}^H + E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} + \sigma_v^2 \mathbf{I} \\ &= \hat{\mathbf{G}} \mathbf{P}^2 \hat{\mathbf{G}}^H + E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} + \sigma_v^2 \mathbf{I}. \end{aligned} \quad (\text{B.6})$$

With (B.4) ~ (B.6) direct manipulation shows

$$\mathbf{R}_{LMMSE} = \left[\mathbf{I} + \mathbf{P}^H \hat{\mathbf{G}}^H (E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} + \sigma_v^2 \mathbf{I})^{-1} \hat{\mathbf{G}} \mathbf{P} \right]^{-1}. \quad (\text{B.7})$$

Based on (B.1) ~ (B.3), and (B.7) we have

$$\begin{aligned} \bar{I}(\mathbf{y}_k; \mathbf{s}_k | \hat{\mathbf{G}}) &\geq \log \det(\pi e \mathbf{I}) - h(\mathbf{s}_k | \hat{\mathbf{G}}, \mathbf{y}_k) \\ &\geq \log \det(\pi e \mathbf{I}) - \log \det(\pi e \mathbf{R}_{LMMSE}) \\ &= \log \det \left(\mathbf{I} + \mathbf{P}^H \hat{\mathbf{G}}^H (E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} + \sigma_v^2 \mathbf{I})^{-1} \hat{\mathbf{G}} \mathbf{P} \right) \\ &= \log \det \left\{ \mathbf{I} + (E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} + \sigma_v^2 \mathbf{I})^{-1} \hat{\mathbf{G}} \mathbf{P}^2 \hat{\mathbf{G}}^H \right\}. \end{aligned} \quad (\text{B.8})$$

The proof is thus completed.

APPENDIX C PROOF OF LEMMA 4.2

When only a sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{y}}$ is available, the outer product of the channel estimate is thus $\text{vec}(\hat{\mathbf{h}} \hat{\mathbf{h}}^H) = (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \text{vec}(\hat{\mathbf{R}}_{\mathbf{y}})$ (cf. (2.10)), or equivalently,

$$\begin{aligned} \hat{\mathbf{h}} \hat{\mathbf{h}}^H &= \left[\mathbf{A}_1 \text{vec}(\hat{\mathbf{R}}_{\mathbf{y}}) \quad \dots \quad \mathbf{A}_{L+1} \text{vec}(\hat{\mathbf{R}}_{\mathbf{y}}) \right], \\ \mathbf{A}_i &:= (\mathbf{e}_i^T \otimes \mathbf{I}_{L+1}) (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T, \quad 1 \leq i \leq L+1. \end{aligned} \quad (\text{C.1})$$

Based on (C.1), the perturbation of the channel outer-product matrix reads

$$\hat{\mathbf{h}} \hat{\mathbf{h}}^H - \mathbf{h} \mathbf{h}^H = \left[\mathbf{A}_1 \text{vec}(\tilde{\mathbf{R}}_{\mathbf{y}}) \quad \dots \quad \mathbf{A}_{L+1} \text{vec}(\tilde{\mathbf{R}}_{\mathbf{y}}) \right]. \quad (\text{C.2})$$

To prove (4.7) we need the next lemma, which characterizes the first-order perturbation of the dominant singular vector [32].

Lemma C.1: Let $\tilde{\mathbf{h}} = \hat{\mathbf{h}} - \mathbf{h}$ be the channel estimation error. Then we have

$$\tilde{\mathbf{h}} = \left[\frac{1}{\|\hat{\mathbf{h}}\|^2} \Sigma_{\mathbf{h}} \Sigma_{\mathbf{h}}^H \sum_{i=0}^L h(i) \mathbf{A}_i \right] \text{vec}(\tilde{\mathbf{R}}_{\mathbf{y}}). \quad (\text{C.3})$$

[Sketch Proof of Lemma C.1]: We observe that the equation (2.12) is exactly (13) in [32] with $J = 1$, $K = 0$, and $M = N$. Hence the analysis shown in [32] applies to the considered scenario. From (32) in [32], we can directly deduce that $\tilde{\mathbf{h}} = \frac{1}{\|\hat{\mathbf{h}}\|^2} \Sigma_{\mathbf{h}} \Sigma_{\mathbf{h}}^H \left[\hat{\mathbf{h}} \hat{\mathbf{h}}^H - \mathbf{h} \mathbf{h}^H \right] \mathbf{h}$, which together with the definition $\mathbf{h} = [h(0) \ \cdots \ h(L)]^T$ then yield (C.3).

APPENDIX D PROOF OF LEMMA 5.1

For a coherent interval of length T , within which the channel realization \mathbf{G} is fixed, the blind estimation algorithm computes a channel estimate $\hat{\mathbf{G}} = \mathbf{G} - \tilde{\mathbf{G}}$. Subject to such a channel acquisition mechanism, the capacity lower bound (5.2) averaged with respect to (w.r.t.) the channel estimate $\hat{\mathbf{G}}$ should be interpreted as the average over both $\tilde{\mathbf{G}}$ and \mathbf{G} . The computations can thus be done via the two-step approach: (I) First fix the true channel \mathbf{G} and perform expectation w.r.t. $\tilde{\mathbf{G}}$; (II) Then perform expectation w.r.t. to \mathbf{G} . Toward this end let us write, for a fixed pair of \mathbf{G} and $\tilde{\mathbf{G}}$ (to simplify notation in what follows the dependency on τ_p is not explicitly shown), the capacity lower bound as in (D.1), shown at the bottom of this page. We shall first compute the average of $f(\mathbf{G}, \tilde{\mathbf{G}})$. For this let us first expand $f(\mathbf{G}, \tilde{\mathbf{G}})$ into

$$f(\mathbf{G}, \tilde{\mathbf{G}}) = \log \det \left\{ \mathbf{I} + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \left(-\mathbf{G} \mathbf{P}^2 \tilde{\mathbf{G}}^H - \tilde{\mathbf{G}} \mathbf{P}^2 \mathbf{G}^H + \tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H \right) \right\}. \quad (\text{D.2})$$

We further note that the assumption $\mathbf{I}/T \ll \sigma_v^2 \tau_p \mathbf{R}_e^{-1}$ is equivalent to

$$\mathbf{R}_e/(T\tau_p) = \mathbf{R}_e/T_p \ll \sigma_v^2 \mathbf{I}. \quad (\text{D.3})$$

Since the quantity \mathbf{R}_e/T_p accounts for the channel error covariance (cf. (4.9)), inequality (D.3) together with the assumption σ_v^2 small thus assert that the channel estimate is quite accurate so that $\tilde{\mathbf{G}}$ is small. Using the approximation [5, p-641]

$$\log \det [\mathbf{X} + \Delta \mathbf{X}] \approx \log \det \mathbf{X} + \text{Tr} [\mathbf{X}^{-1} \Delta \mathbf{X}], \quad \text{for small } \Delta \mathbf{X}, \quad (\text{D.4})$$

we have the following approximation based on (D.2):

$$f(\mathbf{G}, \tilde{\mathbf{G}}) \approx \log \det \left\{ \mathbf{I} + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right\} + \text{Tr} \left\{ \left[\mathbf{I} + [\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I}]^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right]^{-1} \left[\mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right]^{-1} \left(-\mathbf{G} \mathbf{P}^2 \tilde{\mathbf{G}}^H - \tilde{\mathbf{G}} \mathbf{P}^2 \mathbf{G}^H + \tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H \right) \right\}. \quad (\text{D.5})$$

For a fixed \mathbf{G} , we shall first take the expectation with respect to $\tilde{\mathbf{G}}$ of both sides of (D.5) to get

$$f_1(\mathbf{G}) := E \left\{ f(\mathbf{G}, \tilde{\mathbf{G}}) \right\} \approx \log \det \left\{ \mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right\} + \text{Tr} \left\{ \left[\mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right]^{-1} \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{R}_e/(\tau_p T) \right\}, \quad (\text{D.6})$$

where we have used the fact $E\{\mathbf{G}\} = \mathbf{0}$ and $E\{\tilde{\mathbf{G}} \mathbf{P}^2 \tilde{\mathbf{G}}^H\} = \mathbf{R}_e/(\tau_p T)$ (cf. (4.9)). From (D.3), we have

$$\begin{aligned} & \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{R}_e/(\tau_p T) \\ & \approx \left[\sigma_v^2 \mathbf{I} \right]^{-1} \mathbf{R}_e/(\tau_p T) \\ & \ll \mathbf{I} < \mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H, \end{aligned} \quad (\text{D.7})$$

and from (D.6),

$$\begin{aligned} f_1(\mathbf{G}) & \stackrel{(c)}{\approx} \log \det \left\{ \mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \mathbf{R}_e/(\tau_p T) \right\} \\ & \stackrel{(d)}{\approx} \log \det \left\{ \mathbf{I} + \left[\mathbf{I}/T + \sigma_v^2 \tau_p \mathbf{R}_e^{-1} \right]^{-1} \left[\tau_p \mathbf{R}_e^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H + \mathbf{I}/T \right] \right\}, \end{aligned} \quad (\text{D.8})$$

where (c) holds from (D.6), (D.7), and by again using (D.3), and (d) follows via some manipulations. By assumptions $\mathbf{I}/T \ll \sigma_v^2 \tau_p \mathbf{R}_e^{-1}$ and σ_v^2 is small, we have

$$\mathbf{I}/T \ll \sigma_v^2 \tau_p \mathbf{R}_e^{-1} \ll \tau_p \mathbf{R}_e^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H. \quad (\text{D.9})$$

Also, by using the approximation $(\mathbf{X} + \Delta \mathbf{X})^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1} \Delta \mathbf{X} \mathbf{X}^{-1}$ for small $\Delta \mathbf{X}$, we have

$$\begin{aligned} & \left(\sigma_v^2 \tau_p \mathbf{R}_e^{-1} + \mathbf{I}/T \right)^{-1} \\ & \approx \sigma_v^{-2} \tau_p^{-1} \mathbf{R}_e - \sigma_v^{-2} \tau_p^{-1} \mathbf{R}_e (\mathbf{I}/T) \sigma_v^{-2} \tau_p^{-1} \mathbf{R}_e \\ & = \sigma_v^{-2} \tau_p^{-1} \mathbf{R}_e - \sigma_v^{-4} \tau_p^{-2} \mathbf{R}_e^2/T. \end{aligned} \quad (\text{D.10})$$

Based on (D.8), (D.9), and (D.10), we have (D.11) as shown at the bottom of this page, where (e) follows from (D.3) and

$$\begin{aligned} \underline{I}(\mathbf{G}, \tilde{\mathbf{G}}) & = \tau_p \log \det \left[\mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \hat{\mathbf{G}} \mathbf{P}^2 \hat{\mathbf{G}}^H \right] + (1 - \tau_p) \log \det \left[\mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} \hat{\mathbf{G}} \hat{\mathbf{G}}^H \right] \\ & = \tau_p \log \det \underbrace{\left[\mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} (\mathbf{G} - \tilde{\mathbf{G}}) \mathbf{P}^2 (\mathbf{G} - \tilde{\mathbf{G}})^H \right]}_{:=f(\mathbf{G}, \tilde{\mathbf{G}})} \\ & \quad + (1 - \tau_p) \log \det \underbrace{\left[\mathbf{I} + \left\{ \mathbf{R}_e/(\tau_p T) + \sigma_v^2 \mathbf{I} \right\}^{-1} (\mathbf{G} - \tilde{\mathbf{G}}) (\mathbf{G} - \tilde{\mathbf{G}})^H \right]}_{:=g(\mathbf{G}, \tilde{\mathbf{G}})}. \end{aligned} \quad (\text{D.11})$$

(D.4). Starting from $g(\mathbf{G}, \tilde{\mathbf{G}})$ in (D.1) and by following the similar procedures, the expectation of $g(\mathbf{G}, \tilde{\mathbf{G}})$ w.r.t. $\tilde{\mathbf{G}}$ can be verified to be

$$\begin{aligned} g_1(\mathbf{G}) &:= E \left\{ g(\mathbf{G}, \tilde{\mathbf{G}}) \right\} \\ &\approx \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \\ &\quad - \frac{1}{\tau_p T \sigma_v^4} \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{G}^H \right]. \end{aligned} \quad (\text{D.12})$$

Based on (D.1), (D.11), and (D.12), we have (D.13). Equation (5.3) follows from taking expectation of both sides of (D.13) w.r.t. to the true channel \mathbf{G} followed by direct manipulations.

APPENDIX E PROOF OF EQUATION (5.12)

From (5.3) we have

$$\begin{aligned} &E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \right\} - \underline{\mathbf{I}}(1) \\ &\approx f(1) - \frac{1}{T \sigma_v^4} E \left\{ \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{G}^H \right. \right. \\ &\quad \left. \left. - (\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \right\} \\ &= E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] \right\} \\ &\quad - E \left\{ \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \right\} \\ &\quad + \frac{1}{T \sigma_v^4} E \left\{ \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \right\}. \end{aligned} \quad (\text{E.1})$$

Note that

$$\begin{aligned} &\frac{1}{T \sigma_v^4} E \left\{ \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \right\} \\ &= E \left\{ \text{Tr} \left[(\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \left(\frac{\mathbf{R}_e}{T \sigma_v^2} \right) \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \right\} \\ &\stackrel{(f)}{\approx} \log \det \left[\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H + \left(\frac{\mathbf{R}_e}{T \sigma_v^2} \right) \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \\ &\quad - \log \det [\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \\ &= \log \det \left[\sigma_v^2 \mathbf{I} + \left(\mathbf{I} + \frac{\mathbf{R}_e}{T \sigma_v^2} \right) \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \\ &\quad - \log \det [\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \\ &\stackrel{(g)}{\approx} \log \det [\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] - \log \det [\sigma_v^2 \mathbf{I} + \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \\ &= 0, \end{aligned} \quad (\text{E.2})$$

where (f) follows from (D.3) and (D.4), and (g) holds since $\mathbf{I} + \frac{\mathbf{R}_e}{T \sigma_v^2} \approx \mathbf{I}$ (cf. (D.3)). Based on (E.1) and (E.2), equation (5.12) then follows.

REFERENCES

- [1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338-2353, Aug. 2002.
- [2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd edition. John Wiley & Sons, Inc., 2003.
- [3] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Performance analysis of a deterministic channel estimator for block transmission systems with null guard intervals," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 684-695, Mar. 2002.
- [4] H. Bolcskei, R. W. Heath, and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96-109, Jan. 2000.
- [5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [6] A. Chevreuil, P. Loubaton, and L. Vandendorpe, "Performance of general transmitter induced cyclostationarity precoders: analysis based on a MMSE-DF receiver," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3072-3086, Nov. 2000.
- [7] A. Chevreuil, E. Serpedin, P. Loubaton, and G. B. Giannakis, "Blind channel identification and equalization using periodic modulation precoders: performance analysis," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1570-1586, June 2000.
- [8] P. Davis, *Circulant Matrices*. John Wiley & Sons, Inc, 1979
- [9] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3053-3061, Dec. 1997.

$$\begin{aligned} f_1(\mathbf{G}) &\approx \log \det \left\{ \mathbf{I} + [\sigma_v^{-2} \tau_p^{-1} \mathbf{R}_e - \sigma_v^{-4} \tau_p^{-2} \mathbf{R}_e^2 / T] [\tau_p \mathbf{R}_e^{-1} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] \right\} \\ &= \log \det \left\{ \mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H - \frac{1}{\tau_p \sigma_v^4 T} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right\} \\ &\stackrel{(e)}{\approx} \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] - \frac{1}{\tau_p T \sigma_v^4} \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \end{aligned} \quad (\text{D.11})$$

$$\begin{aligned} \underline{\mathbf{I}}(\mathbf{G}) &:= E \left\{ \underline{\mathbf{I}}(\mathbf{G}, \tilde{\mathbf{G}}) \right\} = \tau_p f_1(\mathbf{G}) + (1 - \tau_p) g_1(\mathbf{G}) \\ &= \tau_p \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H] - \frac{1}{T \sigma_v^4} \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{P}^2 \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{P}^2 \mathbf{G}^H \right] \\ &\quad + (1 - \tau_p) \log \det [\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H] - \frac{(1 - \tau_p)}{\tau_p T \sigma_v^4} \text{Tr} \left[(\mathbf{I} + \sigma_v^{-2} \mathbf{G} \mathbf{G}^H)^{-1} \mathbf{R}_e \mathbf{G} \mathbf{G}^H \right] \end{aligned} \quad (\text{D.13})$$

- [10] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, pp. 58-66, Apr. 2002.
- [11] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobil Communication, Vol. I: Trends in Channel Identification and Equalization*. Prentice Hall PTR, 2001.
- [12] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.
- [13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [14] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [15] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
- [16] T. P. Krauss and M. D. Zoltowski, "Bilinear approach to multiuser second-order statistics-based blind channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2473-2486, Sep. 2000.
- [17] C. A. Lin and J. Y. Wu, "Blind identification with periodic modulation: a time-domain approach," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2875-2888, Nov. 2002.
- [18] R. Lin and A. P. Petropulu, "Linear precoding assisted blind channel estimation for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 983-995, May 2005.
- [19] H. G. Myung and D. J. Goodman, *Single-Carrier FDMA: A New Air Interface for Long Term Evolution*. Wiley, 2008.
- [20] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516-525, Feb. 1995.
- [21] S. Nayeb Nazar and I. N. Psaromiligkos, "Performance of blind channel estimation algorithms for space-frequency block coded multi-carrier code division multiple access systems," *IET Commun.*, vol. 2, no. 2, pp. 320-328, Feb. 2008.
- [22] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784-2795, Nov. 2003.
- [23] A. P. Petropulu, R. Zhang, and R. Lin, "Blind simple OFDM channel estimation through simple linear precoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 647-655, Mar. 2004.
- [24] J. R. Schott, *Matrix Analysis for Statistics*, 2nd edition. John Wiley & Sons, Inc., 2005.
- [25] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1930-1944, July 1998.
- [26] C. Shin, R. W. Heath, and E. J. Powers, "Non-redundant precoding-based blind and semi-blind channel estimation for MIMO block transmission with a cyclic prefix," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2509-2523, June 2008.
- [27] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, 1999.
- [28] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340-349, Mar. 1994.
- [29] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12-25, Nov. 2004.
- [30] M. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1785-1794, July 1997.
- [31] J. Y. Wu and T. S. Lee, "Periodic-modulation based blind channel identification for single-carrier block transmission with frequency-domain equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1114-1130, Mar. 2006.
- [32] Z. Xu, "Effects of imperfect blind channel estimation on performance of linear CDMA receivers," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2873-2884, May 2002.
- [33] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1215-1228, May 2002.



Jwo-Yuh Wu received the B. S. degree in 1996, the M. S. degree in 1998, and the Ph. D. degree in 2002, all in Electrical and Control Engineering, National Chiao Tung University, Taiwan. During 2003 and 2007, he was a post doctor research fellow in the Department of Communications Engineering, National Chiao Tung University, Taiwan. Since 2008 he has been an assistant professor in the Department of Electrical Engineering, and the Institute of Communications Engineering, National Chiao Tung University, Taiwan. His general research interests

are in signal processing, wireless communications and networking, control systems, and linear algebra.