Signal Design for Transmitter Diversity Wireless Communication Systems Over Rayleigh Fading Channels

Jiann-Ching Guey, Michael P. Fitz, Mark R. Bell, and Wen-Yi Kuo, *Member, IEEE*

Abstract— **In this paper, transmitter diversity wireless communication systems over Rayleigh fading channels using pilot symbol assisted modulation (PSAM) are studied. Unlike conventional transmitter diversity systems with PSAM that estimate the superimposed fading process, we are able to estimate each individual fading process corresponding to the multiple transmitters by using appropriately designed pilot symbol sequences. With such sequences, special coded modulation schemes can then be designed to access the diversity provided by the multiple transmitters without having to use an interleaver or expand the signal bandwidth. The notion of code matrix is introduced for the coded modulation scheme, and its design criteria are also established. In addition to the reduction in receiver complexity, simulation results are compared to, and shown to be superior to, that of an intentional frequency offset system over a wide range of system parameters.**

Index Terms— **Channel coding, diversity methods, Rayleigh channels.**

I. Introduction

PROVIDING an architecture with diversity is important for maintaining high performance in wireless mobile communications. Diversity can be achieved by using multiple antennas, using interleaved coded modulation, resolving propagation paths in time or spatially, and using multicarrier transmission [1], [2]. Perhaps the most commonly used technique is interleaved coded modulation. The coding adds the redundancy to provide diversity and the interleaving separates the code symbols to (hopefully) provide independent fading distortion for each of the code symbols. The problem with standard interleaved coded modulation is that a tradeoff must be made between decoding delay (a function of the interleaver depth) and demodulation performance. This is especially important in applications where performance is decoding delay sensitive (e.g., voice transmission). For situations with small Doppler spread (e.g., pedestrian or stopped vehicle), either a very long interleaver is needed to achieve quasi-independent distortion on code symbols or else interleaving is not effective.

An effective technique in wireless communications is transmission diversity. The advantage of transmission diversity is that by transmitting from multiple spatially separated antennas (e.g., a base station) diversity can be achieved without greatly increasing the complexity of the receiver (e.g., a portable unit). The simplest idea is to switch between the transmitters at different time instants and allow only one transmitter to be on at a time. Because the transmitters are operated intermittently, their peak power is considerably higher than their average power, which complicates the design of their output amplifiers. Other transmission diversity techniques that do not switch off the transmitter are ones using an intentional time offset [3] or frequency offset [4], phase sweeping [5], frequency hopping [6], and modulation diversity [7]. Most of these techniques use phase or frequency modulation of each transmitter carrier to induce intentional time-varying fading at the receiver.[1] The advantage of these techniques is that the modulation level of the carrier and the interleaving depth can be chosen to achieve near ideal interleaving. In these applications, a shorter interleaver depth is usually only achieved with an expanded signal bandwidth. The focus of this paper is the exposition of a fairly simple alternate system architecture which can provide the diversity inherent in multiple transmissions without requiring interleaving even with low mobility.

In this paper we consider linear modulation on frequency nonselective fading channels. Consequently, applications of this work are in modems using narrowband or multicarrier modulation. Decoding of error control codes in frequency nonselective fading channels requires an estimate of the channel state (or multiplicative distortion), and transmitted reference techniques usually provide the simplest method for channel state estimation. Common transmitted reference techniques are tone-calibration techniques (TCT) [8] and pilot symbol assisted modulation (PSAM) [9]. PSAM is preferred in practice because it typically provides a better peak to average transmitted power ratio without the need to redesign the modulation pulse. Both TCT and PSAM are amenable to a performance analysis for ideal interleaved coded modulation [1] and for correlated fading [10]. In fact, the work in [11] designed and analyzed the performance of a system using interleaved coded modulation and frequency offset diversity

Paper approved by P. H. Witke, the Editor for Communication Theory of the IEEE Communications Society. Manuscript received August 28, 1995; revised March 12, 1998 and July 27, 1998. This work was supported by Purdue Research Foundation and National Science Foundation under Grant NCR-9115820. This paper was presented in part at the 1996 IEEE Vehicular Technology Conference, Atlanta, GA.

J.-C. Guey is with Ericsson Inc., Research Triangle Park, NC 27709 USA.

M. P. Fitz is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210-1272 USA.

M. R. Bell is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285 USA (e-mail: mrb@ecn.purdue.edu).

W.-Y. Kuo is with Lucent Technologies, Whippany, NJ 07981-0903 USA.

Publisher Item Identifier S 0090-6778(99)03318-8.

[1] Note this is not the case for [3] and [7].

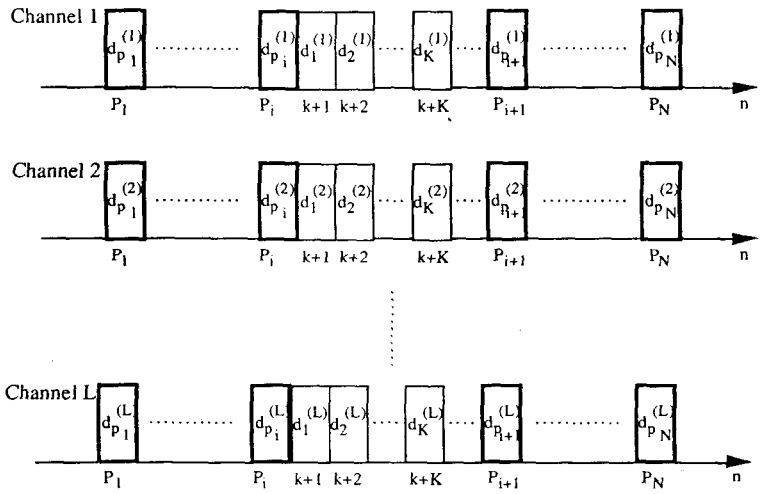


Fig. 1. The transmitted signals.

transmission using PSAM. At the time of our study [12], the system of [11] was the highest performing complete transmitter diversity system and will be used as a benchmark for comparison to our proposed system.

In this paper, we propose a new coded modulation scheme to access the diversity of a multiple transmitters system without having to use interleaving. This method is similar to the modulation diversity method proposed by [7], but does not require an equalizer for decoding. The relationship between the system's error performance and the design of the coded modulation scheme is also thoroughly examined from a more general point of view. Since the proposed coded modulation scheme requires the knowledge of the states on all channels, not just the single state of the superimposed process, pilot symbol sequences need to be appropriately designed. This important issue that has been missing in [7] will be addressed in a solid mathematical framework.

The rest of this paper is organized as follows. The next section formulates the problem and introduces the notation used throughout this paper. A receiver architecture is developed and its error performance is analyzed in Section III. The information provided by this analysis is then used as design criteria in Sections IV and V, where code matrix and pilot symbol matrix designs are considered, respectively. In Section VI, an example is shown and compared to the result of an intentional frequency offset system. Section VII concludes.

II. Problem Description

In the sequel, bold lowercase italic denotes vectors, bold uppercase italic denotes matrices; $\boldsymbol{x}^*$, $\boldsymbol{x}^T$, and $\boldsymbol{x}^H$ will denote the complex conjugate, transpose, and Hermitian transpose of $\boldsymbol{x}$, respectively; $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ denotes the inner product between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, $|C|$ denotes the determinant of C, and $\boldsymbol{I}_K$ denotes an identity matrix of dimension K.

Assuming that m bits per baud are transmitted, a representation of the transmitted signals for a transmitter diversity system is given in Fig. 1 where each channel represents the signal on a different antenna. When a linearly modulated information bearing signal, $s(t) = \Sigma_n \, d_n u(t - nT)$, where $u(t)$ is the unit energy pulse shape and T is the symbol time, is transmitted over a frequency nonselective, time-varying, isotropic scattering, Rayleigh fading channel, the signal at the receive end is modeled by

$$y(t) = c(t)s(t) + n(t) \tag{1}$$

where $c(t)$ is a zero mean complex Gaussian multiplicative distortion (MD) random process and $n(t)$ is a zero mean AWGN process with one-sided spectral density N_0. The processes $c(t)$ and $n(t)$ are assumed independent and the isotropic Rayleigh scattering assumption implies that $c(t)$ is a wide sense stationary random process with autocorrelation function [2]

$$R_c(\tau) \triangleq E[c(t)c^*(t-\tau)] = E_s J_0(2\pi f_D \tau) \tag{2}$$

where f_D is the Doppler spread of the channel, E_s is the average energy per transmitted symbol, and $J_0(\cdot)$ is the Bessel function of zeroth order. Using superscript (l) to index the transmitter, the multiple transmitter case can be described as

$$y(t) = \sum_{l=1}^{L} c^{(l)}(t)s^{(l)}(t) + n(t) \tag{3}$$

where $\{c^{(1)}(t), c^{(2)}(t), \cdots, c^{(L)}(t)\}$ are assumed i.i.d. with an autocorrelation function given in (2). The assumption of independent multiplicative distortions for each antenna implies the antennas are separated appropriately. For indoor network topologies this separation can be a small number of wavelengths. For elevated outdoor antennas, greater care in placement of antennas is important [2].

By assuming the fading is slow enough to be roughly constant over the support of the pulse and that $u(t)$ is appropriately shaped so that intersymbol interference can be ignored, the matched filter outputs are approximate sufficient statistics and given as

$$\begin{aligned} x_k &= \int_{-\infty}^{\infty} y(t)u^*(t - kT)\, dt \\ &= \sum_{l=1}^{L} d_k^{(l)} c_k^{(l)} + n_k \end{aligned} \tag{4}$$

where $c_k^{(l)}$ is complex Gaussian with $E[c_k^{(l)} c_{k-m}^{(l)^*}] = R_c(mT)$. The validity of this assumption can be ascertained by the results in [13].

By arranging each of the transmitted symbols for each antenna in a matrix, we define a code matrix transmitted starting at symbol time k as

$$\boldsymbol{D}(k) = \begin{bmatrix} d_k^{(1)} & d_k^{(2)} & \cdots & d_k^{(L)} \\ d_{k+1}^{(1)} & d_{k+1}^{(2)} & \cdots & d_{k+1}^{(L)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k+K-1}^{(1)} & d_{k+K-1}^{(2)} & \cdots & d_{k+K-1}^{(L)} \end{bmatrix} \tag{5}$$

where $d_j^{(i)}$ is a complex modulation symbol transmitted on the ith antenna and on the jth baud and K is the length of the code. Note that a total of 2^{mK} code matrices are used in this transmission scheme and a block of mK information bits will select a code matrix for transmission. We define the set of all code matrices (the code) as Ξ_D and use a subscript to denote different elements within the code (i.e., $\boldsymbol{D}_\alpha \in \Xi_D$).

One set of observations to be used to demodulate the information symbols is the sequence of K complex values given as

$$\begin{aligned} x_k &= \sum_{l=1}^{L} d_k^{(l)} c_k^{(l)} + n_k \\ x_{k+1} &= \sum_{l=1}^{L} d_{k+1}^{(l)} c_{k+1}^{(l)} + n_{k+1} \\ &\vdots \\ x_{k+K-1} &= \sum_{l=1}^{L} d_{k+K-1}^{(l)} c_{k+K-1}^{(l)} + n_{k+K-1}. \end{aligned} \tag{6}$$

To consider an individual code matrix, we drop the subscript k to minimize confusion and define

$$\begin{aligned} \boldsymbol{x} &= [x_k \quad x_{k+1} \quad \cdots \quad x_{k+K-1}]^T \\ \boldsymbol{n} &= [n_k \quad n_{k+1} \quad \cdots \quad n_{k+K-1}]^T \\ \boldsymbol{\Lambda_D} &= [\boldsymbol{D}^{(1)} \quad \boldsymbol{D}^{(2)} \quad \cdots \quad \boldsymbol{D}^{(L)}] \\ \boldsymbol{c} &= [\boldsymbol{c}^{(1)^T} \quad \boldsymbol{c}^{(2)^T} \quad \cdots \quad \boldsymbol{c}^{(L)^T}]^T \end{aligned} \tag{7}$$

where $\boldsymbol{D}^{(l)} = \text{diag}\,[d_k^{(l)} \quad d_{k+1}^{(l)} \quad \cdots \quad d_{k+K-1}^{(l)}]$ and $\boldsymbol{c}^{(l)} = [c_k^{(l)} \quad c_{k+1}^{(l)} \quad \cdots \quad c_{k+K-1}^{(l)}]^T$. Note that the notation diag $[\boldsymbol{a}]$ indicates a square matrix where the diagonal elements are the elements of $\boldsymbol{a}$ and all other elements are zero. Consequentially the observation for a particular code matrix (6) is written as

$$\boldsymbol{x} = \boldsymbol{\Lambda_D c} + \boldsymbol{n} = \sum_{l=1}^{L} \boldsymbol{D}^{(l)} \boldsymbol{c}^{(l)} + \boldsymbol{n}. \tag{8}$$

Aside from observations associated with the information code matrices, a set of known sequences are inserted every P_{ins} symbols for the purpose of channel state estimation. For channel state estimation at a particular baud, the receiver uses the N nearest pilot symbol observations to acquire the channel state information required for decoding. The pilot symbol sequences can also be collected in an $N \times L$ matrix:

$$\boldsymbol{P} = \begin{bmatrix} d_{p_1}^{(1)} & d_{p_1}^{(2)} & \cdots & d_{p_1}^{(L)} \\ d_{p_2}^{(1)} & d_{p_2}^{(2)} & \cdots & d_{p_2}^{(L)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{p_N}^{(1)} & d_{p_N}^{(2)} & \cdots & d_{p_N}^{(L)} \end{bmatrix} \tag{9}$$

where $d_{p_j}^{(i)}$ are pilot symbols transmitted on the ith antenna and at the p_jth baud. The observations corresponding to the pilot sequences are

$$\begin{aligned} x_{p_1} &= \sum_{l=1}^{L} d_{p_1}^{(l)} c_{p_1}^{(l)} + n_{p_1} \\ x_{p_2} &= \sum_{l=1}^{L} d_{p_2}^{(l)} c_{p_2}^{(l)} + n_{p_2} \\ &\vdots \\ x_{p_N} &= \sum_{l=1}^{L} d_{p_N}^{(l)} c_{p_N}^{(l)} + n_{p_N} \end{aligned} \tag{10}$$

where $p_k = p_1 + P_{\text{ins}} * (k-1)$, or in a matrix form (with subscript p obviously indicating its association with pilot symbols):

$$\boldsymbol{x}_p = \boldsymbol{\Lambda_P c}_p + \boldsymbol{n}_p = \sum_{l=1}^{L} \boldsymbol{P}^{(l)} \boldsymbol{c}_p^{(l)} + \boldsymbol{n}_p \tag{11}$$

where $\boldsymbol{x}_p, \boldsymbol{\Lambda_P}, \boldsymbol{c}_p, \boldsymbol{P}^{(l)}, \boldsymbol{c}_p^{(l)}$, and $\boldsymbol{n}_p$ are all defined in a similar manner as (7) and (8). Note that the ith row of $\boldsymbol{P}$ will be referred to as $\boldsymbol{d}_{p_i}$ and its jth column as $\boldsymbol{d}_{\boldsymbol{P}}^{(j)}$. Also note that the elements of $\boldsymbol{P}$ do not necessarily belong to the information symbol alphabet.

III. Receiver Architecture and Performance Analysis

In this section, a near optimal receiver structure,[2] given in Fig. 2, for the model described in the previous section is established. This demodulator assumes symbol and frame synchronization.

The demodulator takes the output of the sampler and demultiplexes the data samples $\boldsymbol{x}$ and the pilot samples $\boldsymbol{x}_p$. The pilot samples are used to form a channel state estimate, $\hat{\boldsymbol{c}}$, for each code matrix. The observation vector $\boldsymbol{x}$ and the channel estimation $\hat{\boldsymbol{c}}$ are then fused at the maximum-likelihood detector to form the final decision on the transmitted code matrix. This mechanism is discussed and analyzed in the following subsections.

A. Maximum-Likelihood Detector

The maximum-likelihood detector using both the data observation $\boldsymbol{x}$ and the pilot observations $\boldsymbol{x}_p$ computes

$$\hat{\boldsymbol{D}} = \arg \max_{\boldsymbol{D}_\gamma \in \Xi} f(\boldsymbol{x} | \boldsymbol{x}_p, \boldsymbol{D}_\gamma) \tag{12}$$

[2] $\boldsymbol{x}$ and $\boldsymbol{x}_p$ is not a sufficient statistic for demodulation unless $f_D = 0$.

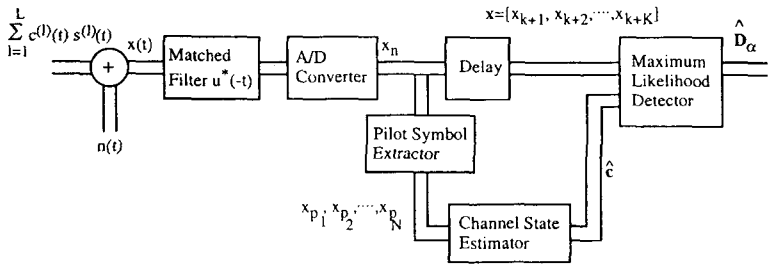


Fig. 2. Block diagram of the receiver.

where since $\boldsymbol{x}$ and $\boldsymbol{x}_p$ are both zero mean and jointly Gaussian

$$f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma) = \frac{1}{\pi^K \det(\boldsymbol{C}_{x|p})} \cdot \exp\left[-(\boldsymbol{x}-\hat{\boldsymbol{x}})\boldsymbol{C}_{x|p}^{-1}(\boldsymbol{x}-\hat{\boldsymbol{x}})^H\right] \quad (13)$$

where $\hat{\boldsymbol{x}} = E[\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma]$ and $\boldsymbol{C}_{x|p} = \mathrm{cov}(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma)$. Note

$$\begin{aligned}\boldsymbol{C}_{x|p} &= E\left[(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})^H|\hat{\boldsymbol{c}}, \boldsymbol{D}_\gamma\right] \\ &= \boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\mathrm{Cov}(\tilde{\boldsymbol{c}})\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}^H + N_0\boldsymbol{I}_K \end{aligned} \quad (14)$$

where

$$\mathrm{Cov}(\tilde{\boldsymbol{c}}) = E\left[(\boldsymbol{c}-\hat{\boldsymbol{c}})(\boldsymbol{c}-\hat{\boldsymbol{c}})^H\right]. \quad (15)$$

$\boldsymbol{C}_{x|p}$ is not a function of the pilot observations but is a function of the postulated transmitted code matrix. To reflect this dependence, we modify notation slightly and denote $\boldsymbol{C}_{x|p}$ by $\boldsymbol{C}_\gamma$. Since

$$\begin{aligned}E[\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma] &= E[\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\boldsymbol{c} + \boldsymbol{n}|\hat{\boldsymbol{c}}, \boldsymbol{D}_\gamma] \\ &= \boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}E[\boldsymbol{c}|\hat{\boldsymbol{c}}] = \boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}} \end{aligned} \quad (16)$$

the sufficient statistic resulting from the pilot observations is the MMSE channel state estimator $\hat{\boldsymbol{c}}$.

Hence, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{x}_p$ and $\boldsymbol{D}_\gamma$ is

$$f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma) = (\pi^K|\boldsymbol{C}_\gamma|)^{-1} \cdot \exp\left(-(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})^H\boldsymbol{C}_\gamma^{-1}(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})\right). \quad (17)$$

It should be noted that this demodulator has a complexity which is $O(LK^2m^K)$ and can be implemented in a recursive fashion as in [14] and [15]. Often in PSAM based systems [1], [9], the following reduced complexity form of the conditional density function is utilized:

$$f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma) = (\pi^K|\tilde{\boldsymbol{C}}_\gamma|)^{-1} \cdot \exp\left(-(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})^H\tilde{\boldsymbol{C}}_\gamma^{-1}(\boldsymbol{x}-\boldsymbol{\Lambda}_{\boldsymbol{D}_\gamma}\hat{\boldsymbol{c}})\right) \quad (18)$$

where

$$\tilde{\boldsymbol{C}}_\gamma = N_0\boldsymbol{I}_K \quad (19)$$

or

$$\tilde{\boldsymbol{C}}_\gamma = \mathrm{diag}\left[(\boldsymbol{C}_\gamma)_{11}, (\boldsymbol{C}_\gamma)_{22}, \cdots, (\boldsymbol{C}_\gamma)_{KK}\right] \quad (20)$$

where $(\boldsymbol{C}_\gamma)_{ii}$ is the ith diagonal element of $\boldsymbol{C}_\gamma$. These reduced complexity demodulators have a complexity of $O(LKm^K)$; and if the transmitted signals have a trellis structure, the complexity can be reduced to a complexity of $O(LK)$. It should be noted that the demodulator using (19) is equivalent to assuming that channel state estimate produced by the pilot symbols is perfect, and that the demodulator using (20) ignores the information about the channel state in $\boldsymbol{x}$. We will investigate in detail the optimal demodulator, but either of the demodulators described by (19) or (20) is a simple extension.

The function of the maximum-likelihood detector in Fig. 2 is then to form the decision statistics $f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma)$ (or, equivalently, $\ln[f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma)]$) for all $\boldsymbol{D}_\gamma \in \Xi$. Decoding is accomplished by selecting the largest $f(\boldsymbol{x}|\boldsymbol{x}_p, \boldsymbol{D}_\gamma)$.

B. Channel State Estimation

This subsection details the finite dimensional channel state estimator using the received pilot symbols. The optimum channel state estimation would obviously use all pilot symbols, but here we only consider the estimator which uses the N closest pilot symbols. After the extraction of the N nearest pilot symbols to a code matrix, the channel state estimator interpolates between these samples to construct an estimate of the distortion at every symbol period for every channel. In this paper we adopt the generalization of the Wiener filter proposed in [9] as specified in (16). Although the Wiener filter was only used to estimate the distortion process of a single channel in [9], it is straightforward to apply it to the multiple-channel case. The minimum mean square error estimator of the state vector $\boldsymbol{c}$ takes the form

$$\hat{\boldsymbol{c}} = E[\boldsymbol{c}|\boldsymbol{x}_p]. \quad (21)$$

Due to the jointly Gaussian nature of the problem, this becomes a linear estimator

$$\hat{\boldsymbol{c}} = E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\mathrm{Cov}(\boldsymbol{x}_p)^{-1}\boldsymbol{x}_p \triangleq \boldsymbol{H}\boldsymbol{x}_p \quad (22)$$

where $\mathrm{Cov}(\boldsymbol{x}_p)$ is the covariance matrix of $\boldsymbol{x}_p$. Using the orthogonal projection theorem [16] of the minimum mean square error estimator, the covariance matrix of the channel state estimation error $\tilde{\boldsymbol{c}} \triangleq \boldsymbol{c} - \hat{\boldsymbol{c}}$ can be written as

$$\begin{aligned}\mathrm{Cov}(\tilde{\boldsymbol{c}}) &= \mathrm{Cov}(\boldsymbol{c}) - \mathrm{Cov}(\hat{\boldsymbol{c}}) \\ &= \mathrm{Cov}(\boldsymbol{c}) - E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\mathrm{Cov}(\boldsymbol{x}_p)^{-1}E\left[\boldsymbol{x}_p\boldsymbol{c}^H\right]. \end{aligned} \quad (23)$$

Note $\mathrm{Cov}(\tilde{\boldsymbol{c}})$ is precomputable.

C. Union Bound

In this subsection, we will analyze the error performance of the receiver architecture. Usually, the information stream being transmitted is binary. An outer encoder is used to map this stream into the code matrix symbols that have been discussed so far. Therefore, our real interest is the bit error probability (BEP), not the code matrix error probability.

Define the pairwise code matrix error event $\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\}$ as the event that the receiver decodes code matrix $\boldsymbol{D}_\beta$ when $\boldsymbol{D}_\alpha$ is actually transmitted. The BEP given that $\boldsymbol{D}_\alpha$ is transmitted is

$$P_B(E|\boldsymbol{D}_\alpha) = \sum_{\boldsymbol{D}_\beta \in \Xi, \boldsymbol{D}_\beta \neq \boldsymbol{D}_\alpha} P\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} \cdot \left(\frac{\text{number of error bits in event } \boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta}{\text{number of information bits per code matrix}} \right) \tag{24}$$

where

$$P\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} = P\Big\{ \ln[f(\boldsymbol{x}_\alpha|\boldsymbol{x}_p, \boldsymbol{D}_\beta)] \geq \max_{\boldsymbol{D}_\gamma \in \Xi, \boldsymbol{D}_\gamma \neq \boldsymbol{D}_\beta} \ln[f(\boldsymbol{x}_\alpha|\boldsymbol{x}_p, \boldsymbol{D}_\gamma)] \Big\} \tag{25}$$

where $\boldsymbol{x}_\alpha = \boldsymbol{\Lambda}_{\boldsymbol{D}_\alpha} \boldsymbol{c} + \boldsymbol{n}$.

Equation (25) is analytically intractable and is usually upper-bounded by a union bound (e.g., [1])

$$P\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} \leq P\{\ln[f(\boldsymbol{x}_\alpha|\boldsymbol{x}_p, \boldsymbol{D}_\beta)] \geq \ln[f(\boldsymbol{x}_\alpha|\boldsymbol{x}_p, \boldsymbol{D}_\alpha)]\} \triangleq P_1\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\}. \tag{26}$$

To evaluate (26), we can rewrite it in a quadratic form

$$P_1\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} = P\{\boldsymbol{z}^H \boldsymbol{Q}_{\alpha\beta} \boldsymbol{z} \leq \delta_{\alpha\beta}\} \tag{27}$$

where $\delta_{\alpha\beta} = \ln(|\boldsymbol{C}_\alpha|/|\boldsymbol{C}_\beta|)$, $\boldsymbol{z} = [\boldsymbol{x}_\alpha^T \hat{\boldsymbol{c}}^T]^T$ and as shown in (28) at the bottom of the page.

Since both $\boldsymbol{x}_\alpha$ and $\hat{\boldsymbol{c}}$ are complex Gaussian vectors, we can use results for the quadratic form of complex Gaussians. Using characteristic functions, contour integration, and some algebra, (27) can be evaluated by the following formula [1]:

$$P_1\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} = \begin{cases} -\sum \text{Residue}[\phi(s)\ \exp(s\delta_{\alpha,\beta})/s]_{\text{Right Plane poles}}, & \text{if } \delta \leq 0 \\ \sum \text{Residue}[\phi(s)\ \exp(s\delta_{\alpha,\beta})/s]_{\text{Left Plane poles} \cup \{0\}}, & \text{if } \delta > 0 \end{cases} \tag{29}$$

where

$$\phi(s) \triangleq E[\exp(-s\boldsymbol{z}^H \boldsymbol{Q}_{\alpha\beta} \boldsymbol{z})] = \frac{1}{|\boldsymbol{I}_{K+KL} + s\,\text{Cov}(\boldsymbol{z})\boldsymbol{Q}_{\alpha\beta}|} \tag{30}$$

is the characteristic function [17] of the random variable $\boldsymbol{z}^H \boldsymbol{Q}_{\alpha\beta} \boldsymbol{z}$. Note this is a slight generalization of the results in [1] since the channel and channel estimate are not assumed to be independent for each code symbol. The union bound on the bit error probability given that $\boldsymbol{D}_\alpha$ is transmitted is

$$P_B(E|\boldsymbol{D}_\alpha) \leq \sum_{\boldsymbol{D}_\beta \in \Xi, \boldsymbol{D}_\beta \neq \boldsymbol{D}_\alpha} P_1\{\boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta\} \cdot \left(\frac{\text{number of error bits in event } \boldsymbol{D}_\alpha \rightarrow \boldsymbol{D}_\beta}{\text{number of information bits per code matrix}} \right). \tag{31}$$

IV. Code Matrix Design

The above analysis provides a very accurate approximation to the real error performance of our modeled system. However, it does not provide much insight on the relationship between the system performance, the design of the coded modulation scheme, and the design of pilot symbol sequences. For the purpose of code matrix design, we will make some further assumptions to simplify the model.

From the analysis of Section III, it is clear that the code design will depend on the fading rate. However, our interest is in a nonadaptive coded modulation scheme. Therefore, we will design our code assuming stationary fading, and then show this provides acceptable performance in time-varying fading. With this assumption, the channel state remains constant over the interval of the entire code matrix, i.e., $c_{k+1}^{(l)} = c_{k+2}^{(l)} = \cdots = c_{k+K}^{(l)} \triangleq c^{(l)}$ for all l and (8) can be simplified as

$$\boldsymbol{x} = \boldsymbol{D}\boldsymbol{c}_s + \boldsymbol{n} \tag{32}$$

where $\boldsymbol{c}_s = [c^{(1)}\ \ c^{(2)}\ \ \cdots\ \ c^{(L)}]^T$ is an $L \times 1$ column vector whose elements are i.i.d. zero mean complex Gaussian random variables with variance E_s, and $\boldsymbol{n}$ is a $K \times 1$ vector whose elements are i.i.d. zero mean complex Gaussian random variables with variance N_0. Note that we have replaced the $K \times KL$ matrix $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ by $\boldsymbol{D}$ in the matrix representation of a code matrix given in (5), and the dimension of $\boldsymbol{c}_s$ is now $L \times 1$, not $KL \times 1$.

For the purposes of simplification and isolating the code design from the pilot symbol sequence design, we next assume that the channel state $\boldsymbol{c}_s$ can be perfectly estimated. With this assumption, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{c}_s$ and $\boldsymbol{D}_\gamma$ is then

$$f(\boldsymbol{x}|\boldsymbol{c}_s, \boldsymbol{D}_\gamma) = (\pi N_0)^{-K} \cdot \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{D}_\gamma \boldsymbol{c}_s)^H (\boldsymbol{x} - \boldsymbol{D}_\gamma \boldsymbol{c}_s)}{N_0} \right). \tag{33}$$

Following the derivation in Section III-C, we can write down the upper bound of the pairwise error event that code matrix

$$\boldsymbol{Q}_{\alpha\beta} = \begin{bmatrix} \boldsymbol{C}_\beta^{-1} - \boldsymbol{C}_\alpha^{-1} & \boldsymbol{C}_\alpha^{-1} \boldsymbol{\Lambda}_{\boldsymbol{D}_\alpha} - \boldsymbol{C}_\beta^{-1} \boldsymbol{\Lambda}_{\boldsymbol{D}_\beta} \\ \boldsymbol{\Lambda}_{\boldsymbol{D}_\alpha}^H \boldsymbol{C}_\alpha^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{D}_\beta}^H \boldsymbol{C}_\beta^{-1} & \boldsymbol{\Lambda}_{\boldsymbol{D}_\beta}^H \boldsymbol{C}_\beta^{-1} \boldsymbol{\Lambda}_{\boldsymbol{D}_\beta} - \boldsymbol{\Lambda}_{\boldsymbol{D}_\alpha}^H \boldsymbol{C}_\alpha^{-1} \boldsymbol{\Lambda}_{\boldsymbol{D}_\alpha} \end{bmatrix} \tag{28}$$

$\boldsymbol{D}_\alpha$ is decoded as code matrix $\boldsymbol{D}_\beta$ as

$$P_1\{\boldsymbol{D}_\alpha \to \boldsymbol{D}_\beta\} = P\{\ln f(\boldsymbol{x}_\alpha|\boldsymbol{D}_\alpha, \boldsymbol{c}_s) < \ln f(\boldsymbol{x}_\alpha|\boldsymbol{D}_\beta, \boldsymbol{c}_s)\} \quad (34)$$

or

$$P_1\{\boldsymbol{D}_\alpha \to \boldsymbol{D}_\beta\} = P\{(\boldsymbol{x}_\alpha - \boldsymbol{D}_\alpha\boldsymbol{c}_s)^H(\boldsymbol{x}_\alpha - \boldsymbol{D}_\alpha\boldsymbol{c}_s) > (\boldsymbol{x}_\alpha - \boldsymbol{D}_\beta\boldsymbol{c}_s)^H(\boldsymbol{x}_\alpha - \boldsymbol{D}_\beta\boldsymbol{c}_s)\}. \quad (35)$$

To simplify (35), we note since $\boldsymbol{D}_\alpha$ is assumed transmitted, we can replace $\boldsymbol{x}_\alpha$ by $\boldsymbol{D}_\alpha\boldsymbol{c}_s + \boldsymbol{n}$ and get

$$P_1\{\boldsymbol{D}_\alpha \to \boldsymbol{D}_\beta\} = P\{2\mathcal{R}[\boldsymbol{n}^H(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)\boldsymbol{c}_s] > \boldsymbol{c}_s^H(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)^H(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)\boldsymbol{c}_s\} \quad (36)$$

where $\mathcal{R}(\cdot)$ denotes the real part of its argument.

Due to the singular value decomposition theorem [18] any $K \times L$ matrix, $\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha$ can be written

$$\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha = \boldsymbol{V}^H\boldsymbol{\Sigma}\boldsymbol{W} \quad (37)$$

where $\boldsymbol{V}$ is a $K \times K$ unitary matrix, $\boldsymbol{W}$ is an $L \times L$ unitary matrix, and $\Sigma = [\sigma_{ij}]$ is a $K \times L$ nonnegative matrix whose elements are given as $\sigma_{ij} = 0$ for all $i \neq j$, and $\sigma_{11} \geq \sigma_{22} \geq \cdots \sigma_{\rho\rho} > \sigma_{\rho+1,\rho+1} = \cdots = \sigma_{qq} = 0$, where $q = \min\{K, L\}$ and ρ is the rank of (37). The numbers σ_{ii} are the nonnegative square roots of the eigenvalues of $(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)^H$. Now (36) is written as

$$P_1\{\boldsymbol{D}_\alpha \to \boldsymbol{D}_\beta\} = P\{2\mathcal{R}[\boldsymbol{n}^H\boldsymbol{V}^H\boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{c}_s] > \boldsymbol{c}_s^H\boldsymbol{W}^H\boldsymbol{\Sigma}^H\boldsymbol{V}\boldsymbol{V}^H\boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{c}_s\}. \quad (38)$$

Since $\boldsymbol{V}$ is a unitary matrix, $\boldsymbol{V}\boldsymbol{V}^H$ is a $K \times K$ identity matrix and by letting $\boldsymbol{n}' = \boldsymbol{V}\boldsymbol{n}$ and $\boldsymbol{c}_s' = \boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{c}_s$, (38) becomes

$$P_1\{\boldsymbol{D}_\alpha \to \boldsymbol{D}_\beta\} = P\{2\mathcal{R}[\boldsymbol{n}'^H\boldsymbol{c}_s'] > \boldsymbol{c}_s'^H\boldsymbol{c}_s'\} \quad (39)$$

where it can be easily shown that $\boldsymbol{n}'$ is a zero mean i.i.d. Gaussian random vector with covariance matrix $N_0\boldsymbol{I}_K$, and $\boldsymbol{c}_s'$ is a zero mean Gaussian random vector with a diagonal covariance matrix $E_s\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H$ having diagonal elements $\{E_s\sigma_{11}^2, E_s\sigma_{22}^2, \cdots, E_s\sigma_{\rho\rho}^2, 0, \cdots, 0\}$, and $\boldsymbol{n}'$ and $\boldsymbol{c}_s'$ are independent. The result in (39), while seemingly innocuous, implies a great deal about good signal design for multiple transmitters. First, the diversity gained in a multiple transmitter system L_+ is

$$L_+ = \text{Rank}[\boldsymbol{D}_\alpha - \boldsymbol{D}_\beta] \leq \min[L, K]. \quad (40)$$

Since it is typically easier to build longer codes than more transmitters, it is appropriate to assume $L_+ \leq L$ in stationary fading. Second, repetition coding across the antennas without a time shift is obviously not effective since then the rank of $[\boldsymbol{D}_\alpha - \boldsymbol{D}_\beta]$ would be 1. The transmitter switching scheme mentioned in the Introduction obviously achieves $L_+ = L$. Furthermore, (39) shows that appropriately designed codes can achieve $L_+ = L$ while simultaneously transmitting from all antennas, and this is our goal. Also note in a stationary fading channel that the full diversity available can be achieved with $K = L$, so the *only* advantage in choosing $K > L$ is to achieve a greater bandwidth efficiency.

Since at most L diversity levels are available for a system with L transmitters in stationary fading, and the error performance of a code depends on the metric between the code matrices defined as the diagonal elements of $\boldsymbol{\Sigma}$, i.e., the nonnegative square roots of the eigenvalues of $(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)^H$, we can summarize the codeword design criteria as follows.

1) An ideal codeword set should have full rank for all $\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha$ (so that the number of nonzero entries in $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H$ is $\min\{K, L\}$).
2) $\sigma_{11}^2, \sigma_{22}^2, \cdots, \sigma_{\rho\rho}^2$ should be evenly distributed and maximized subject to the total available signal energy.

A. An Example for Code Matrix Design

This example presents a simple block code design. More powerful codes can be designed, and subsequent to the completion of this work [12] they *have* been designed [19]. The upper half of Fig. 3 shows the encoder of the design example. In this example, we consider a code matrix of block length $K = L = 3$ with $m = 4/3$ bits per baud.[3] By letting $\boldsymbol{V} = \boldsymbol{I}_K$ and choosing $\boldsymbol{W}$ as a 3×3 constant-magnitude unitary matrix

$$\boldsymbol{W} = \frac{1}{\sqrt{3}}\begin{bmatrix} e^{j0} & e^{j0} & e^{j0} \\ e^{j0} & e^{j2\pi/3} & e^{j4\pi/3} \\ e^{j0} & e^{j4\pi/3} & e^{j8\pi/3} \end{bmatrix} \quad (41)$$

the code matrix can be expressed as

$$\boldsymbol{D} = \boldsymbol{\Gamma}\boldsymbol{W} \quad (42)$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix whose diagonal elements $\{d_1, d_2, d_3\}$ can be drawn from the alphabet of a QAM constellation. If a block of 4 information bits $\{I_1, I_2, I_3, I_4\}$ (assume each of them takes on values ± 1) is to be encoded, a simple scrambled repetition code matrix set is constructed as

$$\begin{aligned} d_1 &= \sqrt{\frac{2}{15}}\left((2I_1 + I_2) + j(2I_3 + I_4)\right) \\ d_2 &= \sqrt{\frac{2}{15}}\left((2I_4 + I_3) + j(2I_2 + I_1)\right) \\ d_3 &= \sqrt{\frac{2}{15}}\left((2I_3 + I_2) + j(2I_4 + I_1)\right) \end{aligned} \quad (43)$$

where a factor of $\sqrt{2/15}$ is introduced to normalize the energy. The resulting code matrix

$$\boldsymbol{D} = \frac{1}{\sqrt{3}}\begin{bmatrix} d_1e^{j0} & d_1e^{j0} & d_1e^{j0} \\ d_2e^{j0} & d_2e^{j2\pi/3} & d_2e^{j4\pi/3} \\ d_3e^{j0} & d_3e^{j4\pi/3} & d_3e^{j8\pi/3} \end{bmatrix} \quad (44)$$

will then have no zero elements and, therefore, no transmitter is shut off at anytime. Also, the metric between any pair of code matrices defined earlier as the set of nonnegative

[3] These parameters are chosen to be comparable to those in the simple block code in [11].

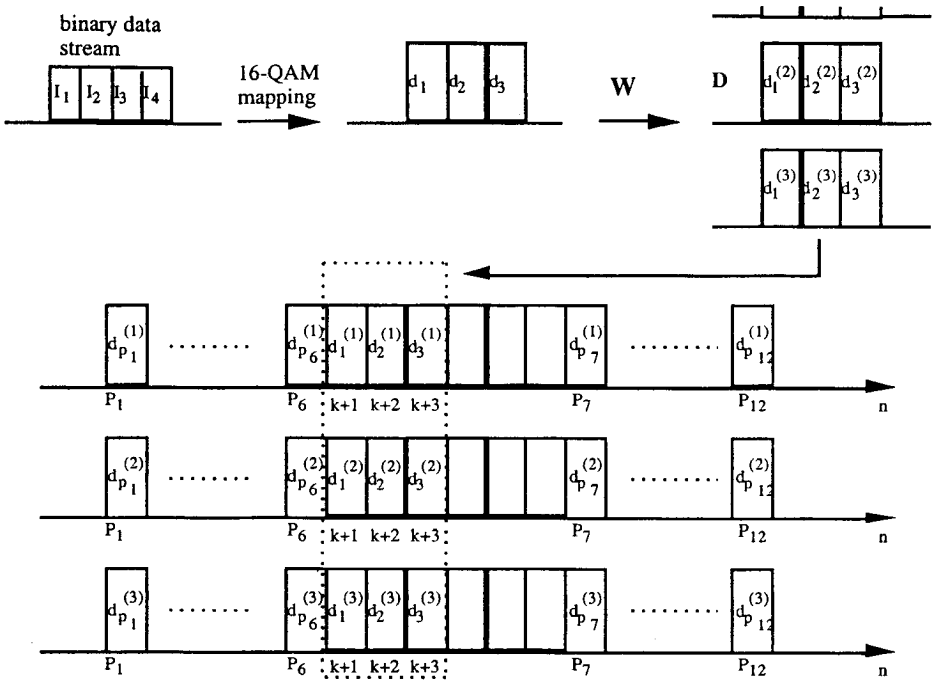


Fig. 3. Codeword design example.

eigenvalues of

$$\begin{aligned}(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)(\boldsymbol{D}_\beta - \boldsymbol{D}_\alpha)^H &= (\boldsymbol{\Gamma}_\beta \boldsymbol{W} - \boldsymbol{\Gamma}_\alpha \boldsymbol{W})(\boldsymbol{\Gamma}_\beta \boldsymbol{W} - \boldsymbol{\Gamma}_\alpha \boldsymbol{W})^H \\ &= (\boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}_\alpha)\boldsymbol{W}\boldsymbol{W}^H(\boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}_\alpha)^H \\ &= (\boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}_\alpha)(\boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}_\alpha)^H \end{aligned} \tag{45}$$

is just $\{|d_{\beta 1} - d_{\alpha 1}|^2, |d_{\beta 2} - d_{\alpha 2}|^2, |d_{\beta 3} - d_{\alpha 3}|^2\}$. It is easy to see that none of the eigenvalues is zero since the mapping in (43) is one to one for each symbol.

V. Pilot Symbol Matrix Designs

Although a direct relationship between the probability of pairwise error event and pilot symbol sequences is established in Section III, the optimization problem of minimizing the overall bit error probability over all possible pilot sequences is extremely difficult if not impossible. However, it is apparent that accurate channel state estimation will lead to better system performance. Therefore, once the code is determined, the error performance of the system will be decided by the accuracy of the channel state estimation, which in turn is determined by the error covariance matrix $\mathrm{Cov}(\tilde{\boldsymbol{c}}) \triangleq E[(\boldsymbol{c}-\hat{\boldsymbol{c}})(\boldsymbol{c}-\hat{\boldsymbol{c}})^H]$. A set of good pilot symbol sequences should then lead to an error covariance matrix that has the following properties.

1) The total estimation error over all channels is minimized.
2) The estimation errors are evenly distributed over all channels.
3) The cross correlation between channel state estimation errors, i.e., the off-diagonal elements of $\mathrm{Cov}(\tilde{\boldsymbol{c}})$, should be kept as small as possible.
4) The above properties should remain shift-invariant from frame to frame.

Furthermore, the associated optimal Wiener filter coefficients (see Section III) should be periodic with a short period so that a large memory is not needed to store the filter coefficients.

Let us first look at the total error variance of channel state estimates over an entire code matrix. This is just the trace of the error covariance matrix of $\hat{\boldsymbol{c}}$ which [using (23)] can be expressed as

$$\begin{aligned}&\mathrm{Tr}[\mathrm{Cov}(\tilde{\boldsymbol{c}})] \\ &\quad = \mathrm{Tr}\left[E\left[\boldsymbol{c}\boldsymbol{c}^H\right] - E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\mathrm{Cov}(\boldsymbol{x}_p)^{-1}E\left[\boldsymbol{x}_p\boldsymbol{c}^H\right]\right] \\ &\quad = \mathrm{Tr}\left[E\left[\boldsymbol{c}\boldsymbol{c}^H\right]\right] - \mathrm{Tr}\left[E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\mathrm{Cov}(\boldsymbol{x}_p)^{-1}E\left[\boldsymbol{x}_p\boldsymbol{c}^H\right]\right] \\ &\quad = \mathrm{Tr}\left[E\left[\boldsymbol{c}\boldsymbol{c}^H\right]\right] - \mathrm{Tr}\left[\mathrm{Cov}(\boldsymbol{x}_p)^{-1}E\left[\boldsymbol{x}_p\boldsymbol{c}^H\right]E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\right]\end{aligned} \tag{46}$$

where in the last equality we have used the property $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$. With some manipulation and noting that all channels are i.i.d., one can show that the elements of $\mathrm{Cov}(\boldsymbol{x}_p)$ are

$$[\mathrm{Cov}(\boldsymbol{x}_p)]_{i,j} = N_0\delta(i-j) + \left[E\left[\boldsymbol{c}_p^{(l)}\boldsymbol{c}_p^{(l)^H}\right]\right]_{i,j}\langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_j}\rangle \tag{47}$$

and the elements of $E[\boldsymbol{x}_p\boldsymbol{c}^H]E[\boldsymbol{c}\boldsymbol{x}_p^H]$ are

$$\begin{aligned}&\left[E\left[\boldsymbol{x}_p\boldsymbol{c}^H\right]E\left[\boldsymbol{c}\boldsymbol{x}_p^H\right]\right]_{i,j} \\ &\quad = \left[E\left[\boldsymbol{c}_p^{(l)}\boldsymbol{c}^{(l)^H}\right]E\left[\boldsymbol{c}^{(l)}\boldsymbol{c}_p^{(l)^H}\right]\right]_{i,j}\langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_j}\rangle\end{aligned} \tag{48}$$

where the superscripts (l) are arbitrary. Equations (47) and (48) therefore imply that (46) is completely determined by the inner products between $\boldsymbol{d}_{p_i}$'s. Hence, a sufficient condition for the pilot symbol sequences to be shift-invariant is (i.e., property 4 above) $\langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_{i+k}}\rangle = \langle \boldsymbol{d}_{p_j}, \boldsymbol{d}_{p_{j+k}}\rangle$ for all i, j, and k. A particular set of sequences with this property is

$$d_{p_i}^{(l)} = \frac{1}{\sqrt{L}}\exp\left(\frac{j2(\pi+\epsilon)(i-1)(l-1)}{L}\right) \tag{49}$$

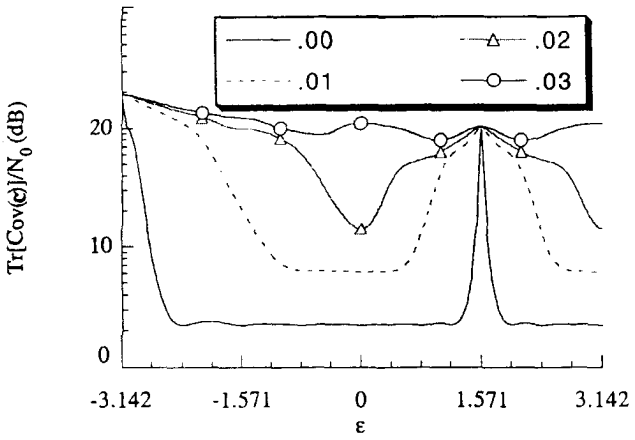


Fig. 4. Total estimation error variance as a function of ϵ for various fading rate $f_DT = 0.0, 0.01, 0.02$ and 0.03. $E_s/N_0 = 15$ dB, $P_{\text{ins}} = 7$, $N = 12$.

with inner product

$$\langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_{i+k}} \rangle = \frac{1}{L} \sum_{l=1}^{L} \exp\left(\frac{-j2(\pi+\epsilon)k(l-1)}{L} \right). \quad (50)$$

With such pilot sequences, (46) becomes a function of a single variable ϵ and can be calculated.

As an example, let $K = L = 3$, $P_{\text{ins}} = 7$, $E_s/N_0 = 15$ dB, and $N = 12$. Since the dimension of $\boldsymbol{c}$ is $KL = 9$, $\text{Tr}[E[\boldsymbol{c}\boldsymbol{c}^H]]/N_0 = 15 + 10 \log 9 = 24.54$ dB. Fig. 4 shows curves of the total normalized estimation error given in (46) as a function of ϵ for various fading rates. For $\epsilon = 0$, the $\boldsymbol{d}_{p_i}$'s in (49) are orthogonal. Therefore, from (46) we know its resulting channel state estimation error is the same as that of the antenna-switching system. The effective sampling period on each individual channel is then $LP_{\text{ins}} = 21$ instead of 7. From Fig. 4 we have noted that when this effective sampling rate is at or above Nyquist rate, $\epsilon = 0$ minimizes the total estimation error for pilot sequences of the form in (49) and achieves property 1 above.

Now, we will inspect the individual error variance $E[|\tilde{c}_k^{(l)}|^2]$ for a particular instant and channel. Recall that $\boldsymbol{P}^{(l)}$ is a diagonal matrix whose diagonal elements are the lth column of $\boldsymbol{P}$. Therefore, $\boldsymbol{P}^{(l)}$'s are invertible if and only if all elements of $\boldsymbol{P}$ are nonzero. If this is the case, then

$$\begin{aligned} E\left[|\tilde{c}_k^{(l)}|^2\right] &= E\left[|c_k^{(l)}|^2\right] - E\left[|\hat{c}_k^{(l)}|^2\right] \\ &= E\left[|c_k^{(l)}|^2\right] - E\left[c_k^{(l)} \boldsymbol{x}_p^H\right] \text{Cov}(\boldsymbol{x}_p)^{-1} E\left[\boldsymbol{x}_p c_k^{(l)^H}\right] \\ &= E\left[|c_k^{(l)}|^2\right] - E\left[c_k^{(l)} \boldsymbol{c}_p^{(l)^H}\right] \\ &\quad \cdot \left(\boldsymbol{P}^{(l)^{-1}} \text{Cov}(\boldsymbol{x}_p) \left(\boldsymbol{P}^{(l)^H} \right)^{-1} \right)^{-1} E\left[\boldsymbol{c}_p^{(l)} c_k^{(l)^H}\right]. \end{aligned} \quad (51)$$

Now let $\boldsymbol{U}^{(l)} = \boldsymbol{P}^{(l)^{-1}} \text{Cov}(\boldsymbol{x}_p)(\boldsymbol{P}^{(l)^H})^{-1}$, then one can show that the elements of $\boldsymbol{U}^{(l)}$ are

$$[\boldsymbol{U}^{(l)}]_{i,j} = \frac{N_0}{|d_{p_i}^{(l)}|^2} \delta(i-j) + \frac{E\left[c_{p_i}^{(l)} c_{p_j}^{(l)^*}\right]}{d_{p_i}^{(l)} d_{p_j}^{(l)^*}} \langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_j} \rangle. \quad (52)$$

A sufficient condition for (52) to be equal for all l and property 2 to be satisfied is

$$|d_{p_n}^{(l)}|^2 \text{ is a constant for all } l$$

and either

$$\langle \boldsymbol{d}_{p_i}, \boldsymbol{d}_{p_j} \rangle = 0 \qquad i \neq j$$

or

$$d_{p_i}^{(l)} d_{p_j}^{(l)^*} \text{ is a constant for all } l \text{ and } i \neq j.$$

That is, $\boldsymbol{d}_{p_i}$'s are either orthogonal or a scalar multiple of one another. The set of sequences in (49) with $\epsilon = 0$ satisfies these conditions.

If L divides N, one can also easily show that when the fading is stationary (i.e., $f_DT = 0$), the error covariance matrix is a block diagonal matrix

$$\text{Cov}(\tilde{\boldsymbol{c}}) = \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Lambda} \end{bmatrix} \quad (53)$$

where the elements of the $K \times K$ diagonal blocks are all

$$\boldsymbol{\Lambda}_{i,j} = \frac{E_s N_0}{\frac{N}{L} E_s + N_0}. \quad (54)$$

As long as the fading process is being sampled at or above Nyquist rate, we have noted that even in relatively fast fading (e.g., $f_DT = 0.01$), the magnitude of the off-diagonal entries of the error covariance matrix are still less than one percent of its diagonal elements. Thus, with appropriately designed pilot symbol sequences, we can obtain nearly independent estimations of the channel states by observing the same set of receiver outputs and satisfy property 3 above.

Therefore, we conclude that a good choice for the pilot symbol matrix can be obtained by collecting L constant-magnitude orthogonal row vectors $\boldsymbol{d}_{p_1}, \boldsymbol{d}_{p_2}, \cdots, \boldsymbol{d}_{p_L}$ and repeating them in the same order:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{d}_{p_1} \\ \vdots \\ \boldsymbol{d}_{p_L} \\ \boldsymbol{d}_{p_1} \\ \vdots \\ \boldsymbol{d}_{p_L} \\ \vdots \end{bmatrix} = \begin{bmatrix} d_{p_1}^{(1)} & d_{p_1}^{(2)} & \cdots & d_{p_1}^{(L)} \\ \vdots & \vdots & & \vdots \\ d_{p_L}^{(1)} & d_{p_L}^{(2)} & \cdots & d_{p_L}^{(L)} \\ d_{p_1}^{(1)} & d_{p_1}^{(2)} & \cdots & d_{p_1}^{(L)} \\ \vdots & \vdots & & \vdots \\ d_{p_L}^{(1)} & d_{p_L}^{(2)} & \cdots & d_{p_L}^{(L)} \\ \vdots & \vdots & & \vdots \end{bmatrix}. \quad (55)$$

The sequence given in (49) is a particular example. Similar ones can also be generated by concatenating binary or multiphase Hadamard matrices for various L. Take $L = 4$, for example, we have

$$\boldsymbol{P}^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdots \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \cdots \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \cdots \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \cdots \end{bmatrix}. \quad (56)$$

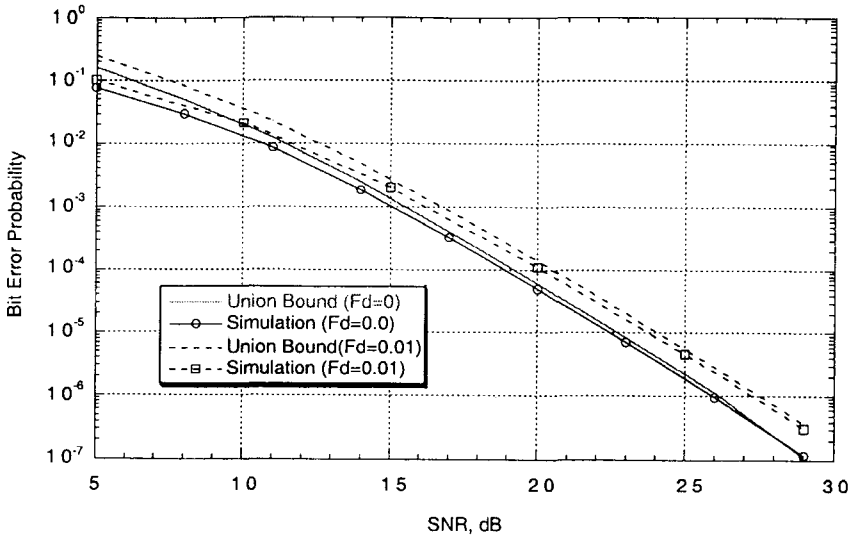


Fig. 5. BEP of the proposed system for $f_DT = 0$ and $f_DT = 0.01$.

VI. Simulation Results and Comparison

In this section, we present some simulation results for a system using the simple codes presented in Section IV-A ($L = 3$) and the pilot matrices for $L = 3$ discussed in the previous section. Note that an effort has been made to choose the system parameters like the simple block code in [11], so that a fair comparison can be made. Each 3×3 code matrix is produced by 4 information bits $\{I_1, I_2, I_3, I_4\}$. Two code matrices are placed between two consecutive pilot symbol slots, i.e., P_{ins} is 7. The pilot symbol sequences $\boldsymbol{P}$ used here take the form of (49) with $\epsilon = 0$ and $L = 3$. $N = 12$ nearest pilot observations are used to get the channel estimations for the code. This is shown in Fig. 3.

Since pilot symbols are inserted every P_{ins} symbols, the normalized information bit energy is

$$E_b = E_s \frac{P_{\text{ins}}}{P_{\text{ins}} - 1}. \tag{57}$$

The bit error probability of this system is then plotted in Fig. 5 for both $f_DT = 0$ and $f_DT = 0.01$ against E_b/N_0. Optimal Wiener filter coefficients and the decoding algorithm derived in Section III are used for both fading rates. That is, we assume an adaptive system exists that can keep updating the channel statistics and control the filter coefficients when channel statistics change. Both union bound results derived in Section III-C and simulation results are shown. The number of Monte Carlo simulations ranged from 10^6 to 10^8 depending on the error rate. Simulated Rayleigh fading is generated by the Jakes model [2]. Consequently, this figure demonstrates that the union bound is a very accurate performance estimator, and the remainder of the discussion will focus on the union bound analytical results.

The best performance occurred when the fading is stationary. Since there is no variation in the channel states with time, the channel state estimation is very accurate, and therefore we have lower error rates. As the fading rate increases, the channel state estimation becomes less accurate (due to a larger interpolation filter bandwidth) and error rates get higher. The difference between these two extremes is about 1 dB. Also note that for stationary fading, the channel state estimation for each channel is completely independent, while at higher fading rate, they are correlated. This also contributes to this larger gap between the two extremes, that is not characteristic of standard PSAM demodulation.

Next a direct comparison is made with the pilot symbol assisted frequency offset transmitter diversity system proposed in [11]. The union bound of the conditional BEP of the information block {1, 1, 1, 1} for the herein proposed system is plotted in Fig. 6. Also plotted in the same figure is the union bound of the conditional BEP of the same information block for the intentional frequency offset system studied in [11], which has the same parameters, except that a frequency offset of $f_0T = 0.005$ between antennas is used. Note the union bound presented here is the 2-D progressive union bound developed in [11]. The intentional frequency offset system requires interleaving to achieve independent fading for each symbol and for $f_0T = 0.005$, this depth is about 57. Both results of $f_DT = 0$ and $f_DT = 0.01$ are given. The former are shown in solid lines, while the latter are indicated by dashed lines.

The comparison between the herein proposed scheme and the intentional frequency offset scheme analyzed in [11] demonstrates several interesting characteristics. Compared to the intentional frequency offset method, our worst case ($f_DT = 0.01$) performance is approximately the same as its performance at stationary fading. The reason is as follows. For our method, since there are three channels that need to be estimated, the sampling period for each individual channel is in effect $3P_{\text{ins}} = 21$ instead of 7 as with the intentional frequency offset system. This is a critical value for $f_DT = 0.01$, and the channel state estimation is expected to be less accurate than that of the intentional frequency offset system. However, as discussed in Section V, the estimation errors of our method are nearly independent, whereas for the intentional frequency offset method, the estimation errors are highly correlated, especially when fading is slow, since the

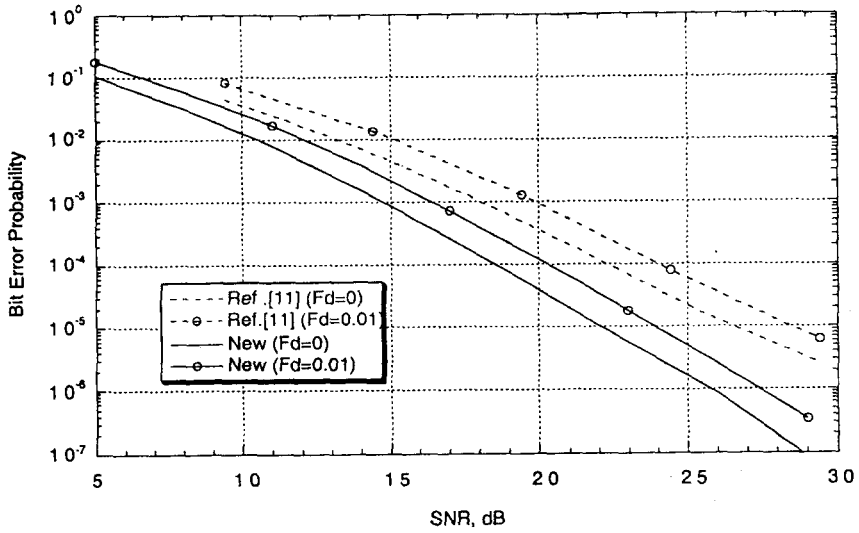


Fig. 6. Conditional BEP (only union bound results are shown) curves of the information block {1, 1, 1, 1} for the proposed scheme and the intentional frequency offset system.

pilot symbol observations are drawn from overlapped samples. This gives a little edge to our method and roughly cancels the degradation caused by the less accurate channel estimation.

One disadvantage of our scheme is, however, that it cannot operate at as high a Doppler spread as the frequency offset scheme, since the equivalent fading rate is L times the fading rate of an individual channel. In an intentional frequency offset system, on the other hand, this equivalent fading rate is approximately the larger one of $f_D T$ and $f_0 T$.

VII. Conclusion

We have considered transmitter diversity in wireless communication systems over Rayleigh fading channels using coded modulation with pilot symbol assisted channel state estimation. A new method of accessing diversity provided by multiple transmit antenna systems is proposed. Unlike the conventional phase-sweeping or frequency offset methods, the diversity is gained at the level of coded modulation instead of carrier phase or frequency, which most often will require some bandwidth expansion. Furthermore, this new method does not require an interleaver to separate symbols of a code to obtain independent fading effects or an equalizer. This significantly reduces the receiver complexity and avoids possible decoding delay.

Because of the special structure of the coded modulation scheme considered, the channel state estimator is required to estimate all individual channel states of the multiple transmitters, unlike the conventional methods where only a single channel is to be estimated. The design of such sequences is carefully studied, and their general constructions are derived. A detailed optimal receiver design and its performance analysis are also presented. Simulation results have shown that our method is superior to the conventional intentional frequency offset system with similar normalized parameters. Improvement is even more significant when the simplification of our system is taken into account.

References

[1] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulation in Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, pp. 74–83, Jan. 1992.

[2] W. C. Jakes, *Microwave Mobile Communication.* New York: Wiley, 1974.

[3] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex transmission system using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, pp. 49–60, 1994.

[4] V. Weerackody, "Characteristics of simulated fast fading indoor radio channel," in *Proc. 43rd IEEE Veh. Tech. Conf.*, Secaucus, NJ, 1993, pp. 231–235.

[5] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 170–176, May 1992.

[6] A. A. Saleh and L. J. Cimini, Jr., "Indoor radio communication using time-division multiple access with cyclical slow frequency hopping and coding," *IEEE J. Select. Areas Commun.*, vol. SAC-7, pp. 59–70, Jan. 1989.

[7] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, 1993, pp. 1630–1634.

[8] F. Davarian, "Mobile digital communication via tone calibration," *IEEE Trans. Veh. Technol.*, vol. VT-36, pp. 55–62, May 1987.

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

[10] C. Tellambura, Q. Wang, and V. K. Bhargava, "A performance analysis of trellis-coded modulation schemes over Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 491–501, Nov. 1993.

[11] W. Y. Kuo and M. P. Fitz, "Design and analysis of transmitter diversity using intentional frequency offset for wireless communications," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 871–881, Nov. 1997.

[12] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. 1996 IEEE Veh. Technol. Conf.*, Atlanta, GA, 1996, pp. 136–140.

[13] J. K. Cavers, "On the validity of the slow and moderate fading models for matched filter detection of Rayleigh fading signals," *Canadian J. Elect. Computer Eng.*, vol. 17, pp. 183–189, Oct. 1992.

[14] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, June 1990.

[15] J. P. Seymour and M. P. Fitz, "Near-optimal symbol-by-symbol demodulation algorithms for flat Rayleigh fading," *IEEE Trans. Commun.*, vol. 43, pp. 1525–1533, Feb. 1995.

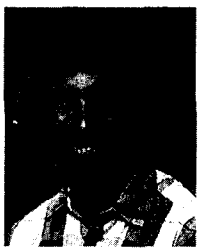
[16] H. W. Sorenson, *Parameter Estimation.* New York: Marcel Dekker, 1980.

[17] G. L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, pp. 199–201, 1960.
[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990, pp. 411–426.
[19] N. Seshadri, V. Tarokh, and A. Calderbank, "Space–time codes for wireless communications: Code construction," in *Proc. IEEE Veh. Technol. Conf.*, Phoenix, AZ, 1997, pp. 637–641.
[20] J. C. Guey and M. R. Bell, "Diversity waveform sets for high-resolution delay-Doppler imaging," *IEEE Trans. Inform. Theory*, to be published.

Jiann-Ching Guey was born in Taichung City, Taiwan, in 1965. He received the B.S. degree from National Sun Yat-Sen University, Taiwan, in 1987, the M.S. degree from Southeastern Massachusetts University (now University of Massachusetts at Dartmouth), North Dartmouth, MA, in 1991, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1995, all in electrical engineering.

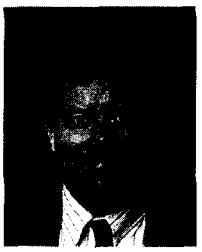
From 1987 to 1989, he served in the Army of Taiwan as an Officer in the communication corps. From 1989 to 1991, he was a Research Assistant at Southeastern Massachusetts University, working on ultrasonic nondestructive evaluation of materials. From 1992 to 1995, he was a Research Assistant at Purdue University, working in the areas of radar signal processing and wireless communication. Since July 1995, he has been with Ericsson Inc., Research Triangle Park, NC, where he is currently a Consulting Engineer in the Advanced Development and Research Group, working on the design and analysis of Ericsson's wireless communication products. His research interests are in the areas of statistical communication, information theory, signal processing and radar.

Dr. Guey was the recipient of a Purdue Research Foundation Fellowship from 1993 to 1995.



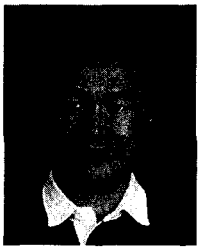
Michael P. Fitz was born on December 1, 1960, in Akron, OH. He received the B.E.E. degree (summa cum laude) from the University of Dayton, Dayton, OH, in 1983, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 1984 and 1989, respectively.

While at USC, he was a Communications System Engineer at both Hughes Aircraft Co., Fullerton, CA, and TRW Inc., Redondo Beach, CA. His responsibilities included the design, analysis, and testing of both satellite and portable communication systems. In 1989, he ventured into academia at Purdue University, West Lafayette, IN, and in 1996, he started his current position as an Associate Professor of Electrical Engineering at The Ohio State University, Columbus. His current research interests are in physical layer communication theory and include synchronization, signal design for fading channels, space-time modems, and equalization.



Mark R. Bell was born in Long Beach, CA, in 1959. He received the B.S. degree in electrical engineering from California State University, Long Beach, in 1981 and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, (Caltech), Pasadena in 1982 and 1988, respectively.

From 1979 to 1989, he was employed by Hughes Aircraft Company, Fullerton, CA. From 1981 to 1989, he was affiliated with the Radar Systems Laboratory at Hughes, where he held the positions of Member of the Technical Staff and Staff Engineering and worked in the areas of radar signal processing, electromagnetic scattering, radar target identification, and radar systems analysis. While at Caltech, he held a Howard Hughes Doctoral Fellowship from 1984 to 1988. Since 1989, he has been on the Faculty of Purdue University, West Lafayette, IN, where he is an Associate Professor in the School of Electrical and Computer Engineering. His research interests are in the areas of radar, information theory, detection and estimation, and communications.



Wen-Yi Kuo (S'88–M'95) was born in 1965 in Taiwan. He received the B.S. degree in communication engineering from National Chiao-Tung University, Taiwan, the M.S.E.E. degree from National Taiwan University, and the Ph.D. degree from Purdue University, West Lafayette, IN, in December 1994.

During his stay at Purdue University, he was engaged in research on fading channel communications. Since January 1995, he has been with Lucent Technologies (formerly AT&T) Bell Laboratories at Whippany, NJ, working on CDMA performance analysis and algorithm design. He is currently involved in third generation (IMT-2000) system design. His research interests include fading channel modeling, synchronization, coded modulation, diversity techniques, system capacity/coverage analysis, and burst control function for high speed data. He has 15 pending patents.

Dr. Kuo is a member of IEEE Technical Committee on Radio Communications and Technical Committee on Personal Communications.