Open access • Journal Article • DOI:10.1109/TWC.2003.809454

Training-based channel estimation for multiple-antenna broadband transmissions — Source link

Christina Fragouli, Naofal Al-Dhahir, William Turin

Institutions: Rutgers University, AT&T

Published on: 01 Mar 2003 - IEEE Transactions on Wireless Communications (IEEE)

Topics: Minimum mean square error, Block code and Phase-shift keying

Training-Based Channel Estimation for Multiple-Antenna Broadband Transmissions

Christina Fragouli, *Member, IEEE*, Naofal Al-Dhahir, *Senior Member, IEEE*, and William Turin, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of training sequence design for multiple-antenna transmissions over quasi-static frequency-selective channels. To achieve the channel estimation minimum mean square error, the training sequences transmitted from the multiple antennas must have impulse-like auto correlation and zero cross correlation. We reduce the problem of designing multiple training sequences to the much easier and well-understood problem of designing a single training sequence with impulse-like auto correlation. To this end, we propose to encode the training symbols with a space–time code, that may be the same or different from the space–time code that encodes the information symbols.**

Optimal sequences do not exist for all training sequence lengths and constellation alphabets. We also propose a method to easily identify training sequences that belong to a standard 2^m-PSK constellation for an arbitrary training sequence length and an arbitrary number of unknown channel taps. Performance bounds derived indicate that these sequences achieve near-optimum performance.

***Index Terms*—Channel estimation, space–time coding (STC), training sequence.**

I. Introduction

SPACE–TIME coding (STC) is a powerful wireless transmission technology that enables joint optimized design of the modulation, coding, and transmit diversity modules on wireless links. STC techniques of the trellis and block types were introduced in [1] and [2], respectively. A key attractive feature of all STC techniques is being *open loop*, i.e., channel knowledge is *not* required at the transmitter. While several noncoherent STC schemes that do not require channel information at the receiver as well have been developed [3]–[5], they suffer a significant performance penalty from coherent techniques. The noncoherent techniques are more suitable for rapidly-fading channels that experience significant variation within the transmission block. For quasi-static or slowly-varying fading channels, training-based channel estimation at the receiver is very common in practice. More specifically, current single-antenna wireless packet communication systems provide for a training sequence to be inserted in each packet[1] to aid in channel estimation at the receiver end. This motivates the need to develop practical high-performance training-based channel estimation algorithms for multiple-antenna systems. This can be easily achieved for narrowband transmissions (that encounter flat fading) by using orthogonal pilot training sequences (see e.g. [6]). For broadband multiple-antenna transmissions, training-based channel estimation presents several challenges and is the subject of this paper.

Consider the multiple-transmit single-receive[2] transmission scenario. The receiver observes the superposition of training sequences transmitted through different channels. The training sequences that achieve the channel estimation minimum mean square error (MMSE)[3] have an impulse-like auto-correlation sequence and zero cross correlation. This last property makes the channel estimation problem different for multiple-antenna systems from single-antenna systems, and has motivated research in this area.

Training-based estimation for a single-input–single-output (SISO) frequency-selective channel has been widely investigated in the literature (see for example [7] and the references therein). For the multiple-transmit-antenna scenario, a straightforward method to achieve zero cross correlation is to transmit training symbols only from one antenna at a time. This approach results in a high peak-to-average power ratio and, hence, is undesirable in practice.

For implementation purposes (to avoid nonlinear amplifier distortion), it is desirable to use constant-amplitude training sequences which can be classified in two main categories according to the training symbol alphabet size N.

The first approach [8] constructs optimal sequences[4] from an N^{th} root-of-unity alphabet $A_N = \{\exp((i2\pi k)/N) | k = 0 \dots N-1\}$, without constraining the alphabet size N. Such sequences are the perfect roots-of-unity sequences (PRUS) or polyphase sequences that have been proposed in the literature for different applications (see [9] and the references therein). For any training sequence length N_t, there exist optimal training sequences that belong to an N^{th} root-of-unity alphabet. The training sequence length N_t determines the smallest possible alphabet size. Chu [10] shows that for any length N_t there exists

Manuscript received May 24, 2001; revised March 14, 2002; accepted March 14, 2002. The editor coordinating the review of this paper and approving it for publication is W. Lu. This paper was presented in part at the WCNC 2002 and ICC 2002.

C. Fragouli was with AT&T Shannon Laboratory, Florham Park, NJ 07932 USA. She is now a Visiting Scholar at DIMACS, Rutgers University, Piscataway, NJ 08854-8018 USA (e-mail: fragouli@dimacs.rutgers.edu).

N. Al-Dhahir and W. Turin are with the AT&T Shannon Laboratory, Florham Park, NJ 07932 USA (email: naofal@research.att.com; wt@research.att.com).

Digital Object Identifier 10.1109/TWC.2003.809454

[1]The terms packet, block, and burst will be used interchangeably in this paper.

[2]The number of receiver antennas does not affect the training setup as long as the antennas are spaced sufficiently apart to ensure uncorrelated received signals.

[3]Throughout this paper, we use the mean square error as a performance metric for channel estimation. Although achieving the MMSE is not always equivalent to the best bit-error rate (BER), typically it gives a very good indication of the expected BER performance and is analytically tractable.

[4]By optimal sequences we mean sequences that achieve the channel estimation MMSE.

a PRUS with alphabet size $N = 2N_t$, and Mow [11] shows that for some N_t, smaller alphabet sizes are possible.

The second approach in the literature constrains the training sequence symbols to belong to a specific constellation, typically binary phase-shift keying (BPSK) or quaternary phase-shift keying (QPSK), to have a simpler transmitter/receiver implementation [12]. In this case, optimal sequences do not exist for all training lengths N_t. Instead, exhaustive searches can identify suboptimal sequences according to some performance criteria. Tables of such sequences from a BPSK alphabet are provided, for example, in [12].

The training sequence best suited to a particular application depends on the training sequence length N_t (which for standardized systems is predetermined), the number of channels taps to estimate, and the signal constellation used. A PRUS of a predetermined length may not belong to a standard constellation, while exhaustive searches are in many cases computationally prohibitive. For a system with M transmit antennas over frequency-selective channels with L taps each, an exhaustive search must identify M training sequences.

As an example, in the third generation TDMA proposal enhanced data for GSM evolution (EDGE) [13], 8-PSK constellation symbols are transmitted in blocks of $N_i = 116$ information symbols, and $N_t = 26$ training symbols. No optimal training sequence exists for this N_t and constellation. For two-transmit antennas, an exhaustive search would involve $8^{2\times 26}$ sequences. Restricting the training sequence alphabet size to BPSK would reduce the search space to $2^{2\times 26}$ sequences, which is still large and would increase the achievable MMSE.

This paper proposes a method to easily identify training sequences for multiple transmit antennas that enjoy the following attractive properties.

1) They belong to a standard constant-amplitude signal constellation of size 2^m, $m = 1, 2, 3, \ldots$, such as BPSK, QPSK, 8-PSK, etc.
2) They can be easily identified or constructed for an arbitrary training sequence length N_t and an arbitrary number of unknown channel taps L.
3) They result in negligible MSE increase from the lower bound.

The main idea is to reduce the problem of designing multiple training sequences with impulse-like auto correlation and zero cross correlation to designing a single training sequence with impulse-like auto correlation. This makes exhaustive searches more practical and, thus, facilitates the identification of good training sequences. In some cases, no search is necessary since optimal sequences are available from published results in the literature. Moreover, when optimal sequences do not exist, instead of exhaustive searches we propose a method that identifies suboptimal sequences from a standard signal constellation with a small MSE increase from the respective lower bound.

This paper is organized as follows. Section II presents the channel model and formally defines the optimal training sequences. Section III proposes three methods to generate multiple training sequences starting from a single one. Section IV introduces "L-perfect" sequences, investigates a method (alternative to exhaustive search) to identify suboptimal training sequences when optimal training sequences do not exist, and derives bounds on the performance loss. Section V presents simulation results and the paper is concluded in Section VI.

II. Channel Model and Optimal Training Sequences

Consider a system that employs two-transmit and one-receive antennas. The analysis can be generalized to multiple transmit/receive antennas. Two signals $\mathbf{s}_1$ and $\mathbf{s}_2$ are simultaneously transmitted over two frequency-selective channels $\mathbf{h}_1 = [h_1(0) \ldots h_1(L-1)]^T$ and $\mathbf{h}_2 = [h_2(0) \ldots h_2(L-1)]^T$, where $(\cdot)^T$ denotes the transpose operation. Each channel is modeled as a finite-impulse response (FIR) filter with L taps. The received signal at time k can be expressed as

$$y(k) = \sum_{i=0}^{L-1} h_1(i) s_1(k-i) + \sum_{i=0}^{L-1} h_2(i) s_2(k-i) + z(k) \quad (1)$$

where $z(k)$ is assumed to be additive white Gaussian noise (AWGN). The input sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ belong to a finite-signal constellation and are transmitted in data blocks where each block consists of N_i information symbols and N_t training symbols. For two-transmit antennas, the receiver uses the $2N_t$ known training symbols to estimate the $2L$ unknown channel coefficients. We assume that the channels $\mathbf{h}_1$ and $\mathbf{h}_2$ remain constant over the transmission of a block and vary independently from block to block (quasi-static assumption).

The observed training sequence output that does not have interference from information or preamble symbols can be expressed as

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{z} = [\mathbf{S}_1(L, N_t)\ \mathbf{S}_2(L, N_t)] \begin{bmatrix} \mathbf{h}_1(L) \\ \mathbf{h}_2(L) \end{bmatrix} + \mathbf{z} \quad (2)$$

where $\mathbf{y}$ and $\mathbf{z}$ are of dimension $(N_t - L + 1) \times 1$, $\mathbf{h}_1$ and $\mathbf{h}_2$ are of dimension $L \times 1$, $\mathbf{S}_1$ and $\mathbf{S}_2$ are Toeplitz matrices of dimension $(N_t - L + 1) \times L$, and

$$\mathbf{S}_i(L, N_t) = \begin{bmatrix} s_i(L-1) & \ldots & s_i(0) \\ s_i(L) & \ldots & s_i(1) \\ \vdots & & \vdots \\ s_i(N_t - 1) & \ldots & s_i(N_t - L) \end{bmatrix}$$

for $i = 1, 2$. The linear least square channel estimates, assuming that $\mathbf{S}$ has full column rank, can be calculated as [14]

$$\hat{\mathbf{h}} = \begin{bmatrix} \hat{\mathbf{h}}_1 \\ \hat{\mathbf{h}}_1 \end{bmatrix} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{y} \quad (3)$$

where $(\cdot)^H$ and $(\cdot)^{-1}$ denote the complex-conjugate (Hermitian) transpose and the inverse, respectively. For zero-mean Gaussian noise, the channel estimator is unbiased (i.e., $E[\hat{\mathbf{h}}] = \mathbf{h}$). The channel estimation MSE is defined as

$$\mathrm{MSE} = E\left[(\mathbf{h} - \hat{\mathbf{h}})^H (\mathbf{h} - \hat{\mathbf{h}})\right] = 2\sigma^2 tr\left((\mathbf{S}^H \mathbf{S})^{-1}\right) \quad (4)$$

where we assume white noise with auto-correlation matrix $R_z = E\mathbf{z}\mathbf{z}^H = 2\sigma^2 \mathbf{I}_{N_t - L + 1}$, $\mathbf{I}_n$ denotes the identity matrix of dimension $n \times n$, and $tr(\cdot)$ denotes the trace of a matrix. The MMSE is equal to

$$\mathrm{MMSE} = \frac{2\sigma^2 L}{(N_t - L + 1)} \quad (5)$$

which is achieved if and only if [8]

$$\mathbf{S}^H\mathbf{S} = \begin{bmatrix} \mathbf{S}_1^H\mathbf{S}_1 & \mathbf{S}_2^H\mathbf{S}_1 \\ \mathbf{S}_1^H\mathbf{S}_2 & \mathbf{S}_2^H\mathbf{S}_2 \end{bmatrix} = (N_t - L + 1)\mathbf{I}_{2L}. \quad (6)$$

The sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ that satisfy (6) are henceforth referred to as *optimal* sequences. Equation (6) implies that the optimal sequences have an impulse-like auto-correlation sequence and zero cross correlation.

III. Complexity Reduction Methods

This section provides three methods to reduce the complexity of designing training sequences for multiple-antenna systems to that of designing a single training sequence.

A. Use of Subsequences

A straightforward method to design two optimal training sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ of length N_t to estimate two channels each of L taps, is to design instead a single training sequence $\mathbf{s}$ of length $N_t' = N_t + L + 1$ to estimate a single channel with $L' = 2L$ taps

$$\mathbf{y} = \mathbf{S}\,(L', N_t')\,\mathbf{h}(L') + \mathbf{z} \quad (7)$$

where $\mathbf{S}(L', N_t')$ is a Toeplitz matrix of dimension $(N_t' - L' + 1) \times L' = (N_t - L + 1) \times 2L$. Again, for optimality, we require that

$$\mathbf{S}^H\,(L', N_t')\,\mathbf{S}\,(L', N_t') = (N_t - L + 1)\mathbf{I}_{2L} \quad (8)$$

and construct the sequences $\mathbf{s}_1$ and $\mathbf{s}_1$ as

$$\mathbf{s}_1 = [s(0) \dots s(N_t)]\,,\ \mathbf{s}_2 = [s(L) \dots s(N_t + L)]\,.$$

Thus, the multiple-training-sequence design problem can now be reduced to designing a single, but longer, optimal sequence $\mathbf{s}$ that achieves the MMSE when estimating the longer channel impulse response with L' taps. A similar approach can be followed for more than two-transmit antennas.

In the case where an optimal sequence of length N_t does not exist, an exhaustive search over all independent sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ may achieve a lower MSE than a search that uses the above described construction method.

B. Block Code for Training Symbols

We propose to encode the training symbols with a simple block code that takes an input sequence $\mathbf{s}$ with impulse-like auto correlation, and produces sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ with zero cross correlation. The code can be described by a block matrix $\mathbf{U}$ applied to the training matrix $\mathbf{S}$ that corresponds to the input training sequence $\mathbf{s}$. The received output can be expressed as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix} \underbrace{\begin{bmatrix} -\mathbf{I}_L & \mathbf{I}_L \\ \mathbf{I}_L & \mathbf{I}_L \end{bmatrix}}_{\mathbf{U}} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \quad (9)$$

where $\mathbf{S}$ is of dimension $(N_t - L + 1) \times L$. Any orthogonal matrix $\mathbf{U}$ such that $\mathbf{U}^H\mathbf{U} = n\mathbf{I}_{2L}$, where n is the number of transmit antennas, leads to an equivalent (in terms of MSE) block code. Multiplying the received output with the transpose-conjugate matrix we get

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \begin{bmatrix} 2\mathbf{S}^H\mathbf{S} & 0 \\ 0 & 2\mathbf{S}_2^H\mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \end{bmatrix}$$

where the noise now becomes

$$\begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \end{bmatrix} = \begin{bmatrix} -\mathbf{S}^H & \mathbf{S}^H \\ \mathbf{S}^H & \mathbf{S}^H \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}. \quad (10)$$

If we choose $\mathbf{S}^H\mathbf{S} = (N_t - L + 1)\mathbf{I}_L$, then

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = 2(N_t - L + 1) \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \end{bmatrix}. \quad (11)$$

The MSE of this scheme achieves the lower bound. The linear processing at the receiver does not color the white noise in (9). If instead of an optimal sequence $\mathbf{s}$ we use a sequence with good auto-correlation properties, the block code structure would still guarantee the orthogonality between $\mathbf{s}_1$ and $\mathbf{s}_2$. However, in this case the noise will not be white.

This method assumes no interference between two consecutive transmitted training blocks.

C. Equivalent Channel Estimation

The training methods in the literature employ independent training sequences $\mathbf{s}_1$ and $\mathbf{s}_2$ to estimate the channels $\mathbf{h}_1$ and $\mathbf{h}_2$. In contrast, in this section we propose to use a single training sequence encoded by the same space–time encoder as the information symbols, to estimate the overall equivalent channel that incorporates the space–time code.

This method is of interest when an exhaustive search is to be used. The code structure imposes a constraint on the possible generated $\mathbf{s}_1$ and $\mathbf{s}_2$ sequences, which amounts to a reduction of the search space from N^{2N_t} to N^{N_t} (assuming equal input and output alphabet size N and two-transmit antennas), making exhaustive searches more practical and, thus facilitating the identification of good training sequences. The search space can be further reduced by exploiting special characteristics of the employed space–time code. In the following, we examine a space–time trellis example and a space–time block code example.

1) Trellis Code Example: For a space–time trellis code with m memory elements over channels with memory $(L - 1)$, the receiver can incorporate the space–time trellis code structure in the channel model to create an equivalent SISO channel $\mathbf{h}_{\text{eq}}$ of length $m + L$.

Consider for example the 8-state 8-PSK space–time trellis code [1] for two-transmit and one-receive antennas. The transmitted signals at time k for this code can be expressed as

$$s_1(k) = s(k),\ s_2(k) = (-1)^{p_{k-1}} s(k-1)$$

where $p_k = u_{1,k} + 2u_{2,k} + 4u_{3,k}$ takes values in $\{0, 1, \dots 7\}$, and $[u_{1,k}\ u_{2,k}\ u_{3,k}]$ denote the code binary inputs. The received signal at time k can be expressed as

$$y(k) = \sum_{i=0}^{L} h_{\text{eq}}(i, k) s(k-1) + z(k)$$

where $h_{\rm eq}(i,k)$ are the taps of the equivalent input-dependent channel given by

$$h_{\rm eq}(i,k)=\begin{cases} h_1(0), & \text{for } i=0\\ (-1)^{p_{k-L}}h_2(L-1), & \text{for } i=L\\ h_1(i)+(-1)^{p_{k-i}}h_2(i-1), & \text{for } 0<i<L.\end{cases} \tag{12}$$

Note that the number of unknowns is $2L$ whether we estimate $\mathbf{h}_{\rm eq}$, or $\mathbf{h}_1$ and $\mathbf{h}_2$. Thus, we can reduce the training sequence search space from 2^{2N_t} to 2^{N_t} without increasing the number of unknowns to estimate. The search space can be further reduced by taking advantage of the special structure of the code.

For a given block and constant $\mathbf{h}_1$ and $\mathbf{h}_2$, the input sequence determines the sequence of $\mathbf{h}_{\rm eq}$ values. By transmitting only "even" training symbols from the constellation subset $C_e=\{0,2,4,6\}$, we observe the $L+1$ taps

$$\mathbf{h}_e=[h_1(0)\ h_1(i)+h_2(i-1)\ h_2(L-1)]$$

while by transmitting only "odd" training symbols from the set $C_o=\{1,3,5,7\}$, we observe the $L+1$ taps

$$\mathbf{h}_o=[h_1(0)\ h_1(i)-h_2(i-1)-h_2(L-1)]$$

where $1\le i\le L-1$. To estimate $\mathbf{h}_e$, we would use training symbols in the C_e subconstellation, while to estimate $\mathbf{h}_o$ we would use training symbols in the C_o subconstellation. To estimate both, we propose to use half of the input training symbols from each subconstellation, on the basis that for L large enough the suboptimality incurred will be negligible. That is, for even training length N_t, we propose to use a training sequence of the form $\mathbf{s}=[\mathbf{s}_e\ \mathbf{s}_o]$ where $\mathbf{s}_e$ has length $N_t/2$ and takes values in the C_e subconstellation and $\mathbf{s}_o$ has length $N_t/2$ and takes values in the C_o subconstellation. Note that $\mathbf{s}_e$ and $\mathbf{s}_o$ related as

$$\mathbf{s}_o=a\,\mathbf{s}_e\ :\ \text{for } a=\exp\left(\frac{i\pi k}{4}\right) \tag{13}$$

and any $k=1,3,5,7$, achieve the same MSE for the estimation of $\mathbf{h}_e$ and $\mathbf{h}_o$, respectively. Thus, instead of searching over all possible 8^{N_t} sequences $\mathbf{s}$, we can restrict the search space to the $4^{N_t/2}$ sequences $\mathbf{s}_e$.

An exhaustive search can identify sequences $\mathbf{s}_e$ and $\mathbf{s}_o=a\,\mathbf{s}_e$ such that the matrix $\mathbf{S}$ corresponding to the overall training sequence achieves minimum MSE. Section V provides a table of such sequences and simulation results.

2) Block Code Example: This section presents training schemes suitable for the space–time block code in [15] which is an extension of the code in [2] for frequency-selective channels. The encoder maps two consecutive input blocks $\mathbf{s}_1$ and $\mathbf{s}_2$ to the blocks $[\mathbf{s}_1\ -\tilde{\mathbf{s}}_2^*]$ and $[\mathbf{s}_2\ \tilde{\mathbf{s}}_1^*]$ to be transmitted from the two antennas. The operation denoted by $\tilde{(\cdot)}$ refers to time-reversing a sequence, that is, if $\mathbf{s}=[s(0)\ s(1)\ \dots\ s(N_t-1)]$, then $\tilde{\mathbf{s}}=[s(N_t-1)\ \dots\ s(1)\ s(0)]$, and $(\cdot)^*$ refers to component-wise complex conjugation. Assume that the block code is applied to the training symbols and that the channels $\mathbf{h}_1$ and $\mathbf{h}_2$ to be estimated remain constant over two blocks. The received signals during the first and second blocks denoted by $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively, can be expressed in matrix notation as

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\underbrace{\begin{bmatrix}-\tilde{\mathbf{S}}_2^* & \tilde{\mathbf{S}}_1^*\\ \mathbf{S}_1 & \mathbf{S}_2\end{bmatrix}}_{\mathbf{S}}\begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{bmatrix}+\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix} \tag{14}$$

where the matrices $\mathbf{S}_i$ and $\tilde{\mathbf{S}}_i$ (for $i=1,2$) are of dimension $(N_t-L+1)\times L$. Then

$$\mathbf{S}^*\mathbf{S}=\begin{bmatrix}\tilde{\mathbf{S}}_2^T\tilde{\mathbf{S}}_2^*+(\mathbf{S}_1^*)^T\mathbf{S}_1 & -\tilde{\mathbf{S}}_2^T\tilde{\mathbf{S}}_1^*+(\mathbf{S}_1^*)^T\mathbf{S}_2\\ -\tilde{\mathbf{S}}_1^T\tilde{\mathbf{S}}_2^*+(\mathbf{S}_2^*)^T\mathbf{S}_1 & \tilde{\mathbf{S}}_1^T\tilde{\mathbf{S}}_1^*+(\mathbf{S}_2^*)^T\mathbf{S}_2\end{bmatrix}.$$

From (6), a necessary condition to achieve the MSE lower bound is to have zero cross-correlation terms, i.e.,

$$-\tilde{\mathbf{S}}_1^T\tilde{\mathbf{S}}_2^*+(\mathbf{S}_2^*)^T\mathbf{S}_1=0=\left(-\tilde{\mathbf{S}}_2^T\tilde{\mathbf{S}}_1^*+(\mathbf{S}_1^*)^T\mathbf{S}_2\right)^T.$$

If $\mathbf{J}_{N_t}=[\delta_{i,N_t-i+1}]$ denotes the $N_t\times N_t$ square reversion ("backward identity") matrix, then we have the equivalent condition

$$\mathbf{J}_L\left((\mathbf{S}_2^*)^T\mathbf{S}_1\right)^T\mathbf{J}_L=(\mathbf{S}_2^*)^T\mathbf{S}_1. \tag{15}$$

Moreover, an additional requirement to achieve the MSE lower bound is that

$$\tilde{\mathbf{S}}_2^T\tilde{\mathbf{S}}_2^*+(\mathbf{S}_1^*)^T\mathbf{S}_1=2(N_t-L+1)\mathbf{I}_L. \tag{16}$$

Two simple choices that satisfy conditions (16) and (15) are the following.

1) $(\mathbf{S}_1^*)^T\mathbf{S}_1=(N_t-L+1)\mathbf{I}_L$, $\tilde{\mathbf{S}}_1=\mathbf{S}_1$, and $\mathbf{S}_2=\mathbf{S}_1$. That is, identify a sequence $\mathbf{s}_1$ symmetric about its center with impulse-like auto correlation and set $\mathbf{s}_2=\mathbf{s}_1$.
2) $(\mathbf{S}_1^*)^T\mathbf{S}_1=(N_t-L+1)\mathbf{I}_L$ and $\mathbf{S}_2=\tilde{\mathbf{S}}_1$. That is, identify a sequence $\mathbf{s}_1$ with impulse-like auto correlation and set $\mathbf{s}_2=\tilde{\mathbf{s}}_1$.

IV. Training Sequences Construction

This section addresses the design of a single training sequence of length N_t used to estimate L channel taps without restricting its use to multiple or single antenna systems.

A. *Perfect and L-perfect Training Sequences*

A root-of-unity sequence with alphabet size N has complex root-of-unity elements of the form $\{\exp((i2\pi x)/N)\}$, with $x=0\dots N-1$. The N roots of unity define a constant-amplitude finite-size constellation.

A sequence is said to be *perfect* if all of its out-of-phase periodic auto correlation terms are equal to zero [11]. The periodic auto correlation of a sequence $\mathbf{s}$ of length N_t at shift t can be calculated as$\theta(t)=\sum_{n=0}^{N_t-1}s^*(n)s((n+t)\bmod N_t)$. A unified construction method in [9] and [11] constructs a PRUS of any length $N_{\rm PRUS}$ but with alphabet size N determined by $N_{\rm PRUS}$. We are only interested in alphabets of size $N=2^m$, $m\ge 1$. The construction method in [11] for this alphabet can only produce PRUS of length $N_{\rm PRUS}$ which is also a power of two.

A sequence $\mathbf{s}$ of length N_t is called $L-perfect$ if the corresponding training matrix $\mathbf{S}$ of dimension $(N_t-L+1)\times L$ [constructed as in (7)] satisfies $\mathbf{S}^H\mathbf{S}=(N_t-L+1)\mathbf{I}_L$. Thus,

TABLE I
L-PERFECT SEQUENCES THAT EXIST FOR THE MINIMUM LENGTH AND FROM SMALL-SIZE CONSTELLATIONS

L=2 taps												
const. N_t	3	2	5	6	7	8	9	10	11	12	13	14
BPSK	4	-	*12	-	40	-	*140	-	504	-	*x	-
QPSK	*16	-	*144	-	*1600	-	*x	-	*x	-	*x	-
8PSK	*64	-	*1344	-	*x	-	*x	-	*x	-	*x	-
L=3 taps												
const. N_t	3	2	5	6	7	8	9	10	11	12	13	14
BPSK				*8	-	-	-	*72	-		-	*x
QPSK				*64	-	192	-	*3776	-		-	*x
8PSK				*512	-	768	-	*x	-		-	*x
L=4 taps												
const. N_t	3	2	5	6	7	8	9	10	11	12	13	14
BPSK					*8	-	-	-	*40	-		-
QPSK					*64	-	-	-	*1216	-		-
8PSK					*512	-			*x	-		-

an L-perfect sequence of length N_t is optimal [i.e., achieves the MMSE in (5)] for a channel with L taps.

Proposition 1: The length N_t of an L-perfect sequence from a 2^m-constellation can only be equal to

$$N_t = \left\{ \begin{array}{ll} 2(L+i), & \text{for } L\text{=odd} \\ 2(L+i)-1, & \text{for } L\text{=even} \end{array} \right\} \tag{17}$$

for i nonnegative integer.

The proof uses the fact that for 2^m roots-of-unity and any $x_1,\ x_2 \in \{2^m\text{roots-of-unity}\}$, there does not exist $x_3 \in \{2^m\text{roots-of-unity}\}$ such that $x_1 + x_2 = x_3$, which implies that the number of rows of $\mathbf{S}$ which is $(N_t - L + 1)$ has to be an even number. Equation (17) is a necessary (but not sufficient) condition for L-perfect sequences of length N_t to exist.

From a perfect sequence of length N_{PRUS}, we can build L-perfect training sequences to estimate up to $L = N_{\text{PRUS}}$ unknowns, that have length $N_t = kN_{\text{PRUS}} + L - 1$ for any $k \geq 1$ integer. These L-perfect sequences can be constructed by repeating k times the perfect sequence and circularly extending it by $L - 1$ symbols.

Table I shows exhaustive search results for L-perfect sequences for some small L and N_t. The "*" indicates that all or some of the existing sequences can be generated from perfect sequences. The "x" shows that such sequences exist but we do not know their exact number. The "$-$" indicates that such sequences do not exist. For our search range, perfect sequences could be used to construct most, but not all, of the L-perfect sequences. L-perfect sequences exist for a broader range of N_t than perfect sequences can provide.

B. Suboptimal Sequences Construction

L-perfect sequences do not exist for all training sequence lengths N_t and alphabets, or may be computationally intensive to identify. For example, for a specific L, if there exists an L-perfect sequence of length N_t, then from (17), there does not exist an L-perfect sequence of length $N_t + 1$. Next, we propose a method to construct suboptimal sequences.

Assume that N_t symbols from a specific alphabet are available to estimate L unknowns. Express N_t as $N_t = kN_{\text{PRUS}} + L - 1 + M$ for a nonnegative integer M and for a PRUS from the desired alphabet. Choose the value of $N_{\text{PRUS}} \geq L$ that minimizes M. Construct the L-perfect sequence of length $kN_{\text{PRUS}} + L - 1$ and extend it by adding M symbols through exhaustive search. If $M = 0$, the solution is optimal. If $M = 1$, no search is needed as Proposition 2 below states.

In the following we assume AWGN with $2\sigma^2 = 1$, i.e., we drop the term $2\sigma^2$ from the MSE which is common for all different training matrices $\mathbf{S}$.

Proposition 2: Consider an L-optimal training sequence of length N_t from a 2^m roots-of-unity alphabet. Adding one training symbol to create a training sequence of length $N_t + 1$ leads to MSE value (which is denoted by MSE_1) equal to

$$\text{MSE}_1 = \frac{L}{N_S} - \frac{L}{N_S(N_S + L)} \tag{18}$$

where $N_S = N_t - L + 1$ is the number of rows of the $\mathbf{S}$ matrix.

The proof is provided in the Appendix. Note that MSE_1 does not depend on the added symbol, and that it is not always the minimum MSE possible for this training sequence length and restricted alphabet.

Proposition 3: Consider an L-optimal training sequence of length N_t from a 2^m root-of-unity alphabet. Adding two training symbols to create a training sequence of length $N_t + 2$ amounts to adding two rows $\mathbf{r}_1$ and $\mathbf{r}_2$ to matrix $\mathbf{S}_{N_S}$, i.e.,

$$S_{N_S+2} = \begin{bmatrix} S_{N_S} \\ \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} \tag{19}$$

and leads to an MSE value, which we denote by MSE_2, equal to

$$\text{MSE}_2 = \text{MSE}_1 - \frac{(L + N_S)^2 L - a(L + 2N_S)}{N_S(L + N_S)\left((L + N_S)^2 - a\right)}$$

where $a = |\mathbf{r}_1 \mathbf{r}_2^*|^2$ and MSE_1 is given by (18).

The proof uses the matrix inversion lemma and is similar to the proof of Proposition 2. Extending an L-perfect training sequence by two symbols leads to a training sequence length $N_t + 2$, for which L-perfect sequences may exist. However, it is easy to show that these L-perfect sequences cannot be created by such an extension method.

Continuing along the same lines, one could derive the MSE_k achieved when extending an L-optimal sequence by k symbols, but the calculations become tedious as k increases. Instead, we give an upper bound on the MSE as a function of the extension length k.

Proposition 4: Consider an L-optimal training sequence of length N_t from a 2^m root-of-unity alphabet. Extending the

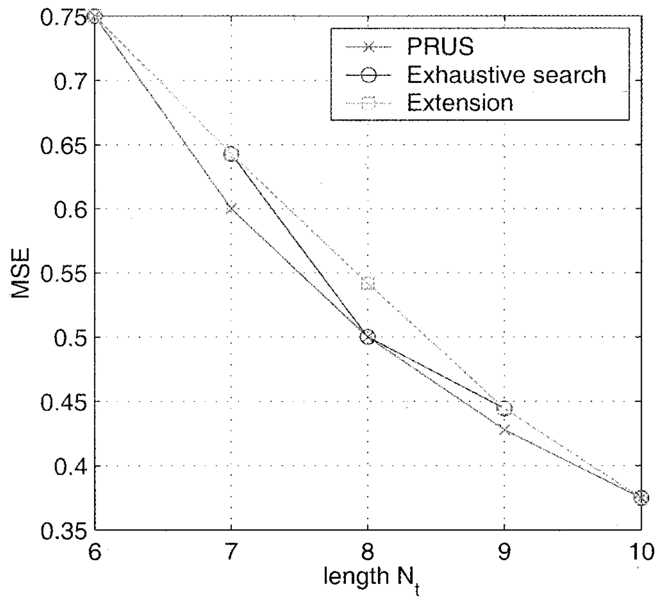

Fig. 1. Achieved MSE with PRUS (any alphabet size), MSE with QPSK-alphabet and exhaustive search, MSE with QPSK-alphabet and the proposed construction method. Assumed $L = 3$ taps and QPSK constellation alphabet.

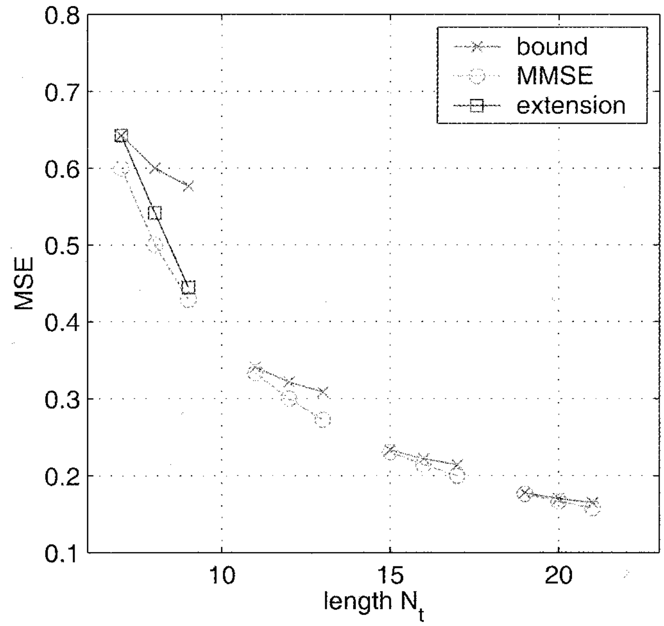

Fig. 2. MSE achieved with PRUS (any alphabet size) and upper bound on MSE achieved through extension. Assumed $L = 3$ taps.

training sequence to length $N_t + k$ by adding k symbols results in MSE_k upper bounded as follows:

$$\mathrm{MSE}_k \leq \frac{L}{N_S}\left(1 - \frac{k}{Lk + N_S}\right). \tag{20}$$

Thus, the maximum increase in MSE from not using an optimal sequence is upper bounded by

$$\mathrm{MSE}_k - \mathrm{MMSE} \leq \frac{L}{N_S}\left(1 - \frac{k}{Lk + N_S}\right) - \frac{L}{k + N_S}.$$

The proof is provided in the Appendix. This bound does not depend on the constellation employed or the extension symbols. It upper bounds the largest MSE we may get by extending an L-perfect sequence by k randomly-chosen symbols. For $k = 1$ the bound becomes equal to MSE_1 in (18).

A different approach would be, instead of extending a perfect sequence by k symbols, to truncate it. This approach has less freedom, since reducing a sequence of length $N_t = kN_{\mathrm{PRUS}} + L - 1$ by M symbols leads to a subset of the sequences we can get by increasing a sequence of length $N_t = (k-1)N_{\mathrm{PRUS}} + L - 1+$ by $N_{\mathrm{PRUS}} - M$ symbols.

V. Simulation Results

A. MSE Performance of Extension Method

Fig. 1 shows the MSE versus training sequence length with QPSK alphabet and L = 3. We plot three curves: the MMSE achieved with PRUS and no restriction on the alphabet size, the minimum MSE for training sequences with QPSK alphabet found through exhaustive search, and the MSE for training sequences with QPSK alphabet identified from the proposed construction method.

For $L = 3$ (Fig. 1), perfect sequences can be used to construct 3-perfect sequences of length $N_t = 6$ and 10. Extending the L-perfect sequences by one symbol leads to the same min-

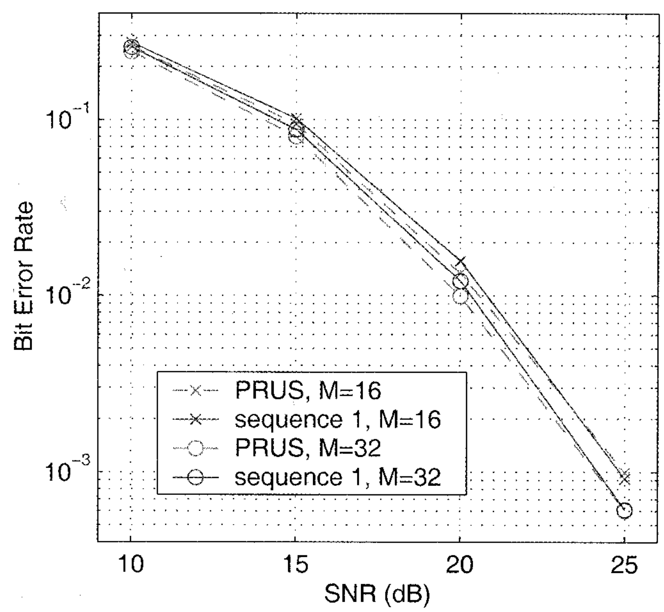

Fig. 3. Performance for the optimal PRUS and a proposed suboptimal training sequence.

imum MSE achieved by an exhaustive search over all possible sequences of this length. Also, there exist perfect sequences of length $N_t = 8$ which cannot be produced by extending the length $N_t = 6$ sequences by two symbols.

Fig. 2 plots the bound in (20) and the optimal MMSE for $L = 3$. The bound predicts the largest MSE we may get by extending an L-perfect sequence by some randomly-chosen symbols, i.e., the worst case scenario. The bound closely approaches MMSE as the training length increases.

B. Trellis Code Example

Fig. 3 compares the bit-error rate (BER) achieved when optimally estimating $\mathbf{h}_1$ and $\mathbf{h}_2$ with a PRUS sequence, and the

BER of the proposed scheme in Section III-C.1 with the sequences $\mathbf{s}_o = \{-1\ 1\ 1\ 1\ 1\ -1\ -i\ -1\ 1\ 1\ -1\ 1\ 1\}$ and $\mathbf{s}_e = e^{(5\pi)/4}\mathbf{s}_o$. These sequences of length $N_t/2 = 13$ are applicable to the EDGE typical urban (TU) environment (where $N_t = 26$ and $L = 4$ [16], [17]) and the eight-state, 8-PSK space–time trellis code. An exhaustive search identified a total of 94 sequences that achieve the minimum MSE, which in this case was 0.0816. The lower bound for MSE achieved by PRUS was 0.0435.

The joint space–time equalizer/decoder employs a prefilter to concentrate the channel energy in a smaller number of taps followed by a reduced-complexity maximum aposteriori (MAP) equalizer/decoder with M active trellis states as described in [18]. The figure shows BER results for $M = 16$ and $M = 32$. The optimal and the suboptimal training sequences achieve similar performance.

C. Block Code Example

For the EDGE TU environment, the lower bound on MSE when using two consecutive training sequences to estimate the channels $\mathbf{h}_1$ and $\mathbf{h}_2$ is 0.1739. Using the method discussed in Section III-C2 and $(\mathbf{s}_2 = \tilde{\mathbf{s}}_1)$ leads to an MSE of 0.175, which is very close to MMSE.

VI. Conclusion

This paper studied various methods to identify good training sequences for systems employing multiple transmit antennas over frequency-selective channels.

We simplified the channel estimation problem from designing multiple training sequences with impulse-like auto correlation and zero cross correlation to designing a single training sequence with impulse-like auto correlation. Furthermore, we proposed a method to identify suboptimal training sequences for an arbitrary sequence length and number of channel taps to be estimated. Upper bounds on the MSE increase with the proposed extension method indicate that achievable performance is close to optimal. Our focus was on training symbols belonging in alphabets of size 2^m such as BPSK, QPSK, and 8-PSK, as they simplify the transmitter/receiver structure and result in negligible MSE increase from MMSE.

Appendix

Proof of Proposition 2: Denote by $\mathbf{S}_{N_S}$ the matrix $\mathbf{S}$ of dimension $N_S \times L$ in (7) where $N_S = N_t - L + 1$. Assume that it is constructed from an optimal sequence $[s(0)\ s(1) \ldots s(N_t - 1)]$ which implies that $\mathbf{S}_{N_S}^H \mathbf{S}_{N_S} = N_S \mathbf{I}_L$. Adding one symbol $s(N_t)$ to the training sequence amounts to adding a row $\mathbf{r}_1$ in matrix $\mathbf{S}_{N_S}$. Then

$$\left(\mathbf{S}_{N_S+1}^H \mathbf{S}_{N_S+1}\right)^{-1} = \frac{1}{N_S}\mathbf{I}_L - \frac{\mathbf{r}_1^H \mathbf{r}_1}{N_S(N_S+L)}$$

where we used the fact that $\mathbf{r}_1 \mathbf{r}_1^H = L$ and the matrix inversion lemma [19]. Taking the trace of both sides leads to the desired result. □

Proof of Proposition 4: Adding k symbols to the training sequence amounts to adding k rows in matrix $\mathbf{S}_{N_S}$, i.e.,

$$\mathbf{S}_{N_S+k} = \begin{bmatrix} \mathbf{S}_{N_S} \\ \mathbf{r} \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_k \end{bmatrix} \tag{21}$$

where $\mathbf{r}$ is of dimension $k \times L$. We assume that the matrix $\mathbf{S}_{N_S}$ is constructed from an optimal sequence, thus

$$\mathbf{S}_{N_S+k}^H \mathbf{S}_{N_S+k} = N_S \mathbf{I}_L + \mathbf{r}^H \mathbf{r}. \tag{22}$$

Applying the matrix inversion lemma [19] we get that

$$tr\left(\left(\mathbf{S}_{N_S+k}^H \mathbf{S}_{N_S+k}\right)^{-1}\right) = \frac{L}{N_S} - \frac{1}{N_S^2} tr\left\{\left(\mathbf{r}^H \left(\mathbf{I}_k + \frac{1}{N_S}\mathbf{r}\mathbf{r}^H\right)^{-1} \mathbf{r}\right)\right. . \tag{23}$$

The Gram matrix $\mathbf{r}\mathbf{r}^H$ can be upper bounded as

$$\mathbf{r}\mathbf{r}^H < (Lk + \epsilon)\mathbf{I}_k \tag{24}$$

for any positive number ϵ. Indeed, the matrix $A = (Lk+\epsilon)\mathbf{I}_k - \mathbf{r}\mathbf{r}^H$ is Hermitian and diagonally dominant. Therefore, all eigenvalues of A are real and positive [19], which implies that A is positive definite. Thus, the following inequalities hold

$$\mathbf{I}_k + \frac{1}{N_S}\mathbf{r}\mathbf{r}^H < \mathbf{I}_k + \frac{Lk+\epsilon}{N_S}\mathbf{I}_k$$

$$\Rightarrow \mathbf{r}^H \left(\mathbf{I}_k + \frac{1}{N_S}\mathbf{r}\mathbf{r}^H\right)^{-1} \mathbf{r} > \frac{N_S}{Lk+\epsilon+N_S}\mathbf{r}^H \mathbf{r}$$

$$\Rightarrow \frac{L}{N_S} - \frac{1}{N_S^2} tr\left(\mathbf{r}^H \left(\mathbf{I}_k + \frac{1}{N_S}\mathbf{r}\mathbf{r}^H\right)^{-1} \mathbf{r}\right) < \frac{L}{N_S}\left(1 - \frac{k}{Lk+\epsilon+N_S}\right).$$

Therefore

$$\mathrm{MSE}_k < \frac{L}{N_S}\left(1 - \frac{k}{Lk+\epsilon+N_S}\right). \tag{25}$$

Since this bound holds for $\epsilon \to 0$ from above, then in the limit we get the desired result. □

References

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data tale wireless communications: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[2] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[3] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1169–1174, July 2000.

[4] B. L. Hughes, "Differential space–time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[5] S. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Differential space–time transmission for frequency-selective channels," *IEEE Commun. Lett.*, vol. 6, pp. 253–255, June 2002.

[6] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space–time coding modem for high data rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.

[7] P. Spasojevic and C. Georghiades, "Complementary sequence for ISI channel estimation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1145–1152, Mar. 2001.

[8] G. Caire and U. Mitra, "Training sequence design for adaptive equalization of multi-user systems," in *Proc. 32nd Asilomar Conf.*, vol. 2, 1998, pp. 1479–1483.

[9] W. H. Mow, *Sequence Design for Spread Spectrum*. Hong Kong: The Chinese University Press: Chinese Univ. Hong Kong, 1995.

[10] D. C. Chu, "Polyphase codes with good correlation properties," *IEEE Trans. Inform. Theory*, vol. IT–18, pp. 531–532, July 1972.

[11] W. H. Mow, "A new unified construction of perfect root-of-unity sequences," in *Proc. Spread-Spectrum Techniques and Applications*, vol. 3, 1996, pp. 955–959.

[12] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," in *Proc. IEEE*, vol. 138, Aug. 1902, pp. 371–378.

[13] A. Furuskar, S. Mazur, F. Muller, and H. Olofsson, "EDGE: Enhanced Data Rates for GSM and TDMA/136 Evolution," *IEEE Pers. Commun.*, pp. 56–66, June 1999.

[14] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[15] E. Lindskog and A. Paulraj, *A Transmit Diversity Scheme for Channels with Intersymbol Interference*: ICC, 2000, vol. 1, pp. 307–311.

[16] *GSM-0.5.02. Digital Cellular Telecommunications System (Phase 2+): Multiplexing and Multiple Access on the Radio Path*: ETSI, July 2000.

[17] *GSM-0.5.04. Digital Cellular Telecommunications System (Phase 2+): Modulation*: ETSI, 1999.

[18] C. Fragouli, N. AI-Dhahir, S. Diggavi, and W. Turin, "Prefiltered space–time M-BCJR equalization for frequency-selective channels," *Trans. Commun.*, pp. 742–753, May 2002.

[19] R. A. Horn and G. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

Christina Fragouli (M'00) received the B.S. degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, in 1998 and 2000, respectively.

From September 2000 to September 2001, she was with AT&T Shannon Laboratory, Florham Park, NJ, as a Consultant. She is currently a Research Associate at the National Capodistrean University of Athens, Athens, Greece. Her research interests are in the area of channel coding and signal processing for communications.

Naofal Al-Dhahir (S'89–M'90–SM'98) received the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1990 and 1994, respectively, in electrical engineering.

He was an Instructor at Stanford University during the Winter of 1993. From August 1994 to July 1999, he was a Member of the technical staff at the Communications Program at GE Corporate R&D Center, Schenectady, NY, where he worked on various aspects of satellite communication systems design and anti-jam GPS receivers. Since August 1999, he has been a Principal Member of technical staff at AT&T Shannon Laboratory, Florham Park, NJ. He has authored over 45 journal papers and holds nine U.S. patents in the areas of satellite communications and digital television. He is coauthor of *Doppler Applications for LEO Satellite Systems* (Norwell, MA: Kluwer, 2001). His current research interests include equalization schemes, space–time coding and signal processing, OFDM, and digital subscriber line technology.

Dr. Al-Dhahir is a Member of the IEEE SP4COM technical committee. He is an Associate Editor for IEEE TRANSACTION ON SIGNAL PROCESSING, IEEE COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON COMMUNICATIONS.

William Turin (SM'83) received the M. S. degree in mathematics from Odessa State University, Odessa, U.S.S.R., in 1958, and the Ph.D. degree in mathematics from the Institute for Problems of Mechanics of the Academy of Sciences of the U.S.S.R., Moscow, in 1966.

From 1960 to 1979, he was associated with the Moscow Institute of Electrical Engineering and Telecommunications, first as Assistant and later an Associate Professor. From 1980 to 1981, he was a Senior Research Scientist in the Department of Psychology, New York University, NY. Since 1981, he has been a Member of Technical Staff at AT&T Bell Laboratories, Holmdel, NJ and Murray Hill, NJ. Currently, he is a Technology Consultant at AT&T Labs-Research, Shannon Laboratory, Florham Park, NJ. He is the author of five books and numerous papers. His research interests include modeling channels with memory, digital signal processing, handwriting and speech recognition and compression, computer simulation, error-correcting codes, microwave radio, and Markov processes.