

# Optimal Training for MIMO Frequency-Selective Fading Channels

Xiaoli Ma, *Member, IEEE*, Liuqing Yang, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

**Abstract**—High data rates give rise to frequency-selective propagation effects. Space-time multiplexing and/or coding offer attractive means of combating fading and boosting capacity of multi-antenna communications. As the number of antennas increases, channel estimation becomes challenging because the number of unknowns increases, and the power is split at the transmitter. Optimal training sequences have been designed for flat-fading multi-antenna systems or for frequency-selective single transmit antenna systems. We design a low-complexity optimal training scheme for block transmissions over frequency-selective channels with multiple antennas. The optimality in designing our training schemes consists of maximizing a lower bound on the ergodic (average) capacity that is shown to be equivalent to minimizing the mean square error of the linear channel estimator. Simulation results confirm our theoretical analysis that applies to both single- and multicarrier transmissions.

**Index Terms**—Channel estimation, ergodic capacity, frequency-selective channels, multi-input multi-output (MIMO) systems, pilot symbol-aided modulation.

## I. INTRODUCTION

HIGH RATE wireless transmissions experience frequency-selective propagation effects. Albeit challenging to mitigate, once acquired, these frequency-selective fading channels offer multipath-diversity gains. Especially when coupled with space-diversity gains that emerge with multi-antenna communications, flexible designs become available to enhance the overall system performance and capacity, provided that the resulting multi-input multi-output (MIMO) channel can be acquired at the receiver.

For channel state information (CSI) acquisition, three classes of methods are available: blind methods, which estimate CSI merely from the received symbols; differential ones that bypass CSI estimation by differential encoding; and input-output methods which rely on training symbols that are known to the receiver. Relative to training-based schemes, differential approaches incur performance loss by design [6], while blind methods typically require longer data records and entail higher

complexity [3], [15], [29]. On the other hand, the insertion of known training symbols can be suboptimal and bandwidth consuming. But training remains attractive especially when it decouples symbol detection from channel estimation and thus simplifies the receiver implementation and relaxes the required identifiability conditions [19].

Training symbols can be placed either at the beginning of each burst (as a preamble) or regularly throughout the burst. In rapidly fading or quasi-static fading channels, preamble-based training may not work well. This motivates embedding training symbols in each transmitted block, instead of concentrating them at the preamble. The so-termed pilot symbol-aided modulation (PSAM) originally developed for time-selective channels [5] has recently been extended to single-antenna frequency- and doubly-selective channels [8], [17], [19], [23] and also to MIMO flat-fading channels [12], [27]. It is known that imperfect channel knowledge has an effect on mutual information [18]. Recently, training sequence design has been linked with channel capacity or its bounds (see, e.g., [1], [12], [17], and [20]). Similar to [12], [17], [19], and [23], our approach in this paper links mutual information with channel estimation and designs training sequences to maximize a lower bound of the average mutual information.

By decoupling the channel estimation from symbol detection, we design optimal PSAM schemes for linear minimum mean-square error (LMMSE) estimation of MIMO frequency-selective fading channels. Training sequences for such channels have so far been designed through exhaustive search for space-time (ST) trellis codes [10], [11] or for ST-OFDM systems [2], [13], [14], [22]. (Semi-)blind channel estimators have also been reported [4], [15], [28]. Here, we start from a general model, in which training symbols are superimposed on information symbols. Our design criteria optimize channel MSE and ergodic (average) capacity bounds to jointly account not only for channel estimation performance but also for transmission rate.

Attractive properties over block transmissions are provided by the widespread applicability of OFDM in wireless standards (see [24]). To enable block-by-block transmissions, we consider removing interblock interference (IBI). The motivation behind IBI avoidance using the insertion of guard times (either cyclic prefix or zero padding) is threefold: 1) each block oftentimes corresponds to a transmission burst followed by a silent period (all-zero guard); 2) in a block fading model the channel changes per block, making channel estimation necessary on a block-by-block basis; and 3) since we deal with frequency-selective channels, every two consecutive received bursts (blocks) entail IBI, and low-complexity receiver processing motivates well block-by-block decoding. For both

Manuscript received November 29, 2002; accepted September 18, 2003. The editor coordinating the review of this paper and approving it for publication is H. Boelcke. The work in this paper was prepared through collaborative participation in the Communications and Networks Consortium supported by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

X. Ma is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: xiaoli@eng.auburn.edu).

L. Yang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: ldyang@ece.ufl.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TWC.2004.842998

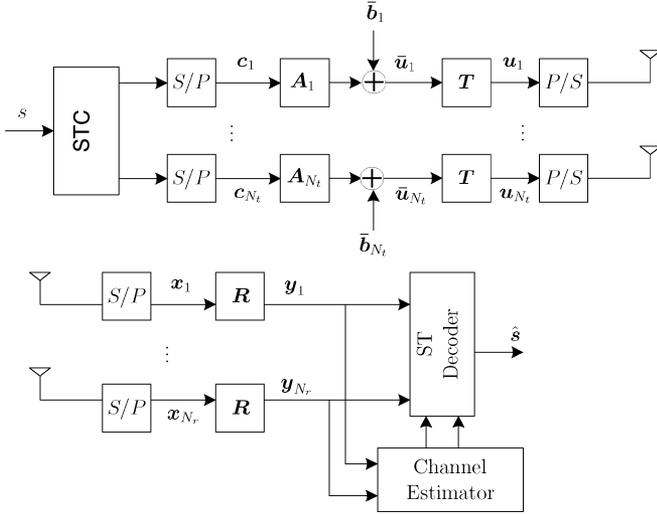


Fig. 1. Discrete-time baseband equivalent Tx-Rx MIMO model.

zero-padded (ZP) *single-carrier* block transmissions as well as for cyclic-prefixed (CP) *multicarrier* systems, we develop optimal training parameters which are flexible to accommodate any ST code.

The rest of the paper is organized as follows. Section II introduces the system model. The optimal training schemes for ZP- and CP-based block transmissions are detailed in Sections III and IV, respectively. Some special cases are discussed in Section V. Numerical examples and simulated performance are presented in Section VI, while Section VII concludes the paper.

*Notation:* Upper (lower) bold face letters will be used for matrices (column vectors). Superscript  $\mathcal{H}$  will denote Hermitian,  $*$  conjugate, and  $T$  transpose. We will reserve  $\otimes$  for the Kronecker product,  $\odot$  for the Hadamard product,  $\lceil \cdot \rceil$  for integer ceiling,  $E[\cdot]$  for expectation, and  $\text{tr}[\mathbf{A}]$  for matrix trace.  $\mathbf{I}_N$  will denote the  $N \times N$  identity matrix and  $\mathbf{F}_N$  the  $N \times N$  normalized (unitary) FFT matrix;  $\text{diag}\{\mathbf{x}\}$  will stand for a diagonal matrix with  $\mathbf{x}$  on its main diagonal; and  $[\mathbf{A}]_{m,n}$  will denote the  $(m, n)$ th entry of the matrix  $\mathbf{A}$ .

## II. SYSTEM MODEL

Our multi-antenna system employs  $N_t$  transmit- and  $N_r$  receive-antennas (see Fig. 1). The information-bearing symbols having a symbol rate  $1/T_s$  are parsed into blocks of size  $N_s \times 1$ :  $\mathbf{s}(k) := [s(kN_s), s(kN_s + 1), \dots, s(kN_s + N_s - 1)]^T$ , where  $k$  denotes the block index. At the transmitter, each block  $\mathbf{s}(k)$  is first encoded and/or multiplexed in space and time. The resulting  $N_t$  blocks  $\{\mathbf{c}_{\mu}(k)\}_{\mu=1}^{N_t}$  have length  $N_c$ , and each is directed to one transmit antenna. At each, e.g.,  $\mu$ th transmit antenna,  $\mathbf{c}_{\mu}(k)$  is processed by a matrix  $\mathbf{A}_{\mu}$  of size  $M \times N_c$ , where  $M \geq N_c$ . Training blocks  $\bar{\mathbf{b}}_{\mu}(k)$  of length  $M$ , which are known to both transmitter and receiver, are then superimposed to  $\mathbf{A}_{\mu}\mathbf{c}_{\mu}(k)$  after the ST mapper to form  $\bar{\mathbf{u}}_{\mu}(k) := \mathbf{A}_{\mu}\mathbf{c}_{\mu}(k) + \bar{\mathbf{b}}_{\mu}(k)$ . This affine model is fairly general; it encompasses linear (pre) coding via  $\mathbf{A}_{\mu}$  as well as inserted training symbols by having nonzero entries of  $\bar{\mathbf{b}}_{\mu}(k)$  where  $\mathbf{A}_{\mu}\mathbf{c}_{\mu}(k)$  has zero entries. To eliminate IBI at the receiver, redundant (guard) symbols are added to  $\bar{\mathbf{u}}_{\mu}(k)$  via the  $N \times M$  matrix  $\mathbf{T}$  to form blocks

$\mathbf{u}_{\mu}(k) := \mathbf{T}\bar{\mathbf{u}}_{\mu}(k)$  of length  $N > M$ . The type and size of the added redundancy will be specified later. The blocks  $\mathbf{u}_{\mu}(k)$  are then parallel-to-serial converted, pulse-shaped, carrier modulated, and transmitted from the  $\mu$ th transmit antenna.

Let  $h^{(\nu,\mu)}(l), l \in [0, L]$  denote the discrete-time baseband equivalent channel that includes transmit–receive filters as well as the frequency-selective propagation effects between the  $\mu$ th transmit antenna and the  $\nu$ th receive antenna. With  $\tau_{\max}^{(\nu,\mu)}$  denoting the delay spread of the channel between the  $\mu$ th transmit antenna and the  $\nu$ th receive antenna, we define the maximum delay spread as  $\tau_{\max} := \max\{\tau_{\max}^{(\nu,\mu)}\}_{(\nu,\mu)=(1,1)}^{(N_r,N_t)}$ . The maximum channel order is then given by  $L := \lceil \tau_{\max}/T_s \rceil$ .

The samples at the  $\nu$ th receive-antenna filter output can be expressed as

$$x_{\nu}(n) = \sum_{\mu=1}^{N_t} \sum_{l=0}^L h^{(\nu,\mu)}(l)u_{\mu}(n-l) + \eta_{\nu}(n) \quad (1)$$

where  $\eta_{\nu}(n)$  is zero-mean, white, complex Gaussian distributed noise with variance  $\sigma^2$ . The sequence is then serial-to-parallel (SP) converted into  $N \times 1$  blocks  $\mathbf{x}_{\nu}(k) := [x_{\nu}(kN), x_{\nu}(kN + 1), \dots, x_{\nu}(kN + N - 1)]^T$ . Selecting  $N \geq L$ , we can write the matrix-vector counterpart of (1) as

$$\mathbf{x}_{\nu}(k) = \sum_{\mu=1}^{N_t} \left[ \mathbf{H}^{(\nu,\mu)} \mathbf{u}_{\mu}(k) + \mathbf{H}_{\text{ibi}}^{(\nu,\mu)} \mathbf{u}_{\mu}(k-1) \right] + \boldsymbol{\eta}_{\nu}(k) \quad (2)$$

where  $\boldsymbol{\eta}_{\nu}(k) := [\eta_{\nu}(kN), \eta_{\nu}(kN + 1), \dots, \eta_{\nu}(kN + N - 1)]^T$ ,  $\mathbf{H}^{(\nu,\mu)}$  is an  $N \times N$  lower triangular Toeplitz matrix with first column  $[h^{(\nu,\mu)}(0), \dots, h^{(\nu,\mu)}(L), 0, \dots, 0]^T$ , and  $\mathbf{H}_{\text{ibi}}^{(\nu,\mu)}$  is an  $N \times N$  upper triangular Toeplitz matrix with first row  $[0, \dots, 0, h^{(\nu,\mu)}(L), \dots, h^{(\nu,\mu)}(1)]$ . The  $\mathbf{u}_{\mu}(k-1)$  dependent term in (2) captures the IBI due to the channel delay spread.

When the blocks have finite length, to enable block-by-block channel estimation and symbol decoding, we need to remove IBI. As shown in Fig. 1, we achieve this by postprocessing  $\mathbf{x}_{\nu}(k)$  by a matrix  $\mathbf{R}$  to obtain

$$\mathbf{y}_{\nu}(k) = \sum_{\mu=1}^{N_t} \left[ \mathbf{R}\mathbf{H}^{(\nu,\mu)}\mathbf{T}\bar{\mathbf{u}}_{\mu}(k) + \mathbf{R}\mathbf{H}_{\text{ibi}}^{(\nu,\mu)}\mathbf{T}\bar{\mathbf{u}}_{\mu}(k-1) \right] + \mathbf{R}\boldsymbol{\eta}_{\nu}(k).$$

There are two prevalent and easy-to-implement guard options for IBI elimination: ZP and CP guards, each requiring a specific  $(\mathbf{T}, \mathbf{R})$  pair. In this paper, we wish to design optimal training strategies *after the ST mapper* for multi-antenna frequency-selective channel estimation. For both ZP- and CP-based block transmissions, we will design optimal pairs  $(\mathbf{A}_{\mu}, \bar{\mathbf{b}}_{\mu}), \forall \mu \in [1, N_t]$ . For both schemes, the optimal designs will guide our selection of the number of training symbols per block, the placement of training symbols, and the power allocation between training and information symbols, using as criteria the conditional mutual information and the MIMO channel's estimation error. We will seek designs that are optimal *for any* block size. Before we start deriving our major results, we outline the major steps of our approach as follows.

- Step 1) Design  $\{\mathbf{A}_\mu, \mathbf{b}_\mu\}_{\mu=1}^{N_t}$  to decouple channel estimation from symbol detection.
- Step 2) Develop the LMMSE estimator for the MIMO channel.
- Step 3) Derive upper and lower bounds on the average (ergodic) capacity of the MIMO channel.
- Step 4) Link the lower bound on capacity with the MIMO channel MSE.
- Step 5) Finalize the parameter choices to optimize the overall design of the training sequence per antenna.

Since the ensuing channel estimators and symbol detectors operate on a block-by-block basis, we will henceforth omit the block index  $k$ .

### III. ZP-BASED BLOCK TRANSMISSIONS

If each block  $\mathbf{u}_\mu$  contains  $L$  trailing zeros, the IBI term vanishes since the nonzero entries of  $\mathbf{H}_{\text{ibi}}^{(\nu, \mu)}$  in (2) are confined to its last  $L$  columns. To achieve this, we select the matrices  $\mathbf{T}$  and  $\mathbf{R}$  as follows:

$$\mathbf{T} = \mathbf{T}_{\text{zp}} := [\mathbf{I}_M \quad \mathbf{0}_{M \times L}]^T, \quad \mathbf{R} = \mathbf{I}_{M+L}$$

where  $\mathbf{T}_{\text{zp}}$  is the zero-padding matrix which appends  $L$  zeros to the  $M \times 1$  vector  $\bar{\mathbf{u}}_\mu$  when multiplying it from the left. It is clear that the length of  $\mathbf{u}_\mu = \mathbf{T}_{\text{zp}} \bar{\mathbf{u}}_\mu$  is now  $N := M + L$ . Defining the  $N \times M$  matrix  $\bar{\mathbf{H}}^{(\nu, \mu)} := \mathbf{H}^{(\nu, \mu)} \mathbf{T}_{\text{zp}}$ , the input-output relationship (2) reduces to

$$\mathbf{y}_\nu = \sum_{\mu=1}^{N_t} \left( \bar{\mathbf{H}}^{(\nu, \mu)} \mathbf{A}_\mu \mathbf{c}_\mu + \bar{\mathbf{H}}^{(\nu, \mu)} \bar{\mathbf{b}}_\mu \right) + \boldsymbol{\eta}_\nu \quad (3)$$

where  $\bar{\mathbf{H}}^{(\nu, \mu)}$  is an  $N \times M$  Toeplitz matrix with first column  $[h^{(\nu, \mu)}(0), \dots, h^{(\nu, \mu)}(L), 0, \dots, 0]^T$ . Recall that a convolution between two vectors can be represented as the product of a Toeplitz matrix with a vector. Because convolution is a commutative operation, we deduce that  $\bar{\mathbf{H}}^{(\nu, \mu)} \bar{\mathbf{b}}_\mu = \bar{\mathbf{B}}_\mu \mathbf{h}^{(\nu, \mu)}$ , where  $\bar{\mathbf{B}}_\mu$  is an  $N \times (L + 1)$  Toeplitz matrix having first column  $[\bar{b}_\mu(1), \dots, \bar{b}_\mu(M), 0, \dots, 0]^T$  with  $\bar{b}_\mu(m)$  being the  $m$ th entry of  $\bar{\mathbf{b}}_\mu$ , and  $\mathbf{h}^{(\nu, \mu)} := [h^{(\nu, \mu)}(0), h^{(\nu, \mu)}(1), \dots, h^{(\nu, \mu)}(L)]^T$ . Equation (3) can be re-expressed as

$$\begin{aligned} \mathbf{y}_\nu &= \sum_{\mu=1}^{N_t} \left( \bar{\mathbf{H}}^{(\nu, \mu)} \mathbf{A}_\mu \mathbf{c}_\mu + \bar{\mathbf{B}}_\mu \mathbf{h}^{(\nu, \mu)} \right) + \boldsymbol{\eta}_\nu \\ &= \left[ \bar{\mathbf{H}}^{(\nu, 1)} \mathbf{A}_1 \quad \dots \quad \bar{\mathbf{H}}^{(\nu, N_t)} \mathbf{A}_{N_t} \right] \mathbf{c} + \bar{\mathbf{B}} \mathbf{h}_\nu + \boldsymbol{\eta}_\nu \end{aligned} \quad (4)$$

where  $\mathbf{h}_\nu := [(\mathbf{h}^{(\nu, 1)})^T, \dots, (\mathbf{h}^{(\nu, N_t)})^T]^T$ ,  $\mathbf{c} := [\mathbf{c}_1^T, \dots, \mathbf{c}_{N_t}^T]^T$ , and  $\bar{\mathbf{B}} := [\bar{\mathbf{B}}_1, \dots, \bar{\mathbf{B}}_{N_t}]$ . Concatenating the  $N_r$  received vectors into a single block  $\mathbf{y} := [\mathbf{y}_1^T, \dots, \mathbf{y}_{N_r}^T]^T$ , we have

$$\mathbf{y} = \Phi \mathbf{c} + (\mathbf{I}_{N_r} \otimes \bar{\mathbf{B}}) \mathbf{h} + \boldsymbol{\eta} \quad (5)$$

where  $\mathbf{h} := [\mathbf{h}_1^T, \dots, \mathbf{h}_{N_r}^T]^T$ , and the  $N_r N \times N_t N_c$  matrix  $\Phi$  is defined as

$$\Phi := \begin{bmatrix} \bar{\mathbf{H}}^{(1,1)} \mathbf{A}_1 & \dots & \bar{\mathbf{H}}^{(1, N_t)} \mathbf{A}_{N_t} \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{H}}^{(N_r, 1)} \mathbf{A}_1 & \dots & \bar{\mathbf{H}}^{(N_r, N_t)} \mathbf{A}_{N_t} \end{bmatrix}. \quad (6)$$

As  $\Phi$  depends on  $\mathbf{h}$ , we infer from (5) that estimating  $\mathbf{h}$  and recovering  $\mathbf{c}$  from  $\mathbf{y}$  is a nonlinear problem, whose joint solution is cumbersome or even impossible (see [19] for an example). Therefore, our first step is to judiciously design  $(\mathbf{A}_\mu, \bar{\mathbf{b}}_\mu)$  so that we can decouple channel estimation from symbol decoding.

#### A. Decoupling Channel Estimation From Symbol Decoding

To estimate  $\mathbf{h}$  from  $\mathbf{y}$ , we resort to the linear MMSE (LMMSE) estimator

$$\hat{\mathbf{h}} := \left( \sigma^2 \mathbf{R}_h^{-1} + \mathbf{I}_{N_r} \otimes (\bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{B}}) \right)^{-1} (\mathbf{I}_{N_r} \otimes \bar{\mathbf{B}}^{\mathcal{H}}) \mathbf{y} \quad (7)$$

where  $\mathbf{R}_h := \mathbb{E}[\mathbf{h} \mathbf{h}^{\mathcal{H}}]$  is the channel covariance matrix, and  $\sigma^2$  denotes the noise variance. It will be clear later that the LMMSE channel estimator will allow us to derive the mutual information bounds. Joint channel estimation and symbol detection is not considered in this paper, because: 1) the nonlinear search required is not only computationally complex but also its convergence to a globally optimum solution is not guaranteed; 2) symbol identifiability cannot be ensured in general; and 3) the designed training sequences will be tailored for only a specific encoder. These reasons motivate us to design training sequences able to decouple channel estimation from symbol detection.

Since  $\mathbf{c}$  is unknown, we cannot obtain  $\hat{\mathbf{h}}$  directly from (5). To facilitate decoupling of channel estimation from symbol detection, we then need

$$\begin{aligned} \left( \sigma^2 \mathbf{R}_h^{-1} + \mathbf{I}_{N_r} \otimes (\bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{B}}) \right)^{-1} (\mathbf{I}_{N_r} \otimes \bar{\mathbf{B}}^{\mathcal{H}}) \Phi \mathbf{c} &= \mathbf{0} \\ \Leftrightarrow (\mathbf{I}_{N_r} \otimes \bar{\mathbf{B}}^{\mathcal{H}}) \Phi \mathbf{c} &= \mathbf{0}. \end{aligned} \quad (8)$$

After noticing that (8) must hold true for all  $\mathbf{c}$ , our selection of  $\bar{\mathbf{B}}$  and  $\mathbf{A}_\mu$ 's should satisfy

$$\begin{bmatrix} \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{H}}^{(1,1)} \mathbf{A}_1 & \dots & \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{H}}^{(1, N_t)} \mathbf{A}_{N_t} \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{H}}^{(N_r, 1)} \mathbf{A}_1 & \dots & \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{H}}^{(N_r, N_t)} \mathbf{A}_{N_t} \end{bmatrix} = \mathbf{0}$$

or equivalently

$$\bar{\mathbf{B}}_{\mu_1}^{\mathcal{H}} \bar{\mathbf{H}}^{(\nu, \mu_2)} \mathbf{A}_{\mu_2} = \mathbf{0}, \quad \forall \mu_1, \mu_2 \in [1, N_t], \quad \nu \in [1, N_r]. \quad (9)$$

Re-expressing  $\mathbf{A}_{\mu_2}$  in terms of its  $N_c$  columns, we have

$$\begin{aligned} \bar{\mathbf{H}}^{(\nu, \mu_2)} \mathbf{A}_{\mu_2} &= \left[ \bar{\mathbf{H}}^{(\nu, \mu_2)} \mathbf{a}_{\mu_2, 1} \dots \bar{\mathbf{H}}^{(\nu, \mu_2)} \mathbf{a}_{\mu_2, N_c} \right] \\ &= \left[ \mathbf{A}_{\mu_2, 1} \mathbf{h}^{(\nu, \mu_2)} \dots \mathbf{A}_{\mu_2, N_c} \mathbf{h}^{(\nu, \mu_2)} \right] \end{aligned} \quad (10)$$

where  $\mathbf{A}_{\mu_2, n}$ 's are  $N \times (L + 1)$  Toeplitz matrices generated by  $\mathbf{a}_{\mu_2, n}$ . Because (9) must hold true for all channel realizations, our decoupling goal mandates  $(\mathbf{A}_\mu, \bar{\mathbf{b}}_\mu)$  pairs  $\forall \mu_1, \mu_2 \in [1, N_t]$ ,  $\nu \in [1, N_r]$ , and  $\forall n \in [1, N_c]$  satisfying

$$\bar{\mathbf{B}}_{\mu_1}^{\mathcal{H}} \mathbf{A}_{\mu_2, n} = \mathbf{0}. \quad (11)$$

*Lemma 1 [Decoupling Training From Information]:* Pairs of  $(\mathbf{A}_\mu, \bar{\mathbf{b}}_\mu)$  that guarantee decoupling of channel estimation from symbol detection, for any block length  $N$ , must have the form

$$\mathbf{A}_\mu = \begin{bmatrix} \Theta \\ \mathbf{0}_{(M-N'_c) \times N_c} \end{bmatrix}, \quad \bar{\mathbf{b}}_\mu = \begin{bmatrix} \mathbf{0}_{N'_c+L} \\ \mathbf{b}_\mu \end{bmatrix} \quad (12)$$

where the  $(M - N'_c - L) \times 1$  vector  $\mathbf{b}_\mu$  contains all the nonzero entries of  $\bar{\mathbf{b}}_\mu$ , and  $\Theta$  is an  $N'_c \times N_c$  matrix that optionally pre-codes (if  $\Theta \neq \mathbf{I}_{N_c}$ ) the information-bearing block  $\mathbf{c}_\mu$  linearly.

*Proof:* See Appendix A.  $\square$

It is worth emphasizing that we neither assumed insertion of training symbols *a fortiori*, nor do we impose time-division multiplexing (TDM) of training with information symbols at the outset. Certainly, TDM ensures decoupling, but the converse we established in Lemma 1 is not obvious. Lemma 1 revealed that in order to separate channel estimation from symbol decoding, we need to insert at least  $L$  zeros between the information symbols and the nonzero training symbols, per antenna. Note that the  $L$  guard zeros can be replaced by  $L$  known symbols. However, this replacement wastes unnecessary power on the guard symbols and causes interference from the training symbols to the information symbols.

As a consequence of Lemma 1, the superimposed model  $\mathbf{u}_\mu = \mathbf{T}_{\text{zp}}(\mathbf{A}_\mu \mathbf{c}_\mu + \bar{\mathbf{b}}_\mu)$  boils down to TDM the ST blocks  $\mathbf{c}_\mu$  with the training symbol blocks  $\mathbf{b}_\mu$  and suggests guarding against their channel-induced IBI using at least  $L$  zeros, per antenna.

Since the information blocks  $\mathbf{c}_\mu$  have been encoded, in order to retain the structure<sup>1</sup> of  $\mathbf{c}_\mu$  and reduce decoding complexity, we simply choose  $N'_c = N_c$  and  $\Theta = \mathbf{I}_{N_c}$ . For this design, we then have

$$\bar{\mathbf{B}}_\mu := \begin{bmatrix} \mathbf{0}_{(N_c+L) \times (L+1)} \\ \mathbf{B}_\mu \end{bmatrix}$$

and

$$\bar{\mathbf{B}} := \begin{bmatrix} \mathbf{0}_{(N_c+L) \times N_t(L+1)} \\ \mathbf{B} \end{bmatrix} \quad (13)$$

where  $\mathbf{B} := [\mathbf{B}_1, \dots, \mathbf{B}_{N_t}]$  is an  $(N_b + L) \times N_t(L + 1)$  matrix.

Using (12) and (13), we can split  $\mathbf{y}_\nu$  in (4) into two parts, as follows:

$$\begin{aligned} \mathbf{y}_\nu &= \begin{bmatrix} \mathbf{y}_{\nu, c} \\ \mathbf{y}_{\nu, b} \end{bmatrix} = \begin{bmatrix} \sum_{\mu=1}^{N_t} \mathbf{H}_c^{(\nu, \mu)} \mathbf{c}_\mu + \boldsymbol{\eta}_{\nu, c} \\ \sum_{\mu=1}^{N_t} \mathbf{B}_\mu \mathbf{h}^{(\nu, \mu)} + \boldsymbol{\eta}_{\nu, b} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H}_c^{(\nu)} \mathbf{c} + \boldsymbol{\eta}_{\nu, c} \\ \mathbf{B} \mathbf{h}_\nu + \boldsymbol{\eta}_{\nu, b} \end{bmatrix} \end{aligned} \quad (14)$$

<sup>1</sup>Different from, e.g., [10] and [11] where training is designed *before* ST mapping and is tailored to ST trellis codes, our approach of designing training *after* ST mapping offers flexibility to accommodate various ST coded multi-antenna systems. It also retains the performance and capacity features of any chosen ST code while facilitating transmitter and receiver design.

where  $\mathbf{H}_c^{(\nu, \mu)} := [\mathbf{I}_{N_c+L}, \mathbf{0}_{(N_c+L) \times (N_b+L)}] \bar{\mathbf{H}}^{(\nu, \mu)} \mathbf{A}_\mu$  extracts the first  $(N_c + L)$  rows and  $N_c$  columns of the channel matrix  $\bar{\mathbf{H}}^{(\nu, \mu)}$ , and  $\mathbf{H}_c^{(\nu)} := [\mathbf{H}_c^{(\nu, 1)}, \dots, \mathbf{H}_c^{(\nu, N_t)}]$ ; the parts  $\mathbf{y}_{\nu, c}$  and  $\mathbf{y}_{\nu, b}$  have lengths  $(N_c + L)$  and  $(N_b + L)$ , respectively. Stacking  $\mathbf{y}_{\nu, b}$ 's from all receive antennas into  $\mathbf{y}_b := [\mathbf{y}_{1, b}^T, \dots, \mathbf{y}_{N_r, b}^T]^T$ , we can now write the LMMSE channel estimator as

$$\hat{\mathbf{h}} = (\sigma^2 \mathbf{R}_h^{-1} + \mathbf{I}_{N_r} \otimes (\mathbf{B}^{\mathcal{H}} \mathbf{B}))^{-1} (\mathbf{I}_{N_r} \otimes \mathbf{B}^{\mathcal{H}}) \mathbf{y}_b. \quad (15)$$

It can be readily verified that (15) is equivalent to (7). Notice that as  $\hat{\mathbf{h}}$  in (15) depends only on the training blocks  $\{\mathbf{b}_\mu\}_{\mu=1}^{N_t}$ , we have indeed accomplished our goal of decoupling channel estimation from symbol decoding. This offers flexibility in designing optimal training for MIMO channels without modifying ST designs having desirable rate performance-complexity trade-offs.

Upon defining the channel estimation error as  $\check{\mathbf{h}} := \mathbf{h} - \hat{\mathbf{h}}$ , we can express its correlation as

$$\mathbf{R}_{\check{\mathbf{h}}} := \mathbb{E}[\check{\mathbf{h}} \check{\mathbf{h}}^{\mathcal{H}}] = \left( \mathbf{R}_h^{-1} + \frac{1}{\sigma^2} \mathbf{I}_{N_r} \otimes (\bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{B}}) \right)^{-1}. \quad (16)$$

The MSE of  $\hat{\mathbf{h}}$  is given by

$$\begin{aligned} \sigma_{\check{\mathbf{h}}}^2 &:= \mathbb{E}[\|\check{\mathbf{h}}\|^2] = \text{tr}[\mathbf{R}_{\check{\mathbf{h}}}] \\ &= \text{tr} \left[ \left( \mathbf{R}_h^{-1} + \frac{1}{\sigma^2} \mathbf{I}_{N_r} \otimes (\mathbf{B}^{\mathcal{H}} \mathbf{B}) \right)^{-1} \right]. \end{aligned} \quad (17)$$

Clearly, the design of training symbols across all transmit antennas affects  $\sigma_{\check{\mathbf{h}}}^2$  through  $\mathbf{B}$ . To facilitate the ensuing analysis on the design of  $\mathbf{B}$ , we assume that:

- A1)** The channel coefficients  $h^{(\nu, \mu)}(l)$  are independently Gaussian distributed, and the covariance matrices  $\mathbf{R}_{h_\nu} := \mathbb{E}[\mathbf{h}_\nu \mathbf{h}_\nu^{\mathcal{H}}]$  are the same  $\forall \nu \in [1, N_r]$ ; i.e.,  $\mathbf{R}_h$  is block diagonal and can be written as  $\mathbf{R}_h = \mathbf{I}_{N_r} \otimes \mathbf{R}_{h_\nu}$  with trace  $\text{tr}[\mathbf{R}_h] = N_t N_r$ .

Note that **A1)** will not affect the optimality of our training design, simply because no CSI is assumed available at the transmitter. Based on **A1)**, (16) becomes

$$\mathbf{R}_{\check{\mathbf{h}}} = \mathbf{I}_{N_r} \otimes \left( \mathbf{R}_{h_1}^{-1} + \frac{1}{\sigma^2} \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{B}} \right)^{-1}. \quad (18)$$

Consequently, using [20, Lemma 7] and **A1)**, it can be shown that  $\sigma_{\check{\mathbf{h}}}^2$  in (17) is lower bounded by

$$\begin{aligned} \sigma_{\check{\mathbf{h}}}^2 &= N_r \text{tr} \left[ \left( \mathbf{R}_{h_1}^{-1} + \frac{1}{\sigma^2} \mathbf{B}^{\mathcal{H}} \mathbf{B} \right)^{-1} \right] \\ &\geq N_r \sum_{m=1}^{N_t(L+1)} \left( [\mathbf{R}_{h_1}^{-1}]_{m, m} + \frac{1}{\sigma^2} [\mathbf{B}^{\mathcal{H}} \mathbf{B}]_{m, m} \right)^{-1} \end{aligned} \quad (19)$$

where the equality holds if and only if  $\mathbf{B}^{\mathcal{H}} \mathbf{B}$  is a diagonal matrix. Therefore, the following condition is required for our training strategy to attain the minimum channel MSE  $\sigma_{\check{\mathbf{h}}}^2$ .

- C1)** For fixed  $N_b$  and  $N_c$ , the training symbols are inserted so that the matrix  $\mathbf{B}^{\mathcal{H}} \mathbf{B}$  is diagonal, or, equivalently, so that the training sequences across transmit antennas are orthogonal.

For single-input single-output (SISO) channels, **C1** coincides with that in [7] and [8], and it is also assumed *a fortiori* by [1]. In addition to TDM orthogonality asserted by Lemma 1 per antenna, **C1** establishes channel MMSE optimality by designing orthogonal training sequences across transmit antennas.

Although we have set our channel estimator in (15), and our optimal guideline **C1**, there are additional training parameters that have not been decided, such as the optimal placement and the optimal number of training symbols. These parameters affect the performance of MIMO channel estimation, the effective transmission rate, the mutual information, as well as the bit-error rate (BER). In the following, we will select these training parameters by optimizing an ergodic (average) capacity bound.

### B. Ergodic Capacity Bounds

Our criterion for optimal training will be a lower bound of the ergodic (average) capacity, due to the difficulty associated with deriving an exact average capacity formula facilitating optimization. The optimal training parameters will be those maximizing this lower bound of the average capacity. But, we will also invoke an upper bound of the average capacity to serve as a benchmark for the maximum rate achievable by single-carrier transmissions over MIMO frequency-selective channels.

Collecting the  $N_r$  received symbol blocks corresponding to all receive antennas, the information-bearing part of (14) becomes

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{c} + \boldsymbol{\eta}_c \quad (20)$$

where  $\boldsymbol{\eta}_c := [\boldsymbol{\eta}_{1,c}^T, \dots, \boldsymbol{\eta}_{N_r,c}^T]^T$ ,  $\mathbf{y}_c := [\mathbf{y}_{1,c}^T, \dots, \mathbf{y}_{N_r,c}^T]^T$ , and

$$\mathbf{H}_c := \begin{bmatrix} \mathbf{H}_c^{(1,1)} & \dots & \mathbf{H}_c^{(1,N_t)} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_c^{(N_r,1)} & \dots & \mathbf{H}_c^{(N_r,N_t)} \end{bmatrix}_{N_r(N_c+L) \times N_t N_c}.$$

Let  $\mathcal{P}$  denote the total transmit-power per block,  $\mathcal{P}_c$  the power allocated to the information part, and  $\mathcal{P}_b$  the power allocated to the training part. Let  $\hat{\mathbf{h}}$  be any estimator of  $\mathbf{h}$ . Since training symbols  $\mathbf{b}$  do not convey information, for a fixed power  $\mathcal{P}_c := \mathbb{E}[\|\mathbf{c}\|^2]$ , the mutual information between transmitted information symbols and received symbols in (20) is given by  $\mathcal{I}(\mathbf{y}_c; \mathbf{c} | \hat{\mathbf{h}})$ . The channel capacity averaged over the random channel  $\mathbf{h}$  is defined as

$$C := \frac{1}{N} \mathbb{E} \left[ \max_{p_c(\cdot), \mathcal{P}_c} \mathcal{I}(\mathbf{y}_c; \mathbf{c} | \hat{\mathbf{h}}) \right] \text{ bits/s/Hz} \quad (21)$$

where  $p_c(\cdot)$  denotes the probability density function of  $\mathbf{c}$ .

When the channel estimate is perfect, i.e.,  $\hat{\mathbf{h}} \equiv \mathbf{h}$ , the upper bound of the capacity can be obtained for a Gaussian distributed  $\mathbf{c}$  with  $\mathbf{R}_c := \mathbb{E}[\mathbf{c}\mathbf{c}^{\mathcal{H}}]$  (see e.g., [1], [17], and [20]) as

$$\bar{C} := \frac{1}{N} \mathbb{E} \left[ \max_{\mathbf{R}_c} \log \det \left( \mathbf{I}_{N_r(N_c+L)} + \frac{1}{\sigma^2} \mathbf{H}_c \mathbf{R}_c \mathbf{H}_c^{\mathcal{H}} \right) \right] \text{ bits/s/Hz.} \quad (22)$$

Recall that  $\mathbf{c}$  is obtained by encoding the information symbol block  $\mathbf{s}$  in space and time. When  $\mathbf{s}$  is Gaussian,  $\mathbf{c}$  will also be approximately Gaussian for many ST mappers, such as block codes and BLAST-type multiplexers [9], [15], [16], [26], [28]. Even if  $\mathbf{s}$  is drawn from a finite alphabet, when the block size is large, with precoding,  $\mathbf{c}$  is also approximately Gaussian distributed. For simplicity, we assume the following.

**A2)** The information-bearing symbol block  $\mathbf{c}$  is zero-mean Gaussian with covariance  $\mathbf{R}_c = \bar{\mathcal{P}}_c \mathbf{I}_{N_t N_c}$ , where  $\bar{\mathcal{P}}_c := \mathcal{P}_c / (N_t N_c)$  is the normalized power.

Note that for some ST codes (e.g., ST orthogonal codes [23]), **A2**) may not hold true. However, going through the same procedure as follows, similar results hold with different  $\mathbf{R}_c$ . With **A2**), the capacity upper bound (22) becomes

$$\bar{C} := \frac{1}{N} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{N_r(N_c+L)} + \frac{\bar{\mathcal{P}}_c}{\sigma^2} \mathbf{H}_c \mathbf{H}_c^{\mathcal{H}} \right) \right] \text{ bits/s/Hz.} \quad (23)$$

When the estimate  $\hat{\mathbf{h}}$  is not perfect, we have [cf., (20)]  $\mathbf{y}_c = \hat{\mathbf{H}}_c \mathbf{c} + \mathbf{v}$ , where  $\mathbf{v} := \check{\mathbf{H}}_c \mathbf{c} + \boldsymbol{\eta}_c$  and  $\check{\mathbf{H}}_c := \mathbf{H}_c - \hat{\mathbf{H}}_c$ . The correlation matrix of  $\mathbf{v}$  is then given by

$$\begin{aligned} \mathbf{R}_v &:= \mathbb{E}[\mathbf{v}\mathbf{v}^{\mathcal{H}}] = \mathbb{E} \left[ \check{\mathbf{H}}_c \mathbf{c} \mathbf{c}^{\mathcal{H}} \check{\mathbf{H}}_c^{\mathcal{H}} \right] + \sigma^2 \mathbf{I}_{N_r(N_c+L)} \\ &= \bar{\mathcal{P}}_c \mathbb{E} \left[ \check{\mathbf{H}}_c \check{\mathbf{H}}_c^{\mathcal{H}} \right] + \sigma^2 \mathbf{I}_{N_r(N_c+L)}. \end{aligned} \quad (24)$$

To express  $\mathbb{E}[\check{\mathbf{H}}_c \check{\mathbf{H}}_c^{\mathcal{H}}]$  in an explicit form, notice that with the diagonal structure of  $\mathbf{R}_h$  in (18), we have

$$\mathbb{E} \left[ \check{\mathbf{h}}_{\nu_1} \check{\mathbf{h}}_{\nu_2}^{\mathcal{H}} \right] = \begin{cases} \left( \mathbf{R}_{h_1}^{-1} + \frac{1}{\sigma^2} \bar{\mathbf{B}}^{\mathcal{H}} \bar{\mathbf{B}} \right)^{-1}, & \text{if } \nu_1 = \nu_2 \\ \mathbf{0}_{N_t(L+1) \times N_t(L+1)}, & \text{if } \nu_1 \neq \nu_2 \end{cases}.$$

It follows that  $\mathbb{E}[\check{\mathbf{H}}_c^{(\nu_1)} \check{\mathbf{H}}_c^{(\nu_2)\mathcal{H}}] = \mathbf{0}, \forall \nu_1 \neq \nu_2$ , and  $\mathbb{E}[\check{\mathbf{H}}_c^{(\nu)} \check{\mathbf{H}}_c^{(\nu)\mathcal{H}}]$  are the same for all  $\nu$ . As a result, we have

$$\mathbb{E} \left[ \check{\mathbf{H}}_c \check{\mathbf{H}}_c^{\mathcal{H}} \right] = \mathbf{I}_{N_r} \otimes \mathbb{E} \left[ \check{\mathbf{H}}_c^{(1)} \check{\mathbf{H}}_c^{(1)\mathcal{H}} \right].$$

Defining  $\check{\psi}_l^{(1,\mu)} := \mathbb{E}[\check{h}^{(1,\mu)}(l) \check{h}^{(1,\mu)*}(l)], \forall l \in [0, L]$ , we can show using [17, Prop. 1] that when  $N_c \gg 2L$

$$\begin{aligned} \mathbb{E} \left[ \check{\mathbf{H}}_c^{(1)} \left( \check{\mathbf{H}}_c^{(1)} \right)^{\mathcal{H}} \right] &= \sum_{\mu=1}^{N_t} \mathbb{E} \left[ \check{\mathbf{H}}_c^{(1,\mu)} \left( \check{\mathbf{H}}_c^{(1,\mu)} \right)^{\mathcal{H}} \right] \\ &\approx \sum_{\mu=1}^{N_t} \sum_{l=0}^L \check{\psi}_l^{(1,\mu)} \mathbf{I}_{N_c+L} = \sigma_{h_1}^2 \mathbf{I}_{N_c+L} \\ &= \frac{\sigma_h^2}{N_r} \mathbf{I}_{N_c+L}. \end{aligned}$$

In other words,  $\mathbb{E}[\check{\mathbf{H}}_c \check{\mathbf{H}}_c^{\mathcal{H}}]$  is approximately a diagonal matrix given by  $(\sigma_{h_1}^2 / N_r) \mathbf{I}_{N_r(N_c+L)}$ . Consequently, the resulting correlation matrix  $\mathbf{R}_v$  in (24) is as follows:

$$\mathbf{R}_v \approx \left( \frac{\bar{\mathcal{P}}_c \sigma_h^2}{N_r} + \sigma^2 \right) \mathbf{I}_{N_r(N_c+L)} \quad (25)$$

	$N_c$	$L$	$1$	$L$	$1$	$L$	$1$
$\bar{\mathbf{u}}_1$	$\mathbf{c}_1$	$\mathbf{0}_{1 \times L}$	$\mathbf{b}$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$
$\bar{\mathbf{u}}_2$	$\mathbf{c}_2$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$	$\mathbf{0}_{1 \times L}$	$\mathbf{b}$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$
$\bar{\mathbf{u}}_3$	$\mathbf{c}_3$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$	$\mathbf{0}_{1 \times L}$	$\mathbf{0}$	$\mathbf{0}_{1 \times L}$	$\mathbf{b}$

Fig. 2. Training scheme example ( $N_t = 3$ ) for ZP-based transmissions.

from which we deduce that as  $\sigma_h^2$  decreases,  $\mathbf{R}_v$  decreases accordingly.

By employing the LMMSE channel estimator and using assumptions A1) and A2), it has been shown in [17, Lemma 2] that the capacity in (21) is lower bounded by

$$C \geq \underline{C} := \frac{1}{N} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{N_r(N_c+L)} + \bar{\mathcal{P}}_c \mathbf{R}_v^{-1} \hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^{\mathcal{H}} \right) \right] \text{ bits/s/Hz.} \quad (26)$$

Our objective is to select training parameters so that  $\underline{C}$  is maximized. Intuitively, the optimal training parameters should improve both the lower capacity bound and the channel estimator together with its associated MMSE. To solidify this intuition, we borrow the following result from [17, Appendix C].

*Lemma 2:* Suppose C1), A1), and A2) hold true, and the information symbol power  $\bar{\mathcal{P}}_c$  and the training (information) block lengths  $N_b(N_c)$  are fixed. Then, maximizing  $\underline{C}$  in (26) is equivalent to minimizing  $\mathbf{R}_v$  in (25), at a high signal-to-noise ratio (SNR).

The implication of Lemma 2 is that  $\underline{C}$  increases as  $\mathbf{R}_v$  decreases, which occurs when  $\sigma_h^2$  decreases, as implied by (25). Our intuition that better channel estimation (lower MMSE) induces a higher average capacity bound has now been solidified.

### C. Optimal Training Parameters

Under C1), the lower bound on  $\sigma_h^2$  in (17) is achieved with

$$\sigma_h^2 = N_r \sum_{m=1}^{N_t(L+1)} \left( [\mathbf{R}_{h_1}^{-1}]_{m,m} + \frac{1}{\sigma^2} [\mathbf{B}^{\mathcal{H}} \mathbf{B}]_{m,m} \right)^{-1}.$$

This bound is further lower bounded by

$$\sigma_h^2 \geq N_r \sum_{m=1}^{N_t(L+1)} \left( [\mathbf{R}_{h_1}^{-1}]_{m,m} + \frac{1}{\sigma^2} \frac{\mathcal{P}_b}{N_t} \right)^{-1}$$

where the equality holds if and only if  $\mathbf{B}^{\mathcal{H}} \mathbf{B} = (\mathcal{P}_b/N_t) \mathbf{I}_{N_t(L+1)}$ . Condition C1) can now be modified as follows.

C1') For fixed  $N_b$  and  $N_c$ , the training symbols are inserted so that  $\mathbf{B}^{\mathcal{H}} \mathbf{B} = (\mathcal{P}_b/N_t) \mathbf{I}_{N_t(L+1)}$ .

By the definition of  $\mathbf{B}$ , the Toeplitz training matrices of every transmit antenna pair should satisfy

$$\mathbf{B}_{\mu_1}^{\mathcal{H}} \mathbf{B}_{\mu_2} = \frac{\mathcal{P}_b}{N_t} \mathbf{I}_{L+1} \delta(\mu_1 - \mu_2), \quad \forall \mu_1, \mu_2 \in [1, N_t]. \quad (27)$$

Since according to C1'),  $\mathbf{B}$  is an  $(N_b + L) \times N_t(L+1)$  matrix with full column rank, the minimum possible number of training symbols is given by  $N_b = N_t(L+1) - L$ , which suggests only one nonzero entry for each transmit antenna. In fact, for fixed

$\mathcal{P}_c$  and  $\mathcal{P}_b$ , as  $N_b$  increases beyond  $N_t(L+1) - L$ , the average capacity bound  $\underline{C}$  decreases. We summarize our results so far in the following.

*Proposition 1:* Suppose A1) and A2) hold true. For fixed  $\mathcal{P}_c$  and  $\mathcal{P}_b$ , the optimal placement of each block from the  $\mu$ th transmit antenna is  $[c_{\mu}^T \mathbf{0}_L^T \mathbf{b}_{\mu}^T]^T$ , where  $\mathbf{b}_{\mu}$  is selected to satisfy (27) with length  $N_b = N_t(L+1) - L$ .

One simple example is to design  $\mathbf{b}_{\mu}$  as  $\mathbf{b}_{\mu} = [\mathbf{0}_{(\mu-1)(L+1)}^T \mathbf{b} \mathbf{0}_{(N_t-\mu)(L+1)}^T]^T$ . With this structure of training blocks, we have

$$\mathbf{B}_{\mu} = \sqrt{\bar{\mathcal{P}}_b} [\mathbf{0}_{(L+1) \times (\mu-1)(L+1)} \mathbf{I}_{L+1} \mathbf{0}_{(L+1) \times (N_t-\mu)(L+1)}]^T.$$

Evidently, the proposed placement satisfies condition C1'). As a result, the channel MMSE is achieved and the average capacity's lower bound  $\underline{C}$  is maximized.

An example of the optimal placement with  $N_t = 3$  transmit antennas and frequency-selective fading channels of order  $L$  is shown in Fig. 2. Notice that for given  $N_t$  and  $L$ , the design of  $\mathbf{b}_{\mu}$  is not unique. With the factor  $\sigma_{\hat{\mathbf{H}}_c}^2 := \text{tr}(\mathbb{E}[\hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^{\mathcal{H}}])$ , we can normalize the estimated channel matrix as

$$\hat{\mathbf{H}}_c := \frac{1}{\sigma_{\hat{\mathbf{H}}_c}} \hat{\mathbf{H}}_c. \quad (28)$$

Substituting (25) and (28) into (26) yields the capacity lower bound when optimal placement is employed

$$\underline{C} = \frac{1}{N} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{N_r(N_c+L)} + \rho_{\text{eff}} \hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^{\mathcal{H}} \right) \right] \text{ bits/s/Hz} \quad (29)$$

where

$$\rho_{\text{eff}} := \frac{N_r \bar{\mathcal{P}}_c \sigma_{\hat{\mathbf{H}}_c}^2}{\bar{\mathcal{P}}_c \sigma_h^2 + N_r \sigma^2}.$$

It can be easily verified that  $\sigma_{\hat{\mathbf{H}}_c}^2 = N_c(N_t N_r - \sigma_h^2)$ . Moreover, with  $\psi_l^{(1,\mu)} := \mathbb{E}[h^{(1,\mu)}(l) h^{(1,\mu)*}(l)]$ , we have

$$\sigma_h^2 = N_r \sum_{\mu=1}^{N_t} \sum_{l=0}^L \frac{1}{\frac{\mathcal{P}_b}{N_t \sigma^2} \psi_l^{(1,\mu)} + 1} \psi_l^{(1,\mu)}. \quad (30)$$

Consider now the total power  $\mathcal{P} = \mathcal{P}_c + \mathcal{P}_b$ , and define the power allocation factor  $\alpha_{\text{zp}} := \mathcal{P}_c/\mathcal{P} \in (0, 1)$ . Equation (30) can then be expressed as

$$\rho_{\text{eff}} = \frac{\alpha_{\text{zp}} N_r \mathcal{P} (N_t N_r - \sigma_h^2)}{\alpha_{\text{zp}} \frac{\mathcal{P}}{N_c} \sigma_h^2 + N_r \sigma^2}. \quad (31)$$

Due to the fact that  $\sigma_h^2$  is dependent on  $\psi_l^{(1,\mu)}$ , it is difficult to find an optimal power allocation factor that does not depend on

any CSI directly from (31). Therefore, we will consider the following three cases: low SNR, high SNR, and identically distributed channel taps.

1) *Case I—Low SNR* ( $\mathcal{P}/N_t \ll \sigma^2$ ): In this case, (30) becomes  $\sigma_h^2 \approx N_t N_r - N_r(1 - \alpha_{\text{zp}})\mathcal{P}/\sigma^2$ . Plugging this result into (31), we obtain

$$\rho_{\text{eff}} \approx \frac{\alpha_{\text{zp}}(1 - \alpha_{\text{zp}})N_c N_r \mathcal{P}^2}{\alpha_{\text{zp}}\mathcal{P}(N_t \sigma^2 - (1 - \alpha_{\text{zp}})\mathcal{P}) + N_c \sigma^4}. \quad (32)$$

Differentiating  $\rho_{\text{eff}}$  with respect to  $\alpha_{\text{zp}}$ , we find that, at low SNR, the optimal power allocation factor is

$$\alpha_{\text{zp}} \approx \frac{1}{2}. \quad (33)$$

2) *Case II—High SNR* ( $\mathcal{P}/N_t \gg \sigma^2$ ): In this case, (30) becomes  $\sigma_h^2 \approx N_t^2(L+1)\sigma^2/\mathcal{P}_b$ . Plugging this result into (31), we obtain

$$\rho_{\text{eff}} \approx \frac{\alpha_{\text{zp}} N_r N_t \mathcal{P} \left(1 - N_t(L+1)\frac{\sigma^2}{\mathcal{P}_b}\right)}{N_t^2(L+1)\frac{\alpha_{\text{zp}}\sigma^2}{N_c(1-\alpha_{\text{zp}})} + \sigma^2}. \quad (34)$$

Differentiating  $\rho_{\text{eff}}$  with respect to  $\alpha_{\text{zp}}$ , we find that, at low SNR, the optimal power allocation factor is

$$\alpha_{\text{zp}} = \frac{1 - \sqrt{\lambda}\sqrt{1 + (1-\lambda)N_c\sigma^2/\mathcal{P}}}{1 - \lambda} \approx \frac{1}{1 + \sqrt{\lambda}} \quad (35)$$

where  $\lambda := N_t^2(L+1)/N_c$ .

3) *Case III—Identically Distributed Channel Coefficient* ( $\psi_b^{(1,\mu)} = 1/(L+1)$ ): In this case, (30) becomes

$$\sigma_h^2 = \frac{N_t N_r}{\frac{(1-\alpha_{\text{zp}})\mathcal{P}}{N_t(L+1)\sigma^2} + 1}.$$

Plugging this result into (31), we obtain

$$\rho_{\text{eff}} = \frac{\alpha_{\text{zp}} N_r N_t \mathcal{P} \left(1 - \frac{1}{\frac{(1-\alpha_{\text{zp}})\mathcal{P}}{N_t(L+1)\sigma^2} + 1}\right)}{N_t \frac{\alpha_{\text{zp}}\mathcal{P}}{N_c} \frac{1}{\frac{(1-\alpha_{\text{zp}})\mathcal{P}}{N_t(L+1)\sigma^2} + 1} + \sigma^2}. \quad (36)$$

Differentiating  $\rho_{\text{eff}}$  with respect to  $\alpha_{\text{zp}}$ , we find that at low SNR the optimal power allocation factor is

$$\alpha_{\text{zp}} = \frac{\gamma - \sqrt{\gamma}\sqrt{\gamma - (1-\lambda)}}{1 - \lambda} \quad (37)$$

where  $\gamma := 1 + N_t(L+1)\sigma^2/\mathcal{P}$ .

In addition to guiding our selection of optimal training parameters for ZP-based single-carrier MIMO transmissions, (23) and (29) are interesting on their own since they offer simple closed forms for bounding the average capacity of MIMO frequency-selective fading channels. The optimal training parameters for ZP-based block transmissions are tabulated in Table I. In fact, the optimal placement and design of training symbols given in Table I and illustrated in Fig. 2 are not unique. The structure of training blocks can be shuffled among the  $N_t$  transmit antennas without affecting either the channel MSE or the capacity lower bound.

TABLE I  
SUMMARY OF DESIGN PARAMETERS FOR ZP-BASED SCHEMES

Parameters	Optimal training
Placement of training symbols	After the information symbols and $L$ trailing zeros
Structure of training sub-blocks	$\mathbf{b}_\mu = [\mathbf{0}_{(\mu-1)(L+1)}^T \ b \ \mathbf{0}_{(N_t-\mu)(L+1)}^T]^T, \forall \mu \in [1, N_t]$
Number of Training symbols	$N_t(L+1)$ per block
Power Allocation	$\alpha_{\text{zp}} = \frac{\sqrt{N_c}}{\sqrt{N_c} + \sqrt{N_t(L+1)}}$

#### IV. CP-BASED BLOCK TRANSMISSIONS

Instead of having  $L$  zeros at the end of each block  $\mathbf{u}_\mu$ , an alternative way to eliminate IBI is by adding a CP of length  $L$  at the transmitter and removing it at the receiver. As in OFDM, the CP-inducing and CP-removing matrices are defined, respectively, as

$$\mathbf{T} = \mathbf{T}_{\text{cp}} := [\mathbf{I}_{\text{cp}} \ \mathbf{I}_M]^T \text{ and } \mathbf{R} = \mathbf{R}_{\text{cp}} := [\mathbf{0}_{M \times L} \ \mathbf{I}_M]$$

where  $\mathbf{I}_{\text{cp}}$  contains the last  $L$  columns of  $\mathbf{I}_M$ . In this case, the channel matrix becomes

$$\tilde{\mathbf{H}}^{(\nu,\mu)} = \mathbf{R}_{\text{cp}} \mathbf{H}^{(\nu,\mu)} \mathbf{T}_{\text{cp}}$$

which is an  $M \times M$  columnwise circulant matrix with first column  $[h^{(\nu,\mu)}(0), \dots, h^{(\nu,\mu)}(L), 0, \dots, 0]^T$ . The input–output relationship in matrix-vector form is then given by

$$\mathbf{y}_\nu = \sum_{\mu=1}^{N_t} \left( \tilde{\mathbf{H}}^{(\nu,\mu)} \mathbf{A}_\mu \mathbf{c}_\mu + \tilde{\mathbf{H}}^{(\nu,\mu)} \tilde{\mathbf{b}}_\mu \right) + \boldsymbol{\xi}_\nu \quad (38)$$

where  $\boldsymbol{\xi}_\nu := \mathbf{R}_{\text{cp}} \boldsymbol{\eta}_\nu$ . As in the ZP case, the roles of  $\tilde{\mathbf{H}}^{(\nu,\mu)}$  and  $\tilde{\mathbf{b}}_\mu$  can be interchanged, i.e.,  $\tilde{\mathbf{H}}^{(\nu,\mu)} \tilde{\mathbf{b}}_\mu = \tilde{\mathbf{B}}_\mu \mathbf{h}^{(\nu,\mu)}$  with  $\tilde{\mathbf{B}}_\mu$  an  $M \times (L+1)$  columnwise circulant matrix with first column  $[\tilde{b}(1), \dots, \tilde{b}(M)]^T$ . Concatenating the  $N_r$  channel vectors and defining  $\tilde{\mathbf{B}} := [\tilde{\mathbf{B}}_1, \dots, \tilde{\mathbf{B}}_{N_t}]$ , we have

$$\mathbf{y} = \tilde{\Phi} \mathbf{c} + (\mathbf{I}_{N_r} \otimes \tilde{\mathbf{B}}) \mathbf{h} + \boldsymbol{\xi}$$

where the  $N_r M \times N_t N_c$  matrix  $\tilde{\Phi}$  has the same structure as the  $N_r N \times N_t N_c$  matrix in (6) but with  $\tilde{\mathbf{H}}^{(\nu,\mu)}$ 's in place of  $\mathbf{H}^{(\nu,\mu)}$ 's. The LMMSE channel estimator for CP-based transmissions is then given by

$$\hat{\mathbf{h}} := \left( \sigma^2 \mathbf{R}_h^{-1} + \mathbf{I}_{N_r} \otimes (\tilde{\mathbf{B}}^{\mathcal{H}} \tilde{\mathbf{B}}) \right)^{-1} (\mathbf{I}_{N_r} \otimes \tilde{\mathbf{B}}^{\mathcal{H}}) \mathbf{y}. \quad (39)$$

Furthermore, for the purpose of decoupling symbol detection from channel estimation, we need

$$\tilde{\mathbf{B}}_{\mu_1}^{\mathcal{H}} \tilde{\mathbf{H}}^{(\nu,\mu_2)} \mathbf{A}_{\mu_2} = \mathbf{0}, \quad \forall \mu_1, \mu_2 \in [1, N_t] \text{ and } \nu \in [1, N_r]. \quad (40)$$

As in (10), we re-express  $\mathbf{A}_{\mu_2}$  in terms of its  $N_c$  columns as follows:

$$\begin{aligned} \tilde{\mathbf{H}}^{(\nu,\mu_2)} \mathbf{A}_{\mu_2} &= \left[ \tilde{\mathbf{H}}^{(\nu,\mu_2)} \mathbf{a}_{\mu_2,1} \cdots \tilde{\mathbf{H}}^{(\nu,\mu_2)} \mathbf{a}_{\mu_2,N_c} \right] \\ &= \left[ \tilde{\mathbf{A}}_{\mu_2,1} \mathbf{h}^{(\nu,\mu_2)} \cdots \tilde{\mathbf{A}}_{\mu_2,N_c} \mathbf{h}^{(\nu,\mu_2)} \right] \end{aligned} \quad (41)$$

where  $\tilde{\mathbf{A}}_{\mu_2,n}$  is an  $M \times (L+1)$  columnwise circulant matrix with first column given by the  $M \times 1$  vector  $\mathbf{a}_{\mu_2,n}$ . For (40) to hold true over *all* channel realizations, the  $(\mathbf{A}_\mu, \tilde{\mathbf{b}}_\mu)$  pairs  $\forall \mu_1, \mu_2 \in [1, N_t], \nu \in [1, N_r]$  and  $\forall n \in [1, N_c]$  must satisfy

$$\tilde{\mathbf{B}}_{\mu_1}^{\mathcal{H}} \tilde{\mathbf{A}}_{\mu_2,n} = \mathbf{0}. \quad (42)$$

*Lemma 3 [Decoupling Training From Information]:* The designs  $(\mathbf{A}_\mu, \bar{\mathbf{b}}_\mu)$  that guarantee decoupling of channel estimation from symbol detection must satisfy (42). The desirable design is given by

$$\mathbf{A}_\mu = \mathbf{F}_M^H \mathbf{P}_A \Theta, \quad \bar{\mathbf{b}}_\mu = \mathbf{F}_M^H \mathbf{P}_b \mathbf{b}_\mu \quad (43)$$

where  $\Theta$  is an  $N'_c \times N_c$  matrix that optionally precodes (if  $\Theta \neq \mathbf{I}_{N_c}$ ) the information-bearing block  $\mathbf{c}_\mu$  linearly,  $\mathbf{b}_\mu$  consists of the  $N_b := M - N'_c$  possibly nonzero training symbols from the  $\mu$ th transmit antenna, and permutation matrices  $\mathbf{P}_A, \mathbf{P}_b$  satisfy  $\mathbf{P}_A^T \mathbf{P}_b = \mathbf{0}_{N'_c \times N_b}$ .

*Proof:* See Appendix B.  $\square$

Notice that the permutation matrices play the role of assigning OFDM carriers to each (possibly precoded) information symbol and training symbol (pilot tones). Lemma 3 reveals that we need to assign different frequency tones to information-bearing and training symbols in order to separate channel estimation from symbol decoding in the frequency domain. As a result, the superimposed model boils down to FDM, the ST codewords  $\mathbf{c}_\mu$  with the training blocks  $\mathbf{b}_\mu$ . Notice that as with the single-carrier ZP-based transmissions of the previous section, the multicarrier FDM separability was not assumed at the outset. It followed naturally from the CP-based IBI removal and our objective of decoupling symbol detection from MIMO channel estimation.

Since our objective is to design an optimal training structure, the precoder choice is out of the scope of this paper. For simplicity, we choose  $N'_c = N_c$  and  $\Theta = \mathbf{I}_{N_c}$ . With IFFT operations at the transmitter, it is natural to employ FFT at the receiver and end up with a MIMO OFDM system. The FFT processed block at the  $\nu$ th receiver is

$$\mathbf{z}_\nu := \mathbf{F}_M \mathbf{y}_\nu = \sum_{\mu=1}^{N_t} \left( \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_A \mathbf{c}_\mu + \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_b \mathbf{b}_\mu \right) + \xi_\nu, \quad (44)$$

With the mutually orthogonal permutation matrices  $\mathbf{P}_A$  and  $\mathbf{P}_b$ , we can now easily decouple the training from the information-bearing parts

$$\begin{aligned} \mathbf{z}_{\nu, b} &:= \mathbf{P}_b^T \mathbf{z}_\nu = \sum_{\mu=1}^{N_t} \mathbf{P}_b^T \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_b \mathbf{b}_\mu + \xi_{\nu, b} \\ \mathbf{z}_{\nu, c} &:= \mathbf{P}_A^T \mathbf{z}_\nu = \sum_{\mu=1}^{N_t} \mathbf{P}_A^T \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_A \mathbf{c}_\mu + \xi_{\nu, c} \end{aligned} \quad (45)$$

where the noise terms are defined as  $\xi_{\nu, b} := \mathbf{P}_b^T \xi_\nu$  and  $\xi_{\nu, c} := \mathbf{P}_A^T \xi_\nu$ . Using the definition of permutation matrices and the definition of  $\mathbf{D}_H^{(\nu, \mu)}$ , it can be verified that

$$\begin{aligned} \mathbf{P}_b^T \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_b \mathbf{b}_\mu &= \text{diag} \left\{ \mathbf{P}_b^T \mathbf{F}_{0:L} \mathbf{h}^{(\nu, \mu)} \right\} \mathbf{b}_\mu \\ &= \mathbf{B}_\mu \mathbf{P}_b^T \mathbf{F}_{0:L} \mathbf{h}^{(\nu, \mu)} \end{aligned}$$

with  $\mathbf{B}_\mu := \text{diag} \{ \mathbf{b}_\mu \}$  being an  $N_b \times N_b$  diagonal matrix. The input–output relationship for training symbols then becomes

$$\mathbf{z}_{\nu, b} = \mathbf{B} \mathbf{h}_\nu + \xi_{\nu, b} \quad (46)$$

where  $\mathbf{B} := \sqrt{M} [\mathbf{B}_1 \mathbf{P}_b^T \mathbf{F}_{0:L}, \dots, \mathbf{B}_{N_t} \mathbf{P}_b^T \mathbf{F}_{0:L}]$ . Notice that as  $\Phi$  in the ZP case,  $\mathbf{B}$  is the same  $\forall \nu \in [1, N_r]$ .

Stacking  $\mathbf{z}_{\nu, b}$ 's from all receive antennas into  $\mathbf{z}_b := [\mathbf{z}_{1, b}^T, \dots, \mathbf{z}_{N_r, b}^T]^T$ , the LMMSE channel estimator is now given by

$$\hat{\mathbf{h}} = (\sigma^2 \mathbf{R}_h^{-1} + \mathbf{I}_{N_r} \otimes (\mathbf{B}^H \mathbf{B}))^{-1} (\mathbf{I}_{N_r} \otimes \mathbf{B}^H) \mathbf{z}_b \quad (47)$$

which can be easily shown to be equivalent to the one in (39) but uses only the training symbols.

Defining the  $N_c \times N_c$  diagonal matrix  $\mathbf{D}_{H_c}^{(\nu, \mu)} := \mathbf{P}_A^T \mathbf{D}_H^{(\nu, \mu)} \mathbf{P}_A$  and stacking  $N_r N_t$  such matrices into

$$\mathbf{D}_{H_c} := \begin{bmatrix} \mathbf{D}_{H_c}^{(1,1)} & \dots & \mathbf{D}_{H_c}^{(1, N_t)} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{H_c}^{(N_r, 1)} & \dots & \mathbf{D}_{H_c}^{(N_r, N_t)} \end{bmatrix}_{N_r N_c \times N_t N_c}$$

we arrive at the overall input–output relationship of information-carrying symbols

$$\mathbf{z}_c = \mathbf{D}_{H_c} \mathbf{c} + \xi_c. \quad (48)$$

Comparing (47) and (48) in CP with (15) and (20) of ZP, we find that the LMMSE estimators and the overall input–output relationships in both cases share the same form. Therefore, the analysis of one can be carried over to the other.

Not surprisingly, we find that  $\mathbf{C}1'$ ) also applies to the CP case, and the minimum channel MSE

$$\sigma_h^2 = N_r \sum_{m=1}^{N_t(L+1)} \left( [\mathbf{R}_h^{-1}]_{m, m} + \frac{1}{\sigma^2} \frac{\mathcal{P}_b}{N_t} \right)^{-1}$$

is the same as in the ZP case and can be achieved if and only if  $\mathbf{B}^H \mathbf{B} = (\mathcal{P}_b / N_t) \mathbf{I}_{N_t(L+1)}$ , i.e.,  $\forall \mu_1, \mu_2 \in [1, N_t]$ ,

$$M \mathbf{F}_{0:L}^H \mathbf{P}_b \mathbf{B}_{\mu_1}^H \mathbf{B}_{\mu_2} \mathbf{P}_b^T \mathbf{F}_{0:L} = \frac{\mathcal{P}_b}{N_t} \mathbf{I}_{L+1} \delta(\mu_1 - \mu_2). \quad (49)$$

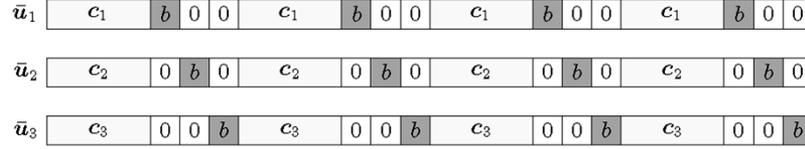
Using the fact that  $\mathbf{P}_b^T \mathbf{P}_b = \mathbf{I}_{N_b}$ , and letting  $\mathbf{D}_{\tilde{B}_\mu} := \sqrt{M} \mathbf{P}_b \mathbf{B}_\mu^H \mathbf{P}_b^T$ , we can re-express (49) as  $\forall \mu_1, \mu_2 \in [1, N_t]$

$$\mathbf{F}_{0:L}^H \mathbf{D}_{\tilde{B}_{\mu_1}} \mathbf{D}_{\tilde{B}_{\mu_2}}^H \mathbf{F}_{0:L} = \frac{\mathcal{P}_b}{N_t} \mathbf{I}_{L+1} \delta(\mu_1 - \mu_2). \quad (50)$$

Clearly, when  $\mu_1 = \mu_2$ , (50) can be satisfied if  $K (> L)$  out of the  $N_b$  pilot tones are equi-spaced and equi-powered with  $\mathcal{P}_b / (K N_t)$ , and the rest of the  $N_b - K$  pilots are zero. When  $\mu_1 \neq \mu_2$ , setting  $\mathbf{D}_{\tilde{B}_{\mu_1}}^H \mathbf{D}_{\tilde{B}_{\mu_2}} = \mathbf{0}$  will satisfy (50). To do so, one pilot tone cannot be shared by more than one nonzero training symbol on different transmit antennas. In other words, the nonzero pilot tones across the  $N_t$  transmit antennas are mutually orthogonal:  $\mathbf{B}_{\mu_1}^H \mathbf{B}_{\mu_2} = \mathbf{0}_{N_b \times N_b}$ ,  $\forall \mu_1 \neq \mu_2$ .

Recall that the number of nonzero pilot tones in each transmit antenna is  $K$  and is lower bounded by  $L + 1$ . For the same channel MSE and a given power allocation ( $\mathcal{P}_c$  and  $\mathcal{P}_b$  fixed), as the number of nonzero pilot tones increases, the capacity lower bound decreases. Therefore, the number of nonzero pilot tones should be chosen to be the minimum possible:  $K = L + 1$ .

*Proposition 2:* Suppose A1) and A2) hold true. For fixed  $\mathcal{P}_c$  and  $\mathcal{P}_b$ , the following placement is optimal: the  $L + 1$  subblocks of pilot tones for the  $\mu$ th transmit antenna are inserted equi-spaced into the ST mapper output; each of the training subblocks


 Fig. 3. Training scheme example ( $L = 3, N_t = 3$ ) for CP-based transmissions.

has structure  $[\mathbf{0}_{\mu-1}^T b \mathbf{0}_{N_t-\mu}^T]^T$  and length  $\bar{N}_b := N_t$ ; all pilot symbols are equi-powered with  $\bar{\mathcal{P}}_b := \mathcal{P}_b/(N_t(L+1))$ .

An example of the optimal placement of pilot tones with  $N_t = 3$  transmit antennas and frequency-selective fading channels of order  $L = 3$  is shown in Fig. 3. As before, this example gives an optimal but not unique placement. For instance, shuffling the plotted structure among the  $N_t$  transmit antennas yields other placements that are also optimal. With the optimal training placement, the capacity lower bound for CP-based block transmissions is given by [cf., (26)]

$$\underline{C} := \frac{1}{N} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{N_r N_c} + \bar{\mathcal{P}}_c \mathbf{R}_v^{-1} \hat{\mathbf{D}}_{H_c} \hat{\mathbf{D}}_{H_c}^{\mathcal{H}} \right) \right] \quad (51)$$

where  $\bar{\mathcal{P}}_c = \mathcal{P}_c/(N_t N_c)$ . With  $\mathcal{P}$  denoting the total power per block, notice that the sum  $\mathcal{P}_b + \mathcal{P}_c$  is the ‘‘total’’ power excluding the CP, i.e.,  $\mathcal{P}_b + \mathcal{P}_c = \mathcal{P}M/N$ . Defining the normalization factor  $\sigma_{\hat{\mathbf{D}}_{H_c}}^2 := \text{tr}(\mathbb{E}[\hat{\mathbf{D}}_{H_c} \hat{\mathbf{D}}_{H_c}^{\mathcal{H}}])$ , we have the normalized channel matrix

$$\hat{\hat{\mathbf{D}}}_{H_c} := \frac{1}{\sigma_{\hat{\mathbf{D}}_{H_c}}} \hat{\mathbf{D}}_{H_c}.$$

Equation (51) can then be re-expressed as

$$\underline{C} = \frac{1}{N} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{N_r N_c} + \rho_{\text{eff}}^{\text{cp}} \hat{\hat{\mathbf{D}}}_{H_c} \hat{\hat{\mathbf{D}}}_{H_c}^{\mathcal{H}} \right) \right] \text{ bits/s/Hz} \quad (52)$$

with effective SNR

$$\rho_{\text{eff}}^{\text{cp}} := \frac{\mathcal{P}_b \bar{\mathcal{P}}_c \sigma_{\hat{\mathbf{D}}_{H_c}}^2}{(N_t^2 (L+1) \bar{\mathcal{P}}_c + \mathcal{P}_b) \sigma^2}. \quad (53)$$

Differentiating  $\rho_{\text{eff}}^{\text{cp}}$  with respect to  $\alpha_{\text{cp}} := \mathcal{P}_c/\mathcal{P} \in (0, 1)$ , we obtain the optimal power allocation factor

$$\alpha_{\text{cp}} = \frac{\sqrt{N_c}}{\sqrt{N_c} + \sqrt{N_t(L+1)}} \frac{M}{N}. \quad (54)$$

The optimal training parameters are summarized in Table II. As in the ZP case, the placement of optimal pilot tones is not unique. The structure of each of the  $L+1$  subblocks can be interchanged among transmit antennas, as long as the pilot tones from different transmit antennas are distinct (and thus orthogonal) and those from the same transmit antenna are equi-spaced.

## V. COMPARISONS AND DISCUSSION

In the preceding sections, we designed two optimal training schemes for MIMO frequency-selective fading channels, which are tailored for ZP- and CP-based block transmissions. The former (ZP) leads to single-carrier MIMO transmissions with TDM between training and information symbols, while the

 TABLE II  
SUMMARY OF DESIGN PARAMETERS FOR CP-BASED SCHEMES

Parameters	Optimal training
Placement of training symbols	Equally long training sub-blocks (length $\bar{N}_t$ )
Structure of training sub-blocks	$\mathbf{b}_\mu = [\mathbf{0}_{(\mu-1)}^T b \mathbf{0}_{(N_t-\mu)}^T]^T, \forall l \in [0, L], \mu \in [1, N_t]$
Number of Sub-blocks	$L+1$ training and $L+1$ information sub-blocks
Number of Training symbols	$N_t(L+1)$ per block
Power Allocation	$\alpha_{\text{cp}} = \frac{\sqrt{N_c}}{\sqrt{N_c} + \sqrt{N_t(L+1)}} \frac{M}{N}$

latter (CP) leads to multicarrier MIMO transmissions with FDM between training and information symbols. Both designs are optimal in terms of minimizing channel MSE (without assuming any CSI besides the channel order  $L$ ), and maximizing the lower bound on average capacity. In both cases, the optimal designs ended up possessing (time- or frequency-domain) orthogonality among training blocks *across* the different transmit antennas.

There are tradeoffs in selecting ZP versus CP in practice. When the ST mapper output features constant-modulus, the ZP-based approach results in a two-level constant-modulus transmission, while CP-OFDM generally results in transmitted blocks with wide dynamic ranges, which increase with increasing block length. Moreover, the training symbols in ZP are simply attached to the head or tail of the transmitted blocks, while the pilot tones in CP are inserted equi-spaced into the transmitted blocks. Furthermore, ZP-based transmissions are more tolerant to frequency offsets, in contrast to CP-OFDM transmissions [25]. On the other hand, CP-OFDM transmissions provide easy implementation as well as low-complexity symbol decoding compared to ZP-based transmissions that involve relatively more complicated equalization.

The  $L$  guard symbols that are needed to enable the block-by-block decoding in either ZP or CP schemes can be also beneficial in the ST mapper design. For instance, the ZP guard we adopted to guarantee decoupling of channel estimation from symbol decoding can be shared and play an instrumental role in the ST code design of [28]. Likewise, the CP guard can be shared with the space–time–frequency code designed in [16]. Our optimal training designs merge well with any ST mapper without sacrificing their achievable diversity gains.

So far, our analysis applies to general MIMO frequency-selective fading channels. We will now consider two special cases, namely SISO frequency-selective and MIMO flat-fading channels.

### A. Special Case 1: SISO Frequency-Selective Fading Channels

With one transmit and one receive antenna, the ZP transmitted block  $\mathbf{u} = \mathbf{T}_{\text{zp}} \mathbf{A} \mathbf{c} + \bar{\mathbf{b}}$  becomes

$$\mathbf{u} = [\mathbf{c}^T \mathbf{0}_L^T b \mathbf{0}_L^T]^T \quad (55)$$

where the subscript  $\mu$  is removed. This design has the same structure as the one in [1, Theorem 4]. In [23, Theorem 1], it was stated that the optimal number of training pilots is equal to the channel length  $L + 1$ . The apparent discrepancy with (55) that requires  $2L + 1$  pilots comes from the fact that the redundancy of length  $L$ , which is needed to achieve IBI-free block transmission, is included in our training sequence but is forgotten in [23]. At high SNR, our power allocation factor  $\alpha_{zp} = \sqrt{N_c}/(\sqrt{N_c} + \sqrt{L+1})$  coincides with the designs in [1, Theorem 5] and [23, Theorem 1].

For CP-based transmissions, when the number of transmit antennas  $N_t = 1$ , our MIMO results corroborate that OFDM transmissions with equi-spaced and equi-powered pilot tones are optimal. The optimal power allocation factor is  $\alpha_{cp} = (M/N)\sqrt{N_c}/(\sqrt{N_c} + \sqrt{L+1})$ . The same conclusions were reached in both [1, Theorems 2 and 3] and [20] for SISO channels.

### B. Special Case 2: MIMO Flat-Fading Channels

When the MIMO channel is flat fading, we have  $L = 0$ , which implies that for both ZP- and CP-based transmissions one needs only a single pilot per block, per antenna. With such channels, the placement of training blocks has no effect on either the MSE or the average capacity, as long as the orthogonality among the  $N_t$  antennas is preserved. For this special case, a similar conclusion was reached in [12] where training signals with so-called “orthonormal columns” is used. At high SNR, our optimal power allocation factor  $\alpha_{cp} = \alpha_{zp} = \sqrt{N_c}/(\sqrt{N_c} + \sqrt{N_t})$  is the same as that in [12, Corollary 1].

## VI. SIMULATIONS

In this section, we present simulations to validate our analyses and designs. In all test cases, we adopt our training schemes summarized in Tables I and II, unless otherwise mentioned. The SNR is defined as the average received symbol power to noise ratio at each receive antenna.

*Test Case 1 (Training Parameters):* In this example, we depict the relationships between the bounds on average capacity and test the effect on average capacity of several training parameters.

- Average capacity versus SNR: We select  $(N_t, N_r) = (2, 2)$ ,  $L = 6$ , and the transmitted block length  $N = 62$ . We plot average capacity bounds versus SNR for both ZP and CP cases in Fig. 4. As expected, the capacity bounds increase monotonically as SNR increases. The bounds for ZP are greater than those for CP, which is mainly due to the power loss incurred by the CP.
- Average capacity versus  $N_t$ : We depict the average capacity bounds versus the number of transmit antennas in Fig. 5. Here, we fix  $N_r = 3$ ,  $L = 6$ ,  $N = 69$ , and the total power per block  $\mathcal{P}$ . When  $1 \leq N_t \leq 3$ , the average capacity increases with the number of transmit antennas. However, when  $N_t > 3$ , the average capacity starts decreasing. This is because training symbols take over larger and larger portion of the whole transmitted block, leaving less and less to be used for transmitting information symbols.

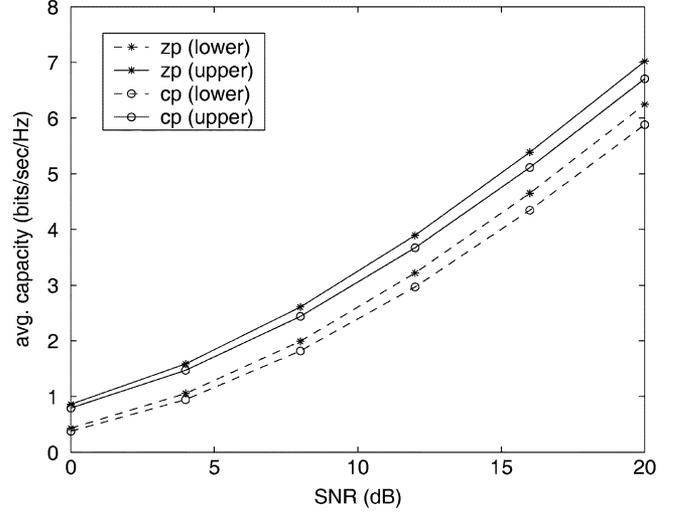


Fig. 4. Average capacity bounds comparison between ZP- and CP-based optimal training with increasing SNR ( $N_t = 2$ ,  $N_r = 2$ ,  $L = 6$ ).

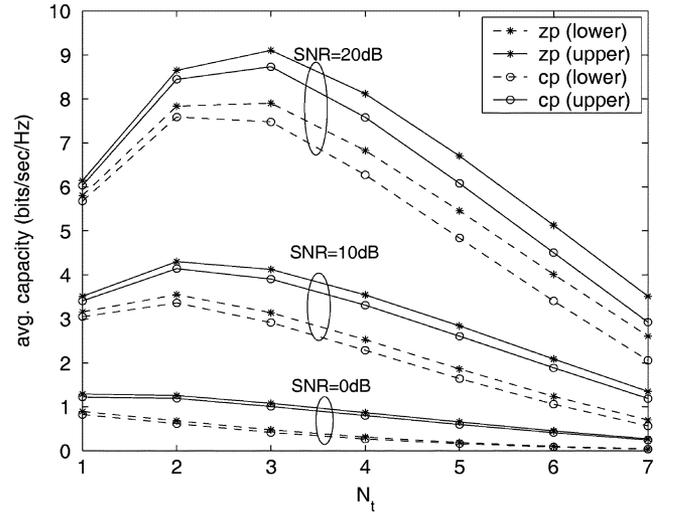


Fig. 5. Average capacity bounds comparison between ZP- and CP-based optimal training with increasing number of transmit antennas ( $N_r = 3$ ,  $L = 6$ ).

- Average capacity versus  $L$ : The relationship between average capacity bounds and channel order  $L$  is depicted in Fig. 6. We fix  $N_t = N_r = 2$ ,  $N = 59$  and the total power per block  $\mathcal{P}$ . Note that the capacity bounds for both ZP and CP decrease monotonically as  $L$  increases, since channel estimation takes an increasing amount of both time and power. It can be observed that as  $L$  increases, the gap between ZP and CP capacity increases, simply because of the power loss due to the CP scheme.
- Average capacity versus  $N_c$ : The last parameter that we tested is the information block length  $N_c$ . Here, we have  $N_t = N_r = 2$ , and  $L = 6$ . As  $N_c$  increases, the total block length  $N$  increases since  $N = N_c + N_t(L+1) + L$ . From Fig. 7, we observe that average capacity bounds increase monotonically with increasing  $N_c$ . For large  $N_c$ , the information transmission dominates the whole block period.

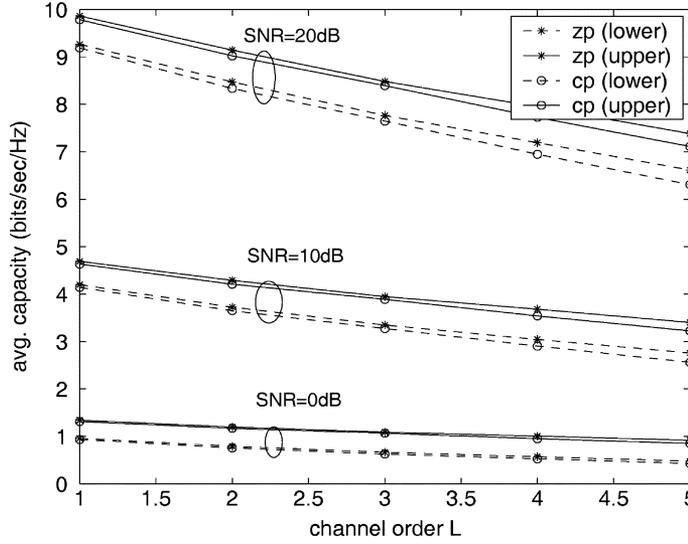


Fig. 6. Average capacity bounds comparison between ZP- and CP-based optimal training with increasing channel order  $L$  ( $N_t = 2, N_r = 2$ ).

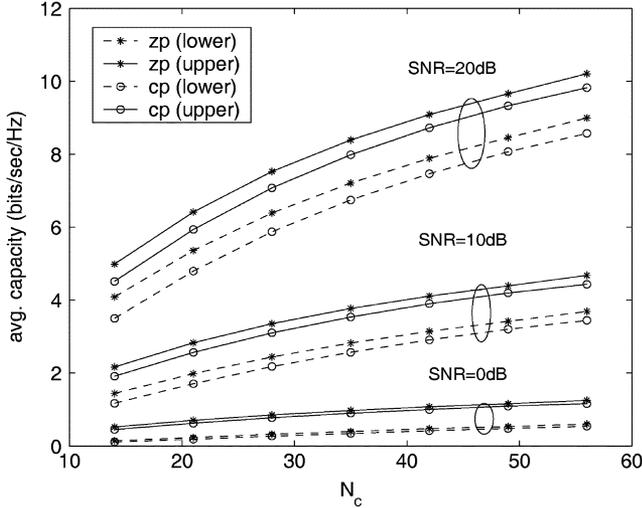


Fig. 7. Average capacity bounds comparison between ZP- and CP-based optimal training with increasing symbol block length ( $N_t = 2, N_r = 2, L = 6$ ).

*Test Case 2: (Comparisons With [13]):* To test our proposed channel estimation scheme, we use as a figure of merit the average channel MSE defined as  $E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2]$ . We compare our schemes with the preamble-based training scheme proposed in [13]. The training blocks for [13] are selected according to [13, eqn. (12)]. Fig. 8 shows that with the same power per training block, [13] achieves a similar channel MSE as our CP case, but lower than that offered by our ZP case. This difference in channel MSE is induced by the power loss of CP that is also used in [13].

We also plot the channel MSE for the three cases when the channel is slowly time varying. Each channel tap is generated by Jakes' model with a terminal speed of 3, 5, 10 m/s and a carrier frequency of 5.2 GHz. The variances of channel taps satisfy the exponential power profile. To maintain the same information rate, every training block is followed by four information blocks for the preamble scheme of [13]. Fig. 9 shows the channel MSE

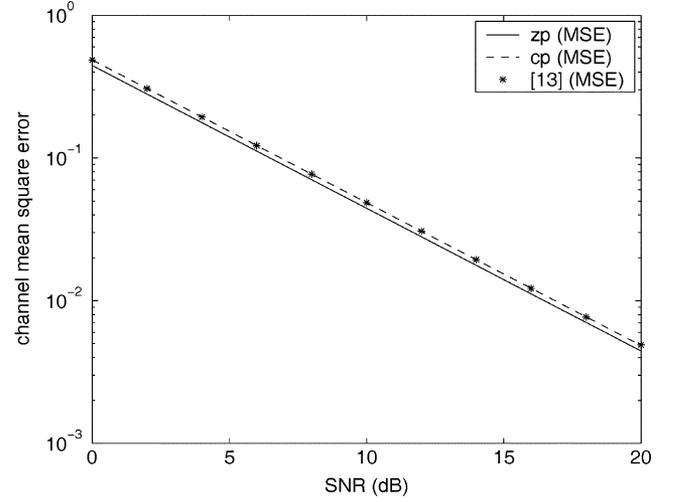


Fig. 8. Channel MSE comparison among ZP-, CP-based optimal training and training scheme in [13] ( $N_t = 2, N_r = 1$ , and  $L = 6$ ).

versus the number of transmitted blocks for the three schemes at SNR = 10 dB. Although the channel is slowly time varying, the MSE of our proposed schemes remains invariant from block to block, while that of [13] increases very fast. Note that [13] yields smaller MSEs at the beginning of each retraining burst, because the total transmitted power is used for training.

*Test Case 3: (BER Performance):* To illustrate the flexibility of our channel estimators with ST codes, we plot in Fig. 10 the BER performance for both ZP and CP cases. Two power allocation schemes are used: one is our optimal  $\alpha$  [cf., (35) and (54)], and the other is the so termed “equi-powered” scheme, where we choose  $\bar{P}_b = \bar{P}_c$ . In the same figure, the ideal cases corresponding to perfect channel estimates are also plotted as benchmarks. We select  $(N_t, N_r) = (2, 1)$ ,  $L = 2$ , and  $N = 128$ . The channel taps are independent complex Gaussian random variables with zero mean and variances that have an exponential power profile. For ZP, we use the ST code of [28]. For CP, we incorporate our training scheme into the ST code design of [16]. Zero-forcing estimators are used for both cases to estimate the information symbols. We observe from Fig. 10 that: 1) for both ZP and CP, our optimal schemes outperform the “equi-powered” schemes; 2) the penalty for inaccurate channel estimation is about 2 dB for both ZP and CP; 3) ZP outperforms CP with both perfect and imperfect LMMSE channel estimates; and 4) from the slopes of the curves, we notice that our training schemes do not alter the achievable diversity of the original ST code.

## VII. CONCLUSION

For LMMSE estimation of MIMO frequency-selective channels, optimal PSAM was designed to decouple the channel estimation from symbol detection and maximize a lower bound on the average capacity. The channel estimation MSE was minimized with no assumption on the channel besides the channel order  $L$ , for both ZP-based single-carrier and CP-based multicarrier block transmissions. Starting from a general superimposed model, the optimal training scheme for both transmission schemes requires  $N_t(L + 1)$  training symbols per transmit antenna. For ZP-based single-carrier systems, the pilot symbols are attached at either the head or tail of the information blocks

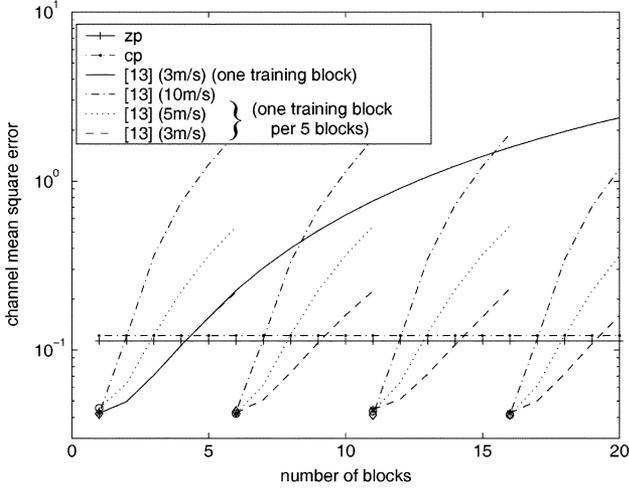


Fig. 9. Channel MSE comparison between ZP- and CP-based optimal training and the training scheme in [13] ( $N_t = 2$ ,  $N_r = 1$ , and  $L = 6$ ).

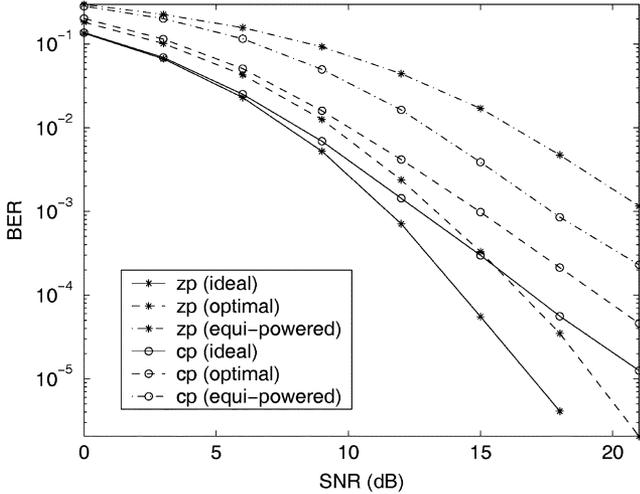


Fig. 10. BER comparison between ZP- and CP-based optimal training ( $N_t = 2$ ,  $N_r = 1$ ,  $L = 2$ ).

(TDM), while for CP-based multicarrier systems, the pilots are placed as equi-spaced pilot tones (FDM) with equal power. For a fixed transmit power per block, we also derived the optimal power allocation factor. It turns out that both training schemes can be combined nicely with ST mappers and coders, with ZP offering better BER performance than CP.

#### APPENDIX A PROOF OF LEMMA 1

Let  $\mathbf{B}$  and  $\mathbf{A}$  be two  $N \times (L+1)$  Toeplitz matrices generated by  $M \times 1$  vectors,  $\mathbf{b}$  and  $\mathbf{a}$ , respectively. The product  $\mathbf{C} := \mathbf{B}^H \mathbf{A}$  will be an  $(L+1) \times (L+1)$  Toeplitz matrix with first column  $[c(0), \dots, c(L)]^T$  and first row  $[c(0), \dots, c(-L)]$ , where

$$c(l) := \sum_{m=\max\{1, 1+l\}}^{\min\{M, M+l\}} a(m)b^*(m-1), \quad \forall l \in [-L, L]. \quad (56)$$

The condition  $\mathbf{C} \equiv \mathbf{0}$  implies  $c(l) \equiv 0, \forall M$ .

When  $M = L+1$ , we have [cf., (56)] a set of linear equations  $\underline{\mathbf{B}}\mathbf{a} = \mathbf{0}$ , where  $\underline{\mathbf{B}}$  is a  $(2L+1) \times (L+1)$  lower triangular Toeplitz matrix with first column  $[b^*(L+1), \dots, b^*(1)]^T$ . From the property of Toeplitz matrices,  $\underline{\mathbf{B}}$  will have full column rank if  $\mathbf{b}$  has at least one nonzero entry. Therefore, when  $M = L+1$ ,  $\mathbf{a}$  will be trivial if  $\mathbf{b}$  is nontrivial and vice versa. Define  $d_{ab}$  as the minimum distance (the position difference in the vectors) between the nonzero entries of  $\mathbf{b}$  and the nonzero entries of  $\mathbf{a}$ . We find that to ensure  $\mathbf{C} \equiv \mathbf{0}, \forall M$ , we need  $d_{ab}$  to be no less than  $L$ .

Notice that (56) only involves the multiplication of  $b^*(m)$  with  $a(m)$  within the distance  $L$ . Therefore,  $d_{ab} = L$  is sufficient and necessary to ensure  $\mathbf{C} \equiv \mathbf{0}, \forall M$ .

With this minimum  $d_{ab} = L$  guaranteed, there could be more than one nonzero cluster in  $\mathbf{b}$ . But, since the channel does not change over one block duration, shuffling the nonzero clusters of  $\mathbf{b}$  and  $\mathbf{a}$  together with their trailing zeros does not affect either the symbol decoding or the channel estimation performance. We can then gather all clusters that contain nonzero  $a(m)$ s at the beginning and move the rest that contain nonzero  $b(m)$ s to the end of the block. Going back to the condition of Lemma 1, we find that for fixed  $\mu_1$ , we need to find  $\bar{\mathbf{b}}_{\mu_1}$  which satisfies  $\bar{\mathbf{B}}_{\mu_1}^H \mathbf{A}_{\mu_2, n} = \mathbf{0}, \forall \mu_2$  and  $n$ . It thus follows readily that  $\bar{\mathbf{b}}_{\mu_1}$  needs to have at least  $L$  common zeros with all  $\mathbf{a}_{\mu_2, n}$ , which establishes (12).

#### APPENDIX B PROOF OF LEMMA 3

Thanks to its special structure, the circulant matrix  $\tilde{\mathbf{B}}_{\mu}$  can be expressed as  $\tilde{\mathbf{B}}_{\mu} = \mathbf{F}_M^H \mathbf{D}_{\tilde{\mathbf{B}}_{\mu}} \mathbf{F}_{0:L}$ , where  $\mathbf{F}_{0:L}$  is a submatrix that consists of the first  $L+1$  columns of  $\mathbf{F}_M$ , and  $\mathbf{D}_{\tilde{\mathbf{B}}_{\mu}}$  is an  $M \times M$  diagonal matrix defined as  $\mathbf{D}_{\tilde{\mathbf{B}}_{\mu}} := \sqrt{M} \text{diag}\{\mathbf{F}_M \bar{\mathbf{b}}_{\mu}\}$ . In the same manner, we have  $\tilde{\mathbf{A}}_{\mu, n} = \mathbf{F}_M^H \mathbf{D}_{\tilde{\mathbf{A}}_{\mu, n}} \mathbf{F}_{0:L}$  with  $\mathbf{D}_{\tilde{\mathbf{A}}_{\mu, n}} := \sqrt{M} \text{diag}\{\mathbf{F}_M \mathbf{a}_{\mu, n}\}$ . Using the preceding expressions, (42) becomes

$$\mathbf{F}_{0:L}^H \mathbf{D}_{\tilde{\mathbf{B}}_{\mu_1}}^H \mathbf{D}_{\tilde{\mathbf{A}}_{\mu_2, n}} \mathbf{F}_{0:L} = \mathbf{F}_{0:L}^H \mathbf{D} \mathbf{F}_{0:L} = \mathbf{0} \quad (57)$$

where the substitution  $\mathbf{D} := \mathbf{D}_{\tilde{\mathbf{B}}_{\mu_1}}^H \mathbf{D}_{\tilde{\mathbf{A}}_{\mu_2, n}}$  is used to simplify notations. It can be verified that any  $\mathbf{D}$  which shifts each column of  $\mathbf{F}_{0:L}$  to the null space of  $\mathbf{F}_{0:L}$  is an eligible solution of (57). Clearly,  $\mathbf{D} = \mathbf{0}_{M \times M}$  is an eligible one. This choice of  $\mathbf{D}$  necessitates

$$\begin{aligned} & \mathbf{D}_{\tilde{\mathbf{B}}_{\mu_1}}^H \mathbf{D}_{\tilde{\mathbf{A}}_{\mu_2, n}} \\ &= M \text{diag}\{(\mathbf{F}_M \bar{\mathbf{b}}_{\mu_1})^H\} \text{diag}\{\mathbf{F}_M \mathbf{a}_{\mu_2, n}\} = \mathbf{0}_{M \times M} \\ &\Leftrightarrow M(\mathbf{F}_M \bar{\mathbf{b}}_{\mu_1})^* \odot (\mathbf{F}_M \mathbf{a}_{\mu_2, n}) = \mathbf{0}_{M \times 1}. \end{aligned}$$

In other words, the  $(\mathbf{A}_{\mu}, \bar{\mathbf{b}}_{\mu})$  pairs must satisfy  $\forall m \in [1, M]$  and  $\forall \mu_1, \mu_2 \in [1, N_t]$

$$\mathbf{f}_m^T \bar{\mathbf{b}}_{\mu_1} = 0, \quad \text{or} \quad \mathbf{f}_m^T \mathbf{a}_{\mu_2} = \mathbf{0}_{1 \times M} \quad (58)$$

where  $\mathbf{f}_m$  denotes the  $m$ th column of  $\mathbf{F}_M$ . It is then natural to introduce the following expressions:

$$\mathbf{P}_A \Theta := \mathbf{F}_M \mathbf{A}_\mu \quad \mathbf{P}_b \mathbf{b}_\mu := \mathbf{F}_M \bar{\mathbf{b}}_\mu \quad (59)$$

where  $\Theta$  is an  $N'_c \times N_c$  matrix that optionally precodes (if  $\Theta \neq \mathbf{I}_{N'_c}$ ) the information-bearing block  $\mathbf{c}_\mu$  linearly,  $\mathbf{b}_\mu$  consists of the  $N_b = M - N'_c$  possibly nonzero training symbols from the  $\mu$ th transmit antenna, and  $\mathbf{P}_A$  and  $\mathbf{P}_b$  are permutation matrices given by

$$\begin{aligned} \mathbf{P}_A &:= [\mathbf{e}_{a_1}, \dots, \mathbf{e}_{a_{N'_c}}], \quad 1 \leq a_k < a_{k+1} \leq M, k \in [1, N'_c] \\ \mathbf{P}_b &:= [\mathbf{e}_{b_1}, \dots, \mathbf{e}_{b_{N_b}}], \quad 1 \leq b_k < b_{k+1} \leq M, k \in [1, N_b] \end{aligned}$$

where  $\mathbf{e}_m$  denotes the  $m$ th column of  $\mathbf{I}_M$ . Evidently, to satisfy (58), we need  $\mathbf{P}_A^T \mathbf{P}_b = \mathbf{0}_{N'_c \times N_b}$ .

Notice that with the choice  $\mathcal{D} = \mathbf{0}$ , the design of training symbols does not depend on the precoding matrix  $\Theta$ . By doing so, the symbol detection performance is decoupled from the channel estimation performance.

Nervertheless, for given  $M$  and  $L$ , any eligible nonzero  $\mathcal{D}$  will require the joint design of not only the placement, but also the contents of training sequence and the precoder. The resulting design will then ruin the independence between the two.

As a result, the desired  $(\mathbf{A}_\mu, \bar{\mathbf{b}}_\mu)$  pairs are given by

$$\mathbf{A}_\mu = \mathbf{F}_M^H \mathbf{P}_A \Theta, \quad \bar{\mathbf{b}}_\mu = \mathbf{F}_M^H \mathbf{P}_b \mathbf{b}_\mu. \quad (60)$$

#### ACKNOWLEDGMENT

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

#### REFERENCES

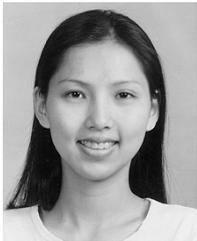
- [1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for unknown channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.
- [2] I. Barhumi, G. Leus, and M. Moonen, "Optimal training sequences for channel estimation in MIMO OFDM systems in mobile wireless channels," in *Proc. Int. Zurich Seminar on Access, Transmission, Networking of Broadband Communications*. Zurich, Switzerland, Feb. 19–21, 2002, pp. 44–1–44–6.
- [3] H. Bölcskei, R. W. Heath Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.
- [4] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.
- [5] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.
- [6] S. N. Diggavi, N. Al-Dahir, A. Stamoulis, and A. R. Calderbank, "Differential space-time transmission for frequency-selective channels," in *Proc. 36th Conf. Information Sciences Systems*. Princeton, NJ, Mar. 20–22, 2002, pp. 859–862.
- [7] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.
- [8] S. A. Fechtel and H. Meyr, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1639–1650, Feb./Mar/Apr. 1994.
- [9] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [10] C. Fragouli, N. Al-Dahir, and W. Turin, "Finite-alphabet constant-amplitude training sequence for multiple-antenna broadband transmissions," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, New York, NY, Apr. 28–May 1 2002, pp. 6–10.
- [11] —, "Reduced-complexity training schemes for multiple-antenna broadband transmissions," in *Proc. Wireless Communications Networking Conf.*, vol. 1, Mar. 17–21, 2002, pp. 78–83.
- [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] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," in *IEEE Trans. Wireless Commun.*, vol. 1, Jan. 2002, pp. 67–75.
- [14] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.
- [15] Z. Liu, G. B. Giannakis, S. Barbarossa, and A. Scaglione, "Transmit-antenna space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Select. Areas Commun.*, vol. 19, no. 7, pp. 1352–1364, Jul. 2001.
- [16] Z. Liu, Y. Xin, and G. B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2465–2476, Oct. 2002.
- [17] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
- [18] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [19] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.
- [20] —, "Capacity maximizing pilots for wireless OFDM over rapidly fading channels," in *Proc. Int. Symp. Signals, Systems Electronics*, Tokyo, Japan, Jul. 24–27, 2001, pp. 246–249.
- [21] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
- [22] T.-L. Tung, K. Yao, and R. E. Hudson, "Channel estimation and adaptive power allocation for performance and capacity improvement of multiple-antenna OFDM systems," in *Proc. 3rd IEEE Workshop Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, R.O.C., Mar. 20–23, 2001, pp. 82–85.
- [23] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. Int. Conf. ASSP*, vol. 4, Salt Lake City, UT, May 7–11, 2001, pp. 2105–2108.
- [24] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where fourier meets Shannon," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
- [25] Z. Wang, X. Ma, and G. B. Giannakis, "Optimality of single-carrier zero-padded block transmissions," in *Proc. Wireless Communications Networking Conf.*, vol. 2, Orlando, FL, Mar. 17–21, 2002, pp. 660–664.
- [26] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.
- [27] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
- [28] S. Zhou and G. B. Giannakis, "Single-carrier space-time block coded transmissions over frequency-selective fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 164–179, Jan. 2003.
- [29] 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, pp. 1215–1228, May 2002.



**Xiaoli Ma** (M'03) received the B.S. degree in automatic control from Tsinghua University, Beijing, China, the M.Sc. degree in electrical engineering from the University of Virginia, Charlottesville, and the Ph.D. degree from the University of Minnesota, Minneapolis, in 1998, 1999, and 2003, respectively.

Since 2003, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL. Her research interests include transmitter and receiver diversity techniques for wireless fading channels, com-

munications over time- and frequency-selective channels, complex-field and space-time coding, channel estimation and equalization algorithms, carrier frequency synchronization for OFDM systems, and wireless sensor networks.



**Liuqing Yang** (S'02–M'04) received the B.S. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1994, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, in 2002 and 2004, respectively.

Since August 2004, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville. Her research interests include communications, signal processing, and networking.

Currently, she has a particular interest in ultrawideband (UWB) communications. Her research encompasses synchronization, channel estimation, multiple access, space-time coding, and multicarrier systems.



**Georgios B. Giannakis** (S'84–M'86–SM'91–F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1981, and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), Los Angeles, in 1983 and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia in 1987, where he became a Professor of electrical engineering in 1997. Since

1999, he has been a Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 200 journal papers, 350 conference papers, and two edited books. His current research interests include transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space-time coding, multicarrier, ultrawide-band wireless communication systems, cross-layer designs, and distributed sensor networks.

Dr. Giannakis is the (co-)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, 2004). He also received the SP Society's Technical Achievement Award in 2000. He served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as member of the SP Publications Board, as member and Vice Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a member of the IEEE Fellows Election Committee. He has also served as a member of the the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.