

Semiblind Channel Estimation and Data Detection for OFDM Systems With Optimal Pilot Design

Tao Cui, *Student Member, IEEE*, and Chintha Tellambura, *Senior Member, IEEE*

Abstract—This paper considers semiblind channel estimation and data detection for orthogonal frequency-division multiplexing (OFDM) over frequency-selective fading channels. We show that the samples of an OFDM symbol are jointly complex Gaussian distributed, where the mean and covariance are determined by the locations and values of fixed pilot symbols. We exploit this distribution to derive a novel maximum-likelihood (ML) semiblind gradient-descent channel estimator. By exploiting the channel impulse response (CIR) statistics, we also derive a semiblind data detector for both Rayleigh and Ricean fading channels. Furthermore, we develop an enhanced data detector, which uses the estimator error statistics to mitigate the effect of channel estimation errors. Efficient implementation of both the semiblind and the improved data detectors is provided via sphere decoding and nulling-canceling detection. We also derive the Cramér-Rao bound (CRB) and design optimal pilots by minimizing the CRB. Our proposed channel estimator and data detector exhibit high bandwidth efficiency (requiring only a few pilot symbols), achieve the CRB, and also nearly reach the performance of an ideal reference receiver.

Index Terms—Channel estimation, data detection, orthogonal frequency-division multiplexing (OFDM).

I. INTRODUCTION

BOTH 4th generation (4G) and beyond 3rd generation (B3G) cellular networks are characterized by the ability to provide high data rates over wireless links. Multiple antennas and orthogonal frequency-division multiplexing (OFDM) are emerging as key technologies for 4G and B3G. It is increasingly believed that OFDM results in an improved downlink performance for 4G [1]. It has high spectral efficiency, robustness to frequency selective fading, and permits one-tap equalization. In practical OFDM systems, pilot-aided channel estimation is widely used to facilitate coherent detection [2]–[4]. Since the number of pilots must be greater than the number of channel taps, the use of cyclic prefix (CP) and pilot symbols entails a significant bandwidth loss, motivating blind methods.

Paper approved by N. Al-Dhahir, the Editor for Space-Time, OFDM, and Equalization of the IEEE Communications Society. Manuscript received April 14, 2006; revised July 17, 2006 and August 17, 2006. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, in part by the Informatics Circle of Research Excellence, and in part by the Alberta Ingenuity Fund. This paper was presented in part at the IEEE International Conference on Communications, Seoul, Korea, May 2005.

T. Cui was with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada. He is now with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: taocui@caltech.edu).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: chintha@ece.ualberta.ca).

Digital Object Identifier 10.1109/TCOMM.2007.895985

For example, in [5], blind channel estimators (BCEs) that exploit the CP are developed. A subspace BCE using virtual carriers is derived in [6]. Such blind estimators need averaging over many OFDM symbols, and hence may be classified as statistical. In mobile wireless environments, where the channel impulse response (CIR) can vary rapidly, statistical techniques may not be effective. On the other hand, deterministic blind techniques eliminate averaging and instead optimize certain metrics. Recently, we have developed several deterministic¹ channel estimators and data detectors [7] and [8]. They require minimizing complex, integer quadratic forms. In [7] and [8], we provide an efficient solution of such problems by developing sphere decoding (SD) and vertical Bell Labs layered space-time (V-BLAST)-based detectors. Recently, another deterministic blind joint channel and data estimator has been proposed in [9], which applies the branch-and-bound principle to solve a nonlinear integer problem associated with finding the least squares (LS) curve that fits a subchannel. However, the resulting blind detector is not optimal and needs several OFDM symbols.

In this paper, unlike [7]–[9], we focus on deterministic semiblind channel estimation and data detection. Conventional pilot-aided channel estimation uses many pilot symbols, and even some OFDM symbols may be reserved for all pilot symbols only. Such designs satisfy the constraint that the number of pilot symbols should exceed the number of channel taps. In contrast, our proposed semiblind channel estimators and data detector perform even when the number of available pilots is smaller than the number of channel taps. In order to derive these estimators, we first provide an extension of the conventional central limit theorem (CLT) arguments as applied to the OFDM signal. Namely, we show that the samples of an OFDM symbol have a joint complex Gaussian distribution, whose mean and covariance are determined by the locations and values of fixed pilot symbols. We exploit this distribution to derive a novel maximum-likelihood (ML) semiblind gradient-descent channel estimator. This channel estimator uses both the pilot and the data statistics, and turns out to be a generalization of the conventional LS channel estimator. Just like the conventional LS estimators, our channel estimator admits the use of the gradient descent algorithms. We next exploit the CIR statistics to derive a semiblind data detector. This detector can further be enhanced by noting that for a given LS channel estimate, the true CIR can be modeled as a complex Gaussian whose mean is the LS channel estimate itself. Thus, averaging the likelihood

¹Semiblind in [7] and [8] emphasizes that an algorithm uses both channel statistics and noise variance. In this paper, semiblind means that an algorithm uses one or more pilots.

function over the prior distribution gives an enhanced detector that mitigates the effect of channel estimation errors. Both our data detectors result in discrete integer quadratic optimization problems. They can be solved optimally by SD [10] and sub-optimally by V-BLAST detection [11] (also known as nulling and canceling), thereby avoiding computationally prohibitive exhaustive search. We also derive the Cramér-Rao bound (CRB) for semiblind channel estimation and design the optimal pilot distribution. We find that the well-known equipower and equispace conditions [12], [13] remain valid for minimizing the CRB. As well, a stronger condition requires that not only the pilot symbols are equipowered but also the data symbols. Our proposed channel estimators work even when the number of pilot symbols is smaller than the number of channel taps, yielding higher bandwidth efficiency than conventional channel estimators.

This paper is organized as follows. Section II reviews the basic baseband OFDM system model. Section III introduces the semiblind channel estimator. In Section IV, we derive the semiblind data detector and the enhanced detector taking into consideration the channel estimation error. Section V discusses mismatch issues and optimal pilot design. Section VI discusses simulation results and Section VII provides the conclusions. Appendices A and B prove the joint Gaussian nature of the time-domain samples of an OFDM symbol and the CRB for semiblind channel estimation.

Notation

Bold symbols denote matrices or vectors. Operators $\text{Re}(\cdot)$, $\text{Im}(\cdot)$, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $\text{tr}(\cdot)$ denote real part, imaginary part, conjugate, transpose, conjugate transpose, and trace, respectively. The space of K -dimensional complex vectors is \mathcal{C}^K . A circularly complex Gaussian random variable (CGRV) of mean μ and variance σ^2 is denoted by $z \sim \mathcal{CN}(\mu, \sigma^2)$. More generally, if $\mathbf{x} \in \mathcal{C}^n$ is a vector of CGRVs with mean vector μ and covariance matrix \mathbf{R} , $\mathbf{x} \sim \mathcal{CN}(\mu, \mathbf{R})$, then the probability density function (pdf) of \mathbf{x} is given by $p(\mathbf{x}) = (1/\pi^n \det(\mathbf{R})) \exp\{-\mathbf{x} - \mu)^H \mathbf{R}^{-1} (\mathbf{x} - \mu)\}$. The $N \times N$ discrete Fourier transform (DFT) matrix is given by $\mathbf{F} = 1/\sqrt{N} [e^{-j(2\pi/N)kl}]$, $k, l \in 0, 1, \dots, N-1$ (where $j = \sqrt{-1}$). A diagonal matrix formed by vector \mathbf{A} is \mathbf{A}_D . We consider the unitary M -ary phase-shift keying (MPSK) constellation $\mathcal{Q} = \{\zeta^0, \zeta, \dots, \zeta^{M-1} | \zeta = e^{j2\pi/M}\}$ only, although some of our results can be extended approximately to nonunitary constellations.

II. OFDM BASEBAND MODEL

In an OFDM system, the source data are grouped and/or mapped into the symbols from a constellation \mathcal{Q} , which are modulated by inverse DFT (IDFT) on N parallel subcarriers. The resulting time domain samples are

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j(2\pi kn/N)}, \quad n = 0, 1, \dots, N-1 \quad (1)$$

where

$$X_k = \begin{cases} d_k, & k \in I_d \\ p_k, & k \in I_p \end{cases} \quad (2)$$

where I_d is the index set of data subcarriers with N_d elements, I_p is the index set of subcarriers reserved for pilot symbols (pilots for brevity) with N_p elements, and $N_d + N_p = N$. Note that X_k , $k = 0, 1, \dots, N-1$ are called OFDM input symbols. We assume that all the d_k 's have the same power $E\{d_k^2\} = E_s$ and all the p_k 's have the same power $E\{p_k^2\} = E_p$. The IDFT output samples are x_n , $n = 0, \dots, N-1$. The term "OFDM symbol" denotes the entire IDFT output $\{x_0, x_1, \dots, x_{N-1}\}$. The input symbol duration is T_s , and the OFDM symbol duration is NT_s . These samples are appropriately pulse-shaped to construct the time-domain signal $x(t)$ for transmission. Typically, pilots X_k , $k \in I_p$, known *a priori* at the receiver, remain fixed from one OFDM symbol to the next. In this pilot arrangement, $N_p \ll N$. Alternatively, entire OFDM symbols of pilots ($N_p = N$) can be transmitted periodically. This paper focuses on the former, as it is more common in applications.

We assume that the composite CIR including transmit and receive pulse-shaping and the physical channel response between the transmitter and receiver is denoted as $h(\tau)$. After sampling and assuming no timing error, we can obtain $h_l = h(lT_s)$ where $h_l \sim \mathcal{CN}(0, \sigma_l^2)$, and $\sigma_l^2 = E\{h_l^2\}$. Typically, it is assumed that h_l is nonzero from $l = 0$ to $l = L-1$, and this results in a finite impulse response filter with an effective length L . The power sum $\sum_{l=0}^{L-1} \sigma_l^2$ is typically normalized to unity.

Assuming that the channel remains constant per OFDM symbol but may vary between OFDM symbols, and that the CP is sufficiently long ($N_g \geq L$), the post-DFT received samples are given as

$$Y_k = H_k X_k + W_k, \quad 0 \leq k \leq N-1 \\ = \frac{1}{\sqrt{N}} \sum_{n=0}^N H_k x_n e^{-j2\pi kn/N} + W_k \quad (3)$$

where $H_k = H(j2\pi k/N)$ is the complex channel frequency response at subcarrier k , $H(j\omega)$ is the Fourier transform of the CIR, and W_k , $k = 0, 1, \dots, N-1$ are independently and identically distributed (i.i.d.) CGRVs with zero mean and variance σ_n^2 . Assuming $\tau_l = lT_s$, we find $\mathbf{H} = \mathbf{F}_L \mathbf{h}$, where $\mathbf{H} = [H_0, H_1, \dots, H_{N-1}]^T$, $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]^T \in \mathcal{C}^L$ is the CIR, and \mathbf{F}_L is a $N \times L$ submatrix of the DFT matrix \mathbf{F} . We vectorize (3) as

$$\mathbf{Y} = \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} \quad (4)$$

or equivalently as

$$\mathbf{Y} = \mathbf{H}_D \mathbf{F} \mathbf{x} + \mathbf{W} \quad (5)$$

where $\mathbf{X}_D = \text{diag}\{X_0, \dots, X_{N-1}\}$ and $\mathbf{H}_D = \text{diag}\{H_0, \dots, H_{N-1}\}$ are diagonal matrices, $\mathbf{x} = [x_0, \dots, x_{N-1}]^T$, and $\mathbf{Y} = [Y_0, \dots, Y_{N-1}]^T$. The output signal-to-noise ratio (SNR) at the data subcarriers can be obtained as

$$\text{SNR}_s = \frac{E_s \sum_{l=0}^{L-1} \sigma_l^2}{\sigma_n^2} = \frac{E_s}{\sigma_n^2} \quad (6)$$

²The equal-power assumption will be justified in Section V-B by minimizing the CRB.

III. SEMIBLIND CHANNEL ESTIMATION

The received symbol vector \mathbf{Y} (5) is Gaussian with mean $\mathbf{H}_D \mathbf{F} \mathbf{x}$ and autocovariance matrix $\sigma_n^2 \mathbf{I}_N$. The likelihood function conditional on the unknown parameters \mathbf{x} and \mathbf{H}_D is given by

$$f(\mathbf{Y}|\mathbf{x}, \mathbf{H}_D) = \frac{1}{(\pi\sigma_n^2)^N} \exp \left\{ -\frac{1}{\sigma_n^2} \|\mathbf{Y} - \mathbf{H}_D \mathbf{F} \mathbf{x}\|^2 \right\}. \quad (7)$$

Given random data $X_k (k \in I_d)$, each transmitted-signal sample x_n (1) can be modeled as a univariate complex Gaussian via the CLT when N is large [14]. Here, we make a much stronger claim that all the samples x_0, \dots, x_{N-1} are jointly Gaussian. The presence of pilots (since their locations and values are fixed—not random) determines the mean and autocorrelation matrix of this process. Thus, we hypothesize that $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \mathbf{R}_x)$, where the mean and the covariance matrix are given by

$$\begin{aligned} \bar{\mathbf{x}} &= E\{\mathbf{x}\} = \mathbf{F}^H \mathbf{X}_p, \\ \mathbf{R}_x &= E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H\} = \mathbf{F}^H \mathbf{\Lambda}_d \mathbf{F} \end{aligned} \quad (8)$$

where

$$[\mathbf{X}_p]_k = \begin{cases} p_k, & k \in I_p \\ 0, & k \notin I_p \end{cases} \quad [\mathbf{\Lambda}_d]_{k,k} = \begin{cases} E_s, & k \in I_d \\ 0, & k \notin I_d. \end{cases} \quad (9)$$

That each sample x_n is univariate complex Gaussian does not necessarily imply that \mathbf{x} is multivariate complex Gaussian (for a proof and evidence, see Appendix A).

Averaging $f(\mathbf{Y}|\mathbf{x}, \mathbf{H}_D)$ over the pdf of \mathbf{x} gives the marginal likelihood function $f(\mathbf{Y}|\mathbf{H}_D)$

$$f(\mathbf{Y}|\mathbf{H}_D) = \int f(\mathbf{Y}|\mathbf{x}, \mathbf{H}_D) p(\mathbf{x}) d\mathbf{x} \quad (10)$$

where $p(\mathbf{x})$ is the pdf of \mathbf{x} . If the characteristic function (CF) of the quadratic form $Q = \|\mathbf{Y} - \mathbf{H}_D \mathbf{F} \mathbf{x}\|^2$ is $\phi(s) = E_Q\{e^{-sQ}\}$, then $f(\mathbf{Y}|\mathbf{H}_D) = (1/(\pi\sigma_n^2)^N) \phi(1/\sigma_n^2)$. The CF of a complex quadratic of form $\mathbf{x}^H \mathbf{G} \mathbf{x}$, where $\mathbf{x} \sim \mathcal{CN}(\mu, \mathbf{R})$, is known [15, App. B]. Using [15, (B-3-20)], we can show that the marginal likelihood function (10) is given by (11)

$$\begin{aligned} f(\mathbf{Y}|\mathbf{H}_D) &= \frac{1}{\pi^N \det(\sigma_n^2 \mathbf{I}_N + \mathbf{\Lambda}_d \mathbf{H}_D^H \mathbf{H}_D)} \\ &\times \exp \left\{ -\left[\mathbf{X}_p^H \left(\mathbf{\Lambda}_d + \sigma_n^2 (\mathbf{H}_D^H \mathbf{H}_D)^{-1} \right)^{-1} \mathbf{X}_p \right. \right. \\ &\quad + \mathbf{Y}^H (\mathbf{H}_D^H \mathbf{H}_D \mathbf{\Lambda}_d + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{Y} \\ &\quad - 2\text{Re} \left(\mathbf{Y}^H \mathbf{H}_D \left(\mathbf{\Lambda}_d \mathbf{H}_D^H \mathbf{H}_D \right. \right. \\ &\quad \left. \left. + \sigma_n^2 \mathbf{I}_N \right)^{-1} \mathbf{X}_p \right) \left. \right\}. \end{aligned} \quad (11)$$

We note that $\det(\sigma_n^2 \mathbf{I}_N + \mathbf{\Lambda}_d \mathbf{H}_D^H \mathbf{H}_D) = \sigma_n^{2N_p} \prod_{k \in I_d} (\sigma_n^2 + E_s |H_k|^2)$. Taking the logarithm of (11) and ignoring constant

terms, we can show that maximizing the log-likelihood function $\Lambda(\mathbf{Y}|\mathbf{H}_D) = \ln f(\mathbf{Y}|\mathbf{H}_D)$ is equivalent to minimizing

$$\begin{aligned} g_h(\mathbf{h}) &= \sigma_n^2 \sum_{k \in I_d} \left(\frac{|Y_k|^2}{E_s |H_k|^2 + \sigma_n^2} + \ln(E_s |H_k|^2 + \sigma_n^2) \right) \\ &\quad + \sum_{k \in I_p} |Y_k - H_k X_k|^2. \end{aligned} \quad (12)$$

Consequently, we propose the new semiblind channel estimator as

$$\hat{\mathbf{h}}_{sb} = \arg \min_{\mathbf{h} \in \mathcal{C}^L} g_h(\mathbf{h}). \quad (13)$$

Before discussing the properties of the new estimator, we briefly identify a generalization. If the channel remains constant during $M > 1$ OFDM symbols and the same pilots are available in each symbol, a common situation in practice, the estimator (13) can exploit all received samples over the M periods, albeit introducing an overall delay of $M - 1$ symbols. So the choice of M would be constrained by the rate of the time variations or Doppler spread of the channel, and the delay tolerance of the OFDM system. Similar to (7)–(12), an M th-order estimator can be derived as

$$\begin{aligned} \hat{\mathbf{h}}_{msb} &= \arg \min_{\mathbf{h} \in \mathcal{C}^L} \sigma_n^2 \\ &\times \sum_{m=1}^M \sum_{k \in I_d} \left(\frac{|Y_k^m|^2}{E_s |H_k|^2 + \sigma_n^2} + \ln(E_s |H_k|^2 + \sigma_n^2) \right) \\ &\quad + \sum_{m=1}^M \sum_{k \in I_p} |Y_k^m - H_k X_k^m|^2 \end{aligned} \quad (14)$$

where X_k^m denotes the k th data symbol in the m th OFDM symbol.

Remarks:

- 1) It is instructive to consider the special cases of (12). If no pilot exists, the second sum in (12) vanishes and we get a blind estimator with both, say, \mathbf{h} and $\mathbf{h}e^{j\theta}$ possible solutions, where θ is an arbitrary phase shift. Thus one or more pilots are indeed needed to solve the phase ambiguity. When an OFDM symbol consists of pilots entirely, (12) reduces to the conventional LS estimator [2]. Since the first sum in (12) is additional information extracted from the data subcarriers, our estimator (13) is a generalization of the conventional LS estimator when $N_d > 0$.
- 2) The CIR \mathbf{h} can be estimated using (13) via the gradient descent and related algorithms, i.e., conjugate gradient and quasi-Newton methods [16]. The partial derivative of $g_h(\mathbf{h})$ with respect to \mathbf{h}^* is given by

$$\begin{aligned} \frac{\partial g_h(\mathbf{h})}{\partial \mathbf{h}^*} &= \sigma_n^2 \sum_{k \in I_d} \frac{E_s |H_k|^2 + \sigma_n^2 - |Y_k|^2}{(E_s |H_k|^2 + \sigma_n^2)^2} \mathbf{f}_k^H \mathbf{f}_k \mathbf{h} \\ &\quad + \sum_{k \in I_p} |X_k|^2 \mathbf{f}_k^H \mathbf{f}_k \mathbf{h} - X_k^* Y_k \mathbf{f}_k^H \end{aligned} \quad (15)$$

where \mathbf{f}_k is the k th row of matrix \mathbf{F}_L and $H_k = \sum_{l=0}^{L-1} h_l e^{-j2\pi lk/N} = \mathbf{f}_k \mathbf{h}$. However, gradient descent needs an initial channel estimate. If sufficient pilots are

available ($N_p \geq L$), the initial CIR can be obtained via the well-known minimum mean-square error (MMSE) estimator [2] as

$$\begin{aligned}\hat{\mathbf{h}} &= (\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p + \sigma_n^2 \mathbf{R}_h^{-1})^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p \\ &= (E_p \mathbf{F}_p^H \mathbf{F}_p + \sigma_n^2 \mathbf{R}_h^{-1})^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p\end{aligned}\quad (16)$$

when σ_n^2 and \mathbf{R}_h^{-1} are available, or the LS estimator as

$$\begin{aligned}\hat{\mathbf{h}} &= (\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p \\ &= (E_p \mathbf{F}_p^H \mathbf{F}_p)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p\end{aligned}\quad (17)$$

where $\mathbf{P}_D = \text{diag}\{p_1, \dots, p_{N_p}\}$, $\mathbf{Y}_p = \mathbf{Y}(I_p)$, $\mathbf{F}_p = \mathbf{F}_L(I_p, :)$ is the $N_p \times L$ submatrix of \mathbf{F} corresponding to the pilots, and $\mathbf{R}_h = E\{\mathbf{h}\mathbf{h}^H\}$ is the covariance matrix of \mathbf{h} . If $N_p < L$, the CIR can be estimated using singular value decomposition (SVD) and simultaneous diagonalization as in [17]. A simpler alternative is to run the gradient descent algorithm several times with random initial estimates, which can be generated using \mathbf{R}_h . The estimate minimizing (12) is chosen as the final estimate.

The gradient-based algorithms, although faster than alternatives such as simulated annealing [16], cannot guarantee the global minimum. Despite this drawback, our approach gains about 3–10 dB over conventional MMSE and LS channel estimators (Section VI) even with the suboptimal solution. Typically, a few iterations guarantee the convergence.

- 3) The function $g_h(\mathbf{h})$ (12) may have several local minima. However, the channel is identifiable only when (12) has a unique global minimum. Although we are unable to prove that (12) has a unique global minimum, since our channel estimator is derived in an ML sense, it is asymptotically unbiased [18, Th. 7.3] if the regularity conditions of the ML estimator (MLE) [18, Th. 7.3] are satisfied. It can be readily verified that our MLE satisfies the regularity conditions. However, when $N_p = 0$, it can be verified that the Fisher information matrix (FIM) $E\{\partial^2 g_h(\mathbf{h})/\partial^2 \mathbf{h}\}$ of our MLE is singular and not invertible. Therefore, from [18, Th. 7.3], it does not have asymptotically unbiasedness. In fact, in this case, our MLE has ambiguity. When $N_p > 1$, the FIM is invertible. From [18, Th. 7.3], we conclude that a global minimum of (12) is asymptotically unbiased regardless of the number of global minima, and all the global minima asymptotically converge to the true solution.
- 4) To further improve the channel estimate and mitigate the effect of local minima, the decision-direct (DD) technique [19] can be used with (13) as follows. We first use a solution of (13) to detect data symbols via one-tap equalization as

$$\hat{\mathbf{X}} = \frac{\mathbf{Y}}{(\mathbf{F}_L \hat{\mathbf{h}}_{sb})} \quad (18)$$

where the division in (18) denotes component-wise division of two vectors. Assuming $\hat{\mathbf{X}}$ is correct, we use the MMSE approach to get the DD channel estimate as

$$\hat{\mathbf{h}}_{dd} = \left(\mathbf{F}_L^H \hat{\mathbf{X}}_D^H \hat{\mathbf{X}}_D \mathbf{F}_L + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-1} \mathbf{F}_L^H \hat{\mathbf{X}}_D^H \mathbf{Y}. \quad (19)$$

Since the quality of (13) is high in the high-SNR regime, $\hat{\mathbf{X}}$ is almost error-free in high SNR. The DD channel estimator then achieves the CRB at high SNR given in Appendix C, but it performs worse than the initial semiblind channel estimate in low SNR due to incorrect feedback. DD also introduces an extra detection delay.

IV. SEMIBLIND DATA DETECTION

A. Semiblind Data Detector for Rayleigh and Ricean Channels

If we use (4) instead of (5), the received symbol vector \mathbf{Y} is also Gaussian but with mean $\mathbf{X}_D \mathbf{F}_L \mathbf{h}$ and covariance matrix $\sigma_n^2 \mathbf{I}_N$. The likelihood function for the unknown CIR \mathbf{h} and \mathbf{X}_D is given by

$$f(\mathbf{Y}|\mathbf{h}, \mathbf{X}_D) = \exp \left\{ -\frac{1}{\sigma_n^2} \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2 \right\}. \quad (20)$$

We assume here that \mathbf{h} is a complex Gaussian vector with zero mean and covariance matrix \mathbf{R}_h . The average of $f(\mathbf{Y}|\mathbf{h}, \mathbf{X}_D)$ over the distribution of \mathbf{h} gives the marginal likelihood function $f(\mathbf{Y}|\mathbf{X}_D)$, which again relates to the CF of the quadratic form $Q = \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2$. Similar to the derivation of (11), we find that

$$\begin{aligned}f(\mathbf{Y}|\mathbf{X}_D) &= \frac{1}{\pi^N \det(\sigma_n^2 \mathbf{I}_N + \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L)} \\ &\times \exp \left\{ -\mathbf{Y}^H (\mathbf{X}_D \mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{Y} \right\}.\end{aligned}\quad (21)$$

If X_k 's belong to a unitary constellation, e.g., M -PSK, then $\mathbf{X}_D^H \mathbf{X}_D = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix with

$$[\mathbf{\Lambda}]_{k,k} = \begin{cases} E_s, & k \in I_d \\ E_p, & k \in I_p. \end{cases} \quad (22)$$

Hence, $\det(\sigma_n^2 \mathbf{I}_N + \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L) = \det(\sigma_n^2 \mathbf{I}_N + \mathbf{F}_L^H \mathbf{\Lambda} \mathbf{F}_L)$ is independent of \mathbf{X}_D . Maximizing (21) is equivalent to minimizing

$$g_x(\mathbf{X}) = \mathbf{X}^T \mathbf{Y}_D^H \left(\mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H + \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^{-1} \mathbf{Y}_D \mathbf{X}^* \quad (23)$$

where $\mathbf{Y}_D = \text{diag}\{Y_0, \dots, Y_{N-1}\}$ and $\mathbf{X} = [X_0, \dots, X_{N-1}]^T$. The third equality comes from the commutativity of the multiplication of a vector and a diagonal matrix. The semiblind data detector is thus given by

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{Q}^{N_d}} g_x(\mathbf{X}) \quad (24)$$

where \mathcal{Q} denotes a finite constellation.

Once the data symbols \mathbf{X} are estimated using (24), the CIR can be estimated via an MMSE estimator as

$$\hat{\mathbf{h}} = \left(\mathbf{F}_L^H \mathbf{F}_L + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-1} \mathbf{F}_L^H \hat{\mathbf{X}}_D^H \mathbf{Y}. \quad (25)$$

When the channel remains constant within several OFDM symbols, a common occurrence in practice, the channel estimate (25) can be used to detect the data in subsequent symbols without repeatedly performing channel estimation. DD techniques can also be used to track channel variations [4].

When the channel is Ricean, $E\{\mathbf{h}\} = \bar{\mathbf{h}}$ and $E\{(\mathbf{h} - \bar{\mathbf{h}})(\mathbf{h} - \bar{\mathbf{h}})^H\} = \mathbf{R}_h$. Similar to the development (29)–(31), we obtain the semiblind detector for Ricean fading channels as

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{Q}^{Nd}} \left\| \left(\mathbf{F}_L \mathbf{R}_h \mathbf{F}_L^H + \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^{-1/2} \times \left(\mathbf{F}_L \bar{\mathbf{h}} - \mathbf{Y}_D \mathbf{X}^* \right) \right\|^2. \quad (26)$$

Remarks:

- 1) Equation (24) is different from (13) in that the search space is continuous space \mathcal{C}^L in (13) while it is discrete space \mathcal{Q}^N in (24). Since \mathcal{Q} is discrete, (24) is a discrete LS problem. It cannot be solved by the gradient descent algorithms used for (13). The detection algorithms for (24) will be given in Section IV-C.
- 2) Semiblind data detection (24) also incurs phase ambiguity, as both \mathbf{X} and $\mathbf{X}e^{j\theta}$ satisfy (24) if $e^{j\theta}$ belongs to the M -PSK constellation. Pilot symbols are thus needed, and the search space hence reduces from \mathcal{Q}^N to \mathcal{Q}^{Nd} .
- 3) It is important to realize that semiblind data detection (24) need not be performed every symbol. For example, if the channel remains constant for M symbols, the channel estimate $\hat{\mathbf{h}}$ obtained in the first symbol using (25) can be used to detect the data symbols in the remaining $M-1$ symbols. As before, DD techniques can also be applied to track the time-variation of the channel.
- 4) The semiblind data detector needs the knowledge of \mathbf{R}_h and σ_n^2 , which may not be known exactly. The resulting mismatch problem is studied in Section V.
- 5) Note that the semiblind data detector in [7], [8] has the same form as (24) but is restricted to Rayleigh fading channels. Another point is that we can extend (24) to nonunitary constellations. The readers can refer to [7], [8] for further details.
- 6) The solution to (26) is unique and no phase ambiguity exists. This may thus also be used as a totally blind estimator.

B. Enhanced Data Detector With Channel Estimation Error

When the LS channel estimator (17) is used with only a few pilots, the estimated CIR becomes

$$\hat{\mathbf{h}} = \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p \right)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p = \mathbf{h} + \tilde{\mathbf{h}} \quad (27)$$

where $\tilde{\mathbf{h}} = \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p \right)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{W}$ and

$$\mathbf{R}_{\tilde{\mathbf{h}}} = E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\} = \sigma_n^2 \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p \right)^{-1} = \frac{\sigma_n^2}{E_p} \left(\mathbf{F}_p^H \mathbf{F}_p \right)^{-1} \quad (28)$$

where E_p is the power of pilot symbols and we assume that all the pilots have the same energy. If $\hat{\mathbf{h}}$ is used to detect the data in the consecutive OFDM symbols with one-tap equalization, performance loss can be high. This motivates enhanced data detection given channel estimation errors.

Given the channel estimate $\hat{\mathbf{h}}$, the true CIR \mathbf{h} has an *a posteriori* Gaussian distribution

$$p(\mathbf{h}) = \frac{1}{\pi^L \det(\mathbf{R}_{\tilde{\mathbf{h}}})} \exp \left\{ -(\mathbf{h} - \hat{\mathbf{h}})^H \mathbf{R}_{\tilde{\mathbf{h}}}^{-1} (\mathbf{h} - \hat{\mathbf{h}}) \right\}. \quad (29)$$

We now average the $f(\mathbf{Y}|\mathbf{h}, \mathbf{X}_D)$ in (20) with respect to $p(\mathbf{h})$ in (29) resulting in the marginal likelihood function $f(\mathbf{Y}|\mathbf{X}_D)$, which again involves the CF of the quadratic form $Q = \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2$. Again, following the derivation of (11), we obtain

$$f(\mathbf{Y}|\mathbf{X}_D) = \frac{1}{\pi^N \det(\sigma_n^2 \mathbf{I}_N + \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}})} \times \exp \left\{ -\hat{\mathbf{h}}^H \left(\mathbf{R}_{\tilde{\mathbf{h}}} + \sigma_n \times \left(\mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L \right)^{-1} \hat{\mathbf{h}} - \mathbf{Y}^H \times \left(\mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}} \mathbf{F}_L^H \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_N \right)^{-1} \mathbf{Y} + 2\text{Re} \left[\mathbf{Y}^H \left(\mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}} \mathbf{F}_L^H \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_L \right)^{-1} \mathbf{X}_D \mathbf{F}_L \hat{\mathbf{h}} \right] \right\}. \quad (30)$$

If X_k 's are unitary, then $\mathbf{X}_D^H \mathbf{X}_D$ is given in (22). Hence, $\det(\sigma_n^2 \mathbf{I}_N + \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}})$ is independent of \mathbf{X}_D .

Ignoring the terms independent of \mathbf{X}_D , maximizing (30) is equivalent to minimizing (31)

$$g_x(\mathbf{X}) = \mathbf{Y}^H \left(\mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}} \mathbf{F}_L^H \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_N \right)^{-1} \mathbf{Y} - 2\text{Re} \times \left[\mathbf{Y}^H \left(\mathbf{X}_D \mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}} \mathbf{F}_L^H \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_L \right)^{-1} \mathbf{X}_D \mathbf{F}_L \hat{\mathbf{h}} \right] = \left\| \left(\mathbf{F}_L \mathbf{R}_{\tilde{\mathbf{h}}} \mathbf{F}_L^H + \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^{-1/2} \left(\mathbf{F}_L \hat{\mathbf{h}} - \mathbf{Y}_D \mathbf{X}^* \right) \right\|^2 \quad (31)$$

where $\mathbf{\Lambda}$ is defined in (22). Substituting (28) into (31), we have

$$g_x(\mathbf{X}) = \left\| \left(\frac{\sigma_n^2}{E_p} \mathbf{F}_L \left(\mathbf{F}_p^H \mathbf{F}_p \right)^{-1} \mathbf{F}_L^H + \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^{-1/2} \times \left(\mathbf{F}_L \hat{\mathbf{h}} - \mathbf{Y}_D \mathbf{X}^* \right) \right\|^2 = \left\| \frac{E_p}{\sigma_n^2} \left(\mathbf{F}_L \left(\mathbf{F}_p^H \mathbf{F}_p \right)^{-1} \mathbf{F}_L^H + E_p \mathbf{\Lambda}^{-1} \right)^{-1/2} \times \left(\mathbf{F}_L \hat{\mathbf{h}} - \mathbf{Y}_D \mathbf{X}^* \right) \right\|^2. \quad (32)$$

Since the scalar E_p/σ_n^2 does not affect the minimum of (32), the enhanced data detector is now given by

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{Q}^N} \left\| \left(\mathbf{F}_L \left(\mathbf{F}_p^H \mathbf{F}_p \right)^{-1} \mathbf{F}_L^H + E_p \mathbf{\Lambda}^{-1} \right)^{-1/2} \times \left(\mathbf{F}_L \hat{\mathbf{h}} - \mathbf{Y}_D \mathbf{X}^* \right) \right\|^2. \quad (33)$$

Remarks:

- 1) Unlike (24), a remarkable advantage of the enhanced detector (33) is that it requires neither \mathbf{R}_h nor σ_n^2 . The only

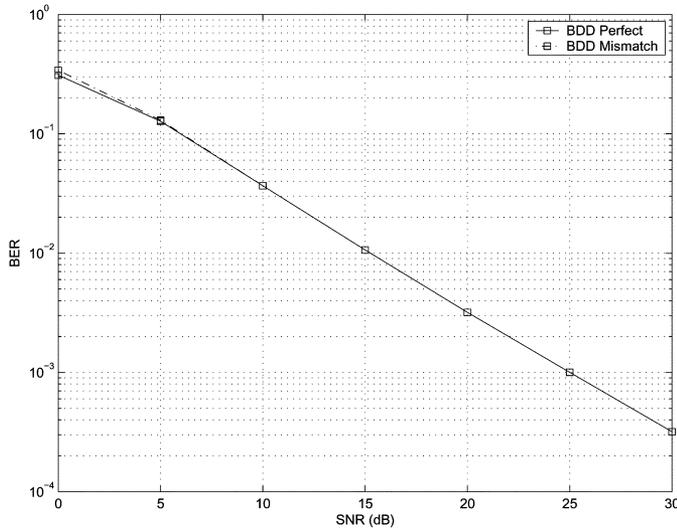


Fig. 1. BER of the mismatched semiblind data detector for a BPSK-OFDM system with $N = 32$ and $N_p = 2$. The channel has an exponential PDP. The correlation matrix for use with the semiblind data detector is computed assuming a uniform PDP.

required parameter is the pilot energy E_p , known at the receiver. Therefore, no parameter mismatch problem exists in (33).

- 2) The performance gain of (33) is obtained by exploiting the statistics of the channel estimation error and by sacrificing the simplicity of one-tap equalization. When more computational power is available, the use of (33) saves the bandwidth and power since the number of pilots is not increased.

C. Detection Algorithms

Straightforwardly, (24), (33) and (26) can be solved by exhaustively searching all feasible candidates $\mathbf{X} \in \mathcal{Q}^{N_d}$, which becomes impractical as N_d becomes large. Note that the cost functions (24), (26), and (33) can be reformulated as integer quadratic forms. We can solve such quadratics using SD [10] and V-BLAST [11], and this approach allows for substantial computational saving over exhaustive search. For brevity, we omit further details and refer the reader to [7] and [8].

V. MISMATCH ANALYSIS AND OPTIMAL PILOT DESIGN

A. Mismatch

The semiblind channel estimator (13) needs the knowledge of σ_n^2 , while blind data detector (BDD) (24) needs both \mathbf{R}_h and σ_n^2 , which may not be known perfectly in practice. However, \mathbf{R}_h can be estimated using $\hat{\mathbf{R}}_h = \sum_{k=1}^K \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H / K$, and σ_n^2 can be obtained by $\sum_{k=1}^K \|\mathbf{Y}_k - \hat{\mathbf{X}}_k \mathbf{F}_L \hat{\mathbf{h}}_k\|^2 / (KN)$, where $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{X}}_k$ are the estimated CIR and data symbols in the k th symbol, respectively. If there still remains a residual mismatch, we have introduced the following robust design criteria [8]: if \mathbf{R}_h is chosen according to a uniform power delay profile (UPDP) and σ_n^2 is chosen for a high SNR, then (24) is robust against the mismatch. Fig. 1 shows how (24) performs against mismatch. The BDD

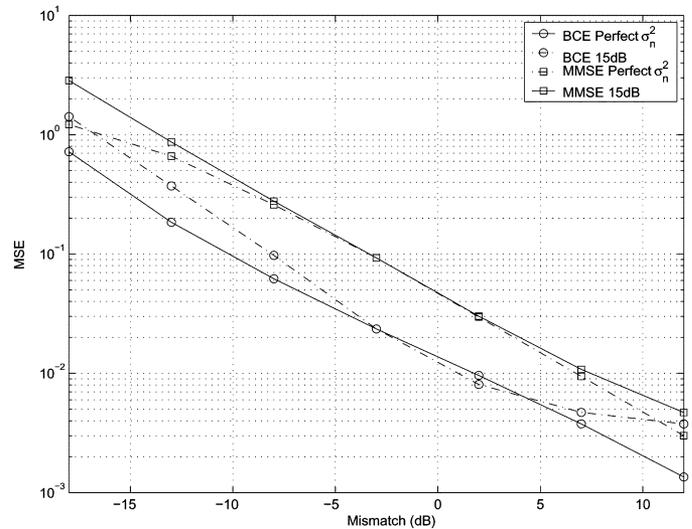


Fig. 2. MSE of the mismatched semiblind channel estimator for a BPSK-OFDM system with $N = 32$, $N_p = 2$. The estimator is designed for a SNR of 15 dB, but functions at other SNRs. Mismatch (dB) is the difference between the actual SNR and the designed SNR.

(24) is designed for a UPDP and an SNR of 20 dB and evaluated for an exponentially decaying power-delay profile. The BER of the robust design is compared with the BER when perfect knowledge of \mathbf{R}_h and σ_n^2 is available. The BERs of the two detectors are almost the same, which confirms the robust design criteria. In Fig. 2, we investigate the mismatch for the BCE (13) designed for an SNR of 15 dB. This estimator (13) too is robust against small SNR mismatch (less than 5 dB).

B. Pilot Design

Optimal pilots design has been given by maximizing the capacity [12] and minimizing the mean-square error (MSE) of channel estimate [13]. But both these results are based on specific channel estimation algorithms. Instead, we next optimize pilots by minimizing the CRB on the MSE of semiblind channel estimators.

The CRB, a lower bound on the variance of any unbiased estimator, has been widely used to optimize the pilots in communications systems [20], [21]. Since our channel estimator is derived in an ML sense, it is asymptotically unbiased [18, Th. 7.3]. The CRB for the MSE of semiblind channel estimation is derived in (48). The optimal pilots are designed by minimizing the CRB with respect to the placement and power constraint of the pilots and data symbols. We assume the power constraint $\sum_{k \in I_d} E_k^d = \mathcal{D}$ and $\sum_{k \in I_p} E_k^p = \mathcal{P}$, where \mathcal{D} and \mathcal{P} are the total power on pilots and data symbols respectively. The power allocation problem is beyond the scope of this paper. Using (48), the problem of CRB minimization becomes

$$(I_p, E_k^p, I_d, E_k^d) = \arg \min_{\substack{\sum_{k \in I_d} E_k^d = \mathcal{D}, \\ \sum_{k \in I_p} E_k^p = \mathcal{P}}} \text{tr} \left\{ \left(\frac{1}{\sigma_n^2} \mathbf{F}_L^H (\mathbf{\Lambda}_d + \mathbf{\Lambda}_p) \mathbf{F}_L + \mathbf{R}_h^{-1} \right)^{-1} \right\}. \quad (34)$$

Using [12, p. 2144, Lemma 1], the MSE of CIR is lower bounded by

$$\begin{aligned} & \text{tr} \left\{ \left(\frac{1}{\sigma_n^2} \mathbf{F}_L^H (\mathbf{\Lambda}_d + \mathbf{\Lambda}_p) \mathbf{F}_L + \mathbf{R}_h^{-1} \right)^{-1} \right\} \\ & \geq \sum_{k=0}^{N-1} \frac{1}{\left[\frac{1}{\sigma_n^2} \mathbf{F}_L^H (\mathbf{\Lambda}_d + \mathbf{\Lambda}_p) \mathbf{F}_L + \mathbf{R}_h^{-1} \right]_{k,k}} \end{aligned} \quad (35)$$

where the equality is attained if and only if $\mathbf{C} = \mathbf{F}_L^H (\mathbf{\Lambda}_d + \mathbf{\Lambda}_p) \mathbf{F}_L$ is diagonal. The (i, j) th $(0 \leq i, j \leq L - 1)$ entry of \mathbf{C} can be written as shown in (36) at the bottom of the page. Therefore, we require

$$\begin{aligned} & \sum_{k \in I_p} E_k^p e^{-j2\pi k(j-i)/N} + \sum_{k \in I_d} E_k^d e^{-j2\pi k(j-i)/N} \\ & = (\mathcal{D} + \mathcal{P})\delta(i - j). \end{aligned} \quad (37)$$

Equation (37) is satisfied if the following conditions are satisfied:

- C1) $E_k^d = \mathcal{D}/N_d \forall k \in I_d$ and $E_k^p = \mathcal{P}/N_p \forall k \in I_p$.
- C2) $d = N/N_p \in \mathcal{Z}$ and $I_p = \{d_0 + kd, k = 0, 1, \dots, N_p - 1\}$, $d_0 \in \{0, 1, \dots, d - 1\}$.

C1) means that both the pilots and the data symbols must be equipowered. C2) means that the pilots must be equispaced. They also agree with the two conditions in [13] by minimizing the MSE of the LS channel estimation, but C1) differs from those in [13] in that the equipower is also required on data symbols. If C1) and C2) are satisfied, we find the MSE of CIR only depends on $\mathcal{D} + \mathcal{P}$ but not on \mathcal{D} and \mathcal{P} separately. This indicates that the power allocation between \mathcal{D} and \mathcal{P} does not affect the CRB. However, the CRB is only asymptotically achievable in high SNR. Therefore, the power allocation affects the required SNR to achieve the CRB as shown in Figs. 3 and 5.

VI. SIMULATION RESULTS

We consider a frequency-selective Rayleigh fading channel, and the zero-mean Gaussian complex coefficients h_l are generated independently with the mean power $\sigma_l^2 = E[|h_l|^2] = \sigma_0^2 e^{-l/5}$ for $l = 1, \dots, L$, where σ_0^2 is the average power of the first path. The channel varies from each OFDM symbol to another. An OFDM system with $N = 32$ subcarriers is simulated, where the carrier frequency is 5 GHz. Pilot and data symbols have the same energy (C1). The BDD (24), which does not use DD, when used with V-BLAST is denoted by BDD-VBLAST, and when used with SD is denoted by BDD-SD. The BCE with DD (DD) is denoted by BCE-DD. The LS with DD is also provided for comparative purposes and is denoted by LS-DD. The DD in all the estimators use LS channel estimator.

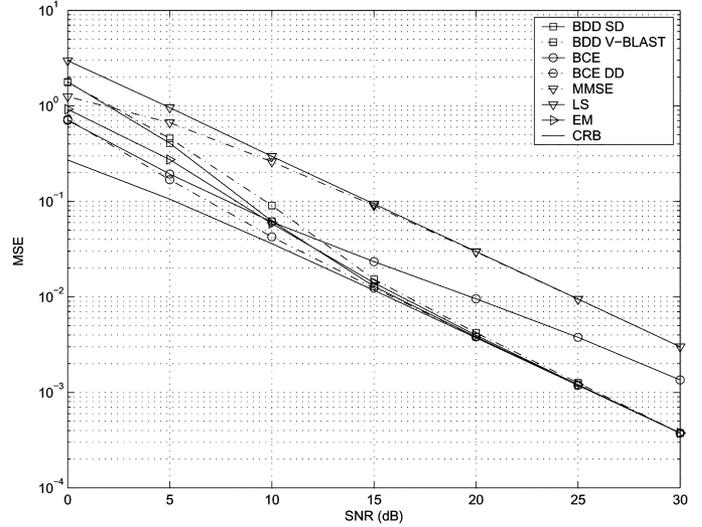


Fig. 3. Comparison of the MSE of the channel estimators for a BPSK-OFDM system with $N = 32$.

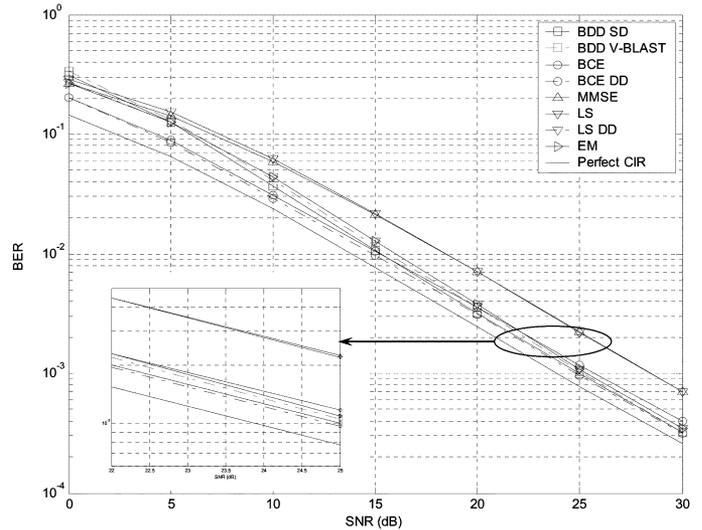


Fig. 4. Comparison of the BER of the data detectors for a BPSK-OFDM system with $N = 32$.

Figs. 3 and 4 show the MSE of channel estimation and BER of a binary phase-shift keying (BPSK) system. A 6-ary channel ($L = 6$) is simulated. The data rate is 500 kb/s. Both the BDD (24) and the BCE (13) uses only two uniformly distributed pilots as per C2). Their performance is compared with those of the MMSE (16) and LS (17) channel estimators using six pilots. We also compare with the expectation-maximization (EM) data detector in [22] and denote it as EM. In Fig. 3, for $\text{MSE} \leq 10^{-2}$, both BDD-VB and BDD-SD perform identically. They perform 3.5 dB better than the BCE and 8.5 dB better than both

$$[\mathbf{C}]_{i,j} = \begin{cases} \mathcal{D} + \mathcal{P}, & i = j \\ \sum_{k \in I_p} E_k^p e^{-j2\pi k(j-i)/N} + \sum_{k \in I_d} E_k^d e^{-j2\pi k(j-i)/N}, & i \neq j \end{cases} \quad (36)$$

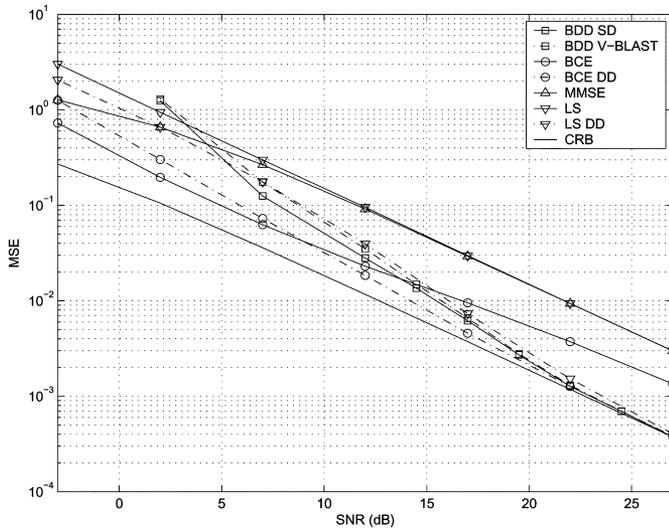


Fig. 5. Comparison of the MSE of the channel estimators for an OFDM system with 8-PSK and $N = 32$.

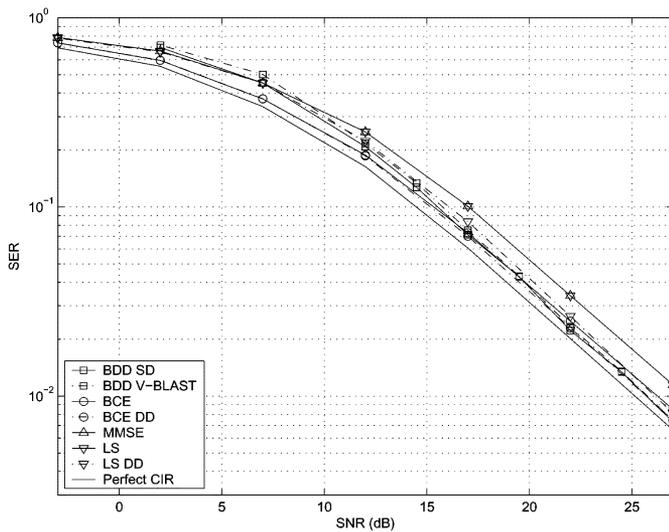


Fig. 6. Comparison of the SER of the data detectors for an OFDM system with 8-PSK and $N = 32$.

the MMSE and LS. The BCE-DD performs even better than the BDD. However, in high SNR ($\text{SNR} \geq 15$ dB), the BDD, LS-DD, EM, and BCE-DD all achieve the CRB (48). The BCE performs 4 dB better than MMSE in high SNR. In Fig. 4, the performance of the EM, BCE and BDD reaches that of the benchmark with perfectly known CIR and one-tap equalization. At $\text{BER} = 10^{-3}$, the BDD-SD gains 0.1 dB over BDD-VB and BCE gains 3.5 dB over MMSE and LS. It performs only 0.8 dB worse than the benchmark. The BCE-DD gains 0.8 dB over the LS-DD. Both BDD and BCE with just two pilots perform close to the benchmark, yet they have higher bandwidth efficiency and better performance than the LS detector.

Figs. 5 and 6 show the MSE of channel estimation and symbol-error rate (SER) of an 8-PSK system for a 3-ary channel. The data rate is 1500 kb/s. Four pilots are used for both BDD and BCE. At $\text{MSE} = 10^{-2}$, the MSE of BDD-VBLAST is also almost the same as that of BDD-SD. They

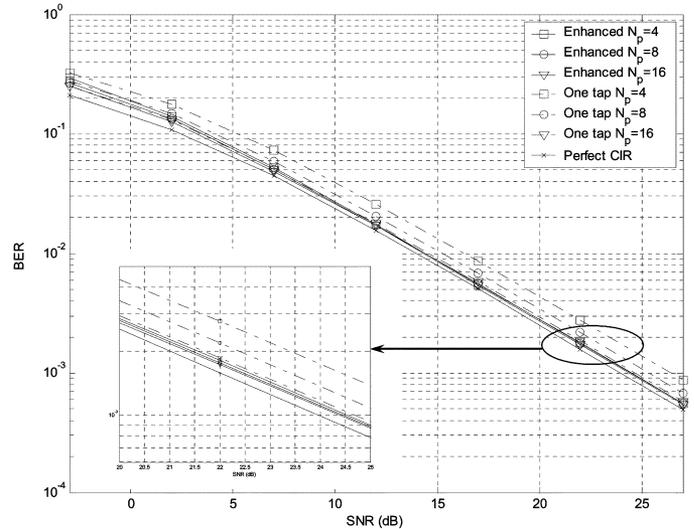


Fig. 7. Performance of the enhanced data detector as a function of the number of pilots for a BPSK-OFDM system with $N = 32$.

perform 7 dB better than both MMSE and LS. The BDD-SD gains about 0.9 dB over the BCE. Unlike BPSK systems, the BCE-DD performs worse than the original BCE because of decision-feedback errors. However, in high SNR ($\text{SNR} \geq 25$ dB), both the BDD and the BCE also achieve the CRB (48). LS-DD needs high SNR to achieve the CRB. In Fig. 6, at $\text{SER} = 2 \times 10^{-2}$, the BDD-SD, BDD-VB, and BCE-DD perform nearly identically. They gain 1.5 dB over MMSE and LS and gain 0.5 dB over BCE and perform 0.5 dB within the benchmark with perfectly known CIR and one-tap equalization in high SNR. Figs. 3 and 5 show that LS-DD can also achieve the CRB at high SNR with low complexity if the pilots and data satisfy C1) and C2). However, for large constellations, LS-DD requires higher SNR to achieve the CRB, and also when the number of pilots N_p is less than the length of channel or $N_p < L$, LS-DD fails to work, which means our BCE and BDD have higher bandwidth efficiency than LS-DD. When $N_p \geq L$, LS-DD has less complexity than both BCE and BDD.

Fig. 7 compares the performance of the enhanced data detector (33) and the one-tap equalization using pilot-aided channel estimates with estimation errors. A 3-ary channel is simulated. We simulate BPSK-OFDM at a data rate of 500 kb/s with different number of pilots. The pilots are uniformly distributed (C2). At low SNR, the enhanced detector performance varies slightly as the number of pilots increases from 4 to 16. At high SNR, the enhanced detector performance is unchanged as the number of pilots increases, and it gains 1.6, 0.75, and 0.2 dB over one-tap equalization with 4, 8, and 16 pilots, respectively.

VII. CONCLUSION

We have developed a novel semiblind channel estimator and semiblind data detectors for OFDM with high bandwidth efficiency. As an aside, we have shown that the samples of an OFDM symbol have a joint complex Gaussian distribution, whose mean and covariance are determined by the locations and values of fixed pilot symbols. This result is a generalization of the conventional CLT reasoning that shows that each individual

OFDM signal sample is univariate complex Gaussian. The semiblind channel estimator is obtained by averaging the likelihood function over the Gaussian transmitted-signal statistics, while the semiblind data detector is obtained by averaging the likelihood function over the CIR. We also derive an enhanced data detector to mitigate the effect of channel estimation error. The data detectors are efficiently implemented by V-BLAST and SD. The semiblind channel estimator achieves the CRB. Moreover, we show that equispaced and equipowered pilot symbols are optimal for minimizing the CRB. Our results may also be extended to multiple-antenna OFDM systems.

APPENDIX A

In this Appendix, we give a simple proof that the time-domain OFDM symbols are jointly complex Gaussian. Let $\mathbf{X} = [X_1, \dots, X_N]^T$ and $X_i \in \mathcal{Q}$, where \mathcal{Q} is a finite set with $E\{X_i\} = 0$ and $E\{|X_i|^2\} = \sigma^2$. We want to show that as N goes to infinity, the vector $\mathbf{x} = \mathbf{A}\mathbf{X}$ distributes as multivariate complex Gaussian, where $\mathbf{A} \in \mathbb{C}^{M \times N}$. Using the CLT, we find that the i th element of \mathbf{x} converges to be complex Gaussian with $x_i \sim \mathcal{CN}(0, \|\mathbf{a}_i\|^2 \sigma^2)$, where \mathbf{a}_i is the i th row of \mathbf{A} . We define another vector $\mathbf{Y} = [Y_1, \dots, Y_N]^T$ and $Y_i \sim \mathcal{CN}(0, \sigma^2)$. We can show that the vector $\mathbf{y} = \mathbf{A}\mathbf{Y}$ is multivariate complex Gaussian, and $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{A}\mathbf{A}^H)$. Since the pdf of x_i converges to the pdf of y_i as $N \rightarrow \infty$, the joint pdf of \mathbf{x} also converges to that of \mathbf{y} if M is finite. Therefore, as $N \rightarrow \infty$, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{A}\mathbf{A}^H)$, which is a stronger result than the CLT.

In addition to the above proof, we have also conducted multivariate normality testing to verify that \mathbf{x} is multivariate complex Gaussian with the mean and covariance (8). Many approaches to testing multivariate normality have been proposed in the literature. Mardia [23] provides two tests based on multivariate generalization of kurtosis and skewness. To test if a set of p -variate observations V_1, V_2, \dots, V_n are multivariate normal with mean \mathbf{m} and the covariance matrix \mathbf{R} , the two test statistics are computed

$$b_{1,p} = \frac{1}{n^2} \sum_{j,k=1}^n ((V_j - \mathbf{m})^T \mathbf{R}^{-1} (V_k - \mathbf{m}))^3$$

$$b_{2,p} = \frac{1}{n} \sum_j ((V_j - \mathbf{m})^T \mathbf{R}^{-1} (V_j - \mathbf{m}))^2. \quad (38)$$

To apply these to our problem, we express the hypothesized complex N -variate Gaussian distribution of \mathbf{x} as a $2N$ -variate real Gaussian distribution [24]. We generate \mathbf{x} assuming data symbols $X_k, k \in I_d$, to be complex Gaussian, $\mathcal{CN}(0, \sigma^2)$ (the hypothesis is true under this assumption), and use Monte Carlo simulation to find the distributions of $b_{1,p}$ and $b_{2,p}$. We next generate \mathbf{x} for $X_k \in \mathcal{Q}, k \in I_d$, and if the resulting $b_{1,p}$ and $b_{2,p}$ are outside the distributional values we reject the hypothesis. Despite extensive trials, the rejection rate is less than 1%; the evidence of non-normality is quite weak.

APPENDIX B

In this appendix, we derive the CRB for semiblind channel estimation. Define

$$\mathbf{X}^d = \begin{cases} d_k, & k \in I_d \\ 0, & \text{otherwise} \end{cases} \quad \mathbf{X}^p = \begin{cases} p_k, & k \in I_p \\ 0, & \text{otherwise} \end{cases} \quad (39)$$

and $\mathbf{X}_D^d = \text{diag}\{\mathbf{X}^d\}, \mathbf{X}_D^p = \text{diag}\{\mathbf{X}^p\}$. Substituting them into (4) and (5), we have

$$\mathbf{Y} = (\mathbf{X}_D^d + \mathbf{X}_D^p) \mathbf{F}_L \mathbf{h} + \mathbf{W} \quad (40)$$

$$\mathbf{Y} = \mathbf{H}_D \mathbf{F} (\mathbf{X}^d + \mathbf{X}^p) + \mathbf{W}. \quad (41)$$

Let $\theta = [(\mathbf{X}^d)^H, \mathbf{h}^H]$. Under regularity conditions, the FIM associated with a complex stochastic parameter vector θ is defined as [20]

$$\mathbf{J}(\theta) = E \left\{ \left[\frac{\partial \ln p(\mathbf{Y}, \theta)}{\partial \theta^*} \right] \left[\frac{\partial \ln p(\mathbf{Y}, \theta)}{\partial \theta^*} \right]^H \right\}$$

$$= E \left\{ E \left\{ \left[\frac{\partial \ln p(\mathbf{Y}|\theta)}{\partial \theta^*} \right] \left[\frac{\partial \ln p(\mathbf{Y}|\theta)}{\partial \theta^*} \right]^H \middle| \theta \right\} \right\}$$

$$+ E \left\{ \left[\frac{\partial \ln p(\theta)}{\partial \theta^*} \right] \left[\frac{\partial \ln p(\theta)}{\partial \theta^*} \right]^H \right\} \quad (42)$$

where $p(\mathbf{Y}, \theta)$ is the joint distribution of \mathbf{Y} and θ . The first term of (42) can be obtained as

$$E \left\{ E \left\{ \left[\frac{\partial \ln p(\mathbf{Y}|\theta)}{\partial \theta^*} \right] \left[\frac{\partial \ln p(\mathbf{Y}|\theta)}{\partial \theta^*} \right]^H \middle| \theta \right\} \right\}$$

$$= E \left\{ \frac{1}{\sigma_n^2} \begin{pmatrix} \mathbf{H}_D^H \Delta_d \mathbf{H}_D & \mathbf{H}_D^H \Delta_d \mathbf{X} \\ \mathbf{X}^H \Delta_d \mathbf{H}_D & \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L \end{pmatrix} \right\}$$

$$= \frac{1}{\sigma_n^2} \begin{pmatrix} E \{ \mathbf{H}_D^H \Delta_d \mathbf{H}_D \} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_L^H (\Lambda_d + \Lambda_p) \mathbf{F}_L \end{pmatrix} \quad (43)$$

where

$$\Delta_d = \text{diag}\{s_1, \dots, s_N\}, \quad s_k = \begin{cases} 1, & k \in I_d \\ 0, & k \notin I_d \end{cases}$$

$$\Lambda_d = \text{diag}\{s_1, \dots, s_N\}, \quad s_k = \begin{cases} E_k^d, & k \in I_d \\ 0, & k \notin I_d \end{cases} \quad (44)$$

and

$$\Lambda_p = \text{diag}\{s_1, \dots, s_N\}, \quad s_k = \begin{cases} E_k^p, & k \in I_p \\ 0, & k \notin I_p. \end{cases} \quad (45)$$

The second equality comes from $E\{\mathbf{H}_D^H \Delta_d \mathbf{X}\} = 0$, and $E\{\mathbf{X}_D^H \mathbf{X}_D\} = E\{(\mathbf{X}^d + \mathbf{X}^p)^H (\mathbf{X}^d + \mathbf{X}^p)\} = \Lambda_d + \Lambda_p$. Note we do not assume here that all the signal powers are the same. Since $p(\theta) = p_x(\mathbf{X}^d) p_h(\mathbf{h})$, the second term in (42) becomes (46), shown at the top of the next page. Therefore, the complex FIM is shown in (47) at the top of the next page. The CRB for the MSE of semiblind channel estimation is then given by

$$\text{CRB}_h = \text{tr} \left\{ \left(\frac{1}{\sigma_n^2} \mathbf{F}_L^H (\Lambda_d + \Lambda_p) \mathbf{F}_L + \mathbf{R}_h^{-1} \right)^{-1} \right\}. \quad (48)$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their critical comments that greatly improved this paper. Especially, they thank one of the anonymous reviewers for the comments on the global minimum of (12).

$$\begin{aligned}
 E \left\{ \begin{bmatrix} \frac{\partial \ln p(\theta)}{\partial \theta^*} \\ \frac{\partial \ln p(\theta)}{\partial \theta^*} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p(\theta)}{\partial \theta^*} \\ \frac{\partial \ln p(\theta)}{\partial \theta^*} \end{bmatrix}^H \right\} &= \begin{pmatrix} E \left\{ \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix}^H \right\} & \mathbf{0} \\ \mathbf{0} & E \left\{ \begin{bmatrix} \frac{\partial \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*} \\ \frac{\partial \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*} \\ \frac{\partial \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*} \end{bmatrix}^H \right\} \end{pmatrix} \\
 &= \begin{pmatrix} E \left\{ \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix}^H \right\} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_h^{-1} \end{pmatrix} \quad (46)
 \end{aligned}$$

$$\mathbf{J}(\theta) = \begin{pmatrix} \frac{1}{\sigma_n^2} E \{ \mathbf{H}_D^H \Delta_d \mathbf{H}_D \} + E \left\{ \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p_x(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \\ \frac{\partial \ln p(\mathbf{X}^d)}{\partial (\mathbf{X}^d)^*} \end{bmatrix}^H \right\} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_n^2} \mathbf{F}_L^H (\boldsymbol{\Lambda}_d + \boldsymbol{\Lambda}_p) \mathbf{F}_L + \mathbf{R}_h^{-1} \end{pmatrix} \quad (47)$$

REFERENCES

- [1] M. Quan and P. Mannion, "AT&T, Nortel Plot 4G Wireless Nets," 2000 [Online]. Available: <http://www.eet.com/story/OEG20000915S0034/>
- [2] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.
- [3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 3, pp. 1207–1215, Jul. 2000.
- [4] Y. Li, L. J. Cimini, Jr., and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [5] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semiblind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.
- [6] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141–150, Jan. 2003.
- [7] T. Cui and C. Tellambura, "Joint channel estimation and data detection for OFDM systems via sphere decoding," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2004, pp. 3656–3660.
- [8] T. Cui and C. Tellambura, "Joint data detection and channel estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 670–679, Apr. 2006.
- [9] M.-X. Chang and Y. T. Su, "Blind and semiblind detections of OFDM signals in fading channels," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 744–754, May 2004.
- [10] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.
- [11] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–15, Jan. 1999.
- [12] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2138–2145, Sep. 2004.
- [13] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [14] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 89–101, Jan. 2001.
- [15] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*. New York: McGraw-Hill, 1966.
- [16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2002.
- [17] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.
- [18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [19] P. K. Frenger, N. Arne, and B. Svensson, "Decision-directed coherent detection in multicarrier systems on Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 2, pp. 490–498, Mar. 1999.
- [20] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.
- [21] L. Berriche, K. Abed-Meraim, and J. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *Proc. ICASSP*, 2004, pp. 17–21.
- [22] X. Ma, H. Kobayashi, and S. Schwartz, "An EM-based estimation of OFDM signals," in *Proc. WCNC*, Mar. 2002, vol. 1, pp. 228–232.
- [23] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.
- [24] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–177, 1963.



Tao Cui (S'04) received the B.Eng. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2003, and the M.Sc. degree from the University of Alberta, Edmonton, AB, Canada, in 2005. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA.

His research interests are in communication theory, broadband wireless communications, space-time coding, MIMO systems and wireless networks. He is a past recipient of postgraduate scholarships

from the Alberta Ingenuity Fund and the Alberta Informatics Circle of Research Excellence (iCORE).



Chintha Tellambura (M'97–SM'02) received the B.Sc. degree (with first-class honors) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Postdoctoral Research Fellow with the University of Victoria (1993–1994) and the University of Bradford (1995–1996). He was with Monash University, Melbourne, Australia, from 1997 to 2002.

Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His research interests include coding, communication theory, modulation, equalization, and wireless communications.

Prof. Tellambura is an Associate Editor for both the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a Co-Chair of the Communication Theory Symposium of Globecom'05, held in St. Louis, MO.