Error Probability Minimizing Pilots for OFDM With M-PSK Modulation Over Rayleigh-Fading Channels

Xiaodong Cai, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Orthogonal frequency division multiplexing (OFDM) with pilot symbol assisted channel estimation is a promising technique for high rate transmissions over wireless frequency-selective fading channels. In this paper, we analyze the symbol error rate (SER) performance of OFDM with M-ary phase-shift keying (M-PSK) modulation over Rayleigh-fading channels, in the presence of channel estimation errors. Both least-squares error (LSE) and minimum mean-square error (MMSE) channel estimators are considered. For prescribed power, our analysis not only yields exact SER formulas, but also quantifies the performance loss due to channel estimation errors. We also optimize the number of pilot symbols, the placement of pilot symbols, and the power allocation between pilot and information symbols, to minimize this loss, and thereby minimize SER. Simulations corroborate our SER performance analysis, and numerical results are presented to illustrate our optimal claims.**

***Index Terms*—Channel estimation, error probability, orthogonal frequency division multiplexing (OFDM), pilots.**

I. INTRODUCTION

ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING (OFDM) provides an effective and low-complexity means of eliminating intersymbol interference for transmissions over frequency-selective fading channels [1], [2]. Channel state information (CSI) is required for the OFDM receiver to perform coherent detection, or diversity combining, if multiple transmit and receive antennas are deployed. In practice, CSI can be reliably estimated at the receiver by inserting training (a.k.a. pilot) symbols at the transmitter. Pilot symbol assisted channel estimation is especially attractive for wireless links [3], where the channel is time-varying. In [4]–[7] the channel correlation in the time and frequency domains was exploited for pilot-based channel estimation in OFDM systems. Channel estimation using pilot symbols in only one OFDM block was advocated in [8] and [9]. Interpolating schemes were investigated in [10], [11], and joint multipath delay and tap estimation of OFDM channels was studied in [12].

While many channel estimators have been developed for OFDM, error probability analysis in the presence of channel estimation errors has received relatively less attention. Only recently, BER approximations for M-ary phase shift keying (M-PSK) and M-ary quadrature amplitude modulation (QAM) were provided for OFDM with channel estimation errors [13], [14]. In this paper, we will derive *exact* symbol error rate (SER) expressions for pilot assisted OFDM transmissions with M-PSK modulation over Rayleigh-fading channels. Our SER analysis also quantifies the performance loss due to channel estimation error and the transmit pilot power. Based on this SER analysis, we will optimize the design of pilots to minimize the performance loss caused by channel estimation errors. For prescribed power, this will lead us to pilots that minimize error probability. Optimizing pilots for wireless OFDM systems has been considered recently, based on: maximizing a lower bound on ergodic capacity [15], [16], or, minimizing the channel mean-square error (MSE) [17]–[19]. Pilot optimization for single-carrier transmissions has also been investigated in [15], [20]–[22] based on these two criteria, and in [23] by minimizing the Cramér–Rao bound on the channel MSE. As error probability directly determines the reliability of a communication link, our use of SER as a criterion is certainly of practical interest.

The rest of this paper is organized as follows. Section II describes the system model, and analyzes the average SER performance in the presence of channel estimation errors. Pilot symbols are optimized to minimize SER in Section III. Simulations and numerical results are presented in Section IV, and conclusions are drawn in Section V.

Notation: Superscripts ^{T}, *, and $^\mathcal{H}$ stand for transpose, conjugate, and Hermitian transpose, respectively; $E[\cdot]$ denotes expectation. Column vectors (matrices) are denoted by boldface lower (upper) case letters; $\mathbf{I}_N$ represents the $N \times N$ identity matrix; $\boldsymbol{\mathcal{D}}(\mathbf{x})$ stands for a diagonal matrix with $\mathbf{x}$ on its diagonal; and $\mathrm{Tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. We use $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$ to denote that $\mathbf{x}$ is a complex Gaussian distributed vector with mean $\boldsymbol{\mu}$, and covariance $\mathbf{R}$.

II. MODELING AND ERROR PROBABILITY ANALYSIS

In this section, we will present the signal model, and analyze the SER performance of OFDM in the presence of channel estimation errors.

A. Signal Model

The OFDM transmission system under consideration is depicted in Fig. 1. Information and pilot symbols are modulated on a set of subcarriers, and transmitted over a frequency-selective fading channel through a single transmitter antenna. After demodulation at the receiver end, where we allow for multiple

Manuscript received May 20, 2003; revised September 23, 2003. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: caixd@ece.umn.edu; georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TVT.2003.819624

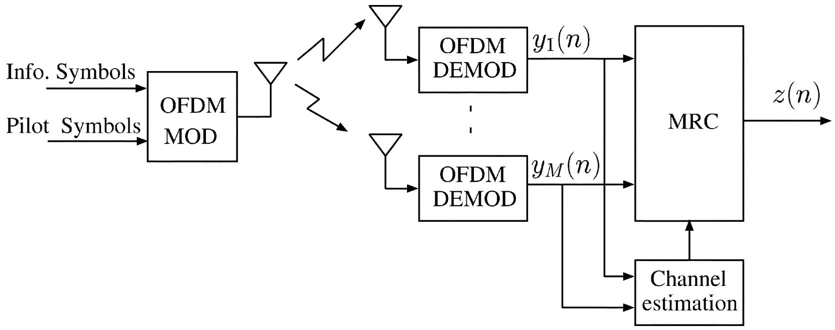


Fig. 1. OFDM transmission system.

antennas, the channel per receive-antenna is estimated using pilots. Based on the estimated channels, the maximum ratio combiner (MRC) is employed to yield decision statistics.

Suppose that the frequency-selective channels remain invariant over an OFDM block, and the length of the cyclic prefix exceeds the channel order. After demodulation, the received signal at the mth receive-antenna on the nth subcarrier corresponding to pilot symbols can be written as

$$y_m[n] = \sqrt{\mathcal{E}_p} H_m(n)s(n)+w_m(n),\ m=1,\ldots,M,\ n\in\mathcal{I}_p \quad (1)$$

where $\mathcal{I}_p$ denotes the set of subcarriers on which pilot symbols are transmitted, $\mathcal{E}_p$ is the transmitted power per pilot symbol, $H_m(n)$ is the channel frequency response of the mth antenna at the nth subcarrier, $s(n)$, $n\in\mathcal{I}_p$, is the pilot symbol, and $w_m(n)$ is complex additive white Gaussian noise (AWGN) with zero-mean and variance $N_0/2$ per dimension; and M is the number of receive-antennas. We select pilot symbols of constant modulus, i.e., $|s(n)|=1, \forall n\in\mathcal{I}_p$. We omitted the block OFDM symbol index in (1) since block-by-block channel estimation and symbol detection will be considered throughout the paper.

The received samples corresponding to information symbols can be expressed as

$$y_m[n] = \sqrt{\mathcal{E}_s} H_m(n)s(n)+w_m(n),\ m=1,\ldots,M,\ n\in\mathcal{I}_s \quad (2)$$

where $\mathcal{E}_s$ is the transmitted power per information symbol, and $\mathcal{I}_s$ denotes the set of subcarriers on which information symbols are transmitted. Suppose that the total number of subcarriers is N, and the size of $\mathcal{I}_p$ is $|\mathcal{I}_p|=P$. For simplicity, we assume that the size of $\mathcal{I}_s$ is $|\mathcal{I}_s|=N-P$, although it is possible that $|\mathcal{I}_s|< N-P$, when null subcarriers are inserted for spectrum shaping. Selecting information symbols from M-PSK constellations, we have also that $|s(n)|=1, \forall n\in\mathcal{I}_s$. The frequency-selective channel is assumed to be Rayleigh-fading, with channel impulse response $\mathbf{h}_m := [h_m(0),\ldots,h_m(L-1)]^T$ corresponding to the mth receive-antenna, and L denoting the number of taps; i.e., $h_m(l)$, $\forall m\in[1,M]$, $\forall l\in[0,L-1]$, are uncorrelated complex Gaussian random variables with zero-mean. We assume that channels associated with different antennas have identical power delay profiles specified by the variance: $\sigma_h^2(l)$, the same $\forall m\in[1,M]$. Channels are normalized so that $\sum_{l=0}^{L-1}\sigma_h^2(l)=1$. Define the $L\times N$ matrix $[\mathbf{F}]_{l,n} := \exp(j2\pi(l-1)(n-1)/N)$, and let $\mathbf{f}_n$ be the nth column of $\mathbf{F}$. Then, $H_m(n)=\mathbf{f}_n^{\mathcal{H}}\mathbf{h}_m$, is a complex Gaussian random variable with zero-mean and unit variance. The average signal-to-noise ratio (SNR) per pilot (information) symbol at each antenna is $\mathcal{E}_p/N_0$ $(\mathcal{E}_s/N_0)$. The AWGN variables $w_m(n)$ are assumed to be uncorrelated, $\forall m$, and $\forall n$.

Suppose that the set of pilot subcarriers is given by $\mathcal{I}_p = \{n_i\}_{i=1}^P$. Letting $\tilde{\mathbf{h}}_m := [H_m(n_1),\ldots,H_m(n_P)]^T$ contain the channel frequency response on pilot subcarriers, and defining $\mathbf{F}_p := [\mathbf{f}_{n_1},\ldots,\mathbf{f}_{n_P}]$, we can relate the fast Fourier transform (FFT) pair via: $\tilde{\mathbf{h}}_m = \mathbf{F}_p^{\mathcal{H}}\mathbf{h}_m$. Let the $P\times 1$ vector $\mathbf{y}_m = [y_m(n_1),\ldots,y_m(n_P)]^T$ consist of the received pilot samples per block, and define $\mathbf{s}_p := [s(n_1),\ldots,s(n_P)]^T$, and $\mathbf{w}_m := [w(n_1),\ldots,w(n_P)]^T$. From (1), we have

$$\mathbf{y}_m = \sqrt{\mathcal{E}_p}\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\tilde{\mathbf{h}}_m + \mathbf{w}_m = \sqrt{\mathcal{E}_p}\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{F}_p^{\mathcal{H}}\mathbf{h}_m + \mathbf{w}_m. \quad (3)$$

Given $\mathbf{s}_p$ and $\mathbf{y}_m$, we wish to estimate $\mathbf{h}_m$ based on (3). While it may be possible to use pilot samples from different OFDM blocks to estimate the channel as advocated in [6], we will rely on pilots from only one block to estimate the channel on a per block basis as in [8] and [9]. This is particularly suitable for packet data transmission, where the receiver may receive different blocks with unknown delays.

B. SER With LSE Channel Estimation

If we define $\mathbf{G} := \left(\mathcal{E}_p\mathbf{F}_p\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{s}_p)\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\left(\sqrt{\mathcal{E}_p}\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{F}_p^{\mathcal{H}}\right)^{\mathcal{H}}$, then the least-squares error (LSE) estimate of the channel impulse response is given by [24, p. 225]

$$\hat{\mathbf{h}}_m = \mathbf{G}\mathbf{y}_m = \mathbf{h}_m + \boldsymbol{\eta}_m \quad (4)$$

where $\boldsymbol{\eta}_m = \mathbf{G}\mathbf{w}_m$. Using the fact that $\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{s}_p)\boldsymbol{\mathcal{D}}(\mathbf{s}_p) = \mathbf{I}_P$, it follows readily that $\boldsymbol{\eta}_m \sim \mathcal{CN}(\mathbf{0},\ (\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}})^{-1}N_0/\mathcal{E}_p)$. The estimated channel frequency response on the nth subcarrier can then be obtained from (4) as

$$\hat{H}_m(n) = \mathbf{f}_n^{\mathcal{H}}\hat{\mathbf{h}}_m = H_m(n) + \nu_m(n) \quad (5)$$

where $\nu_m(n) \sim \mathcal{CN}(0,\ \sigma^2_{\nu(n)})$ with $\sigma^2_{\nu(n)} := \mathbf{f}_n^{\mathcal{H}}(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}})^{-1}\mathbf{f}_n N_0/\mathcal{E}_p$. Since the variance of $\nu_m(n)$ does not depend on the antenna index m, we omitted the index m in $\sigma^2_{\nu(n)}$. For notational brevity, we also define

$$a_n := \mathbf{f}_n^{\mathcal{H}}\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{f}_n \quad (6)$$

and then $\sigma^2_{\nu(n)} = a_n N_0/\mathcal{E}_p$. Since $H_m(n)$ and $\nu_m(n)$ are uncorrelated Gaussian random variables with zero-mean, $\hat{H}_m(n)$

is Gaussian distributed with zero-mean, and variance $\sigma^2_{\hat{H}_m(n)} = \sigma^2_{H_m(n)} + \sigma^2_{\nu(n)} = 1 + \sigma^2_{\nu(n)}$. From (5), we see that $\hat{H}_m(n)$ is correlated with $\nu_m(n)$. Hence, $\nu_m(n)$ can be written as $\nu_m(n) = \bar{\nu}_m(n) + \tilde{\nu}_m(n)$, where $\bar{\nu}_m(n) := E[\nu_m(n)|\hat{H}_m(n)]$, and $\tilde{\nu}_m(n)$ is a complex Gaussian random variable with zero-mean, which is uncorrelated with $\bar{\nu}_m(n)$. Clearly, $\bar{\nu}_m(n)$ is the linear minimum mean-square error (MMSE) estimate of $\nu_m(n)$: $\bar{\nu}_m(n) = \hat{H}_m(n)E[\hat{H}_m(n)\nu^*_m(n)]/\sigma^2_{\hat{H}_m(n)} = \hat{H}_m(n)\sigma^2_{\nu(n)}/(1 + \sigma^2_{\nu(n)})$, and the variance of $\tilde{\nu}_m(n)$ is the corresponding MMSE, which can be found as $\tilde{\sigma}^2_{\nu_m(n)} = \sigma^2_{\nu(n)}/(1 + \sigma^2_{\nu(n)})$.

The output of the mth MRC branch for $s(n)$, $n \in \mathcal{I}_s$, can be expressed as

$$\begin{aligned} z_m(n) &:= \hat{H}^*_m(n)y_m(n) \\ &= \sqrt{\mathcal{E}_s}\left|\hat{H}_m(n)\right|^2 s(n) - \sqrt{\mathcal{E}_s}\hat{H}^*_m(n)s(n)\nu_m(n) \\ &\quad + \hat{H}^*_m(n)w_m(n). \end{aligned} \tag{7}$$

Substituting $\nu_m(n) = \bar{\nu}_m(n) + \tilde{\nu}_m(n)$ into (7), we obtain

$$\begin{aligned} z_m(n) = \frac{\sqrt{\mathcal{E}_s}\left|\hat{H}_m(n)\right|^2}{1+\sigma^2_{\nu(n)}} s(n) - \sqrt{\mathcal{E}_s}\hat{H}^*_m(n)s(n)\tilde{\nu}_m(n) \\ + \hat{H}^*_m(n)w_m(n). \end{aligned} \tag{8}$$

Since $w_m(n)$, $\tilde{\nu}_m(n)$, are uncorrelated, $\forall m$, the instantaneous SNR of the MRC output $z(n) := \sum_{m=1}^{M} z_m(n)$ can be found from (8) as

$$\gamma(n) = \frac{\mathcal{E}_s \sum_{m=1}^{M}\left|\hat{H}_m(n)\right|^2}{\left(1+\sigma^2_{\nu(n)}\right)\left[N_0\left(1+\sigma^2_{\nu(n)}\right) + \mathcal{E}_s\sigma^2_{\nu(n)}\right]}. \tag{9}$$

Since $\{\hat{H}_m(n)\}_{m=1}^{M}$ are independent, identically, (Gaussian) distributed (i.i.d.), the average SER of $s(n)$, $n \in \mathcal{I}_s$, that we denote as $P_e(n)$, can be found in closed form using the SNR $\gamma(n)$ [25, eq. (21)], [26, eqs. (9.19), (5A.17)]. The overall SER is then given by

$$P_e = \frac{1}{N-P}\sum_{n\in\mathcal{I}_s} P_e(n). \tag{10}$$

To quantify the performance degradation caused by channel estimation errors, we define an SNR as

$$\tilde{\gamma}(n) := \frac{\mathcal{E}_s \sum_{m=1}^{M}|H_m(n)|^2}{N_0\left(1+\sigma^2_{\nu(n)}\right) + \mathcal{E}_s\sigma^2_{\nu(n)}}. \tag{11}$$

Since $\sigma^2_{\hat{H}_m(n)} = 1 + \sigma^2_{\nu(n)}$, and $\sigma^2_{H_m(n)} = 1$, $\tilde{\gamma}(n)$ in (11) is equivalent to $\gamma(n)$ in (9) in the sense that the average SER calculated from $\tilde{\gamma}(n)$ is equal to that calculated from $\gamma(n)$. If $N\mathcal{E}$ denotes the total transmitted power per block, then $N\mathcal{E} = (N-P)\mathcal{E}_s + P\mathcal{E}_p$. Accounting for pilot power, the average power per information symbol is $\tilde{\mathcal{E}}_s = N\mathcal{E}/(N-P)$, and (11) can be written as $\tilde{\gamma}(n) = G_L(n)\tilde{\mathcal{E}}_s\sum_{m=1}^{M}|H_m(n)|^2/N_0$, where

$$G_L(n) = \frac{\mathcal{E}_s N_0}{\tilde{\mathcal{E}}_s\left[N_0\left(1+\sigma^2_{\nu(n)}\right) + \mathcal{E}_s\sigma^2_{\nu(n)}\right]}. \tag{12}$$

$G_L(n)$ in (12) quantifies the performance degradation caused by channel estimation errors, and by the power reduction needed for channel estimation. Substituting $\sigma^2_{\nu(n)}$ into (12), we have

$$G_L(n) = \frac{\mathcal{E}_s}{\tilde{\mathcal{E}}_s\left[1 + (N_0/\mathcal{E}_p + \mathcal{E}_s/\mathcal{E}_p)a_n\right]} \tag{13}$$

where a_n is defined in (6). In the ideal case where no pilot symbols are transmitted, and the receiver has perfect CSI, the transmitted power per symbol is $\mathcal{E}$, and the SNR at the MRC output is $\mathcal{E}\sum_{m=1}^{M}|H_m(n)|^2/N_0$. Compared to this ideal case, the performance degradation is

$$G_{L,I}(n) = \frac{\mathcal{E}_s}{\mathcal{E}\left[1 + (N_0/\mathcal{E}_p + \mathcal{E}_s\mathcal{E}_p)a_n\right]}. \tag{14}$$

While $G_L(n)$ in (13) reflects the performance degradation caused by channel estimation errors, and accounts for the power reduction allocated to pilots, $G_{L,I}(n)$ in (14) captures the performance loss only due to channel estimation errors. Since $\tilde{\mathcal{E}}_s > \mathcal{E}_s$, we see from (13) that $G_L(n) < 1$, which implies that there is always performance loss. On the other hand, it may be interesting to compare the SER performance of pilot symbol assisted channel estimation with that of the ideal case. If equal power is allocated to pilot and information symbols ($\mathcal{E}_s = \mathcal{E}_p = \mathcal{E}$), then we can increase P to decrease the variance of channel estimation error $\sigma^2_{\nu(n)} = a_n N_0/\mathcal{E}_p$, and thereby increase $G_{L,I}(n)$. However, with this equal power allocation, we see from (14) that $G_{L,I}(n) < 1$. If on the other hand, power is optimally distributed between pilots and information symbols, it will be shown later that $G_{L,I}(n)$ can be greater than one, which implies that performance may improve relative to the ideal case. Because $G_L(n)$ depends on this power allocation, but also on the number and placement of pilot symbols, we will optimize these parameters in Section III to maximize $G_L(n)$, and thus minimize SER.

C. SER With MMSE Channel Estimation

The LSE channel estimator does not depend on the fading channel's power delay profile. If this knowledge is available, we can use the MMSE channel estimator to further improve SER performance. From (3), the covariance matrix of $\mathbf{y}_m$ is given by $\mathbf{R}_{yy} := E[\mathbf{y}_m\mathbf{y}^{\mathcal{H}}_m] = \mathcal{E}_p\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{F}^{\mathcal{H}}_p\mathbf{R}_{hh}\mathbf{F}_p\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{s}_p) + N_0\mathbf{I}_P$, where $\mathbf{R}_{hh} := E[\mathbf{h}_m\mathbf{h}^{\mathcal{H}}_m] = \mathrm{diag}(\sigma^2_h(0), \ldots, \sigma^2_h(L-1))$. The cross-correlation between $\mathbf{y}_m$ and $\mathbf{h}_m$ is $\mathbf{R}_{yh} := E[\mathbf{y}_m\mathbf{h}^{\mathcal{H}}_m] = \sqrt{\mathcal{E}_p}\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{F}^{\mathcal{H}}_p\mathbf{R}_{hh}$. Then, the MMSE estimator of $\mathbf{h}_m$ is given by[1] $\hat{\mathbf{h}}_m = \mathbf{R}^{\mathcal{H}}_{yh}\mathbf{R}^{-1}_{yy}\mathbf{y}_m$ [24, p. 391]. The channel estimation error is given by $\boldsymbol{\varepsilon}_m = \mathbf{h}_m - \hat{\mathbf{h}}_m$, which is Gaussian distributed with zero-mean, and covariance [24, p. 391]

$$\mathbf{R}_\varepsilon := E\left[\varepsilon_m\varepsilon^{\mathcal{H}}_m\right] = \left(\mathbf{R}^{-1}_{hh} + \mathcal{E}_p\mathbf{F}_p\mathbf{F}^{\mathcal{H}}_p/N_0\right)^{-1} \tag{15}$$

[1]We will use the same notation for LSE and MMSE channel estimation, when there is no confusion.

where $\sigma_h^2(l) \neq 0, \forall l$, so that $\mathbf{R}_{hh}$ is invertible. When there are zero taps in $\mathbf{h}_m$, we can remove these taps from (3), to guarantee invertibility of $\mathbf{R}_{hh}$. The estimated channel frequency response on the nth subcarrier can be obtained as $\hat{H}_m(n) = \mathbf{f}_n^{\mathcal{H}}\hat{\mathbf{h}}_m = H_m(n) - \epsilon_m(n)$, where $\epsilon_m(n) := \mathbf{f}_n^{\mathcal{H}}\boldsymbol{\varepsilon}_m$ with $\epsilon_m(n) \sim \mathcal{CN}(0, \sigma_{\epsilon(n)}^2)$, and $\sigma_{\epsilon(n)}^2 := \mathbf{f}_n^{\mathcal{H}}\mathbf{R}_\varepsilon\mathbf{f}_n$. The estimator $\hat{H}_m(n)$ is Gaussian distributed with zero-mean. Since the orthogonality principle renders $\boldsymbol{\varepsilon}_m$ uncorrelated with $\hat{\mathbf{h}}_m$, $\epsilon_m(n)$ and $\hat{H}_m(n)$ are also uncorrelated. Thus, the variance of $\hat{H}_m(n)$ can be found as $\sigma_{\hat{H}_m(n)}^2 = \sigma_{H_m(n)}^2 - \sigma_{\epsilon(n)}^2 = 1 - \sigma_{\epsilon(n)}^2$.

The output of the mth MRC branch for $s(n), n \in \mathcal{I}_s$, can be written as

$$z_m(n) = \sqrt{\mathcal{E}_s}\left|\hat{H}_m(n)\right|^2 s(n) + \sqrt{\mathcal{E}_s}\left[\hat{H}_m(n)\right]^* s(n)\epsilon_m(n) + \left[\hat{H}_m(n)\right]^* w_m(n). \tag{16}$$

Using the fact that $\epsilon_m(n)$ and $\hat{H}_m(n)$ are uncorrelated, the instantaneous SNR at the MRC output can be found from (16) as

$$\gamma(n) = \frac{\mathcal{E}_s \sum_{m=1}^M \left|\hat{H}_m(n)\right|^2}{N_0 + \mathcal{E}_s\sigma_{\epsilon(n)}^2}. \tag{17}$$

Similar to (11), we define an SNR equivalent to $\gamma(n)$ in (17) as

$$\tilde{\gamma}(n) = \frac{\mathcal{E}_s \sum_{m=1}^M |H_m(n)|^2 \left(1 - \sigma_{\epsilon(n)}^2\right)}{N_0 + \mathcal{E}_s\sigma_{\epsilon(n)}^2}. \tag{18}$$

From the SNR in (18), and the independent and identical Gaussian distributions of $\{H_m(n)\}_{m=1}^M$, we can calculate the average SER in closed form [25, eq. (21)], [26, eqs. (9.19), (5A.17)]. Similar to (12), the performance degradation caused by MMSE channel estimation can be found from (18) as

$$G_M(n) = \frac{\mathcal{E}_s N_0 \left(1 - \sigma_{\epsilon(n)}^2\right)}{\tilde{\mathcal{E}}_s \left(N_0 + \mathcal{E}_s\sigma_{\epsilon(n)}^2\right)}. \tag{19}$$

Compared to the ideal case, the performance loss is given by

$$G_{M,I}(n) = \frac{\mathcal{E}_s N_0 \left(1 - \sigma_{\epsilon(n)}^2\right)}{\mathcal{E} \left(N_0 + \mathcal{E}_s\sigma_{\epsilon(n)}^2\right)}. \tag{20}$$

In the ensuing section, we will optimize pilot symbol parameters to maximize $G_M(n)$, and thus minimize the average SER.

III. Error Probability Minimizing Pilots

Based on the SER performance in Section II, we will optimize here the power allocation between pilot and information symbols, the number of pilots, and the placement of pilot symbols to minimize the average SER.

A. *Optimal Pilots for LSE Channel Estimation*

As we mentioned in Section II-B, the total transmitted power per block is $N\mathcal{E}$. If the power allocated to information symbols is $(N-P)\mathcal{E}_s = \alpha N\mathcal{E}$, where $\alpha \in (0, 1)$, then we have $\mathcal{E}_s = \alpha N\mathcal{E}/(N-P)$. The transmitted power allocated to pilot symbols is $P\mathcal{E}_p = (1-\alpha)N\mathcal{E}$; thus, the transmitted power per pilot symbol is $\mathcal{E}_p = (1-\alpha)N\mathcal{E}/P$. Note that $\alpha = (N-P)/N$ corresponds to having $\mathcal{E}_s = \mathcal{E}_p = \mathcal{E}$, which we refer to as equal power allocation. Since $\tilde{\mathcal{E}}_s = N\mathcal{E}/(N-P)$, then $\mathcal{E}_s/\tilde{\mathcal{E}}_s = \alpha$, and the performance degradation in (13) becomes

$$G_L(n) = \frac{\alpha}{1 + \left(\frac{P}{N(1-\alpha)\bar{\gamma}} + \frac{\alpha P}{(N-P)(1-\alpha)}\right) a_n} \tag{21}$$

where $\bar{\gamma} := \mathcal{E}/N_0$; and from (14), we have

$$G_{L,I}(n) = \frac{N G_L(n)}{N-P}. \tag{22}$$

If we fix the power allocation α and the number of pilot symbols P, it is seen from (21) that $G_L(n)$ is determined by a_n. Since a_n depends on the placement of pilot symbols [cf. (6)], we next find the optimal pilot locations to maximize $G_L(n)$.

But first, let us define the equi-spaced pilot symbols as follows:

Definition 1 (Equispaced Pilot Symbols): If $J = N/P$ is an integer, then we say that the pilot symbols are equispaced, if and only if the pilot subcarrier index set is $\mathcal{I}_p = \{j + Jp : p \in [0, P-1]\}$ for some $j \in [0, J-1]$.

For equispaced pilot symbols, it is easy to verify that $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}} = P\mathbf{I}_L$. Thus, $a_n = L/P, \forall n$, and $G_L(n), \forall n$ are identical. For arbitrarily located pilot symbols however, $G_L(n)$ may be different for different n. Since the average SER P_e in (10) is dominated by the largest $P_e(n)$, it is desirable to maximize the $\min_{n\in\mathcal{I}_s} G_L(n)$ in order to minimize the worst performance loss. It turns out that the equispaced pilot symbols are optimal among all possible pilot placements, which is precisely described by the following lemma (see Appendix for the proof)

Lemma 1: If $J = N/P$ is an integer, then for any given power allocation specified by α, the equispaced pilot symbols are optimal in the sense that the $\min_{n\in\mathcal{I}_s}(G_L(n))$ is maximized.

Because equispaced placement of pilot symbols is impossible when $J = N/P$ is not an integer, we will later develop a suboptimal pilot placement for this case, which will be shown to have almost identical performance to the equispaced one. Equispaced and equipowered pilot symbols were shown necessary and sufficient to minimize $E[\|\boldsymbol{\eta}_m\|^2]$ in [19]. However, minimizing $E[\|\boldsymbol{\eta}_m\|^2]$ may not lead to the minimum average SER, because average SER may be dominated by the SER of subcarriers with large estimation errors. Although our result in Lemma 1 is the same as that in [19], we obtain this result by directly minimizing the worst SER.

The number of pilot symbols P in (21) affects the performance loss. The following lemma, which is proved in the Appendix, characterizes the optimal number of pilot symbols.

Lemma 2: Suppose that N/L is an integer. If the equispaced pilot symbols are employed when $P = L$, then for any power allocation specified by α, $P = L$ is optimal in the sense that the minimum of $G_L(n), \forall n \in \mathcal{I}_s$, and $\forall P \in [L, N)$ is maximized.

When N/L is not an integer, $P = L$ may be no longer optimal. However, the suboptimal pilot placement scheme we will develop later has almost the same performance as its optimal equispaced counterpart. It was shown in [19] that $P = L$ equispaced pilot symbols minimize the channel MSE $E[\|\boldsymbol{\eta}_m\|^2]$ under a transmit power constraint. We here show that $P = L$

equispaced pilot symbols actually minimize the SER, and our proof is different from that in [19].

Having derived the optimal placement and number of pilot symbols in Lemmas 1 and 2, we next determine the optimal power allocation between pilot and information symbols. We first prove the following lemma in the Appendix:

Lemma 3: Given P and a_n, $G_L(n)$ in (21) has a unique maximum over $\alpha \in (0,\, 1)$. The maximum of $G_L(n)$ is achieved at $\alpha_{opt}(n) = 1/2$, if $N - P(1 + a_n) = 0$, and at

$$\alpha_{opt}(n) = \frac{\sqrt{(N-P)(N\bar{\gamma}+Pa_n)(N-P+N\bar{\gamma})Pa_n}}{N\bar{\gamma}(P+Pa_n-N)} - \frac{(N-P)(N\bar{\gamma}+Pa_n)}{N\bar{\gamma}(P+Pa_n-N)} \tag{23}$$

if $N - P(1 + a_n) \neq 0$.

Note that for equispaced pilot symbols, $\alpha_{opt}(n)$ are identical, $\forall n$. If pilot symbols are not equispaced however, $\alpha_{opt}(n)$ is generally different for different n. Since our goal is to maximize the $\min_{n\in\mathcal{I}_s}(G_L(n))$, the optimal power allocation is summarized in the following lemma:

Lemma 4: If pilot symbols are not equispaced, then the optimal power allocation is specified by $\alpha_{\rm opt} = \alpha_{\rm opt}(n_0)$, where $n_0 := \arg\max_{n\in\mathcal{I}_s}\, a_n$. If pilot symbols are equispaced, the optimal power allocation is given by (23) with $a_n = L/P, \forall n$.

This optimal power allocation is applicable to any P, and any placement of pilot symbols. Note that our SER minimizing power allocation here is different from the power allocation that minimizes the normalized symbol MSE in [19]. Combining Lemmas 1, 2, 3, and 4, we summarize our optimal training results in the following proposition:

Proposition 1: If N/L is an integer, then the SER minimizing pilots are specified by the following conditions: the number of pilot symbols is $P = L$; pilot symbols are equispaced; and the power allocation between pilot and information symbols is given by (23) with $a_n = L/P, \forall n$.

If N/L is not an integer, our numerical results in Section IV show that: setting $P = L$, using the suboptimal pilot symbol placement developed in Section III-C, and the power allocation given by Lemma 4, we can achieve almost the same performance as the optimal pilots specified by Proposition 1.

B. Optimal Pilots for MMSE Channel Estimation

Substituting the expression of $\mathcal{E}_s$ and $\bar{\mathcal{E}}_s$ into (19), we obtain

$$G_M(n) = \frac{\alpha\left(1-\sigma^2_{\epsilon(n)}\right)}{1+\alpha\bar{\gamma}\sigma^2_{\epsilon(n)}N/(N-P)} \tag{24}$$

and from (20), we have

$$G_{M,I}(n) = \frac{NG_M(n)}{N-P}. \tag{25}$$

Our objective is again to find the optimal pilot placement, number of pilots, and power allocation to minimize the worst performance loss, or equivalently, to maximize the minimum of $G_M(n)$. The following lemma, which is proved in the Appendix, shows that equispaced pilots also maximize the $\min_{n\in\mathcal{I}_s} G_M(n)$:

Lemma 5: If $J = N/P$ is an integer, then for any power allocation specified by α, the equispaced pilot symbols are optimal in the sense that the $\min_{n\in\mathcal{I}_s}(G_M(n))$ is maximized.

With MMSE channel estimation, it is shown in [15] and [16] that equispaced pilot symbols maximize a lower bound on ergodic capacity. Here, we prove that equispaced pilot symbols also minimize the worst SER.

If pilot symbols are equispaced, it is shown in the proof of Lemma 5 that $\sigma^2_{\epsilon(n)} = \sum_{i=0}^{L-1}\left(1/\sigma_i^2 + P\mathcal{E}_p/N_0\right)^{-1}, \forall n$. If pilot symbols are not equispaced, using $\mathbf{R}_\varepsilon$ in (15), we have

$$\begin{aligned}\sigma^2_{\epsilon(n)} &= \mathbf{f}_n^{\mathcal{H}}\left(\mathbf{R}_{hh}^{-1} + \mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathcal{E}_p/N_0\right)^{-1}\mathbf{f}_n\\ &= \mathbf{f}_n^{\mathcal{H}}\mathbf{R}_{hh}^{1/2}\left(\mathbf{I}_L + \mathbf{R}_{hh}^{1/2}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathbf{R}_{hh}^{1/2}\mathcal{E}_p/N_0\right)^{-1}\\ &\quad\times \mathbf{R}_{hh}^{1/2}\mathbf{f}_n\end{aligned} \tag{26}$$

where $\mathbf{R}_{hh}^{1/2} = \mathrm{diag}(\sigma_h(0), \ldots, \sigma_h(L-1))$. Performing singular value decomposition on $\mathbf{R}_{hh}^{1/2}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\,\mathbf{R}_{hh}^{1/2}$, we obtain $\mathbf{R}_{hh}^{1/2}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathbf{R}_{hh}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathcal{H}}$, where $\boldsymbol{\Lambda} := \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$ contains singular values, and the unitary matrix $\mathbf{U}$ consists of the corresponding singular vectors; and then (26) becomes

$$\sigma^2_{\epsilon(n)} = \sum_{i=1}^{L}\frac{|\mathbf{g}_n(i)|^2}{1+\lambda_i\mathcal{E}_p/N_0} = \sum_{i=1}^{L}\frac{|\mathbf{g}_n(i)|^2}{1+\lambda_i(1-\alpha)\bar{\gamma}N/P} \tag{27}$$

where the vector $\mathbf{g}_n := \mathbf{U}^{\mathcal{H}}\mathbf{R}_{hh}^{1/2}\mathbf{f}_n$. Substituting the expression for $\sigma^2_{\epsilon(n)}$ into (24), we can write $G_M(n)$ as a function of α and P. As it is difficult to find the optimal α and P to maximize $G_M(n)$ analytically, we resort to numerical search to find the maximum of $G_M(n)$. Specifically, letting $n_1 := \arg\max_{n\in\mathcal{I}_s}\sigma^2_{\epsilon(n)}$, $\forall P \in [L, N)$, we can use a one dimensional search, e.g., Golden section search [27, p. 397], to maximize $G_M(n_1)$ with respect to α; and searching over all values of P, we can find the optimal P and α. Our numerical results in Section IV will show that the optimal P is usually equal to L.

If the channel taps are i.i.d and the pilot symbols are equispaced, it is possible to find the optimal α and P in closed form. When the channel taps are i.i.d, we have $\sigma_i^2 = 1/L$, $\forall i \in [0, L-1]$; and using $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}} = P\mathbf{I}_L$ for equispaced pilot symbols, we obtain from (26) $\sigma^2_{\epsilon(n)} = L/(L+P\mathcal{E}_p/N_0)$. Then, $G_M(n)$ in (24) becomes

$$G_M^{(iid)}(n) = \frac{\alpha}{1+(N_0/\mathcal{E}_p+\mathcal{E}_s/\mathcal{E}_p)L/P}. \tag{28}$$

Comparing $G_M^{(iid)}(n)$ with $G_L(n)$ in (13) with $\mathcal{E}_s/\tilde{\mathcal{E}}_s = \alpha$, and $a_n = L/P$ for equispaced pilot symbols, we deduce that LSE and MMSE channel estimation incurs the same SER performance loss, even though the two channel estimators come with different estimation errors. We summarize this result in the following lemma:

Lemma 6: When the pilot symbols are equispaced, and the channel taps are i.i.d., MMSE and LSE channel estimators lead to identical SER performance.

The optimal power allocation in this case can be found from (23) by setting $a_n = L/P$.

C. Suboptimal Placement of Pilots

For LSE channel estimation, we show in Lemma 2 that when N/L is an integer, $P = L$ equispaced pilot symbols are optimal. However, it may not be possible to guarantee that N/L is an integer in practice, since the channel order may change depending on the operating environment. For a fixed power per pilot or information symbol, we can always increase P to reduce channel estimation error, and thereby increase $G_{L,I}(n)$ and $G_{M,I}(n)$ as seen from (14) and (20). This in turn improves SER performance relative to the ideal case, at the price of reducing transmission rate, which may be affordable in some scenarios. When we increase P to reach a desirable trade off between SER performance and transmission rate, it may be difficult to ensure that N/P is an integer. This motivates our suboptimal scheme for the placement of pilot symbols, when N/P is not an integer.

In our channel estimation, we first use a linear LSE or MMSE estimator to obtain an estimate of the channel impulse response, $\hat{\mathbf{h}}_m$, from pilot samples on different subcarriers; and then obtain an estimate of the channel frequency response on the nth subcarrier, $\hat{H}_m(n)$. Note that this two step channel estimator does not sacrifice optimality of linear LSE or MMSE estimation [9]. We can also view this channel estimator as a linear interpolator of pilot samples in the frequency domain. Thus, it is reasonable to place pilot symbols uniformly across subcarriers, which ensures that the estimated channel frequency response at each subcarrier has almost the same error. For this reason, we place P pilot symbols with two values of pilot spacings: S and $S+1$, when N/P is not an integer. This leads to the following two equations to be solved for S

$$P_1 + P_2 = P, \quad P_1 S + P_2(S+1) = N \tag{29}$$

where we have P_1 pilot spacings equal to S, and P_2 pilot spacings equal to $S+1$. Solving these two equations, we obtain $S = \lfloor N/P \rfloor$, and $P_2 = N - PS$, $P_1 = P - P_2$, where $S = \lfloor x \rfloor$ denotes the largest integer less than x. We then uniformly interleave these two pilot spacings. For example, supposing $N = 16$, $P = 6$, we have $S = 2$, $P_1 = 2$, $P_2 = 4$, which places pilots on subcarrier with indexes in $\mathcal{I}_p = \{1, 4, 7, 9, 12, 15\}$. It will be shown in Section IV that this suboptimal placement of pilot symbols has almost the same performance as the equispaced pilot symbols.

IV. Simulations and Numerical Results

We consider an OFDM system with $N = 2,048$ subcarriers corresponding to the $2k$-mode in terrestrial digital video broadcasting (DVB-T) [7]. The frequency selective channel has $L = 40$ zero-mean uncorrelated complex Gaussian random taps. We adopt an exponential power delay profile, with each tap having variance $\sigma_l^2 = \exp(-l/10)/\sum_{i=0}^{L-1}\sigma_i^2$, $l \in [0,\ L)$, as in [9].

Figs. 2 and 3 depict both simulated and analytical SER versus SNR per antenna for LSE and MMSE channel estimation, respectively. The number of pilot symbols is $P = 64$, QPSK constellation is adopted, and the SNR is defined as in the ideal case: $\text{SNR} = \mathcal{E}/N_0$. Simulation and analytical results match very well. The optimal power allocation between pilot and information symbols shows about 1 dB advantage relative to the equal power allocation. Based on the analytical results in Figs. 2 and 3, Fig. 4 compares the performance of LSE and MMSE estimators. We confirm that both have almost identical performance

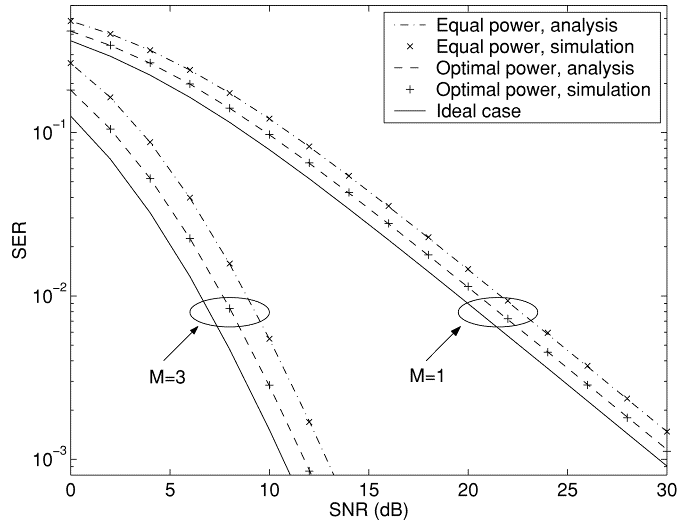


Fig. 2. SER versus SNR (LSE channel estimation).

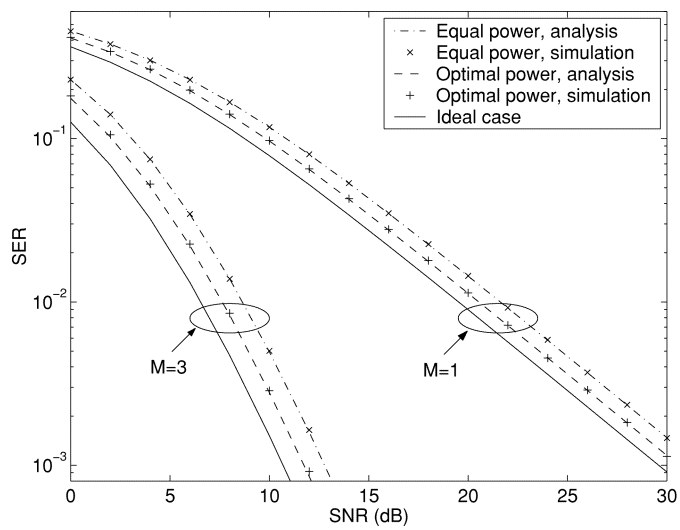


Fig. 3. SER versus SNR (MMSE channel estimation).

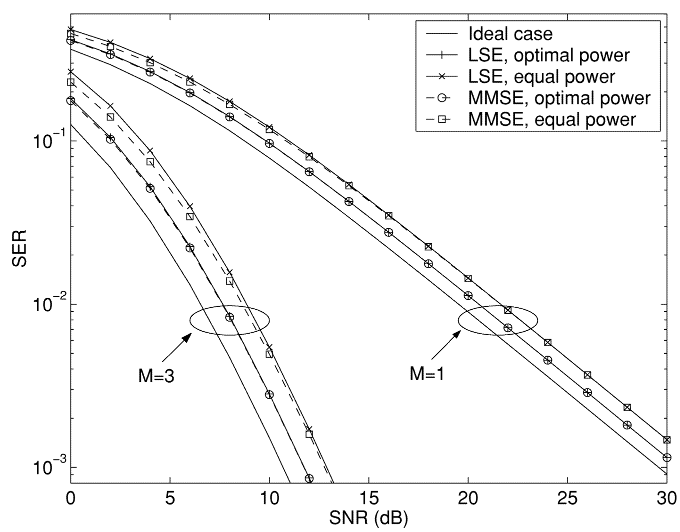


Fig. 4. SER comparison between LSE and MMSE channel estimation.

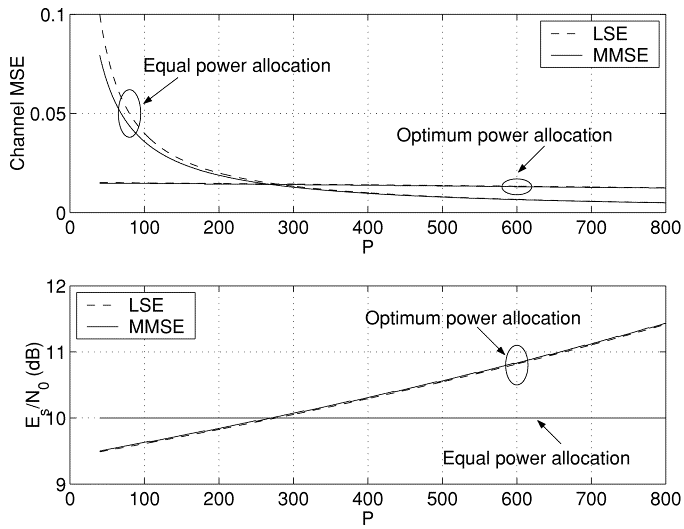


Fig. 5. Channel MSE and transmitted SNR per information symbol versus number of pilot symbols.

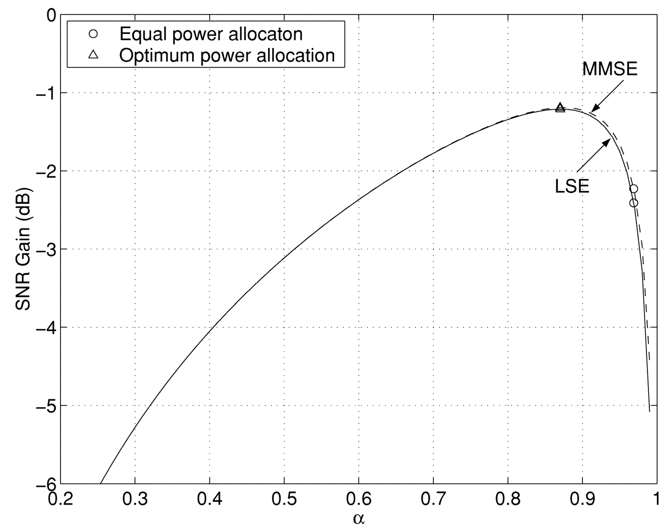


Fig. 6. Effect of power allocation on performance loss.

across the SNR region, when the transmitted power is optimally allocated. With equal power allocation, MMSE estimation has slightly better performance than LSE estimation in the low SNR region.

Fig. 5 shows how the number of pilot symbols affect the channel MSE and the transmitted SNR per information symbol. In deriving the channel MSE, we assumed equispaced pilot symbols, even when N/P is not an integer. Hence, the channel MSE here provides a lower bound when N/P is not an integer. In Fig. 5 and all remaining figures, we use $\mathrm{SNR} = 10$ dB. With optimal power allocation, we see that the channel MSE is almost constant; whereas, with equal power allocation, the channel MSE decreases with P, and is much larger than that of optimal power allocation when P is small. With equal power allocation, the transmitted SNR per information symbol is constant; on the other hand, with optimal power allocation, the transmitted power per information symbol increases with P, and thus SER performance improves. Fig. 6 describes the performance gain[2] under variable power allocation. We use $P = 64$, and equispaced pilot symbols. We see that both LSE and MMSE channel estimators exhibit a unique maximum at optimal power allocation. Figs. 7 and 8 depict the performance gain versus P for LSE and MMSE channel estimation, respectively. When N/P is not an integer, we use the suboptimal scheme of Section III-C to place pilot symbols, and plot $\min_{n\in\mathcal{I}_s} G_L(n)$, $\min_{n\in\mathcal{I}_s} G_{L,I}(n)$, $\min_{n\in\mathcal{I}_s} G_M(n)$, and $\min_{n\in\mathcal{I}_s} G_{M,I}(n)$. We also plot the performance gain for equispaced pilot symbols, which serves as an upper bound on the performance gain, if N/P is not an integer. We draw several conclusions from Figs. 7 and 8. First, the nonequispaced placement of pilots using the suboptimal scheme has almost the same performance as its optimal equispaced counterpart. Second, with optimal power allocation, the performance gain is maximized at $P = L$, and is about 2.3 dB larger than that of equal power allocation. Note that for LSE channel estimation, we proved in Lemma 2 that $P = L$ is optimal, when N/P is an integer. For MMSE channel estimation, $P = L$ is not provably optimal, but the numerical results in Fig. 8 illustrate that $P = L$ is also optimal. Third, the performance gain relative to the

[2]We here and later use performance *gain* in dB, with negative gain indicating performance loss.

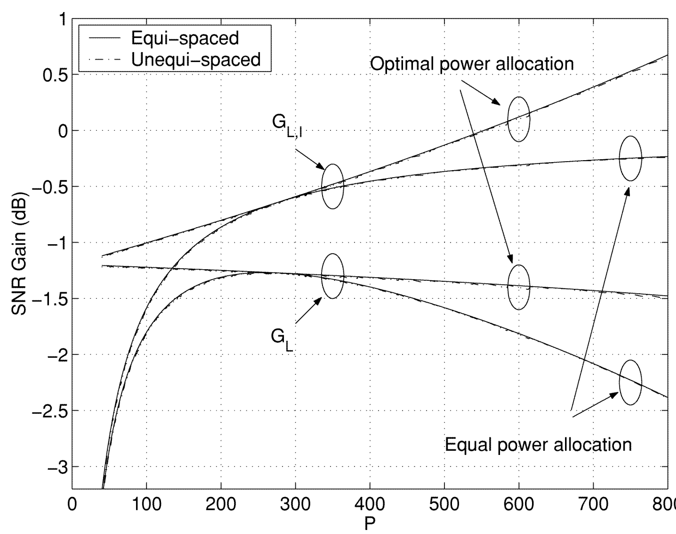


Fig. 7. Performance loss (LSE channel estimation).

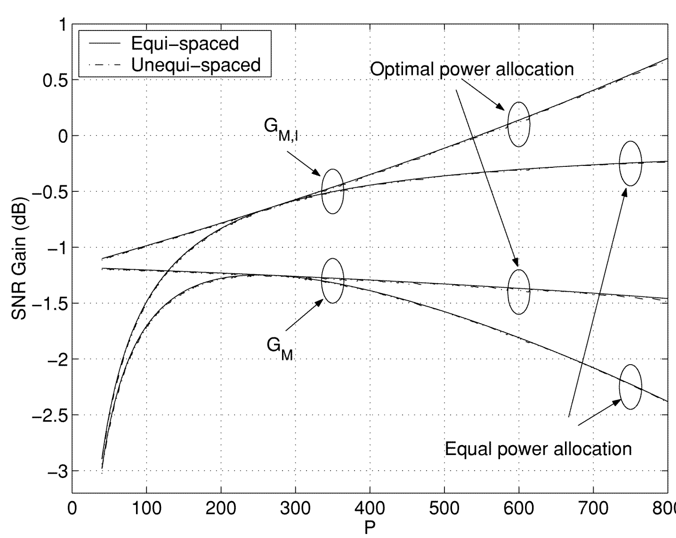


Fig. 8. Performance loss (MMSE channel estimation).

ideal case ($G_{L,I}(n)$ and $G_{M,I}(n)$) can be greater than one with optimal power allocation, which implies that the SER is smaller than that of the ideal case even in the presence of channel estimation errors. This is due to the increase in transmitted power per information symbol.

V. Conclusion

We have analyzed error probability performance of OFDM with M-PSK modulation over Rayleigh-fading channels, in the presence of channel estimation errors. We derived exact SER formulas, and quantified the performance loss due to channel estimation error and transmitted pilot power. Since the number and placement of pilot symbols, as well as the power allocation between the pilot and information symbols affect SER performance, we optimized these parameters for both LSE and MMSE channel estimation to minimize SER. The optimal pilot symbols result in about 2.3 dB performance gain relative to the pilot symbols with equal power allocation for a system with $N = 2048$ subcarriers, and a channel with $L = 40$ taps having an exponentially decaying power profile.[3]

Appendix
Proof of Lemma 1

From (21), we see that $G_L(n)$ is a monotonically decreasing function of a_n. Hence, to prove Lemma 1, it is sufficient to prove that $\max_n(a_n)$ is minimized. By the definition of a_n, we have

$$\begin{aligned}\sum_{n\in\mathcal{I}_s} a_n &= \sum_{n=1}^{N} \mathbf{f}_n^{\mathcal{H}} \left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{f}_n - \sum_{n\in\mathcal{I}_p} \mathbf{f}_n^{\mathcal{H}} \left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{f}_n \\ &= \mathrm{Tr}\left(\mathbf{F}^{\mathcal{H}}\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{F}\right) - \mathrm{Tr}\left(\mathbf{F}_p^{\mathcal{H}}\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{F}_p\right) \\ &= \mathrm{Tr}\left(\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{F}\mathbf{F}^{\mathcal{H}}\right) - \mathrm{Tr}\left(\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right) \\ &= N\mathrm{Tr}\left(\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\right) - L. \end{aligned} \tag{30}$$

At this point, we need the following lemma proved in [19]:

Lemma 7: For an $N \times N$ positive definite matrix $\mathbf{A}$ with (m, n)th entry $a_{m,n}$, it holds that

$$\mathrm{Tr}(\mathbf{A}^{-1}) \geq \sum_{n=1}^{N} \frac{1}{a_{n,n}} \tag{31}$$

where the equality is attained if and only if $\mathbf{A}$ is diagonal.

Since $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}$ is positive definite, and $[\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}]_{n,n} = P$, using (30) and Lemma 7, we have

$$\frac{1}{N-P}\sum_{n\in\mathcal{I}_s} a_n \geq \frac{NL/P - L}{N-P} = \frac{L}{P} \tag{32}$$

where the equality holds if and only if $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}$ is diagonal. If $a_n, \forall n \in \mathcal{I}_s$, are not identical, then there exits an n such that $a_n > L/P$. For equispaced pilot symbols, we have $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}} = P\mathbf{I}_L$, and $a_n = L/P, \forall n$. Hence, equispaced placement of pilot symbols minimizes the $\max_n(a_n)$. This completes the proof.□

[3]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.

VI. Proof of Lemma 2

To prove the lemma, we let $\beta := P/N$, assume equispaced pilot symbols for all possible values of P, and then show that $G_L(n)$ is a decreasing function of β. Since N/L is an integer, for $P = L$, equispaced placement of pilot symbols is feasible. For those values of P that N/P is not an integer, equispaced placement of pilot symbols is impossible; however, assuming equispaced pilot symbols yields an upper bound on the $\min_{n\in\mathcal{I}_s}(G_L(n))$. Thus, it is sufficient to prove the lemma when $G_L(n)$ is a decreasing function of β under the assumption of equispaced pilot symbols.

For equispaced pilots, we have $a_n = L/P, \forall n$. Then, $G_L(n)$ in (21) can be simplified to

$$G_L(n) = \frac{b(1-\beta)}{c+d-c\beta} \tag{33}$$

where $b := \alpha(1-\alpha)$, $c := 1 + L/(N\bar{\gamma}) - \alpha$, and $d := L\alpha/N$. Note that $G_L(n)$'s are equal, $\forall n$. Differentiating $G_L(n)$ with respect to β, we obtain

$$\frac{dG_L(n)}{d\beta} = -\frac{bd}{(c+d-c\beta)^2}. \tag{34}$$

Since $b > 0$ and $d > 0$, we have $d\, G_L(n)/d\beta < 0$, and thus, $G_L(n)$ is a decreasing function of β. □

VII. Proof of Lemma 3

From (21), we can also write $G_L(n)$ as

$$G_L(n) = \frac{\alpha(1-\alpha)}{A+B\alpha} \tag{35}$$

where $A := 1 + Pa_n/(N\bar{\gamma})$, $B := Pa_n/(N-P) - 1$. Since $\log(x)$ is a monotonically increasing function, the value of α that maximizes $\log(G_L(n))$ also maximizes $G_L(n)$. From (35), we have

$$\log(G_L(n)) = \log(\alpha) + \log(1-\alpha) - \log(A+B\alpha). \tag{36}$$

Taking derivative of $\log((G_L(n))$ with respect to α, we obtain

$$\frac{d\log(G_L(n))}{d\alpha} = \frac{1}{\alpha} - \frac{1}{1-\alpha} - \frac{B}{A+B\alpha}. \tag{37}$$

The second derivative of $\log(G_L(n))$ with respect to α can be obtained from (37) as

$$\frac{d^2\log(G_L(n))}{d\alpha^2} = -\frac{1}{\alpha^2} - \frac{1}{(1-\alpha)^2} + \frac{B^2}{(A+B\alpha)^2}. \tag{38}$$

If $B = 0$, it is clear from (37) that $d^2\log(G_L(n))/d\alpha^2 < 0$. If $B > 0$, then $-1/\alpha^2 + B^2/(A+B\alpha)^2 < 0$, since $A > 0$; thus, $d^2\log(G_L(n))/d\alpha^2 < 0$. If $B < 0$, from the definition of B, we have $0 < -B < 1$. Since $A > 1$, we obtain $-A/B > 1$. Hence, $d^2\log(G_L(n))/d\alpha^2 < -1/(1-\alpha)^2 + 1/(-A/B-\alpha)^2 < 0$. Therefore, $d^2\log(G_L(n))/d\alpha^2$ is less than zero $\forall\alpha \in (0,\ 1)$, which implies that $\log(G_L(n))$ is a concave function of α in the interval $\alpha \in (0,\ 1)$, and a unique maximum exists. Setting the first derivative of $\log(G_L(n))$ in (37) equal to zero, we obtain the optimal $\alpha \in (0, 1)$ at which $\log(G_L(n))$ is maximized

$$\alpha_{opt}(n) = \begin{cases} \frac{1}{2} & B = 0 \\ \frac{-A+\sqrt{A^2+AB}}{B} & B \neq 0. \end{cases} \quad (39)$$

Substituting A and B into (39), we obtain (23). □

VIII. Proof of Lemma 5

It is seen from (24) that $G_M(n)$ is a decreasing function of $\sigma^2_{\epsilon(n)}$; thus, maximizing $G_M(n)$ amounts to minimizing $\sigma^2_{\epsilon(n)}$. Hence, to prove the lemma, it is sufficient to prove that $\max_{n\in\mathcal{I}_s} \sigma^2_{\epsilon(n)}$ is minimized. If pilot symbols are not equispaced, similar to derivation of (30), we have

$$\begin{aligned}\sum_{n\in\mathcal{I}_s} \sigma^2_{\epsilon(n)} &= N\mathrm{Tr}(\mathbf{R}_\varepsilon) - \mathrm{Tr}\left(\mathbf{R}_\varepsilon \mathbf{F}_p \mathbf{F}_p^{\mathcal{H}}\right) \\ &= N\mathrm{Tr}\left[\left(\mathbf{R}_{hh}^{-1} + \mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathcal{E}_p/N_0\right)^{-1}\right] \\ &\quad - \mathrm{Tr}\left[\left(\left(\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\right)^{-1}\mathbf{R}_{hh}^{-1} + \mathcal{E}_p/N_0\mathbf{I}_L\right)^{-1}\right]. \end{aligned} \quad (40)$$

Defining $\mathbf{D} := ((\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}})^{-1}\mathbf{R}_{hh}^{-1} + \mathcal{E}_p/N_0\mathbf{I}_L)^{-1}$, and using the matrix inversion lemma [28, p. 19], we obtain

$$\mathbf{D} := N_0/\mathcal{E}_p\mathbf{I}_L - N_0/\mathcal{E}_p\left(\mathbf{R}_{hh}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathcal{E}_p/N_0 + \mathbf{I}_L\right)^{-1}. \quad (41)$$

Using Lemma 7, we have

$$\begin{aligned}&\mathrm{Tr}\left[\left(\mathbf{R}_{hh}\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathcal{E}_p/N_0 + \mathbf{I}_L\right)^{-1}\right] \\ &\qquad \geq \mathrm{Tr}\left[\left(\mathbf{R}_{hh}P\mathcal{E}_p/N_0 + \mathbf{I}_L\right)^{-1}\right]\end{aligned} \quad (42)$$

where equality holds if and only if $\mathbf{F}_p\mathbf{F}_p$ is diagonal. Combining (41) and (42), the second term in (40) becomes

$$\begin{aligned}\mathrm{Tr}(\mathbf{D}) &\leq N_o/\mathcal{E}_p \sum_{i=0}^{L-1}\left[1 - (1+\sigma_i P\mathcal{E}_p/N_0)^{-1}\right] \\ &= P\sum_{i=0}^{L-1}\left(\frac{1}{\sigma_i^2} + P\mathcal{E}_p/N_0\right)^{-1}.\end{aligned} \quad (43)$$

From Lemma 7, the first term in (40) becomes

$$N\mathrm{Tr}\left[\left(\mathbf{R}_{hh}^{-1} + \mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}\mathcal{E}_p/N_0\right)^{-1}\right] \geq N\sum_{i=0}^{L-1}\frac{1}{1/\sigma_i^2 + P\mathcal{E}_p/N_0} \quad (44)$$

where equality holds if and only if $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}$ is diagonal. Combining (40), (43), and (44), we obtain

$$\frac{1}{N-P}\sum_{n\in\mathcal{I}_s}\sigma^2_{\epsilon(n)} \geq \sum_{i=0}^{L-1}\left(1/\sigma_i^2 + P\mathcal{E}_p/N_0\right)^{-1} \quad (45)$$

where equality holds if and only if $\mathbf{F}_p\mathbf{F}_p^{\mathcal{H}}$ is diagonal. If $\sigma^2_{\epsilon(n)}$, $\forall n \in \mathcal{I}_s$, are not identical, then there exits an n such that $\sigma^2_{\epsilon(n)} > \sum_{i=0}^{L-1}(1/\sigma_i^2 + P\mathcal{E}_p/N_0)^{-1}$. If pilot symbols are equispaced, from (15), we have $\mathbf{R}_\varepsilon = (\mathbf{R}_{hh}^{-1} + P\mathcal{E}_p/N_0\mathbf{I}_L)^{-1}$, and then, $\sigma^2_{\epsilon(n)} = \mathrm{Tr}(\mathbf{R}_\varepsilon) = \sum_{i=0}^{L-1}(1/\sigma_i^2 + P\mathcal{E}_p/N_0)^{-1}$, $\forall n$. Hence, equispaced placement of pilot symbols minimizes $\max_{n\in\mathcal{I}_s}\sigma^2_{\epsilon(n)}$. This completes the proof. □

References

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 4–14, May 1990.

[2] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 47, no. 3, pp. 29–48, May 2000.

[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] P. Höeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by wiener filtering," in *Proc. Int. Conf. Acoust., Speech and Signal Processing*, Munich, Germany, Apr. 1997, pp. 1845–1848.

[5] Y. G. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207–1215, July 2000.

[6] Y. G. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902–914, July 1998.

[7] F. Sanzi and J. Speidel, "An adaptive two-dimensional channel estimation for wireless OFDM with application to mobile dvb-t," *IEEE Trans. Broadcasting*, vol. 46, pp. 128–133, June 2000.

[8] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, July 1998.

[9] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, pp. 3065–3073, Dec. 2001.

[10] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcasting*, vol. 48, pp. 223–229, Sept. 2002.

[11] M. Hsieh and C. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 217–225, Feb. 1998.

[12] B. Yang, K. B. Letaief, R. S. Chen, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, pp. 467–478, Mar. 2001.

[13] M.-X. Chang and Y. T. Su, "Performance analysis of equalized OFDM system in Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 721–732, Oct. 2002.

[14] H. Cheon and D. Hong, "Effect of channel estimation error in OFDM-based WLAN," *IEEE Commun. Lett.*, vol. 6, pp. 190–192, May 2002.

[15] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1338–2353, Aug. 2002.

[16] S. Ohno and G. B. Giannakis, "Capacity maximizing pilots and precoders for wireless OFDM over rapidly fading channels," in *Proc. Int. Symp. Signals, Syst. Electron.*, Tokyo, Japan, July 24–27, 2001, pp. 246–249.

[17] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1615–1624, June 2003.

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

[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, pp. 2113–2123, Dec. 2002.

[20] M. Dong, S. Adireddy, and L. Tong, "Optimal pilot placement for semi-blind channel tracking of packetized transmission over time-varying channels," *IEICE Trans. Fundamentals of Electron., Commun. Comput. Sci.*, vol. E86-A, no. 3, pp. 550–563, Mar. 2003.

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

[22] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 712–720, Oct. 2002.

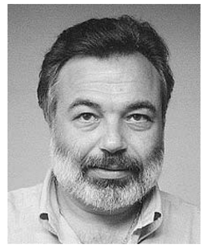
[23] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 3055–3069, Dec. 2002.

[24] S. T. Kay, *Fundamentals of Statistical Signal Processing, Volum I: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[25] S. Chennakeshu and J. B. Anderson, "Error rates for Rayleigh fading multi-channel reception of MPSK signals," *IEEE Trans. Commun.*, vol. 43, pp. 338–346, Feb./Mar./Apr. 1995.
[26] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2000.
[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge: Cambridge Univ., 1995.
[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ., 1985.

Xiaodong Cai (M'01) received the B.S. degree from Zhejiang University, China, the M.Eng. degree from the National University of Singapore, Singapore, and the Ph.D. degree from the New Jersey Institute of Technology, Newark, in 2001, all in electrical engineering.

From February 2001 to June 2001, he was a Member of Technical Staff with Lucent Technologies, NJ, working on WCDMA project. From July 2001 to October 2001, he was a Senior System Engineer with the Sony Technology Center, San Diego, CA, where he was involved in developing high data rate wireless modem. Since November 2001, he has been a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, University of Minnesota. His research interests include communication theory, signal processing, and wireless networks.



Georgios B. Giannakis (F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1981. He received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering, also in 1986, from the University of Southern California (USC).

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 at the University of Minnesota, 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 175 journal papers, 325 conference papers, and two edited books. Current research focuses on 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 four Best Paper Awards from the IEEE Signal Processing (SP) Society (1992, 1998, 2000, 2001). He also received the Society's Technical Achievement Award in 2000. He served as Editor-in-Chief of 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 is currently a member of the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and chairs the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.