Optimal Training and Redundant Precoding for Block Transmissions With Application to Wireless OFDM

Shuichi Ohno, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—The adoption of orthogonal frequency-division multiplexing by wireless local area networks and audio/video broadcasting standards testifies to the importance of recovering block precoded transmissions propagating through frequency-selective finite-impulse response (FIR) channels. Existing block transmission standards invoke bandwidth-consuming error control codes to mitigate channel fades, and training sequences to identify the FIR channels. To enable block-by-block receiver processing, we design redundant precoders with cyclic prefix and superimposed training sequences for optimal channel estimation and guaranteed symbol detectability, regardless of the underlying frequency-selective FIR channels. Numerical results are presented to access the performance of the designed training and precoding schemes.**

***Index Terms*—Block transmissions, channel estimation, multipath, orthogonal frequency-division multiplexing (OFDM), pilot tones.**

I. Introduction

BLOCK transmissions relying on linear redundant filterbank precoding with cyclic prefixed (CP) or zero-padded (ZP) blocks have gained increasing interest recently for mitigating frequency-selective multipath effects (see, e.g., [2], [11], [17], [19] and references therein). Redundancy removes interblock interference (IBI), and facilitates (even blind) acquisition of channel state information (CSI) at the receiver. It also leads to data efficient, low-complexity linear equalizers [zero-forcing (ZF) or minimum mean-squared error (MMSE)] with guaranteed symbol detectability regardless of the zero locations of the underlying finite-impulse response (FIR) channel [17].

When CSI is available at the transmitter (e.g., through a feedback channel), optimal precoders and decoders become available under various criteria [16], [17]. However, rapid variations of the wireless channel render CSI feedback to the transmitter outdated, and motivate *channel-independent* precoders. On the other hand, because CSI is indispensable at the receiver, training sequences are needed to acquire it. Blind schemes offer bandwidth-efficient alternatives, when frequent retraining is required, but they are often more complex and require longer data records than training-based approaches.

Instead of long training sequences at the beginning of the transmitted record, inserting training symbols during the transmission is known as pilot symbol-aided modulation (PSAM), and was originally developed for time-selective channel estimation [3], [7] and synchronization [11], [15]. The inserted pilot symbols in PSAM are separated from the information symbols in the time domain [3], while the so-termed pilot tones (complex exponentials in time) are separated from the information symbols in the frequency domain [13], [14]. In the superimposed (or spread-spectrum) pilot schemes of [7], [8], and [18], a pseudonoise sequence is added to the information sequence. PSAM is also useful for decision-feedback (DF) equalization of block transmissions [10] (see also [1], where optimum allocation of pilot symbols is pursued for DF equalization). However, jointly optimal design of pilot tones and precoders for mitigation of frequency-selective channels with block-by-block processing has not been addressed.

This paper deals with linearly precoded symbol blocks with superimposed training blocks that can be jointly modeled as an *affine* precoder. Affine precoding was also discussed in [12], but the type of affine precoders suitable for optimal channel estimation and guaranteed symbol detection was not specified. After developing our unifying block-modeling framework (Section II), we specify the design constraints for IBI cancellation, and block-by-block reception which enables linear channel estimation that is decoupled from symbol detection (Section III). Affine precoders are then designed to decouple channel estimation from symbol detection and optimize the least-squares (LS) channel estimator. Decoupling channel estimation from symbol detection naturally leads to linearly precoded (LP) orthogonal frequency-division multiplexing (OFDM) systems with pilot tones. Subsequently, we design pilot tones for optimal LS channel estimation in the presence of white or colored noise (Section IV). We also investigate optimal power loading on information bearing and pilot symbols, and design LP-OFDM systems that ensure symbol detectability regardless of channel nulls (Section V). LP-OFDM is found to exhibit enhanced frequency diversity in mitigating channel and noise color effects *deterministically*. Simulations are presented to corroborate LP-OFDM's improved performance over conventional OFDM, and compare LP-OFDM against coded OFDM.

II. Block Modeling and Preliminaries

We consider the block transmission model depicted in Fig. 1. The information-bearing sequence $s(n)$ is parsed into blocks

Paper approved by P. Y. Kam, the Editor for Modulation and Detection for Wireless Systems of the IEEE Communications Society. Manuscript received November 22, 2000; revised October 28, 2001 and June 28, 2002. The work of G. B. Giannakis was supported by the National Science Foundation Wireless Initiative under Grant 9979442. This paper was presented in part at the International Conference on Communications, Helsinki, Finland, June 11–15, 2001.

S. Ohno is with the Department of Artificial Complex Systems Engineering, Hiroshima University, Higashi-Hiroshima 739-8527 Japan (e-mail: o.shuichi@ieee.org).

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/TCOMM.2002.806547

0090-6778/02$17.00 © 2002 IEEE

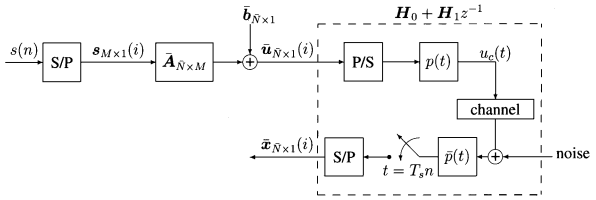


Fig. 1. Schematic diagram for block transmissions.

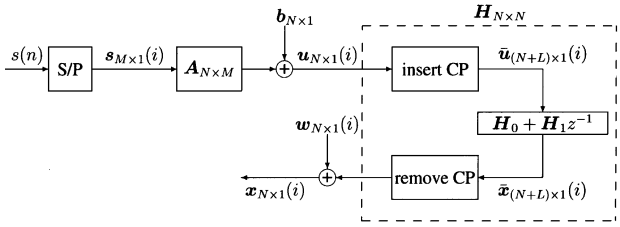


Fig. 2. Discrete-time baseband equivalent block model.

of size M. Each $M\times 1$ block $\boldsymbol{s}(i)$ is precoded by a tall $\bar{N}\times M$ precoding matrix $\bar{\boldsymbol{A}}$ with generally complex-valued entries. Selecting $\bar{N}>M$ introduces redundancy, which will turn out to be beneficial in mitigating the effects of frequency-selective propagation. An $\bar{N}\times 1$ block of training symbols $\bar{\boldsymbol{b}}$, which are also known to the receiver, is added to the precoded information block to obtain

$$\bar{\boldsymbol{u}}(i)=\bar{\boldsymbol{A}}\boldsymbol{s}(i)+\bar{\boldsymbol{b}} \tag{1}$$

which is parallel-to-serial (P/S) to $\bar{u}(k)$, digital-to-analog (D/A) converted, and pulse shaped to yield the continuous-time signal $u_c(t)=\sum_{k=-\infty}^{\infty}\bar{u}(k)p(t-kT_s)$, where $p(t)$ denotes the transmit filter, and T_s stands for the symbol duration.

The transformation in (1) with $\bar{\boldsymbol{b}}=\boldsymbol{0}$ models linearly precoded block transmissions such as those considered in [2], [10], [16], and [17]. On the other hand, with precoder $\bar{\boldsymbol{A}}=\boldsymbol{I}$, (1) describes a block transmission with superimposed training (or pilot) symbols that in serial (nonredundant) form has been used by [3], [7]. Note that $\bar{\boldsymbol{b}}$ can describe both superimposed as well as inserted training symbols (the latter is implemented when $\bar{\boldsymbol{b}}$ has nonzero entries, where $\bar{\boldsymbol{A}}\boldsymbol{s}(i)$ has zero entries). To capture the generality of the redundant precoding in (1), we will, henceforth, borrow the terminology of [12] and call it *affine* precoding. For the mapping in (1) to be invertible, we will choose the redundant precoder so that:

C1. *The $\bar{N}\times M$ precoding matrix $\bar{\boldsymbol{A}}$ is tall, and has full column rank M.*

At the receiver, we assume perfect timing and carrier synchronization, and sample the output of the front-end filter $\bar{p}(t)$ (that is matched to the transmit pulse) at the symbol rate $1/T_s$. As detailed in [11] and [15], preamble-based training can be used for acquisition of timing and carrier frequency offsets at the beginning of each transmission burst, e.g., by setting the information-bearing symbols to zero in (1). During the transmission, these offsets can be tracked using existing synchronization schemes (see [11], [15], and references therein).

Let $h_c(t)$ denote the overall impulse response of the transmit filter, the continuous-time channel, and the receive filter. With $\tau_{\max}$ denoting the maximum delay spread of $h_c(t)$, our discrete-time baseband equivalent FIR channel, $h(n):=h_c(nT_s)$, has order $L=\lceil\tau_{\max}/T_s\rceil$, where $\lceil\ \rceil$ stands for integer ceiling. The channel $h(n)$ is considered linear time invariant over one received block, but is allowed to vary from block to block. Because we will only consider block-by-block receiver processing, we will omit time dependence and express the impulse response of the discrete-time baseband equivalent channel as $\{h(n)\}$.

The information block size M is chosen to satisfy $M\gg L$. We collect $\bar{N}$ noisy samples in an $\bar{N}\times 1$ received vector $\bar{\boldsymbol{x}}(i)$ that can be expressed as (see, e.g., [19])

$$\bar{\boldsymbol{x}}(i)=\boldsymbol{H}_0\bar{\boldsymbol{u}}(i)+\boldsymbol{H}_1\bar{\boldsymbol{u}}(i-1)+\boldsymbol{\eta}(i) \tag{2}$$

where $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$ are square Toeplitz channel convolution matrices with first column $[h(0),h(1),\ldots,h(L),0,\ldots,0]^T$, and first row $[0,\ldots,0,h(L),h(L-1),\ldots,h(1)]$, respectively; and $\boldsymbol{\eta}(i)$ is a zero-mean additive noise. Presence of two successive transmitted blocks in each received block arises due to the channel of length $L+1$ ($\leq\bar{N}$); the second term in (2) captures the IBI.

To enable low-complexity block-by-block processing at the receiver, we first eliminate IBI by utilizing the so-called CP, which is also employed by OFDM, the basic multicarrier modulation that has been adopted by many standards [4], [9].

Fig. 2 shows our discrete-time baseband equivalent model. Discarding the CP at the receiver removes the IBI, provided that the following design condition holds at the transmitter.

A1. *The length $\bar{N}$ of the redundant transmitted block is chosen to satisfy $\bar{N}\geq L+M$.*

Let N be defined as $N:=\bar{N}-L$. To describe the CP insertion in matrix form, consider placing a cyclic replica of the last L entries of an $N\times 1$ vector $\boldsymbol{v}$ at its top to create an $\bar{N}\times 1$ vector $\bar{\boldsymbol{v}}$. This augmentation can be represented by an $\bar{N}\times N$ CP-inducing matrix $\boldsymbol{T}_{\rm cp}$ as $\bar{\boldsymbol{v}}=\boldsymbol{T}_{\rm cp}\boldsymbol{v}$ with $\boldsymbol{T}_{\rm cp}:=[\boldsymbol{I}_{\rm cp}^T,\boldsymbol{I}_N^T]^T$, where $\boldsymbol{I}_{\rm cp}:=[\boldsymbol{0}_{L\times(N-L)},\boldsymbol{I}_L]^T$; $\boldsymbol{I}_L$ denotes the identity matrix of size L; and $\boldsymbol{0}_{L\times(N-L)}$ stands for the $L\times(N-L)$ zero matrix. To enable insertion and removal of the CP, and thus, cancellation of IBI, we select our general design in (1) such that:

A2. *Matrix $\bar{\boldsymbol{A}}$ and vector $\bar{\boldsymbol{b}}$ incorporate the CP; i.e., $\bar{\boldsymbol{A}}=\boldsymbol{T}_{\rm cp}\boldsymbol{A}$, $\bar{\boldsymbol{b}}=\boldsymbol{T}_{\rm cp}\boldsymbol{b}$, where $\boldsymbol{A}$ is an $N\times M$ matrix and $\boldsymbol{b}:=[b(0),\ldots,b(N-1)]^T$ is an $N\times 1$ vector.*

IBI can be removed also by ZP [19], but in order to adhere to existing OFDM standards, this paper focuses on block transmissions with CP.

Taking **A2** into account, we can rewrite the transmitted block in (1) as $\bar{\boldsymbol{u}}(i)=\boldsymbol{T}_{\rm cp}\boldsymbol{u}(i)$, where

$$\boldsymbol{u}(i)=\boldsymbol{A}\boldsymbol{s}(i)+\boldsymbol{b}. \tag{3}$$

Discarding the CP from $\bar{\boldsymbol{x}}(i)$ can be described by the matrix $\boldsymbol{R}_{\rm cp}:=[\boldsymbol{0}_{N\times L},\boldsymbol{I}_N]_{N\times\bar{N}}$, whose leading zeros cancel the nonzero entries of $\boldsymbol{H}_1$, and remove the IBI since $\boldsymbol{R}_{\rm cp}\boldsymbol{H}_1=\boldsymbol{0}$ [cf. (2)]. Specifically, we form $\boldsymbol{x}(i):=\boldsymbol{R}_{\rm cp}\bar{\boldsymbol{x}}(i)$, and use (2) to arrive at

$$\boldsymbol{x}(i)=\boldsymbol{H}\boldsymbol{u}(i)+\boldsymbol{w}(i)=\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}(i)+\boldsymbol{B}\boldsymbol{h}+\boldsymbol{w}(i) \tag{4}$$

where $\boldsymbol{w}(i) := \boldsymbol{R}_{\rm cp}\boldsymbol{\eta}(i)$; $\boldsymbol{h} := [h(0), h(1), \ldots, h(L)]^T$; $\boldsymbol{H}$ is an $N \times N$ circulant matrix with first column $[\boldsymbol{h}^T, 0, \ldots, 0]^T$; $\boldsymbol{B}$ is an $N \times (L+1)$ column-wise circulant matrix with first column $\boldsymbol{b}$; and in deriving (4), we used the commutativity of circular convolution to obtain $\boldsymbol{Hb} = \boldsymbol{Bh}$.

The model (4) can be viewed as a *virtual* two-user model, where one user transmits $\boldsymbol{s}(i)$, and the other one $\boldsymbol{h}$ through equivalent channels $\boldsymbol{HA}$ and $\boldsymbol{B}$, respectively. They interfere with each other and, thus, it is desirable to decouple them, which in our single-user model amounts to separating channel estimation from symbol detection. Two questions arise at this point. What are the classes of precoders that decouple channel from symbol estimation? And what are their degrees of freedom and bandwidth characteristics? We address these questions in the next section, where we design affine precoders $(\boldsymbol{A}, \boldsymbol{b})$ that enable such a separation using only linear operations *regardless* of the underlying FIR channel $\boldsymbol{H}$.

III. Decoupling Symbol From Channel Estimation

Estimating $\boldsymbol{H}$ and recovering $\boldsymbol{s}(i)$ from $\boldsymbol{x}(i)$ in (4) is a nonlinear detection-estimation problem, and its optimal (e.g., in the maximum-likelihood (ML) sense) solution is often computationally prohibitive, and sometimes even impossible. As a counterexample, consider $\boldsymbol{A} = \boldsymbol{I}$, a channel with impulse response having sample mean equal to zero ($[1, \ldots, 1]\boldsymbol{h} = 0$), and a block $\boldsymbol{s}(i)$ with all entries equal to one. Since $\boldsymbol{HAs}(i) = \boldsymbol{0}$, symbol detection is impossible. We will see that judicious design of $\boldsymbol{A}$ can not only prevent this case, but also convert the nonlinear problem to two low-complexity (albeit suboptimum) linear estimation problems.

With reference to (4), let us suppose temporarily that $\boldsymbol{\nu}(i) := \boldsymbol{HAs}(i) + \boldsymbol{w}(i)$ acts as noise in the linear model $\boldsymbol{x}(i) = \boldsymbol{Bh} + \boldsymbol{\nu}(i)$. To be able to estimate $\boldsymbol{h}$ from $\boldsymbol{x}(i)$ using linear LS, we should select our training vector $\boldsymbol{b}$ such that:

C2. *The $N \times (L+1)$ tall training circulant matrix $\boldsymbol{B}$ has full column rank $L+1$.*

We wish to characterize the class of pilots satisfying **C2**. Let us first recall that circulant matrices (like $\boldsymbol{B}$) can be diagonalized by the fast Fourier transform (FFT) matrix and its inverse (IFFT), with the diagonal entries being the frequency-response values of $\boldsymbol{b}$ evaluated at the FFT grid [6, p. 202]. Specifically, let $\boldsymbol{F}$ be the $N \times N$ FFT matrix with (m,n)th entry $[\boldsymbol{F}]_{m,n} = N^{-1/2}W^{-mn}$, where $W := \exp(j2\pi/N)$. Let $B(z)$ be the $\mathcal{Z}$ transform of $\boldsymbol{b}$ defined as $B(z) := \sum_{n=0}^{N-1} b(n)z^{-n}$, and $\boldsymbol{D}_B$ a diagonal matrix defined as $\boldsymbol{D}_B := \mathrm{diag}[B(1), B(W), \ldots, B(W^{N-1})]$. Since $\boldsymbol{B}$ is a tall (column-wise) circulant matrix, its FFT-based diagonalization yields

$$\boldsymbol{B} = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{D}_B\boldsymbol{F}_{0:L} \tag{5}$$

where $\boldsymbol{F}_{0:L}$ is a submatrix of $\boldsymbol{F}$, corresponding to the first $L+1$ columns of $\boldsymbol{F}$, and $^{\mathcal{H}}$ denotes conjugated transposition. Equation (5) links the time-domain pilot matrix $\boldsymbol{B}$ with its frequency-domain counterpart $\boldsymbol{D}_B$.

Suppose $\boldsymbol{D}_B$ has $K \leq N$ nonzero entries, and $\check{\boldsymbol{D}}_B$ denotes the corresponding $K \times K$ submatrix of $\boldsymbol{D}_B$ with all nonzero diagonal entries. Since $\boldsymbol{F}^{\mathcal{H}}$ is full rank, we have from (5) that $\mathrm{rank}(\boldsymbol{B}) = \mathrm{rank}(\boldsymbol{D}_B\boldsymbol{F}_{0:L}) = \mathrm{rank}(\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L})$, where $\check{\boldsymbol{F}}_{0:L}$ is the $K \times (L+1)$ matrix formed by the K rows of $\boldsymbol{F}_{0:L}$ corresponding to the nonzero entries of $\boldsymbol{D}_B$. Sylvester's inequality now implies that $\mathrm{rank}(\check{\boldsymbol{D}}_B) + \mathrm{rank}(\check{\boldsymbol{F}}_{0:L}) - K \leq \mathrm{rank}(\boldsymbol{B}) \leq \min(K, L+1)$, or equivalently, that $L+1 \leq \mathrm{rank}(\boldsymbol{B}) \leq \min(K, L+1)$. To specify and interpret the $\mathrm{rank}(\boldsymbol{B})$, consider $K \geq L+1$, and define the frequency-domain training block as $\tilde{\boldsymbol{b}} := \boldsymbol{Fb} = (1/N^{1/2})[B(1), B(W), \ldots, B(W^{N-1})]^T$. The following can be verified readily.

*Lemma 1: The N-point FFT vector $\tilde{\boldsymbol{b}}$ of the training block $\boldsymbol{b}$ has $K \geq L+1$ nonzero values if and only if the training-based matrix $\boldsymbol{B}$ satisfies the design condition **C2**.*

When $\boldsymbol{b}$ is selected so that $\boldsymbol{B}$ satisfies **C2**, the LS channel estimator is given by [cf. (4)]

$$\hat{\boldsymbol{h}} = \boldsymbol{B}^{\dagger}\boldsymbol{x}(i) = \boldsymbol{h} + \boldsymbol{B}^{\dagger}\left[\boldsymbol{HAs}(i) + \boldsymbol{w}(i)\right] \tag{6}$$

where $\boldsymbol{B}^{\dagger} := (\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathcal{H}}$ is the (minimum norm) pseudo-inverse of $\boldsymbol{B}$. Equation (6) contains a symbol-dependent noise term which must be eliminated for the LS channel estimation error to be minimized. Eliminating this term amounts to designing the affine precoder $(\boldsymbol{A}, \boldsymbol{b})$, or equivalently, $(\boldsymbol{A}, \boldsymbol{B})$, such that:

C3. *Matrix $\boldsymbol{B}^{\dagger}\boldsymbol{HA} = \boldsymbol{0}$ for any FIR channel of order L.*

Interestingly, if one is able to design a class of affine precoders satisfying **C3** for *all* FIR channels up to a given order, then LS-optimal channel estimation becomes possible; and more important, channel estimation is no longer coupled to symbol detection.

To design affine precoders possessing the decoupling property **C3** while also satisfying **C1** and **C2**, we will start with a characterization of the class using properties of circulant matrices. Let $\boldsymbol{A}_i$ be a column-wise circulant matrix with first column the ith column of $\boldsymbol{A}$, which we denote as $\boldsymbol{a}_i$. Then, we can write $\boldsymbol{HA} = [\boldsymbol{Ha}_0, \boldsymbol{Ha}_1, \ldots, \boldsymbol{Ha}_{M-1}]$; and since $\boldsymbol{H}$ is circulant, it follows from the commutativity of circular convolution that $\boldsymbol{Ha}_i = \boldsymbol{A}_i\boldsymbol{h}$; hence, $\boldsymbol{HA} = [\boldsymbol{A}_0\boldsymbol{h}, \boldsymbol{A}_1\boldsymbol{h}, \ldots, \boldsymbol{A}_{M-1}\boldsymbol{h}]$. Based on this, we have that $\boldsymbol{B}^{\dagger}\boldsymbol{HA} = \boldsymbol{0}$ if and only if $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{A}_i\boldsymbol{h} = \boldsymbol{0}, \forall i \in [0, M-1]$. Because the latter is to hold true for any $\boldsymbol{h}$, we arrive at the following lemma.

*Lemma 2: If matrix $\boldsymbol{B}$ satisfies **C2**, then it also satisfies **C3** if and only if*

$$\boldsymbol{B}^{\mathcal{H}}\boldsymbol{A}_i = \boldsymbol{0}, \qquad \forall i \in [0, M-1]. \tag{7}$$

Lemma 2 shows that for decoupling channel from symbol estimation, not only $\boldsymbol{b}$ but also its circular shifts should be orthogonal to $\boldsymbol{A}$.

As per *Lemma 2*, whether **C3** is satisfied or not depends on the number of the nonzero diagonal entries of $\boldsymbol{D}_B$. Letting $N-K$ be the number of zero $B(W^i)$'s, we define the set of ordered integer indexes

$$\mathcal{I}_0 := \{i_k | B(W^i) = 0, i_k < i_{k+1}, k \in [0, N-K-1]\} \tag{8}$$

and its complement $\mathcal{I}_0^{\perp}$ containing these K indexes i, for which $B(W^i) \neq 0$. If we define the $K \times 1$ vector $\check{\boldsymbol{b}}$ to contain the K

nonzero entries of the FFT pilot vector $\tilde{\boldsymbol{b}}$, then the time-domain training vector can be written as

$$\boldsymbol{b} = \boldsymbol{F}^{\mathcal{H}}\tilde{\boldsymbol{b}} = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0}\begin{bmatrix}\check{\boldsymbol{b}}\\ \boldsymbol{0}\end{bmatrix} = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0^{\perp}}\begin{bmatrix}\boldsymbol{0}\\ \check{\boldsymbol{b}}\end{bmatrix} \tag{9}$$

where $\boldsymbol{P}_{\mathcal{I}_0}(\boldsymbol{P}_{\mathcal{I}_0^{\perp}})$ is a permutation matrix collecting the K possibly dispersed nonzero entries $\check{\boldsymbol{b}}$ of $\tilde{\boldsymbol{b}}$ at the top (bottom). Based on these notational conventions, we can state our first result as follows.

*Theorem 1: Consider transmissions of information blocks of length M through an FIR channel of order L using CP to avoid ISI as per **A1** and **A2**. Let K be the number of nonzero entries of the FFT pilot vector $\tilde{\boldsymbol{b}}$ (note that K corresponds to the redundancy added for training). For $K \in [L+1, 2L+1]$, symbol detection from precoded symbols can be decoupled from linear channel estimation based on training symbols (**C2**), regardless of the FIR channel (**C3**), if and only if the precoded blocks have length $N \geq M+K$, and the affine precoders $(\boldsymbol{A}, \boldsymbol{b})$ are selected from the class*

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0^{\perp}}\begin{bmatrix}\boldsymbol{\Theta}_{(N-K)\times M}\\ \boldsymbol{0}_{K\times M}\end{bmatrix}\\ \boldsymbol{b} &= \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0}\begin{bmatrix}\check{\boldsymbol{b}}\\ \boldsymbol{0}\end{bmatrix}\end{aligned} \tag{10}$$

where $\boldsymbol{\Theta}$ is any full column rank matrix, $\boldsymbol{P}_{\mathcal{I}_0}$ and $\boldsymbol{P}_{\mathcal{I}_0^{\perp}}$ are the permutation matrices, and $\check{\boldsymbol{b}}$ is the nonzero vector defined in (9). Clearly, the minimum redundancy choice corresponds to $N = M+L+1$, and after accounting for the CP, the transmitted block must have length $\bar{N} = M + 2L + 1$.

Proof: See Appendix A. ■

With $N - K = M$ and $\boldsymbol{\Theta} = \boldsymbol{I}$ in (10), the fact that $\mathcal{I}_0$ and $\mathcal{I}_0^{\perp}$ contain nonoverlapping indexes implies that the precoded symbols $\boldsymbol{A}\boldsymbol{s}(i)$ are loaded on M subcarriers (the columns of $\boldsymbol{F}^{\mathcal{H}}$) that are distinct from the K subcarriers which correspond to the pilots used for channel estimation; i.e., (10) includes, as a special case, OFDM with inserted pilot tones. Notwithstanding, OFDM was not assumed *a priori*; we reached it starting from block transmissions with CP. Interestingly, separation of information from pilot subcarriers in OFDM is necessary and sufficient for channel-irrespective decoupling of symbols from linear channel estimation with minimum redundancy. Although training is inserted in the frequency domain (via pilot tones), the training vector $\boldsymbol{b}$ has generally nonzero entries, and it is, thus, superimposed on $\boldsymbol{A}\boldsymbol{s}(i)$ in the time domain. This can be thought of as the dual of PSAM [3], where pilot symbols are inserted periodically in the time domain to cope with time-selective channels.

But *Theorem 1* goes beyond OFDM ($\boldsymbol{\Theta} = \boldsymbol{I}$). It introduces the class of LP-OFDM, which generalizes OFDM in three directions of practical importance.

1) It allows $\mathcal{I}_0$ (and accordingly, $\mathcal{I}_0^{\perp}$) to vary from block to block, which corresponds to shifting (or hopping) information and pilot subcarriers that will turn out to improve performance without CSI knowledge at the transmitter.
2) It provides degrees of freedom to optimize the power allocation (via $\check{\boldsymbol{b}}$), and the pilots' location (via $\boldsymbol{P}_{\mathcal{I}_0}$) for reliable channel estimation decoupled from symbol detection.
3) It also offers degrees of freedom in choosing $\boldsymbol{\Theta}$ to optimize performance and guarantee symbol detection without CSI knowledge.

The bandwidth efficiency (that depends on the relative redundancy K/M) is given by

$$\mathcal{E}(M,K) := \frac{M}{\bar{N}} = \frac{M}{N+L} = \frac{M}{M+K+L}. \tag{11}$$

By selecting $M' > M$ and $K' > K$ so that $M'/(M'+K'+L) \equiv M/(M+K+L)$, we can have $\mathcal{E}(M,K) = \mathcal{E}(M',K')$. However, we will henceforth focus on the minimum possible block lengths, because they minimize complexity and decoding delay.[1] It follows readily from *Theorem 1* and (11) that:

Corollary 1: The precoders $(\boldsymbol{A}, \boldsymbol{b})$ of Theorem 1 achieve maximum bandwidth efficiency and minimum decoding delay when $K = L+1$. In this case, $\mathcal{E}_{\max} = M/(M+2L+1)$.

With assured decoupling, we next address optimum placement and power loading of pilots for channel estimation.

IV. Optimal Placement and Power Loading of Pilots

We want to optimize the location (i.e., select the set $\mathcal{I}_0^{\perp}$) as well as the power (i.e., choose $\check{\boldsymbol{b}}$) of pilots in (10). Our criterion will be to minimize the channel mean-square error (MSE) for a given transmit power budget $\|\boldsymbol{b}\|^2 = \mathcal{P}_b = (1/N)\sum_{l_i \in \mathcal{I}_0^{\perp}} |B(W^{l_i})|^2$, where $\|\cdot\|$ denotes the Euclidean norm, and the second equality follows from Parseval's Theorem. Considering (6) with $\boldsymbol{B}^{\dagger}\boldsymbol{H}\boldsymbol{A} = \boldsymbol{0}$, the channel MSE is given by

$$\sigma^2_{\hat{h}} := E\{\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2\} = \mathrm{tr}(\boldsymbol{B}^{\dagger}\boldsymbol{R}_w\boldsymbol{B}^{\dagger\mathcal{H}}) \tag{12}$$

where $E\{\cdot\}$ and $\mathrm{tr}(\cdot)$ denote the expectation and the trace operator, respectively, and $\boldsymbol{R}_w$ is the correlation matrix of $\boldsymbol{w}(i)$ in (4). In addition to additive white noise (AWN) with $\boldsymbol{R}_w = \sigma_w^2\boldsymbol{I}_N$, we will allow $\boldsymbol{w}(i)$ to be correlated in order to account for structured (e.g., adjacent channel) interference. But let us first study the white noise case.

Selecting $\boldsymbol{R}_w = \sigma_w^2\boldsymbol{I}_N$ in (12), we obtain $\sigma^2_{\hat{h}} = \sigma_w^2\mathrm{tr}(\boldsymbol{B}^{\dagger}\boldsymbol{B}^{\dagger\mathcal{H}}) = \sigma_w^2\mathrm{tr}[(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-1}]$. It is easy to show that for an $(L+1)\times(L+1)$ positive definite matrix $\boldsymbol{C}$, we have $\mathrm{tr}(\boldsymbol{C})\mathrm{tr}(\boldsymbol{C}^{-1}) \geq (L+1)^2$, where the equality holds if and only if $\boldsymbol{C} = c\boldsymbol{I}_{L+1}$ for some nonzero constant c. Using this and the fact that $(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-1}$ is positive definite, we find that $\sigma^2_{\hat{h}} \geq \sigma_w^2(L+1)^2/\mathrm{tr}(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})$, with equality if and only if $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = c\boldsymbol{I}_{L+1}$ for some nonzero constant c. It follows from the constraint $\|\boldsymbol{b}\|^2 = \mathcal{P}_b$ that $\mathrm{tr}(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B}) = (L+1)\mathcal{P}_b$, and the equality in $\sigma^2_{\hat{h}} \geq \sigma_w^2(L+1)/\mathcal{P}_b$ holds if and only if $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \mathcal{P}_b\boldsymbol{I}_{L+1}$. But, from $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \check{\boldsymbol{F}}^{\mathcal{H}}_{0:L}\check{\boldsymbol{D}}^{*}_B\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L} = \mathcal{P}_b\boldsymbol{I}_{L+1}$ with * denoting complex conjugation, we obtain $\check{\boldsymbol{D}}^{*}_B\check{\boldsymbol{D}}_B = \mathcal{P}_b(\check{\boldsymbol{F}}_{0:L}\check{\boldsymbol{F}}^{\mathcal{H}}_{0:L})^{-1}$.

[1] Although not dealt with in this paper, we have shown that for $K' > K_{\max} = 2L+1$, the class of necessary and sufficient precoders for **C1**-**C3** is different from that of *Theorem 1*. Different precoders result also when the CP is replaced by ZP.

Since $\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B$ is diagonal, $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \mathcal{P}_b\boldsymbol{I}_{L+1}$ if and only if $\check{\boldsymbol{F}}_{0:L}\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}$ is diagonal. Recalling that $\check{\boldsymbol{F}}_{0:L}$ is an $(L+1)\times(L+1)$ submatrix of the $N \times N$ FFT matrix, we deduce that $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \mathcal{P}_b\boldsymbol{I}_{L+1}$ if and only if $\check{\boldsymbol{F}}_{0:L}\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}} = [(L+1)/N]\boldsymbol{I}_{L+1}$, and hence, $\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B = [N\mathcal{P}_b/(L+1)]\boldsymbol{I}_{L+1}$. Thus, the minimum channel MSE is attained if and only if we select:

1) $N = (L+1)J$ for some nonzero integer J;
2) the spacing of pilot tones to satisfy $l_i = l_0 + Jl (l_i \in \mathcal{I}_0^{\perp})$ for some $l_0 \in [0, J-1]$, and $l \in [1, L]$;
3) the same power is loaded on each pilot tone.

Design rule 2) suggests choosing $\mathcal{I}_0^{\perp}$ in (10) so that the pilots (entries of $\check{\boldsymbol{b}}$) are equispaced nonzero entries of $\tilde{\boldsymbol{b}}$ (when $l_0 \neq 0$ equispaced should be understood in a circular sense; i.e., after periodically repeating $\tilde{\boldsymbol{b}}$).

This shows that in AWN, the optimal pilots in the sense of minimizing the channel MSE [or in the ML sense when $\boldsymbol{w}(i)$ is additive white Gaussian noise (AWGN)] are those that are equispaced and equipowered. This is also corroborated intuitively if we think of probing harmonics that are to be resolved in the output of a system identification experiment. In AWN, these harmonics are easier to retrieve when they are maximally separated over the finite-length observation record. But maximum separation over a finite length is a necessary and sufficient condition for equidistant spacing. In summary, we proved the following.

Theorem 2: If $\boldsymbol{w}(i)$ is AWN with variance σ_w^2, then the minimum channel MSE for a fixed power constraint $\|\boldsymbol{b}\|^2 = \mathcal{P}_b$ is attained for $K = L+1$ if and only if the pilot tones are equispaced and equipowered. Then, the minimum channel MSE is given by

$$\sigma_{\hat{h},\text{opt}}^2 := \frac{L+1}{\mathcal{P}_b}\sigma_w^2. \tag{13}$$

Recalling that the channel has $L+1$ taps, *Theorem 2* implies that the minimum channel MSE is achieved even with the minimum number of pilot tones required for channel estimation. It also asserts that equispaced as well as equipowered pilot tones are necessary and sufficient to attain the minimum MSE. In contrast, [13, Th. 1] *assumed* equipowered pilots in order to establish optimality of equispaced pilots, in the presence of AWN. Interestingly, it will turn out that if the noise power spectral density (psd) is available at the transmitter, white noise is the worst-case scenario in terms of channel estimation accuracy, which motivates looking into the colored noise case as well (also not addressed by [13]).

Let the number of pilot tones be minimum for block-by-block channel estimation, i.e., $K = L+1$. Recalling the discussion following (5), we can write the pseudoinverse of $\boldsymbol{B}\check{\boldsymbol{F}}^{\mathcal{H}}\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L}$ as $\boldsymbol{B}^{\dagger} = (\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathcal{H}} = (\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-1}\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{D}}_B^*\check{\boldsymbol{F}}$. Substituting this into (12), we obtain $\sigma_{\hat{h}}^2 = \text{tr}[(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-2}\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{D}}_B^*\check{\boldsymbol{F}}\boldsymbol{R}_w\check{\boldsymbol{F}}^{\mathcal{H}}\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L}]$. For N sufficiently large, $\boldsymbol{R}_w$ can be diagonalized by FFT and IFFT matrices; hence, we have $\sigma_{\hat{h}}^2 \cong \text{tr}[(\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})^{-2}\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{D}}_B^*\check{\boldsymbol{S}}_w\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L}]$, where $\check{\boldsymbol{S}}_w$ is an $(L+1)\times(L+1)$ diagonal matrix with l_ith entry $S_w(W^{l_i})$ equal to the psd of the noise evaluated at W^{l_i} for $l_i \in \mathcal{I}_0^{\perp}$. With $\check{\boldsymbol{S}}_w$ diagonal and $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B\check{\boldsymbol{F}}_{0:L}$, the channel MSE can be rewritten as

$$\sigma_{\hat{h}}^2 \cong \text{tr}[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}(\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B)^{-1}\check{\boldsymbol{S}}_w]. \tag{14}$$

Since $(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}$ is positive definite and $(\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B)^{-1}\check{\boldsymbol{S}}_w$ is diagonal, $\sigma_{\hat{h}}^2$ is bounded as follows:

$$\sigma_{\hat{h}}^2 \leq \text{tr}\left[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}\right] \cdot \max_{l_i \in \mathcal{I}_0^{\perp}} \frac{S_w(W^{l_i})}{|B(W^{l_i})|^2}. \tag{15}$$

Notice that $\text{tr}[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}]$ does not depend on the channel MSE. Targeting designs that do not require knowledge of the noise statistics at the transmitter, we design our pilot tones so that they minimize $\text{tr}[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}]$. From $\text{tr}(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L}) = (L+1)^2/N$, we deduce that $\text{tr}[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}] \geq (L+1)^2/\text{tr}(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L}) = N$. Hence, $\text{tr}[(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1}]$ is minimized if and only if $\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L} = [(L+1)/N]\boldsymbol{I}_{L+1}$, i.e., when the pilot tones are equispaced.

If the pilot tones are equispaced, then we have from (14) that

$$\begin{aligned}\sigma_{\hat{h}}^2 &\cong \frac{N}{L+1}\sum_{l_i \in \mathcal{I}_0^{\perp}} \frac{S_w(W^{l_i})}{|B(W^{l_i})|^2} \\ &\leq \frac{N}{L+1} \cdot \max_{l_i \in \mathcal{I}_0^{\perp}} S_w(W^{l_i}) \cdot \sum_{l_i \in \mathcal{I}_0^{\perp}} \frac{1}{|B(W^{l_i})|^2}.\end{aligned} \tag{16}$$

Under the constraint $\sum_{l_i \in \mathcal{I}_0^{\perp}} |B(W^{l_i})|^2 = N\mathcal{P}_b$, this upper bound is minimized if and only if we set $|B(W^{l_i})|^2$ to be constant, such that $|B(W^{l_i})|^2 = N\mathcal{P}_b/(L+1)$ for $l_i \in \mathcal{I}_0^{\perp}$, i.e., when the pilot tones are equipowered. For equispaced and equipowered pilot tones, with $(\check{\boldsymbol{F}}_{0:L}^{\mathcal{H}}\check{\boldsymbol{F}}_{0:L})^{-1} = [N/(L+1)]\boldsymbol{I}_{L+1}$ and $(\check{\boldsymbol{D}}_B^*\check{\boldsymbol{D}}_B)^{-1} = [(L+1)/N\mathcal{P}_b]\boldsymbol{I}_{L+1}$, we obtain from (14) that

$$\sigma_{\hat{h}}^2 \cong \frac{1}{\mathcal{P}_b}\text{tr}(\check{\boldsymbol{S}}_w) = \frac{1}{\mathcal{P}_b}\sum_{l_i \in \mathcal{I}_0^{\perp}} S_w(W^{l_i}). \tag{17}$$

We can summarize our results in the presence of colored noise as follows.

Theorem 3: Let the number of pilot tones be minimum for channel estimation; i.e., $K = L+1$. Suppose that the additive noise is stationary with zero mean and psd $S_w(z)$. Under the power constraint $\|\boldsymbol{b}\|^2 = \mathcal{P}_b$, the upper bound (15) on the channel MSE is asymptotically (as $N \to \infty$) minimized if the pilot tones are equispaced and equipowered, and the minimum channel MSE is given by (17).

Theorem 3 implies that without knowing the noise psd at the transmitter, equispaced and equipowered pilot tones are optimal in the sense of minimizing the upper bound of the channel MSE. If we set $\boldsymbol{\Theta} = \boldsymbol{I}_M$, *Theorems 2* and *3* show the optimality of conventional OFDM transmissions with equispaced and equipowered pilot tones in terms of bandwidth efficiency and MMSE channel estimation performance. Recall, however, that (10) offers degrees of freedom in designing $\boldsymbol{\Theta}$.

According to our pilot design rules 1) and 2), there are J possible sets of indexes for equispaced pilot tones, defined as $\mathcal{I}_{0,j}^{\perp} := \{j + Jl | l \in [0, L]\}$ for $j \in [0, J-1]$. For colored noise, these sets yield different channel MSE in general (cf. (17)). For colored noise with variance σ_w^2, one readily finds that $\min_j(1/\mathcal{P}_b)\sum_{l_i \in \mathcal{I}_{0,j}^{\perp}} S_w(W^{l_i}) \leq (L+1)/\mathcal{P}_b\sigma_w^2$. This implies that if the noise psd is available at the transmitter, one can select the optimal set to obtain the minimum channel MSE, which is always less than the channel MSE for the white noise case. Otherwise, shifting (or hopping) these sets from block to block is well motivated, because the *average* channel MSE will be lowered.

If we randomly choose one set with equal probability (or shift the set from block to block), then the average channel MSE is found after scaling (17) with $J = N/(L+1)$

$$\sigma_{\hat{h}}^2 \approx \frac{1}{\mathcal{P}_b J}\sum_{j=0}^{J-1}\sum_{l_i \in \mathcal{I}_{0,j}^{\perp}} S_w(W^{l_i}) = \frac{L+1}{\mathcal{P}_b N}\sum_{n=0}^{N-1} S_w(W^n). \tag{18}$$

For N sufficiently large, the average channel MSE is asymptotically given by $\sigma_w^2(L+1)/\mathcal{P}_b$. Interestingly, it is equal to the minimum channel MSE for AWN with the same variance [cf. (13)]. Recall that the precoder must also shift (or hop) according to the set of pilot tones through $\boldsymbol{P}_{\mathcal{I}_0^{\perp}}$ in (10), and hence, the resulting bit error rate (BER) will be also averaged.

V. Optimal Power Allocation and Precoder Selection

The ultimate goal is to allocate power and design affine precoders that minimize BER for a given transmit-power budget. However, we will rely here on an approximate (but tractable) expression for the symbol MSE at the output of the ZF equalizer constructed from the LS channel estimate given by (6). We assume that the information block $\boldsymbol{s}(i)$ is white with zero mean, and correlation $\sigma_s^2\boldsymbol{I}_M$, and that the AWN has variance σ_w^2. Because MSE (or BER) performance depends critically on the channel estimation accuracy, we will confine ourselves to equispaced and equipowered pilot tones that we have found optimal for channel estimation (*Theorem 2*).

For the moment, let us consider the ZF equalizer based on the exact CSI which generates $\hat{\boldsymbol{s}}_{\text{zf}}(i) := (\boldsymbol{H}\boldsymbol{A})^{\dagger}\boldsymbol{x}(i)$, and let us bound the symbol MSE using (4) and **C3** ($(\boldsymbol{H}\boldsymbol{A})^{\dagger}\boldsymbol{B} = \boldsymbol{0}$) as follows:

$$E\{\|\hat{\boldsymbol{s}}_{zf}(i) - \boldsymbol{s}(i)\|^2\} = \sigma_w^2\|(\boldsymbol{H}\boldsymbol{A})^{\dagger}\|_F^2 \leq \sigma_w^2\|\boldsymbol{H}^{\dagger}\|_F^2\,\|\boldsymbol{A}^{\dagger}\|_F^2 \tag{19}$$

where $\|\boldsymbol{A}\|_F := [\text{tr}(\boldsymbol{A}^{\mathcal{H}}\boldsymbol{A})]^{1/2}$. If CSI is available at the transmitter, then minimizing the symbol MSE amounts to minimizing $\|(\boldsymbol{H}\boldsymbol{A})^{\dagger}\|_F^2$ over $\boldsymbol{A}$. However, targeting a channel-independent precoder, it is reasonable to look for an $\boldsymbol{A}$ that minimizes $\|\boldsymbol{A}^{\dagger}\|_F^2$ in (19).

Under the power constraint $\mathcal{P}_s := E\{\|\boldsymbol{A}\boldsymbol{s}\|^2\} = \sigma_s^2\|\boldsymbol{A}\|_F^2$, and using the fact that $\|\boldsymbol{A}\|_F^2\|\boldsymbol{A}^{\dagger}\|_F^2 \geq M^2$, we find that $\|\boldsymbol{A}^{\dagger}\|_F^2$ is minimized if and only if $\boldsymbol{A}$ is orthogonal, i.e., $\boldsymbol{A}^{\mathcal{H}}\boldsymbol{A} = c_s\boldsymbol{I}_M$, where $c_s := \mathcal{P}_s/(M\sigma_s^2)$. If the channel estimation error is sufficiently small, the same argument applies asymptotically to the symbol MSE of the ZF equalizer output that is based on the channel estimate. Thus, we henceforth restrict $\boldsymbol{A}$ in (10) to be orthogonal.

The ZF equalizer output based on the channel estimate $\hat{\boldsymbol{H}}$ is given by $\hat{\boldsymbol{s}}(i) = (\hat{\boldsymbol{H}}\boldsymbol{A})^{\dagger}\boldsymbol{x}(i)$. Define the normalized symbol MSE as

$$\sigma_{\hat{s}}^2 := \frac{E\{\|\hat{\boldsymbol{s}}(i) - \boldsymbol{s}(i)\|^2\}}{E\{\|\boldsymbol{s}(i)\|^2\}} = \frac{E\{\|\hat{\boldsymbol{s}}(i) - \boldsymbol{s}(i)\|^2\}}{M\sigma_s^2}. \tag{20}$$

Since $N \gg L$ and $\boldsymbol{A}$ is orthogonal, we may approximate $\boldsymbol{A}\boldsymbol{A}^{\mathcal{H}}$ as $\boldsymbol{A}\boldsymbol{A}^{\mathcal{H}} \approx c_s\boldsymbol{I}_N$. Based on this approximation, we show in Appendix B that if $\boldsymbol{H}$ is invertible and well conditioned, then $\sigma_{\hat{s}}^2$ can be approximated as

$$\sigma_{\hat{s}}^2 \approx \frac{1}{M}\|\boldsymbol{H}^{-1}\|_F^2\left(\frac{\sigma_w^2}{c_s\sigma_s^2} + \sigma_{\hat{h}}^2\right). \tag{21}$$

We will allocate power on information bearing and pilot symbols so that the symbol MSE in (21) is minimized. For simplicity, and without loss of generality, we can set $E\{\|\boldsymbol{u}(i)\|^2\} = E\{\|\boldsymbol{A}\boldsymbol{s}(i)\|^2\} + \|\boldsymbol{b}\|^2 = 1$ with $E\{\|\boldsymbol{A}\boldsymbol{s}(i)\|^2\} = \mathcal{P}_s = \alpha\,(= Mc_s\sigma_s^2)$, and $\|\boldsymbol{b}\|^2 = \mathcal{P}_b = 1-\alpha$ for $0 < \alpha < 1$. The value of α dictates the power allocated to information symbols and can be considered as the effective symbol power (rate).

Rewriting (13) as $\sigma_{\hat{h}}^2 = (L+1)\sigma_w^2/(1-\alpha)$, we can express (21) as

$$\sigma_{\hat{s}}^2 \approx \|\boldsymbol{H}^{-1}\|_F^2\,\sigma_w^2\left[\frac{1}{\alpha} + \frac{L+1}{M(1-\alpha)}\right]. \tag{22}$$

By differentiating (22) with respect to α, the minimum $\sigma_{\hat{s},\text{opt}}^2$ and α_{opt} are found as

$$\sigma_{\hat{s},\text{opt}}^2 := \|\boldsymbol{H}^{-1}\|_F^2\,\sigma_w^2\left(1 + \sqrt{\frac{L+1}{M}}\right)^2 \tag{23}$$

$$\alpha_{\text{opt}} := \frac{1}{1 + \sqrt{\frac{(L+1)}{M}}}. \tag{24}$$

Thus, we arrive at the following theorem.

Theorem 4: Let the information block $\boldsymbol{s}(i)$ be white with zero mean and correlation $\sigma_s^2\boldsymbol{I}_M$ and the noise be AWN with variance σ_w^2. Suppose that the number of pilot tones is $K = L+1$ and that the precoder is orthogonal. If we set the effective symbol power (rate) as in (24), then the minimum symbol MSE of the ZF equalizer constructed from the LS channel estimator is approximately attained and is given by (23), provided that the channel matrix $\boldsymbol{H}$ is invertible and well conditioned.

Theorem 4 holds true for most practical constellations, which provide white information blocks. Interestingly, α_{opt} depends only on the ratio of the number $L+1$ of pilot tones over the number M of information symbols. As M increases, α_{opt} increases, and more power is loaded on information symbols. Conversely, as the channel order L increases, α_{opt} decreases, and more power is allocated to pilot tones to obtain reliable channel estimates.

The minimum number of $K = L+1$ pilot tones per block is dictated by our objective of coping with fast-fading channels that may change from block to block. However, there are wireless settings (e.g., low-speed HIPERLAN2) where the channel

does not change as fast. Certainly, if the Lth-order channel remains invariant over a frame of at least $L+1$ OFDM symbols (blocks), then we can have only one pilot per block which should shift (hop) from block to block, so that $L+1$ equipowered pilots are equidistant within the frame. Similarly, if the channel changes, say, every two blocks, then we can have (for L odd) $(L+1)/2$ pilots in each block that are equipowered and remain equidistant over the two blocks. One could argue that these options are subsumed by our analysis if the frame of blocks is viewed as a single superblock. However, further research is needed to delineate the following tradeoffs that appear in such cases: smaller block sizes imply shorter and faster FFTs whose output exhibits lower peak-to-average power ratio (PAR) that leads to higher power efficiency; furthermore, having to allocate power to less than $L+1$ pilots per subblock increases the percentage of power dedicated to the information-bearing symbols. On the other hand, one increases the decoding delay (since the channel estimator requires $L+1$ pilots anyway) and also reduces bandwidth efficiency because more than one CP of length L is needed to remove IBI.

So far, we assumed that the circulant channel matrix $\boldsymbol{H}$ in (6) is invertible. If the channel has nulls on the FFT grid, then $\boldsymbol{H}$ becomes singular, and hence, symbol detection is not guaranteed [17], [19]. To assure *channel-independent* invertibility of $\boldsymbol{HA}$, we rely on the judicious selection of $\boldsymbol{\Theta}$ suggested by [5], [19], and [20] for generalized multicarrier (GMC) CDMA. With this $\boldsymbol{\Theta} \neq \boldsymbol{I}$ choice, our block transmission in (10) can be viewed as an LP-OFDM modulation. LP-OFDM uses $L+1$ subcarriers for pilots and adopts an $(M+L)\times M$ orthogonal matrix $\boldsymbol{\Theta}$ in (10), which distributes each symbol across subcarriers by assigning different linear combinations of symbols per subcarrier. Matrix $\boldsymbol{\Theta}$ introduces redundancy of length L, such that:

A3. *Any M rows of $\boldsymbol{\Theta}$ are linearly independent.*

The channel matrix $\boldsymbol{H}$ can be diagonalized by $\boldsymbol{F}$ to obtain $\boldsymbol{HA} = \breve{\boldsymbol{F}}^{\mathcal{H}}\breve{\boldsymbol{D}}_H\boldsymbol{\Theta}$, where $\breve{\boldsymbol{D}}_H := \mathrm{diag}[H(W^{i_0}),\ldots,H(W^{i_{M+L-1}})]$ for $i_n \in \mathcal{I}_0$ ($i_n < i_m$ if $n<m$), $H(z)$ is the channel transfer function defined as $H(z) := \sum_0^L h(n)z^{-n}$, and $\breve{\boldsymbol{F}}^{\mathcal{H}}$ is an $N\times(M+L)$ matrix with nth column equal to the i_nth column of $\boldsymbol{F}^{\mathcal{H}}$ for $n \in [0, M+L-1]$. Because the channel has order L, at most, L values of $H(W^{i_n})$ are zero. Under **A3**, $\breve{\boldsymbol{D}}_H\boldsymbol{\Theta}$, and hence, $\boldsymbol{HA}$ have full column rank, regardless of the channel nulls. Therefore, channel-independent symbol detectability is guaranteed, while bandwidth efficiency and optimal symbol power are given, respectively, by

$$\mathcal{E}_{\mathrm{opt}}^{\mathrm{LP-OFDM}} := \frac{M}{M+3L+1}$$
$$\alpha_{\mathrm{opt}}^{\mathrm{LP-OFDM}} := \frac{1}{1+\sqrt{\frac{(L+1)}{(M+L+1)}}}. \tag{25}$$

For a fixed block size N, LP-OFDM incurs a small efficiency loss compared to OFDM. In addition, LP-OFDM necessitates an extra multiplication by $\boldsymbol{\Theta}$ at the transmitter as well as matrix inversion and multiplication for ZF equalization, which requires $O(N^3)$ computations. The latter requires $O(N\log N)$ computations due to two FFTs at the receiver. For hard-decision decoding, as OFDM entails only one IFFT at the transmitter and one FFT at the receiver, it is computationally more efficient than LP-OFDM. The bandwidth efficiency loss along with the extra computations is the price we have to pay for the universal property of assuring symbol detectability over all FIR channels of order L.

Precoders ensuring channel-independent symbol detectability enable the full multipath diversity that becomes available with the underlying frequency-selective channel, hence, for channels with $L+1$ independent identically distributed (i.i.d.) complex Gaussian taps, LP-OFDM unleashes the maximum possible diversity of order $L+1$ [20]. Traditionally, error control codes have been used to increase the diversity of uncoded OFDM over fading channels, from order one, to an order equal to $d_{\min}$, where $d_{\min}$ is the Hamming distance of the code used. For a given $(M+L, M)$ code, the Singleton bound asserts that $d_{\min} \leq M+L-M+1 \leq L+1$. However, as codes with $d_{\min} = L+1$ are not available $\forall M$, L sizes, LP-OFDM can enjoy higher diversity gain than coded OFDM (COFDM) [20]. Certainly, to collect the full diversity gain that LP-OFDM enables, ML decoding is required, which can be computationally more demanding than ML decoding of convolutionally coded OFDM that typically relies on the soft Viterbi algorithm (SOVA). However, there are LP-OFDM precoders that allow application of SOVA decoding, and others that can afford near-ML (soft) decoding with complexity as low as cubic in the block size (those include the sphere decoding, semidefinite programming, or probability data association algorithms) [20]. For detailed derivations on diversity and coding-gain analyses of LP-OFDM, batch and iterative decoding options, complexity versus performance tradeoffs, and comparisons between LP-OFDM and COFDM, we refer the reader to [20].

One special $\boldsymbol{\Theta}$ satisfying **A3** is the submatrix formed by M columns of the $(M+L)\times(M+L)$ FFT matrix; e.g., $\boldsymbol{\Theta} = \boldsymbol{F}_{M+L,0:M-1}$ that has the first M columns of the $(M+L)\times(M+L)$ FFT matrix. This selection generates a simple precoding matrix $\boldsymbol{A}$ with 0 and 1 entries, and results in nothing but a CP, single carrier, block transmission.

Suppose, for simplicity, that the channel estimate is perfect, and that the channel has no nulls on the FFT grid. Then, the symbol estimate is $\hat{\boldsymbol{s}}_{\mathrm{zf}}(i) = \boldsymbol{A}^{\dagger}\boldsymbol{H}^{-1}\boldsymbol{x}(i) = \boldsymbol{s}(i) + \breve{\boldsymbol{w}}(i)$, where $\breve{\boldsymbol{w}}(i) := \boldsymbol{F}^{\mathcal{H}}_{M+L,0:M-1}\breve{\boldsymbol{D}}_H^{-1}\breve{\boldsymbol{F}}^{\mathcal{H}}\boldsymbol{w}(i)$. For M sufficiently large, the correlation matrix of $\breve{\boldsymbol{F}}^{\mathcal{H}}\boldsymbol{w}(i)$ is asymptotically diagonal with nth entry $S_w(W^{i_n})$. It follows that the diagonal entries of $\boldsymbol{R}_{\breve{w}} := E\{\breve{\boldsymbol{w}}(i)\breve{\boldsymbol{w}}^{\mathcal{H}}(i)\}$ are equal, and are given by $\sum_{i_n\in\mathcal{I}_0} S_w(W^{i_n})/|H(W^{i_n})|^2$; that is, all entries of $\hat{\boldsymbol{s}}_{\mathrm{zf}}(i)$ are corrupted by (scalar) i.i.d. noise. In other words, the channel and noise effects on each entry of $\hat{\boldsymbol{s}}_{\mathrm{zf}}(i)$ are *averaged over one block*. This implies that the affine precoder with FFT-based $\boldsymbol{\Theta}$ is robust to the channel frequency response and the noise color. It enables one to control the BER averaged over one (even uncoded) information block only by selecting the appropriate transmit power without knowledge of the channel and the noise spectrum at the transmitter. In contrast, for the conventional OFDM with $\boldsymbol{\Theta} = \boldsymbol{I}$, the nth entry of $\hat{\boldsymbol{s}}_{\mathrm{zf}}(i)$ is contaminated by a noise term with variance $S_w(W^{i_n})/|H(W^{i_n})|^2$. This means that the BER for the nth entry of the symbol block degrades when $S_w(W^{i_n})$ is large and/or when $|H(W^{i_n})|$ is small. Thus,

TABLE I
CHANNEL MSE (IN DECIBELS) FOR DIFFERENT EQUISPACED PILOT TONES ($\mathrm{SNR} = 10$ dB)

	white noise	colored noise		
		best set	hopping	worst set
empirical	-14.0	-18.1($\mathcal{I}_{0,5}^{\perp}$)	-14.1	-10.1($\mathcal{I}_{0,0}^{\perp}$)
theoretical	-14.0	-18.5($\mathcal{I}_{0,5}^{\perp}$)	-14.0	-9.46 ($\mathcal{I}_{0,0}^{\perp}$)

error control coding and interleaving are the only remedies for conventional wireless OFDM transmissions. Unlike wireless, wireline (e.g., X-DSL) modems have the extra option of power and bit-loading strategies to control the average BER across each information block.

VI. SIMULATED PERFORMANCE

A. COFDM and LP-OFDM

We set the block size to be $N = 70$, and the guard interval length (channel order) to be $L = 6$. With $M = N-(2L+1) = 57$, we compared two OFDM transmission systems over frequency-selective channels; namely, COFDM with $\boldsymbol{\Theta} = \boldsymbol{I}_{63}$ and (63,57) Bose–Chaudhuri–Hocquengem (BCH) codes, against LP-OFDM with $\boldsymbol{\Theta} = \boldsymbol{F}_{63,0:56}$. We randomly generated binary messages of length 57. For COFDM, messages were coded by BCH (63,57) with minimum Hamming distance $d_{\min} = 3 < L+1 = 7$ and then binary coded, while for LP-OFDM, they were only binary coded. As LP-OFDM enables the full diversity of order $L+1 = 7$, while COFDM can enable diversity only up to order $d_{\min} = 3$, we expect LP-OFDM to outperform COFDM with soft ML decoding. With superior diversity gains, LP-OFDM can afford hard (or even linear) suboptimum decoding. For this reason, we employed ZF equalization and suboptimal hard-decision decoding, which also reduces complexity and decoding delay.

For both OFDM systems, $L+1 = 7$ equispaced pilot tones were inserted. Thus, both OFDM systems have the same bandwidth efficiency $M/(M+3L+1) = 0.75$. We generated 10^4 Rayleigh channels of sixth order with i.i.d. complex zero-mean Gaussian taps, where each channel was normalized to have unit norm. We tested 10^2 OFDM symbols for each channel realization, and averaged the results. The noise was either white or colored first-order Markov with coefficient -0.9. We used the received signal-to-noise ratio (SNR) defined as $E\{\|\boldsymbol{H}\boldsymbol{u}(i)\|^2\}/E\{\|\boldsymbol{w}(i)\|^2\}$.

1) Channel MSE: Table I lists channel MSE (in dB) for white and colored noise at $\mathrm{SNR} = 10$ dB. Empirical channel MSE was computed by averaging. For equipowered and equispaced pilots, the corresponding theoretical channel MSE was evaluated by (13) for white noise, and by (17) for colored noise. We also hopped equipowered pilot tones in different sets $\{\mathcal{I}_{0,0}^{\perp}, \mathcal{I}_{0,1}^{\perp}, \ldots, \mathcal{I}_{0,9}^{\perp}\}$ with equal probability. For a fixed set, the empirically best and worst sets were $\mathcal{I}_{0,5}^{\perp}$ and $\mathcal{I}_{0,0}^{\perp}$, which coincide with the theoretical ones. The difference between channel MSE of the empirically best and worst sets is 8 dB. The channel MSE of hopped pilot tones lies between them, and is almost equal to the channel MSE for white noise. From Table I, we infer that hopping pilots leads to moderate channel MSE without requiring knowledge of the noise psd at the transmitter.

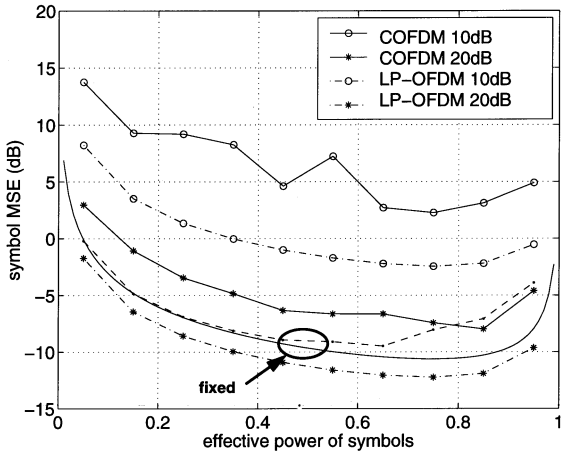


Fig. 3. COFDM over a fixed channel: theoretical symbol MSE (solid line) and empirical symbol MSE (dashed–dotted) at $\mathrm{SNR} = 10$ dB. For Rayleigh channels ($\mathrm{SNR} = 10$ dB, 20 dB): COFDM (solid lines); LP-OFDM (dashed–dotted lines).

2) Power Allocation Between Information-Bearing and Pilot Symbols: To validate our approximate expression (22) for the symbol MSE, we evaluated and compared it with empirical values for COFDM with a fixed channel and white noise. The symbol MSE was computed from the difference of the transmitted and the ZF-equalized information block. The channel matrix $\boldsymbol{H}$ is invertible with $\|\boldsymbol{H}^{-1}\|_F^2 \approx 342$. The optimum α was found from (23) to be $\alpha_{\mathrm{opt}} = 0.75$. Fig. 3 shows agreement between the theoretical symbol MSE and the empirical symbol MSE. It is observed that the symbol MSE is flat around its minimum. Since $\|\boldsymbol{H}^{-1}\|_F^2$ is a scale factor in (22), the symbol MSE is robust to the choice of α for any channel not having nulls on (or close to) the FFT grid.

Fig. 3 also compares the symbol MSE of LP-OFDM and COFDM as a function of α at $\mathrm{SNR} = 10$ dB and $\mathrm{SNR} = 20$ dB for Rayleigh channels. The fluctuation of the symbol MSE with COFDM is due to some realizations having ill-conditioned channel matrices. Except for the fluctuation, we infer that both have minimum symbol MSE around α_{opt}, and that the symbol MSE is flat around α_{opt}. LP-OFDM enjoys a 5-dB gain over conventional OFDM across all SNR values. However, it should be noted that the MSE of COFDM in Fig. 3 could not exhibit frequency diversity gain, because the MSE are computed based on undecoded outputs of ZF equalizers, and uncoded OFDM does not have any frequency diversity gain.

Fig. 4 illustrates the BER corresponding to Fig. 3. The BER curves for COFDM are smooth compared to its symbol MSE curves in Fig. 3. In uncoded OFDM, this is due to channel nulls located on (or close to) the FFT grid. The symbol error probability for these "bad subchannels" can be very high. Even when such a high error probability yields only a few detected symbols in error, the corresponding symbol MSE that is averaged over the information block is high. In other words, there is no exact one-to-one correspondence between symbol MSE and BER. However, as we can see from from Fig. 4, there exists a strong correspondence between symbol MSE and BER, and the minimum BER was also attained around α_{opt}, which validates our design criterion.

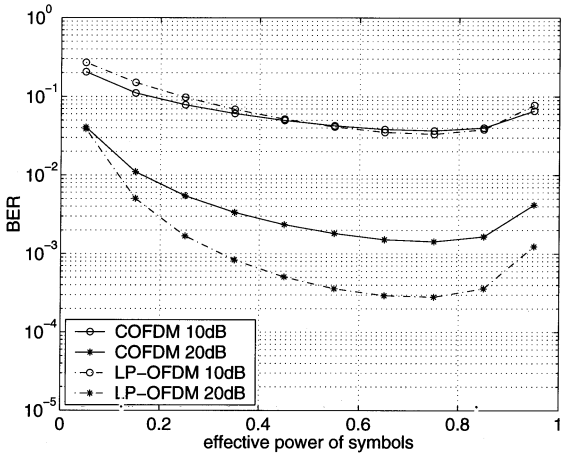


Fig. 4. BER versus α for Rayleigh channels ($\mathrm{SNR} = 10$ dB, 20 dB): COFDM (solid lines); LP-OFDM (dashed–dotted lines).

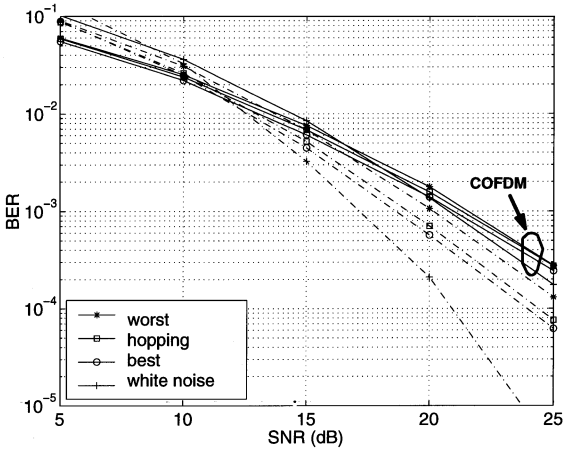


Fig. 5. BER of COFDM and LP-OFDM for Rayleigh channels ($\alpha = 0.75$): COFDM (solid lines), LP-OFDM (dashed–dotted lines). For colored noise, best shift (with ∘), worst shift (with *), hopping with equal probability (with ■). For white noise, hopping with equal probability (with +).

The BER difference between COFDM and LP-OFDM can not be discerned at 10 dB. The diversity gain can be deduced, however, from the MSE of COFDM in Fig. 3. At 20 dB, LP-OFDM has smaller BER, which confirms that LP-OFDM enjoys higher diversity gain than COFDM. Further comparisons between LP-OFDM and COFDM with respect to complexity, diversity, and coding gains, as well as various decoding options (hard, soft, and iterative) can be found in [20].

3) Colored Noise Case: For Rayleigh channels, Fig. 5 depicts the BER of COFDM and LP-OFDM as a function of SNR. LP-OFDM outperforms COFDM above $\mathrm{SNR} = 15$ dB, and the performance gain increases as SNR increases. For COFDM, there is no significant difference between BER performance with different pilot tones. For LP-OFDM, hopping pilot tones leads to improved BER. Since hopping does not require knowledge of the noise psd at the transmitter and is easily implemented, it emerges as an attractive technique for wireless OFDM. Particularly above $\mathrm{SNR} = 15$ dB, LP-OFDM with hopping pilot tones outperforms COFDM with all the sets of pilot tones considered here.

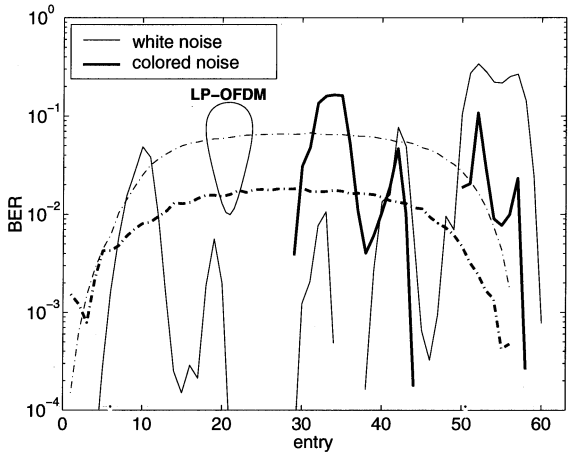


Fig. 6. BER of OFDM and LP-OFDM for each symbol in one block for a fixed channel ($\mathrm{SNR} = 10$ dB).

B. Comparison Between Uncoded OFDM and LP-OFDM

To gain further insight into the performance of LP-OFDM, we compared it with uncoded OFDM for $N = 64$ and $L = 7$. We set $\boldsymbol{\Theta} = \boldsymbol{I}_{56}$ for OFDM, and $\boldsymbol{\Theta} = \boldsymbol{F}_{56,0:48}$ for LP-OFDM. For both OFDM systems, $L + 1 = 8$ equispaced pilot tones were utilized. We dealt with two cases, a fixed FIR channel of seventh order with impulse response vector

$$\begin{aligned}\boldsymbol{h} = [&-0.3431 + 0.2715i,\\ &-0.0603 - 0.2021i, 0.1458 + 0.2076i,\\ &-0.2907 + 0.4070i, -0.2043 + 0.0989i,\\ &0.1959 + 0.2704i, -0.4199 + 0.0405i,\\ &0.3295 + 0.0052i]^T\end{aligned}$$

and Rayleigh channels of seventh order with i.i.d. complex zero-mean Gaussian taps. We conducted 10^4 Monte Carlo simulations and averaged the results. The symbols in information blocks were drawn from a binary phase-shift keying (BPSK) constellation, and the noise was either white or colored first-order Markov with coefficient -0.9.

Fig. 6 reports the BER of each symbol in the information block for OFDM and LP-OFDM with hopping pilot tones, respectively. Recall that the nth entry of the information block for OFDM is loaded on the nth subcarrier. It can be observed from Fig. 6 that for white noise, the BER of the nth symbol depends on the frequency response of the channel. To mitigate the channel effects, interleavers have been employed, traditionally to average the channel effects *statistically* over multiple blocks. On the other hand, the BER of each entry of the information symbol for LP-OFDM is relatively flat regardless of the channel, and the noise color except for the ends of each block. In LP-OFDM, the channel effects are averaged *deterministically* over one block by $\boldsymbol{\Theta}$.

Fig. 7 plots the BER of each entry of the information block for Rayleigh channels. Since the channel effects were averaged, the BER dependence of OFDM on the noise psd can be clearly seen. On the contrary, the BER curves of LP-OFDM are similar to those for the fixed channel shown in Fig. 6, as discussed in the last paragraph of Section V.

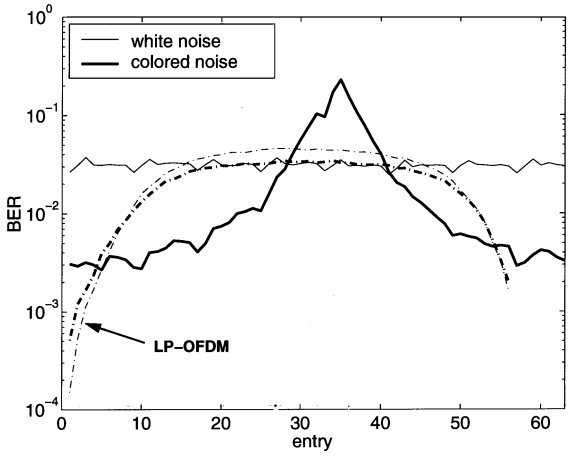


Fig. 7. BER of OFDM and LP-OFDM for each symbol in one block for Rayleigh channels ($\mathrm{SNR} = 10$ dB).

VII. Conclusion

To enable low-complexity block-by-block processing, redundant precoders with CP and superimposed training symbols have been designed to decouple symbol detection from LS-based channel estimation. Optimal pilot tones were designed for channel estimation in the presence of white or colored noise. Optimal power loading scheme on information symbols and pilot tones was derived based on an approximate expression for the symbol MSE of the ZF equalizer constructed from the channel estimate. To assure channel-independent symbol detection, an LP-OFDM modulation was developed, and was shown to be robust to channel effects and correlated noise. Simulations corroborated LP-OFDM's improved performance over coded OFDM with the same bandwidth efficiency.

Appendix I
Proof of *Theorem 1*

To prove *Theorem 1*, we utilize the following lemma.

Lemma 2': Condition ***C3****, or equivalently, (7) holds true if and only if there exists an* $(N-L-1)\times M$ *matrix* $\boldsymbol{\Phi}$ *satisfying* $\boldsymbol{D}_B^*\boldsymbol{F}\boldsymbol{A}=\boldsymbol{F}_{L+1:N-L-1}\boldsymbol{\Phi}$*, where * denotes complex conjugation and* $\boldsymbol{F}_{L+1:N-L-1}$ *is formed by the* $(L+1)$*st to the* $(N-L-1)$*st columns of* $\boldsymbol{F}$*.*

Proof: Similar to (5), the circulant matrix $\boldsymbol{A}_m$ satisfies $\boldsymbol{A}_m = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{D}_{A_m}\boldsymbol{F}_{0:L}$, where $\boldsymbol{D}_{A_m}$ is a diagonal matrix defined as $\boldsymbol{D}_{A_m} := \mathrm{diag}[A_m(1), A_m(W), \ldots, A_m(W^{N-1})]$ with $A_m(z) := \sum_{n=0}^{N-1} a_{m,n}z^{-n}$. Along with (5), this decomposition implies that (7) holds if and only if $\boldsymbol{F}_{0:L}^{\mathcal{H}}\boldsymbol{D}_B^*\boldsymbol{D}_{A_m}\boldsymbol{F}_{0:L} = \boldsymbol{0}$, where we also used $\boldsymbol{F}^{\mathcal{H}}\boldsymbol{F} = \boldsymbol{I}$. If $\boldsymbol{f}_l$ denotes the lth column of $\boldsymbol{F}$, then $\boldsymbol{D}_{A_m}\boldsymbol{f}_l = \boldsymbol{D}(\boldsymbol{f}_l)\tilde{\boldsymbol{\alpha}}_m$ where $\boldsymbol{D}(\boldsymbol{f}_l) := \mathrm{diag}[1, W^l, \ldots, W^{(N-1)l}]$, and $\tilde{\boldsymbol{\alpha}}_m := [A_m(1), A_m(W), \ldots, A_m(W^{N-1})]^T$. Because diagonal matrices commute, we arrive at

$$\boldsymbol{F}_{0:L}^{\mathcal{H}}\boldsymbol{D}(\boldsymbol{f}_l)\boldsymbol{D}_B^*\tilde{\boldsymbol{\alpha}}_m = \boldsymbol{0}, \quad \forall m \in [0, M-1], \quad \forall l \in [0, L]. \tag{26}$$

By direct substitution, one can verify that $\boldsymbol{D}(\boldsymbol{f}_l)$ circularly shifts the rows of $\boldsymbol{F}^{\mathcal{H}}$ so that $\boldsymbol{F}_{0:L}^{\mathcal{H}}\boldsymbol{D}(\boldsymbol{f}_l)$ for $l \in [0, L]$ yields $L+1$ matrices that have FFT rows $\boldsymbol{f}_n^{\mathcal{H}}$ with $n \in \{[0, L] \cup [N-L, N-1]\}$. We note that (26) holds if and only if $\boldsymbol{D}_B^*\tilde{\boldsymbol{\alpha}}_m$ lies in the null space of these matrices, which is spanned by the remaining orthogonal FFT rows of $\{\boldsymbol{f}_n^{\mathcal{H}}\}_{n=L+1}^{N-L-1}$. Thus, (26) is true if and only if $\boldsymbol{D}_B^*\tilde{\boldsymbol{\alpha}}_m = \boldsymbol{F}_{L+1:N-L-1}\boldsymbol{\phi}_m$ for some vector $\boldsymbol{\phi}_m$. With $\boldsymbol{\Phi} := (1//N^{1/2})[\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{M-1}]$, it follows from $\tilde{\boldsymbol{\alpha}}_m = N^{1/2}\boldsymbol{F}\boldsymbol{a}_m$ that (26) and holds if and only if $\boldsymbol{D}_B^*\boldsymbol{F}\boldsymbol{A} = \boldsymbol{F}_{L+1:N-L-1}\boldsymbol{\Phi}$, which completes the proof of the lemma. ■

Because $\boldsymbol{F}^{\mathcal{H}}$ and $\boldsymbol{P}_{\mathcal{I}_0^\perp}$ are full-rank matrices, we can express without loss of generality our precoders as $\boldsymbol{A} = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0^\perp}[\boldsymbol{\Theta}^{\mathcal{H}}, \check{\boldsymbol{\Theta}}^{\mathcal{H}}]^{\mathcal{H}}$, where $\boldsymbol{\Theta}$ and $\check{\boldsymbol{\Theta}}$ are $(N-K)\times M$ and $K\times M$ submatrices to be determined. Using this decomposition of $\boldsymbol{A}$ in *Lemma 2'*, we infer that **C3** holds true if and only if $\boldsymbol{D}_B^*\boldsymbol{P}_{\mathcal{I}_0^\perp}[\boldsymbol{\Theta}^{\mathcal{H}}, \check{\boldsymbol{\Theta}}^{\mathcal{H}}]^{\mathcal{H}} = \boldsymbol{F}_{L+1:N-L-1}\boldsymbol{\Phi}$, or equivalently, $\boldsymbol{P}_{\mathcal{I}_0^\perp}^T\boldsymbol{D}_B^*\boldsymbol{P}_{\mathcal{I}_0^\perp}[\boldsymbol{\Theta}^{\mathcal{H}}, \check{\boldsymbol{\Theta}}^{\mathcal{H}}]^{\mathcal{H}} = \boldsymbol{P}_{\mathcal{I}_0^\perp}^T\boldsymbol{F}_{L+1:N-L-1}\boldsymbol{\Phi}$. Choosing $\boldsymbol{P}_{\mathcal{I}_0^\perp}^T(=\boldsymbol{P}_{\mathcal{I}_0^\perp})$ to be the permutation matrix that moves the $N-K$ all-zero rows of $\boldsymbol{D}_B$ at the top, the last equation leads to

$$\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \check{\boldsymbol{D}}_B^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\Theta}\\ \check{\boldsymbol{\Theta}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{F}_0\boldsymbol{\Phi}\\ \check{\boldsymbol{F}}_0\boldsymbol{\Phi}\end{bmatrix} \Leftrightarrow \left\{\begin{array}{l}\boldsymbol{0} = \boldsymbol{F}_0\boldsymbol{\Phi}\\ \check{\boldsymbol{D}}_B^*\check{\boldsymbol{\Theta}} = \check{\boldsymbol{F}}_0\boldsymbol{\Phi}\end{array}\right\} \tag{27}$$

where $\boldsymbol{F}_0$ is formed by the top $N-K$ rows of $\boldsymbol{P}_{\mathcal{I}_0^\perp}^T\boldsymbol{F}_{L+1:N-L-1}$ and $\check{\boldsymbol{F}}_0$ by its last K rows. Because $\boldsymbol{F}_0$ is full column rank if and only if it is square or tall (i.e., $N-K \geq N-2L-1 \Leftrightarrow K \leq 2L+1$), we find that $\boldsymbol{F}_0\boldsymbol{\Phi} = \boldsymbol{0} \Leftrightarrow \boldsymbol{\Phi} \equiv \boldsymbol{0}$ for $K \leq 2L+1$. Noting that $\check{\boldsymbol{D}}_B$ is square and full rank, we infer from (27) that $\boldsymbol{\Phi} \equiv \boldsymbol{0} \Leftrightarrow \check{\boldsymbol{\Theta}} \equiv \boldsymbol{0}$, or equivalently, from *Lemma 2* and *Lemma 2'* that **C3** holds if and only if $\boldsymbol{A} = \boldsymbol{F}^{\mathcal{H}}\boldsymbol{P}_{\mathcal{I}_0^\perp}[\boldsymbol{\Theta}^{\mathcal{H}}, \boldsymbol{0}]^{\mathcal{H}}$. Recalling that $K \geq L+1$ is necessary and sufficient for **C2**, since $N-K \geq M$ is necessary for $\boldsymbol{\Theta}$ (and hence, $\boldsymbol{A}$) to be full column rank, we deduce that **C2** and **C3** are satisfied by the class of precoders given by (10) if and only if $N \geq M+K$ and $L+1 \leq K \leq 2L+1$, which completes the proof of *Theorem 1*.

Appendix II
Derivation of (21)

For notational simplicity, we omit time indexes (in one block) and define $\boldsymbol{\Psi} := \boldsymbol{H}\boldsymbol{A}$, $\hat{\boldsymbol{\Psi}} := \hat{\boldsymbol{H}}\boldsymbol{A}$, $\Delta\boldsymbol{H} := \hat{\boldsymbol{H}} - \boldsymbol{H}$, and $\Delta\boldsymbol{\Psi} := \hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = \Delta\boldsymbol{H}\boldsymbol{A}$. We assume that $\boldsymbol{\Psi}$ is full column rank, and that $\Delta\boldsymbol{H}$ is sufficiently small so that $\hat{\boldsymbol{\Psi}}^\dagger$ can be approximated as $\hat{\boldsymbol{\Psi}}^\dagger = \boldsymbol{\Psi}^\dagger - \boldsymbol{\Psi}^\dagger\Delta\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger$. This approximation is valid if the channel matrix $\boldsymbol{H}$ is well conditioned.

By using $\boldsymbol{\Psi}^\dagger\boldsymbol{B} = (\boldsymbol{B}^\dagger\boldsymbol{\Psi})^\dagger = \boldsymbol{0}$, the symbol estimated by the ZF equalizer (constructed from channel estimates) is given by $\hat{\boldsymbol{s}} := (\hat{\boldsymbol{\Psi}}^\dagger - \boldsymbol{\Psi}^\dagger\Delta\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger)(\boldsymbol{\Psi}\boldsymbol{s} + \boldsymbol{w}) = \boldsymbol{s} - \boldsymbol{\Psi}^\dagger(\Delta\boldsymbol{\Psi}\boldsymbol{s} - \boldsymbol{w}) - \boldsymbol{\Psi}^\dagger\Delta\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger\boldsymbol{w}$. If $\boldsymbol{w}$ and $\Delta\boldsymbol{\Psi}$ are sufficiently small, so that $\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger\boldsymbol{w} \approx \boldsymbol{0}$, then the error can be approximated as $\hat{\boldsymbol{s}} - \boldsymbol{s} = -\boldsymbol{\Psi}^\dagger(\Delta\boldsymbol{\Psi}\boldsymbol{s} - \boldsymbol{w})$. It follows that

$$E\{\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|^2\} = E\{\|\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}\|^2\} + 2\mathrm{Re}E\{\boldsymbol{w}^{\mathcal{H}}\boldsymbol{\Psi}^{\dagger\mathcal{H}}\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}\} + E\{\|\boldsymbol{\Psi}^\dagger\boldsymbol{w})\|^2\}. \tag{28}$$

Since $\boldsymbol{w}$ and $\Delta\boldsymbol{H}$ are statistically independent of $\boldsymbol{s}$ with zero mean, the second term vanishes because $E\{\boldsymbol{w}^{\mathcal{H}}\boldsymbol{\Psi}^{\dagger\mathcal{H}}\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}\} = E\{\boldsymbol{w}^{\mathcal{H}}\boldsymbol{\Psi}^{\dagger\mathcal{H}}\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{H}\boldsymbol{A}\}E\{\boldsymbol{s}\} = \boldsymbol{0}$.

From $\boldsymbol{A}^{\mathcal{H}}\boldsymbol{A} = c_s\boldsymbol{I}_M$, we obtain

$$\|\boldsymbol{\Psi}^\dagger\|_F^2 = \mathrm{tr}[\boldsymbol{H}^{-1}\boldsymbol{H}^{-\mathcal{H}}\boldsymbol{A}(\boldsymbol{A}^{\mathcal{H}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathcal{H}}] = \frac{\mathrm{tr}(\boldsymbol{H}^{-1}\boldsymbol{H}^{-\mathcal{H}}\boldsymbol{A}\boldsymbol{A}^{\mathcal{H}})}{c_s^2}.$$

Substituting $\boldsymbol{A}\boldsymbol{A}^{\mathcal{H}} \approx c_s \boldsymbol{I}_N$, we can approximate $\|\boldsymbol{\Psi}^\dagger\|_F^2$ as

$$\|\boldsymbol{\Psi}^\dagger\|_F^2 \approx \frac{1}{c_s}\|\boldsymbol{H}^{-1}\|_F^2. \tag{29}$$

Thus, the last term in (28) is given by $E\{\|\boldsymbol{\Psi}^\dagger \boldsymbol{w})\|^2\} = \sigma_w^2\|\boldsymbol{\Psi}^\dagger\|_F^2 \approx \sigma_w^2\|\boldsymbol{H}^{-1}\|_F^2/c_s$. On the other hand, the first term can be approximated as

$$E\{\|\boldsymbol{\Psi}^\dagger \Delta\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}\|^2\} \approx c_s\sigma_s^2 \mathrm{tr}\left(\boldsymbol{\Psi}^\dagger E\{\Delta\boldsymbol{H}\Delta\boldsymbol{H}^{\mathcal{H}}\}\boldsymbol{\Psi}^{\dagger\mathcal{H}}\right) \tag{30}$$

where we used the statistical independence between $\boldsymbol{s}$ and $\Delta\boldsymbol{H}$. Defining the channel error vector as $\Delta\boldsymbol{h} := \hat{\boldsymbol{h}} - \boldsymbol{h} = \boldsymbol{B}^\dagger \boldsymbol{w}$, from $E\{\Delta\boldsymbol{h}\Delta\boldsymbol{h}^{\mathcal{H}}\} = \sigma_w^2 E\{\boldsymbol{B}^\dagger\boldsymbol{B}^{\dagger\mathcal{H}}\} = (\sigma_w^2/\mathcal{P}_b)\boldsymbol{I}_{L+1}$, each entry of the channel error vector is found to be mutually independent. Since $\Delta\boldsymbol{H}$ is circulant with first column $\Delta\boldsymbol{h}$, it turns out that $E\{\Delta\boldsymbol{H}\Delta\boldsymbol{H}^{\mathcal{H}}\} = \sigma_{\hat{h}}^2\boldsymbol{I}_N$. Substituting this into (30) and using (29), we arrive at $E\{\|\boldsymbol{\Psi}^\dagger\Delta\boldsymbol{H}\boldsymbol{A}\boldsymbol{s}\|^2\} \approx \sigma_s^2\sigma_{\hat{h}}^2\|\boldsymbol{H}^{-1}\|_F^2$. It follows that:

$$\begin{aligned} E\{\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|^2\} \approx& \sigma_s^2\sigma_{\hat{h}}^2\|\boldsymbol{H}^{-1}\|_F^2 + \frac{\sigma_w^2\|\boldsymbol{H}^{-1}\|_F^2}{c_s} \\ =& \|\boldsymbol{H}^{-1}\|_F^2\left(\frac{\sigma_w^2}{c_s} + \sigma_s^2\sigma_{\hat{h}}^2\right) \end{aligned}$$

and after dividing by $M\sigma_s^2$, we obtain (21).

REFERENCES

[1] S. Adireddy and L. Tong, "Detection with embedded known symbols: Optimal symbol placement and equalization," in *Proc. Int. Conf. ASSP*, vol. 5, Istanbul, Turkey, June 2000, pp. 2541–2543.

[2] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. Inform. Theory*, vol. 42, pp. 137–160, Jan. 1996.

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

[4] "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2 Technical Specification Part 1—Physical Layer," ETSI, DTS/BRAN030 003–1, 1999.

[5] G. B. Giannakis, Z. Wang, A. Scaglione, and S. Barbarossa, "AMOUR — Generalized multicarrier transceivers for blind CDMA regardless of multipath," *IEEE Trans. Commun.*, vol. 48, pp. 2064–2076, Dec. 2000.

[6] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[7] P. Höeher and F. Tufvesson, "Channel estimation with super imposed pilot sequence," in *Proc. GLOBECOM*, Rio de Janeiro, Brazil, Dec. 1999, pp. 2162–2166.

[8] T. P. Holden and K. Feher, "A spread spectrum-based system technique for synchronization of digital mobile communication systems," *IEEE Trans. Broadcast.*, vol. 36, pp. 185–194, Sept. 1990.

[9] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE 802.11, Nov. 1997.

[10] G. K. Kaleh, "Channel equalization for block transmission systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1728–1736, Jan. 1995.

[11] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *IEEE Commun. Mag.*, vol. 88, pp. 611–640, May 2000.

[12] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. Int. Conf. ASSP*, vol. 5, Istanbul, Turkey, June 2000, pp. 2741–2744.

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

[14] M. Sandell and O. Edfors, "A Comparative Study of Pilot-Based Channel Estimators for Wireless OFDM," Lulea Univ. Technol., Lulea, Sweden, Res. Rep. TULEA 1996, 1996.

[15] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, pp. 100–109, Feb. 1995.

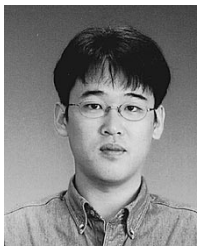
[16] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inform. Theory*, vol. 5, pp. 1019–1032, Apr. 1999.

[17] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers—Parts I and II," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2022, July 1999.

[18] F. Tufvesson, M. Faulkner, P. Höeher, and O. Edfors, "OFDM time and frequency synchronization by spread spectrum pilot technique," in *Proc. 8th Communication Theory Mini-Conf.*, Vancouver, BC, Canada, June 1999, pp. 115–119.

[19] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 47, pp. 29–48, May 2000.

[20] ——, "Linearly precoded or coded OFDM against wireless channel fades?," in *Proc. 3rd IEEE Workshop Signal Processing Advances in Wireless Communications*, Mar. 2001, pp. 267–270.



Shuichi Ohno (M'95) received the B.E., M.E., and Dr. Eng. degrees in applied mathematics and physics from Kyoto University, Kyoto, Japan, in 1990, 1992, and 1995, respectively.

From 1995 to 1999 he was a Research Associate in the Department of Mathematics and Computer Science at Shimane University, Shimane, Japan, where he became an Assistant Professor. He spent 14 months in 2000 and 2001 at the University of Minnesota, Minneapolis, as a Visiting Researcher. Since 2002, he has been an Associate Professor with the Department of Artificial Complex Systems Engineering, Hiroshima University, Hiroshima, Japan. His current interests are in the areas of signal processing in communication, wireless communications, adaptive signal processing, and multirate signal processing.

Dr. Ohno is a member of IEICE. He has served as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS since 2001.

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. He received the MSc. degree in electrical engineering in 1983, M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986, from the University of Southern California (USC), Los Angeles.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, 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 at the 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 140 journal papers, 270 conference papers, and two books. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, precoding and space-time coding for block transmissions, multicarrier, and wide-band wireless communication systems.

Dr. Giannakis is the corecipient of four Best Paper Awards from the IEEE Signal Processing (SP) Society (1992, 1998, 2000, and 2001). He also received the Society's Technical Achievement Award in 2000. He co-organized three IEEE-SP Workshops, and guest co-edited four special issues. He has served as Editor in Chief for the IEEE SIGNAL PROCESSING L ETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING L ETTERS, 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, and as chair of the SP for Communications Technical Committee. He is a member of the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a member of the IEEE Fellows Election Committee, the IEEE-SP Society's Board of Governors, and is a frequent consultant for the telecommunications industry.