

Multuser Channel Estimation for Detection of Cochannel Signals

Stephen J. Grant, *Associate Member, IEEE*, and James K. Cavers, *Senior Member, IEEE*

Abstract—Estimation of the channel impulse responses of multiple cochannel users is a key requirement of all multuser detection and interference cancellation techniques, though little attention has been paid to the subject in the context of time-division multiple-access (TDMA) systems. This paper addresses a pilot-based MMSE technique for multuser channel estimation in a TDMA system, and makes two key contributions. Firstly, it allows for time variation of the channels within and between training sequences, an essential feature in a multuser environment even at moderate fading rates. Secondly, it addresses the design of training sequences for the multiple cochannel users.

Index Terms—Dispersive channels, estimation, fading channels, interference suppression, multidimensional signal detection, multuser channels, sequences, time division multiaccess.

I. INTRODUCTION

MULTIUSER detection and interference cancellation techniques have received much attention recently due to their potential for increasing system capacity. For code-division multiple-access (CDMA) systems, where much of the research has been focused, e.g., [1] and [2], the interest has been motivated by the poor performance of the conventional detector, which ignores multiple access interference. For time-division multiple-access (TDMA) systems, where a steadily increasing amount of research is being directed, e.g., [3]–[6], the interest is motivated by the desire to allow frequency reuse within cell and/or achieve intercell interference cancellation. One aspect common to all of the multuser detection and interference cancellation techniques, however, is the necessity of having reliable channel estimates for all of the cochannel users.

The use of pilot symbols is a well-known method for obtaining good channel impulse response estimates in single-user systems, e.g., [7]–[11]. For the case of multuser systems, pilot-based channel estimation has been studied extensively only for CDMA, e.g., [12] and [13], where processing gain suppresses interference in the channel estimator. However, little is known about multuser channel estimation in TDMA systems. The sole

Paper approved for publication by M. Z. Win, the Editor for Equalization and Diversity of the IEEE Communications Society. Manuscript received November 15, 1998; revised December 28, 1999, and July 25, 2000. This work was supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada. This paper was presented at the IEEE International Conference on Communications, Vancouver, BC, Canada, June 6–10, 1999.

S. J. Grant was with Simon Fraser University, Burnaby, BC, Canada. He is now with the Advanced Development and Research Group of Ericsson Inc., Research Triangle Park, NC 27709 USA (e-mail: stephen.grant@ericsson.com).

J. K. Cavers is with Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: cavers@sfu.ca).

Publisher Item Identifier S 0090-6778(01)09110-3.

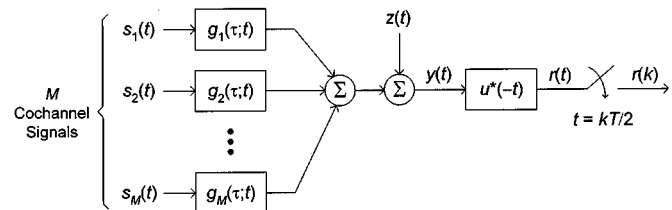


Fig. 1. Cochannel signal model. The M time-variant, dispersive channels are assumed to fade independently.

prior study appears to be [14], in which the channels are treated as time-invariant.

This paper addresses a pilot-based MMSE technique for estimating the channel impulse responses of multiple cochannel users in a TDMA system and makes two key contributions. First, it allows time variation of the channels within and between the training sequences. This is essential in a multuser environment where the training sequences are necessarily longer than in a single user environment [14], resulting in significant time variation during the training periods. Previously, this has been unaddressed.

Second, it addresses the design of appropriate training sequences. In contrast to the single-user case, for which it is often possible to select a single training sequence with perfect auto-correlation properties (as done in [8] and [9]), it is difficult, in general, to select multiple training sequences of arbitrary length with both perfect auto and cross-correlation properties. To overcome these difficulties, reasonable selection criteria are presented for designing good, suboptimal training sequences with training symbols constrained to lie within the modulation alphabet.

II. SIGNAL AND CHANNEL MODELS

Throughout this paper, the following conventions are used: variables in italics are scalars, lowercase boldface variables are vectors, and uppercase boldface variables are matrices. Furthermore, \dagger and T denote, respectively, the complex-conjugate transpose and regular transpose of a vector or matrix, and $*$ denotes complex conjugate. Since all signals are represented by their complex baseband equivalents, the average power (or variance) of the bandpass signal $\tilde{s}(t)$, with baseband equivalent $s(t)$, is $P_s = (1/2)E[|s(t)|^2]$, where $E[\cdot]$ is the expectation operator.

Fig. 1 shows a diagram of the transmission of M cochannel signals through independently fading, dispersive channels. Each channel is represented by its time-variant channel impulse

response (CIR) $g_m(\tau; t)$ where τ represents the memory of the impulse response, and t represents time variation. The m th user's transmitted signal is given by

$$s_m(t) = A_m \sum_n c_m(n) u(t - nT - \tau_m) \quad (1)$$

where the data/training symbol $c_m(n)$ is normalized such that $E[|c_m(n)|^2] = 1$, the pulse shape $u(t)$ is root-Nyquist with deterministic autocorrelation function $x(\alpha) = \int u(t)^*(t-\alpha) dt$ and energy $x(0) = 1$, the symbol period is T , and the transmit amplitude A_m is related to the average power P_m in $\tilde{s}_m(t)$ by $A_m = \sqrt{2P_m}$. The relative delay τ_m appears in (1) since, in a typical TDMA system, the received signals at the base station may not be synchronized by symbol due to differing propagation delays between each user's terminal and the base.

The received signal $y(t)$ consists of the sum of the M filtered cochannel signals and an additive white Gaussian noise component $z(t)$ with double-sided power spectral density N_o . Assuming that the time variations of the channels are slow enough such that $g_m(\tau; t)$ does not vary significantly over the memory of the transmit pulse (a few symbols), the output of the matched filter $u^*(-t)$ is

$$\begin{aligned} r(t) &= \sum_{m=1}^M A_m \sum_n c_m(n) \int g_m(\tau; t) \\ &\quad \cdot x(t - nT - \tau - \tau_m) d\tau + n(t) \\ &= \sum_{m=1}^M \sum_n c_m(n) h_m(t - nT; t) + n(t) \end{aligned} \quad (2)$$

where $h_m(\tau; t)$ is the m th user's composite impulse response given by

$$h_m(\tau; t) = A_m g_m(\tau - \tau_m; t) \otimes x(\tau). \quad (3)$$

The operator \otimes denotes convolution. Notice that the relative delay τ_m is now considered to be part of the channel impulse response. In (2), $n(t)$ is the noise component of the matched filter output. Since $z(t)$ is white, the autocorrelation function of $n(t)$ is $\phi_n(\alpha) = N_o x(\alpha)$.

A. Channel Statistics

In this paper, we adopt an MMSE estimation technique which leads naturally to considering time variation of the channels within and between training periods. MMSE estimation of the users' channels requires knowledge of the second-order statistics of $h_m(\tau; t)$ summarized by the correlation function

$$\begin{aligned} R_{h_m}(\tau_1, \tau_2, \alpha) &= \frac{1}{2} E[h_m(\tau_1; t) h_m^*(\tau_2; t - \alpha)] \\ &= A_m^2 \iint \frac{1}{2} E[g_m(\lambda_1 - \tau_m; t) g_m^*(\lambda_2 - \tau_m; t - \alpha)] \\ &\quad \cdot x(\tau_1 - \lambda_1) x(\tau_2 - \lambda_2) d\lambda_1 d\lambda_2. \end{aligned} \quad (4)$$

Assuming a wide sense stationary uncorrelated scattering (WSSUS) channel as well as a separable scattering function, the correlation function reduces to

$$\begin{aligned} R_{h_m}(\tau_1, \tau_2, \alpha) &= A_m^2 \frac{R_{g_m}(\alpha)}{\sigma_{g_m}^2} \cdot \int G_m(\tau) x(\tau_1 - \tau) \\ &\quad \cdot x^*(\tau_2 - \tau) d\tau. \end{aligned} \quad (5)$$

In this expression, $R_{g_m}(\alpha)$ is the temporal autocorrelation function of the channel. For example, with isotropic scattering, $R_{g_m}(\alpha) = \sigma_{g_m}^2 J_0(2\pi f_{D_m} \alpha)$, where $J_0(\cdot)$ is the Bessel function of the first kind and f_{D_m} is the maximum Doppler shift. The function $G_m(\tau)$ is given by

$$\begin{aligned} G_m(\tau) &= \frac{1}{2} E[|g_m(\tau - \tau_m; t)|^2] \\ &= \int P_{g_m}(\tau - \tau_m) f_m(\tau_m) d\tau_m \end{aligned} \quad (6)$$

where $f_m(\tau_m)$ is the probability density function (PDF) of the relative delay τ_m (often uniform over $[-T/2, T/2]$), and $P_{g_m}(\tau)$ is the power-delay profile (PDP) of the channel. One common example is the exponential PDP $P_{g_m}(\tau) = (\sigma_{g_m}^2 / \tau_{\text{rms}_m}) \exp[-\tau / \tau_{\text{rms}_m}]$, where τ_{rms_m} is the rms delay spread. Another is the single-spike profile for flat fading given by $P_{g_m}(\tau) = \sigma_{g_m}^2 \delta(\tau)$ where $\delta(\tau)$ is the Dirac delta function. Observing (6), $G_m(\tau)$ is given by the convolution of the PDP and the PDF of τ_m . Because the uncertainty in timing is often much larger than the rms delay spread, the shape of $G_m(\tau)$, and thus the performance of the MMSE estimator, is not very sensitive to τ_{rms_m} .

As can be seen from (5) and (6), $R_{h_m}(\tau_1, \tau_2, \alpha)$ does not depend on τ_m itself—only on its PDF. In other words, explicit timing recovery is unnecessary; the relative delays are simply estimated as part of the channels. However, if the relative delays happen to be known, the PDF of τ_m becomes an impulse and $G_m(\tau) = P_{g_m}(\tau - \tau_m)$.

B. SNR Definition

The signal-to-noise ratio (SNR) of user m is defined as $\Gamma_m = E_{s_m} / N_o$, where E_{s_m} is the average received energy per symbol from user m given by (see Fig. 1)

$$\begin{aligned} E_{s_m} &= \int_{-T/2}^{T/2} \frac{1}{2} E[|g_m(\tau; t) \otimes s_m(t)|^2] dt \\ &= \int P_{g_m}(\tau) \int_{-T/2}^{T/2} E[|s_m(t - \tau)|^2] dt d\tau \end{aligned} \quad (7)$$

where, again, a WSSUS channel has been assumed. The inner integral of the latter expression is simply equal to twice the average power in $\tilde{s}_m(t)$. Furthermore, since the area under $P_{g_m}(\tau)$ is $\sigma_{g_m}^2$, the m th user's SNR is simply

$$\Gamma_m = \frac{2P_m \sigma_{g_m}^2}{N_o} = \frac{A_m^2 \sigma_{g_m}^2}{N_o}. \quad (8)$$

C. Channel Vectors

Samples of the matched filter output $r(t)$ are taken at times $t = kT/2$ yielding the discrete-time sequence

$$r(k) = \sum_{m=1}^M \mathbf{c}_m^T(k) \mathbf{h}_m(k) + n(k),$$

$$k = \dots, -2, -1, 0, 1, 2, \dots \quad (9)$$

where the symbol vector $\mathbf{c}_m(k)$ is given by (10) at the bottom of the page and m th user's channel vector—to be estimated—is

$$\mathbf{h}_m(k) = \begin{bmatrix} h_m\left(L_1 T; \frac{kT}{2}\right) \\ h_m\left(\left(L_1 + \frac{1}{2}\right)T; \frac{kT}{2}\right) \\ \vdots \\ h_m\left(0; \frac{kT}{2}\right) \\ h_m\left(\frac{T}{2}; \frac{kT}{2}\right) \\ \vdots \\ h_m\left(L_2 T; \frac{kT}{2}\right) \\ h_m\left(\left(L_2 + \frac{1}{2}\right)T; \frac{kT}{2}\right) \end{bmatrix} \quad (11)$$

where L_1 and L_2 are integers. Clearly, $\mathbf{h}_m(k)$ consists of samples of $h_m(\tau; t)$ at $T/2$ -spaced delays evaluated at time $t = kT/2$. It is assumed that $h_m(\tau; t)$ is generally noncausal such that $L_1 \leq 0$ and $L_2 \geq 0$.

The second-order statistics of $\mathbf{h}_m(k)$ are summarized by the autocorrelation matrix $\mathbf{R}_{\mathbf{h}_m}(j) = (1/2)E[\mathbf{h}_m(k)\mathbf{h}_m^\dagger(k-j)]$. Using (5), the u, v th element of this matrix is

$$\{\mathbf{R}_{\mathbf{h}_m}(j)\}_{u,v} = R_{h_m}\left(\frac{uT}{2}, \frac{vT}{2}, \frac{jT}{2}\right) \quad (12)$$

where $u, v \in \{2L_1, \dots, 0, \dots, 2L_2 + 1\}$. Evidently, the tap gains [elements of $\mathbf{h}_m(k)$] are correlated, in general, even though we have assumed a WSSUS channel [10]. This is due to the convolution of $g_m(\tau - \tau_m; t)$ with the pulse autocorrelation function as shown in (3).

Observing (11), the length of $\mathbf{h}_m(k)$ is $2(L_c + 1)$ where $L_c = L_2 - L_1$. To maintain computational complexity as low as possible, it is desirable to choose L_1 and L_2 as small as possible, keeping only those channel taps with significant variance. Since

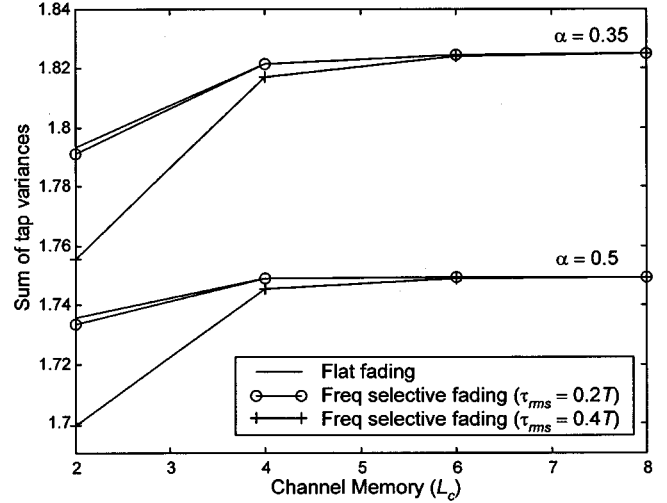


Fig. 2. Sum of channel tap variances versus the channel memory length L_c for flat and frequency-selective fading conditions. The variable α is the rolloff parameter of the root-Nyquist transmit pulse $u(t)$.

the impulse response $h_m(\tau; t)$ decays to zero for large $|\tau|$, the sum of the tap variances, given by $\text{trace}[\mathbf{R}_{\mathbf{h}_m}(0)]$, saturates for large L_c .

As a guide for the selection of the minimum L_c required, Fig. 2 shows a plot of $\text{trace}[\mathbf{R}_{\mathbf{h}_m}(0)]$ vs. L_c for several combinations of rms delay spread and pulse rolloff—the two factors that directly influence the duration of $h_m(\tau; t)$. In this graph, $G_m(\tau)$ in (6) is calculated using an exponential power delay profile and a uniform distribution of relative delay over the interval $[-T/2, T/2]$. As can be seen, $L_c = 4$ is sufficient for capturing most of the energy in the impulse response for $\tau_{rms} \leq 0.4T$. To minimize estimation error for higher values of delay spread, it may be desirable to choose $L_c = 6$; however, the cost incurred when using MLSE-type detection schemes is higher complexity due to the increased state set.

III. JOINT CHANNEL ESTIMATION

In this section, an MMSE technique for jointly estimating the M users' channels is discussed. Other techniques (such as least squares) may also be used, provided they incorporate a model of the temporal variation of the channels. However, we do not investigate them in this study, and restrict our attention to MMSE which enables a consistent treatment of time variation through the temporal autocorrelation function of the channels.

A. General Structure of the Estimator

MMSE estimation of the users' channels relies upon the periodic insertion of a unique training sequence into each user's data sequence. The design of the training sequences is addressed

$$\mathbf{c}_m(k) = \begin{cases} \left[c_m\left(\left\lfloor \frac{k}{2} \right\rfloor - L_1\right) \ 0 \ \cdots \ c_m\left(\left\lfloor \frac{k}{2} \right\rfloor\right) \ 0 \ \cdots \ c_m\left(\left\lfloor \frac{k}{2} \right\rfloor - L_2\right) \ 0 \right]^T, & k \text{ even} \\ \left[0 \ c_m\left(\left\lfloor \frac{k}{2} \right\rfloor - L_1\right) \ \cdots \ 0 \ c_m\left(\left\lfloor \frac{k}{2} \right\rfloor\right) \ \cdots \ 0 \ c_m\left(\left\lfloor \frac{k}{2} \right\rfloor - L_2\right) \right]^T, & k \text{ odd} \end{cases} \quad (10)$$

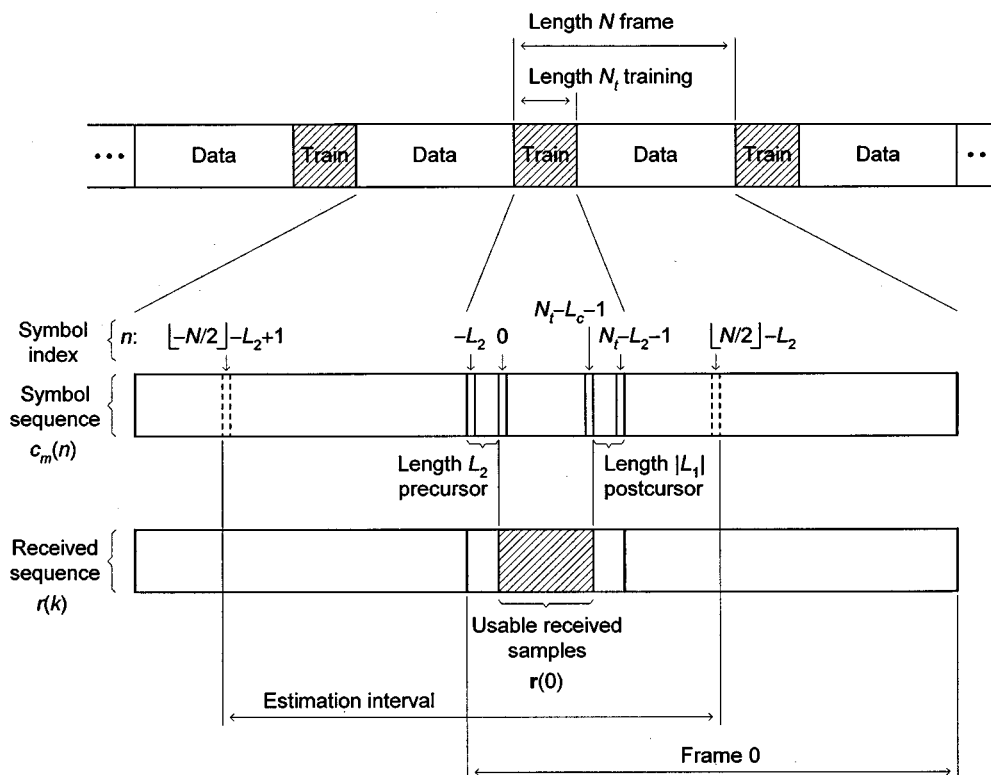


Fig. 3. Frame structure and indexing conventions.

in Section IV. Unique training sequences are required for each user so that the cochannel signals may be distinguished (e.g., see [14]). This is in contrast to CDMA systems where spreading codes are used to distinguish users and suppress interference in the channel estimator.

It is assumed that the asynchronous users are slot-synchronous such that their training sequences are inserted at the same time, although different propagation delays make their arrivals symbol-asynchronous as discussed previously. The received samples during the training periods are then used to derive estimates of the channels which are interpolated between training periods. In this way, time variations of the channels are tracked. The frame structure, along with the symbol and frame indexing conventions used throughout this paper, is shown in Fig. 3. In this structure, the length of each frame is N symbols, and the length of each training sequence is N_t symbols. Note that n indexes symbols, and k indexes samples; thus, $n = \lfloor k/2 \rfloor$. To provide detail, a single training period and the two adjacent data blocks are shown in exploded view. The exploded view shows the zeroth frame which starts at the beginning of the training period and extends to the end of the subsequent data block. The estimation interval extends from mid-frame to mid-frame either side of the training period.

Since the users' channels are to be estimated jointly, we define the length- $2M(L_c + 1)$ vector $\mathbf{h}(k)$ as the concatenation of the M users' individual channel vectors

$$\mathbf{h}(k) = [\mathbf{h}_1^T(k) \quad \mathbf{h}_2^T(k) \quad \cdots \quad \mathbf{h}_M^T(k)]^T. \quad (13)$$

The autocorrelation matrix of the joint channel vector is $\mathbf{R}_{\mathbf{h}}(j) = (1/2)E[\mathbf{h}(k)\mathbf{h}^\dagger(k+j)]$. Because the users' channels fade independently, $\mathbf{R}_{\mathbf{h}}(j)$ is block diagonal and is given by

$$\mathbf{R}_{\mathbf{h}}(j) = \begin{bmatrix} \mathbf{R}_{\mathbf{h}_1}(j) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{h}_2}(j) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_{\mathbf{h}_M}(j) \end{bmatrix} \quad (14)$$

where $\mathbf{R}_{\mathbf{h}_m}(j)$ is defined in (12).

Consider the MMSE estimation of $\mathbf{h}(k)$ in the estimation interval $[-N/2]-L_2+1 \leq n \leq [N/2]-L_2$ shown in Fig. 3. The channel estimator uses the received samples from the training blocks of each of the $2Q+1$ frames centered about frame-0 to form its estimate. These samples are contained in the vector

$$\mathbf{z} = [\mathbf{r}^T(-Q) \quad \cdots \quad \mathbf{r}^T(0) \quad \cdots \quad \mathbf{r}^T(Q)]^T \quad (15)$$

where

$$\begin{aligned} \mathbf{r}(q) &= [r(2qN) \quad r(2qN+1) \quad \cdots \quad r(2qN+2(N_t-L_c)-1)]^T. \end{aligned} \quad (16)$$

The length- $2(N_t - L_c)$ vector $\mathbf{r}(q)$ contains a subset of the received samples during the q th training block called the "usable samples." For example, Fig. 3 shows the usable samples for the zeroth frame ($q = 0$). With use of this subset, $\mathbf{r}(q)$ depends only on training symbols—not on unknown data symbols—due to the

length- L_2 precursor and the length- $|L_1|$ postcursor inserted in each training sequence.

Since \mathbf{z} contains samples of a bandlimited process sampled at a rate greater than the Nyquist rate, the covariance matrix of \mathbf{z} , given by $\mathbf{R}_z = (1/2)E[\mathbf{z}\mathbf{z}^\dagger]$, becomes ill-conditioned as N_t increases (due to an increasing number of users). This suggests the use of rank reduction to remove dependencies in \mathbf{z} as well as to avoid explicit inversion of \mathbf{R}_z . Accordingly, we use eigen-decomposition to write the covariance matrix of \mathbf{z} as

$$\mathbf{R}_z = [\mathbf{M}_1 \quad \mathbf{M}_2] \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^\dagger \\ \mathbf{M}_2^\dagger \end{bmatrix}. \quad (17)$$

The diagonal matrix $\mathbf{\Lambda}_1$ contains the dominant eigenvalues of \mathbf{R}_z , and $\mathbf{\Lambda}_2$ contains those eigenvalues that fall below some very small threshold, e.g., $10^{-9}\lambda_{\max}$, where λ_{\max} is the maximum eigenvalue of \mathbf{R}_z . The nonsquare matrices \mathbf{M}_1 and \mathbf{M}_2 contain the normalized eigenvectors, arranged as columns, corresponding to the eigenvalues in $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ respectively. Now, base the estimate of $\mathbf{h}(k)$ on the reduced dimensionality vector $\mathbf{w} = \mathbf{M}_1^\dagger \mathbf{z}$ (instead of \mathbf{z} itself) which has covariance matrix $\mathbf{R}_w = \mathbf{M}_1^\dagger \mathbf{R}_z \mathbf{M}_1 = \mathbf{\Lambda}_1$.

The optimal (MMSE) estimate of $\mathbf{h}(k)$ based on \mathbf{w} is given by the conditional mean $\mathbf{v}(k) = E[\mathbf{h}(k)|\mathbf{w}]$. Since $\mathbf{h}(k)$ and \mathbf{w} are jointly Gaussian, the conditional mean is linear in \mathbf{w} , and is given by

$$\begin{aligned} \mathbf{v}(k) &= \left(\frac{1}{2} E[\mathbf{h}(k)\mathbf{w}^\dagger]\right) \mathbf{R}_w^{-1} \mathbf{w} \\ &= \mathbf{P}(k) \mathbf{R}_z^\# \mathbf{z} \end{aligned} \quad (18)$$

where $\mathbf{P}(k) = (1/2)E[\mathbf{h}(k)\mathbf{z}^\dagger]$ and $\mathbf{R}_z^\# = \mathbf{M}_1 \mathbf{\Lambda}_1^{-1} \mathbf{M}_1^\dagger$. The latter quantity is recognized as the pseudoinverse, or Moore–Penrose generalized inverse, of \mathbf{R}_z [15]. Note that, for short training sequences (small N_t), \mathbf{R}_z may not be ill-conditioned; in this case $\mathbf{R}_z^\# = \mathbf{R}_z^{-1}$. Evidently, the conditional mean depends upon k , giving a different interpolation matrix $\mathbf{P}(k)\mathbf{R}_z^\#$ for each position within the estimation interval. However, recall that $\mathbf{h}(k)$ is WSS and $r(k)$ is cyclostationary; thus, it is sufficient to calculate the interpolation matrix for each position in only the estimation interval shown in Fig. 3. The same matrix repeats in subsequent intervals.

B. Quality of Channel Estimates

Let $\hat{\mathbf{v}}(k) = \mathbf{U}(k)\mathbf{z}$ denote an arbitrary, not necessarily optimal, estimate of $\mathbf{h}(k)$ with associated channel estimation error

$$\mathbf{e}(k) = \mathbf{h}(k) - \hat{\mathbf{v}}(k). \quad (19)$$

The estimation error covariance matrix is

$$\begin{aligned} \mathbf{R}_e(k) &= \mathbf{R}_h(0) - \mathbf{R}_{\hat{\mathbf{v}}, \mathbf{h}}(k) - \mathbf{R}_{\mathbf{h}, \hat{\mathbf{v}}}(k) + \mathbf{R}_{\hat{\mathbf{v}}}(k) \\ &= \mathbf{R}_h(0) - \mathbf{U}(k)\mathbf{P}^\dagger(k) - \mathbf{P}(k)\mathbf{U}^\dagger(k) \\ &\quad + \mathbf{U}(k)\mathbf{R}_z\mathbf{U}^\dagger(k) \end{aligned} \quad (20)$$

where $\mathbf{R}_{\hat{\mathbf{v}}}(k)$ is the covariance matrix of $\hat{\mathbf{v}}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}}, \mathbf{h}}(k)$ is the cross-covariance matrix of $\hat{\mathbf{v}}(k)$ and $\mathbf{h}(k)$. For the optimal

channel estimate, $\hat{\mathbf{v}}(k) = \mathbf{v}(k)$, $\mathbf{U}(k) = \mathbf{P}(k)\mathbf{R}_z^\#$ and the error covariance matrix is

$$\mathbf{R}_e^{opt}(k) = \mathbf{R}_h(0) - \mathbf{P}(k)\mathbf{R}_z^\#\mathbf{P}^\dagger(k). \quad (21)$$

In this paper, one measure of channel estimation quality for the m th user is the sum of tap error variances, normalized by the sum of the tap variances for that user, that is,

$$\sigma_{e_m}^2(k) = \frac{\text{trace}[\mathbf{R}_{e_m}(k)]}{\text{trace}[\mathbf{R}_{h_m}(0)]} \quad (22)$$

where $\mathbf{R}_{e_m}(k)$ is the m th block along the main diagonal of $\mathbf{R}_e(k)$.

Another measure of channel estimation quality for the m th user that is a major determinant of system performance is the set of correlation coefficients between corresponding taps of estimated channel vector $\hat{\mathbf{h}}_m(k)$ and the true channel vector $\mathbf{h}_m(k)$. Since each of the $2(L_c + 1)$ channel taps has its own correlation coefficient, the following average is defined:

$$\bar{\rho}_m(k) = \frac{1}{2(L_c + 1)} \sum_j \frac{\{\mathbf{R}_{\hat{\mathbf{v}}_m, \mathbf{h}_m}(k)\}_{j,j}}{\sqrt{\{\mathbf{R}_{\hat{\mathbf{v}}_m}(k)\}_{j,j}} \sqrt{\{\mathbf{R}_{h_m}(0)\}_{j,j}}} \quad (23)$$

where j indexes the individual channel taps, and $\{\cdot\}_{j,j}$ denotes the diagonal elements of the bracketed matrix. In this expression, $\mathbf{R}_{\hat{\mathbf{v}}_m}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}}_m, \mathbf{h}_m}(k)$ are the m th blocks along the main diagonals of $\mathbf{R}_{\hat{\mathbf{v}}}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}}, \mathbf{h}}(k)$, respectively. Although both $\sigma_{e_m}^2(k)$ and $\bar{\rho}_m(k)$ depend on k , we found very little variation across the frame. Note that for perfect channel estimation, $\sigma_{e_m}^2(k) = 0$ and $\bar{\rho}_m(k) = 1$.

C. Details of Optimal Estimator

We now examine the optimal estimator in (18) in more detail and obtain the required matrices. It is convenient to first introduce the following data matrix:

$$\mathbf{A} = [\mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \quad \mathbf{A}_M] \quad (24)$$

where the m th submatrix of \mathbf{A} is given by

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{c}_m^T(0) \\ \mathbf{c}_m^T(1) \\ \vdots \\ \mathbf{c}_m^T(2(N_t - L_c) - 1) \end{bmatrix} \quad (25)$$

and $\mathbf{c}_m(\cdot)$ is given in (10). Due to the precursor and postcursor inserted during each training period, \mathbf{A}_m consists only of symbols from the m th user's training sequence and no unknown symbols from the adjacent data sequences. Using (9), (13), (24), and (25), the j th component of $\mathbf{r}(q)$ can be written as

$$r(2qN + j) = \mathbf{a}_j \mathbf{h}(2qN + j) + n(2qN + j) \quad (26)$$

where \mathbf{a}_j is the j th row of the data matrix \mathbf{A} and $j \in \{0, 1, \dots, 2(N_t - L_c) - 1\}$. With this expression

in hand, the elements of the matrices $\mathbf{P}(k)$ and \mathbf{R}_z in (18) may be easily determined.

Using (15),

$$\mathbf{P}(k) = \left[\frac{1}{2} E[\mathbf{h}(k)\mathbf{r}^\dagger(-Q)] \quad \cdots \quad \frac{1}{2} E[\mathbf{h}(k)\mathbf{r}^\dagger(Q)] \right] \quad (27)$$

where $\mathbf{P}(k)$ is of dimension $2M(L_c+1) \times 2(2Q+1)(N_t-L_c)$. Now using (16) and (26), and assuming the noise and channel fading process are uncorrelated, the j th column of the q th submatrix of $\mathbf{P}(k)$ is simply

$$\left\{ \frac{1}{2} E[\mathbf{h}(k)\mathbf{r}^\dagger(q)] \right\}_j = \mathbf{R}_h(k-2qN-j)\mathbf{a}_j^\dagger \quad (28)$$

where $q \in \{-Q, \dots, Q\}$.

Using (15) again,

$$\mathbf{R}_z = \begin{bmatrix} \frac{1}{2} E[\mathbf{r}(-Q)\mathbf{r}^\dagger(-Q)] & \cdots & \frac{1}{2} E[\mathbf{r}(-Q)\mathbf{r}^\dagger(Q)] \\ \vdots & \ddots & \vdots \\ \frac{1}{2} E[\mathbf{r}(Q)\mathbf{r}^\dagger(-Q)] & \cdots & \frac{1}{2} E[\mathbf{r}(Q)\mathbf{r}^\dagger(Q)] \end{bmatrix} \quad (29)$$

where \mathbf{R}_z is of dimension $2(2Q+1)(N_t-L_c) \times 2(2Q+1)(N_t-L_c)$. Using (16) and (26) again, the i, j th element of the q, p th submatrix of \mathbf{R}_z is simply

$$\begin{aligned} & \left\{ \frac{1}{2} E[\mathbf{r}(q)\mathbf{r}^\dagger(p)] \right\}_{i,j} \\ &= \mathbf{a}_i \mathbf{R}_h(2t(q-p)N+i-j)\mathbf{a}_j^\dagger + \phi_n(2(q-p)N+i-j) \end{aligned} \quad (30)$$

where $q, p \in \{-Q, \dots, Q\}$ and $i, j \in \{0, 1, \dots, 2(N_t-L_c)-1\}$. $\phi_n(j) = N_o x(jT/2)$ is the autocorrelation function of the (colored) noise sequence.

Observing (27)–(30), one can see that the optimal interpolation matrix $\mathbf{P}(k)\mathbf{R}_z^\#$ depends only on the data matrix \mathbf{A} , the channel autocorrelation matrix $\mathbf{R}_h(j)$, and the noise autocorrelation function $\phi_n(j)$. For a given scattering environment, the latter two depend on the Doppler fade rate f_{D_m} , the rms delay spread τ_{rms_m} , and the SNR Γ_m for each user. These parameters may not be known at design time and may be different for each user. However, as shown in [7] for the case of a single user and flat fading, a worst-case design methodology may be adopted whereby the interpolator is designed assuming worst-case fading conditions and a typical operating SNR. This is discussed further in Section V-D.

For the special case of a single user ($M=1$), a single pilot symbol ($N_t=1$ and $L_c=0$), one sample per symbol, and flat Rayleigh fading with $G(\tau) = \sigma_g^2 \delta(\tau)$, (27) and (29) give identical interpolator coefficients to those derived in [7]. For the case of a single user and frequency-selective fading, on the other hand, the results here provide an extension to those contained in [8], since we have considered the coloration of the sampled noise sequence and the correlation between channel taps which

was ignored in [8]. As will be shown in Section V-D, the consideration of these effects leads to a significant reduction in estimation error.

IV. TRAINING SEQUENCE DESIGN

Optimal selection of the users' training sequences requires testing all possible combinations of M length- N_t symbol sequences in order to minimize each user's channel estimation error $\sigma_{\epsilon_m}^2(k)$ defined in (22). For several users and practical training sequence lengths, the resulting search space is prohibitively large; furthermore, the amount of computation required to test each candidate sequence is high. In order to overcome these difficulties, a simplified, suboptimal search strategy is developed below which not only yields good training sequences, but offers more insight than an exhaustive computer search.

In the development of this suboptimal search strategy, several assumptions are made: first, the users' channels are assumed to vary slowly enough that they may be considered constant over the duration of each training period; second, the matrix \mathbf{R}_z in (18) is assumed to be nonsingular, so that $\mathbf{R}_z^\# = \mathbf{R}_z^{-1}$; and third, the noise sequence $n(k)$ is assumed to be white. Under these assumptions, (18) may be expressed in an alternate form allowing for a simplified selection criterion. It must be emphasized, though, that these assumptions are made for the purposes of training sequence design only. The resulting sequences are then used to calculate the optimal channel estimate vector $\mathbf{v}(k)$ using (18), (27), and (29) which do not depend on the simplifying assumptions.

Using the slow time-variation assumption, the channel vector $\mathbf{h}(2qN+j)$ in (26) may be approximated by $\mathbf{h}(2qN)$ for all j . Thus, the received vector $\mathbf{r}(q)$ may be written as

$$\mathbf{r}(q) = \mathbf{A}\mathbf{h}(2qN) + \mathbf{n}(q) \quad (31)$$

where $\mathbf{n}(q)$ is the vector of noise samples in the q th training period, which has covariance matrix $\mathbf{R}_n = N_o\mathbf{I}$ under the assumption of white noise.

Using the above expression for $\mathbf{r}(q)$, the q th submatrix of $\mathbf{P}(k)$ may be rewritten as

$$\frac{1}{2} E[\mathbf{h}(k)\mathbf{r}^\dagger(-Q)] = \mathbf{R}_h(k-2qN)\mathbf{A}^\dagger. \quad (32)$$

The q, p th submatrix of \mathbf{R}_z may be rewritten as

$$\frac{1}{2} E[\mathbf{r}(q)\mathbf{r}^\dagger(p)] = \mathbf{A}\mathbf{R}_h(2(q-p)N)\mathbf{A}^\dagger + N_o\mathbf{I}\delta(q-p) \quad (33)$$

where $\delta(l) = 1$ if $l=0$ and zero otherwise. By using these simplified expressions to form the matrices $\mathbf{P}(k)$ and \mathbf{R}_z , and then by using the matrix inversion lemma twice to rewrite the product $\mathbf{P}(k)\mathbf{R}_z^{-1}\mathbf{z}$ in (18) (see [8, Appendix] for the single-user case), the channel estimate vector $\mathbf{v}(k)$ may be expressed as

$$\mathbf{v}(k) = \mathbf{W}(k) \begin{bmatrix} (\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger\mathbf{r}(-Q) \\ \vdots \\ (\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger\mathbf{r}(Q) \end{bmatrix} \quad (34)$$

where

$$\mathbf{W}(k) = [\mathbf{R}_h(k + 2QN) \dots \mathbf{R}_h(k - 2QN)] \cdot \left(\begin{bmatrix} \mathbf{R}_h(0) & \dots & \mathbf{R}_h(-4QN) \\ \vdots & \ddots & \vdots \\ \mathbf{R}_h(4QN) & \dots & \mathbf{R}_h(0) \end{bmatrix} + \begin{bmatrix} N_o(\mathbf{A}^\dagger \mathbf{A})^{-1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & N_o(\mathbf{A}^\dagger \mathbf{A})^{-1} \end{bmatrix} \right)^{-1}. \quad (35)$$

Notice that the term $(\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{r}(q) \triangleq \hat{\mathbf{h}}_{LS}(q)$ in (34) is the least-squares (LS) estimate of the channel vector $\mathbf{h}(2qN)$ during the q th training period, with estimation error $\mathbf{e}_{LS}(q) = (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{n}(q)$, and associated error covariance matrix

$$\mathbf{R}_e^{LS} = N_o(\mathbf{A}^\dagger \mathbf{A})^{-1}. \quad (36)$$

Clearly, \mathbf{R}_e^{LS} is the same for each training block. The channel estimate $\mathbf{v}(k)$ is then an interpolation [using $\mathbf{W}(k)$] of the LS estimates made during the $2Q + 1$ training periods centered about frame-0. In [14], LS estimation of the users' channels in each of the training periods is also performed; however, since the channels are assumed to be time-invariant, no interpolation between training periods is performed. Furthermore, [14] does not address the design of appropriate training sequences.

Equations (34) and (35) immediately suggest that the training sequences have a minimum required length. In order to form the LS estimates, the matrix $\mathbf{A}^\dagger \mathbf{A}$ must be nonsingular. This occurs if the $2(N_t - L_c) \times 2M(L_c + 1)$ matrix \mathbf{A} is of full column rank, which can only occur if the number of rows of \mathbf{A} is greater than or equal to the number of columns. Consequently, the minimum training sequence length is $N_t = M(L_c + 1) + L_c$.

Equation (34) also suggests a simplified criterion for designing good training sequences. Rather than choosing the sequences to minimize $\sigma_{e_m}^2(k)$ in (22) for each user (the optimal criterion), in this paper, the sequences are chosen to minimize $\text{trace}[\mathbf{R}_e^{LS}]$ —an easier task. This is reasonable, since one would expect that minimizing the error variance of the acquired LS estimates during each training block would also lead to a low interpolation error between training blocks. Note, though, that the same minimum sequence length applies to both the optimal and simplified criteria, except that for the optimal criterion, a rank deficiency in \mathbf{A} causes $\sigma_{e_m}^2(k)$ to be excessively high rather than causing an explicit singularity as for the simplified criterion.

Minimization of $\text{trace}[\mathbf{R}_e^{LS}]$ is made easier by defining the $(N_t - L_c) \times M(L_c + 1)$ matrix \mathbf{B} formed by deleting the odd numbered rows of \mathbf{A} and removing the zeros from the even numbered rows. No information is lost here, since, observing (10) and (25), one can see that an even numbered row of \mathbf{A} and its associated odd numbered row contain the same symbols interspersed with zeros. The odd numbered row is just a right shift by one position of the even numbered row. As a consequence, the even and odd numbered rows are linearly independent such that if \mathbf{B} is of full column rank, then so is \mathbf{A} . Furthermore,

due to the sparse nature of \mathbf{A} , one can show that $\text{trace}[\mathbf{R}_e^{LS}] = 2N_o \text{trace}[\mathbf{G}^{-1}]$, where the Gram matrix \mathbf{G} is given by

$$\mathbf{G} = \mathbf{B}^\dagger \mathbf{B} = \begin{bmatrix} \mathbf{B}_1^\dagger \mathbf{B}_1 & \mathbf{B}_1^\dagger \mathbf{B}_2 & \dots & \mathbf{B}_1^\dagger \mathbf{B}_M \\ \mathbf{B}_2^\dagger \mathbf{B}_1 & \mathbf{B}_2^\dagger \mathbf{B}_2 & \dots & \mathbf{B}_2^\dagger \mathbf{B}_M \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_M^\dagger \mathbf{B}_1 & \mathbf{B}_M^\dagger \mathbf{B}_2 & \dots & \mathbf{B}_M^\dagger \mathbf{B}_M \end{bmatrix}. \quad (37)$$

Evidently, selection of the users' training sequences is accomplished simply by minimizing $\text{trace}[\mathbf{G}^{-1}]$. Although we have simplified the preceding analysis by assuming that the sampled noise sequence is white, we found that the training sequences that minimize $\text{trace}[\mathbf{G}^{-1}]$ also minimize the LS estimation error variance for the case of noise coloration due to the matched filter.

In [16] for the case of a single user ($M = 1$), it is shown that $\text{trace}[\mathbf{G}^{-1}]$ is minimized by choosing a single training sequence such that $\mathbf{G} = \mathbf{B}_1^\dagger \mathbf{B}_1$ is diagonal. This implies that the training sequence must have perfect autocorrelation properties, that is, zero autocorrelation for all lags except zero. The design of such a sequence of arbitrary length is not difficult; for example, see [17].

In the multiuser case, on the other hand, a diagonal \mathbf{G} matrix implies that the M different sequences have not only perfect autocorrelation properties, but perfect crosscorrelation properties as well, that is, zero crosscorrelation for all lags. This is generally very difficult to achieve for arbitrary M and L_c if the training symbols are constrained to lie within the modulation alphabet. In this paper, BPSK training sequences are selected such that the off-diagonal elements of \mathbf{G} (autocorrelation values for nonzero lags and crosscorrelation values for all lags) all fall below a certain threshold, which is chosen to be as low as possible for a given M and L_c . Since the diagonal elements of \mathbf{G} are all equal to $N_t - L_c$, this procedure makes \mathbf{G} strongly diagonal.

Training sequence design is made somewhat easier if the first L_c symbols of each user's sequence are constrained to be the same as the last L_c symbols. With this constraint, the m th user's (modified) data matrix is

$$\mathbf{B}_m = [\mathbf{b}_m \quad T^{-1} \mathbf{b}_m \quad \dots \quad T^{-L_c} \mathbf{b}_m] \quad (38)$$

where \mathbf{b}_m is a length- $(N_t - L_c)$ column vector, and the operator $T^l \mathbf{b}_m$ denotes a circular shift of \mathbf{b}_m by l positions. The shift is up if l is positive and down if l is negative. Note that the m th user's training sequence is the concatenation of the last L_c symbols of \mathbf{b}_m and \mathbf{b}_m itself. The m, n th submatrix of \mathbf{G} can be now be written as

$$\mathbf{B}_m^\dagger \mathbf{B}_n = \begin{bmatrix} \theta_{\mathbf{b}_n, \mathbf{b}_m}(0) & \theta_{\mathbf{b}_n, \mathbf{b}_m}(1) & \dots & \theta_{\mathbf{b}_n, \mathbf{b}_m}(L_c) \\ \theta_{\mathbf{b}_n, \mathbf{b}_m}(-1) & \theta_{\mathbf{b}_n, \mathbf{b}_m}(0) & \dots & \theta_{\mathbf{b}_n, \mathbf{b}_m}(L_c - 1) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{\mathbf{b}_n, \mathbf{b}_m}(-L_c) & \theta_{\mathbf{b}_n, \mathbf{b}_m}(-L_c + 1) & \dots & \theta_{\mathbf{b}_n, \mathbf{b}_m}(0) \end{bmatrix} \quad (39)$$

TABLE I
MINIMUM LENGTH BINARY TRAINING SEQUENCES (INCLUDING PRECURSORS
AND POSTCURSORS) FOUND BY SEQUENTIAL SEARCH. THE SEQUENCES
ARE IN HEXADECIMAL FORM AND MUST BE ZERO PADDED TO THE
LEFT TO MAKE UP THE FULL LENGTH N_t

M	L_c	N_t	Threshold	Training Sequence(s)
1	0	1	0	1
	1	3	0	1
	2	5	1	9
	3	7	0	11
	4	9	1	21
2	0	2	0	0,1
	1	5	0	11,4
	2	8	2	41,C7
	3	11	2	30B,111
	4	14	2	2C0B,461
3	0	3	1	0,1,2
	1	7	2	41,4,4B
	2	11	3	209,62F,69F
	3	15	4	3053,5065,30BB
	4	19	5	18053,5818B,8361
4	0	4	0	0,3,5,6
	1	9	4	101,4,10B,20
	2	14	4	1009,3047,206A,312F
	3	19	4	30053,50335,10581,70BDF
	4	24	4	B0032B,302DC3,D0657D,E1506E

where $\theta_{\mathbf{x}, \mathbf{y}}(l)$ is the periodic crosscorrelation function of the column vectors \mathbf{x} and \mathbf{y} [18] defined as

$$\theta_{\mathbf{x}, \mathbf{y}}(l) = \langle \mathbf{x}, T^l \mathbf{y} \rangle = (T^l \mathbf{y})^\dagger \mathbf{x}. \quad (40)$$

The design of the training sequences now involves selecting a set of M different \mathbf{b}_m s that satisfy three criteria: 1) $|\theta_{\mathbf{b}_m, \mathbf{b}_m}(l)|$ is less than the threshold for $l \in \{1, 2, \dots, L_c\}$ for all m ; 2) $|\theta_{\mathbf{b}_n, \mathbf{b}_m}(l)|$ is less than the threshold for $l \in \{-L_c, \dots, 0, \dots, L_c\}$ for all $m \neq n$; and 3) \mathbf{B} is full column rank. Since the number of combinations of M different \mathbf{b}_m s is huge for several users and typical channel memory lengths, a sequential search is used, rather than an exhaustive one, to build up a set of M training sequences one-by-one that satisfy the three criteria.

Table I shows the results of a computer search for minimum length binary training sequences that meet these criteria. For compactness, the training sequences are listed in hexadecimal form. The most significant bit corresponds to the first symbol to be transmitted in the training sequence, that is $c_m(-L_2)$, and the least significant bit to the last symbol, that is $c_m(N_t - L_2 - 1)$. BPSK symbols are derived from the bits of the training sequences using the mapping $\{0, 1\} \rightarrow \{-1, +1\}$. Note that for the case of $M = 1$, $L_c = 1$ where the best training sequence is 001, the constraint of first and last symbols being equal is lifted in order to avoid a singular Gram matrix.

The table also lists the threshold used for the off-diagonal elements of \mathbf{G} . For each case, the threshold was first set to zero and then increased until a full set of training sequences was found. A value of zero indicates that \mathbf{G} is diagonal, implying that the M training sequences have perfect auto and crosscorrelation properties. This occurs when $N_t - L_c = 1, 2$, or 4. It is interesting to note that the resulting data matrix \mathbf{B} is equivalent to the Hadamard matrix of order 1, 2, or 4, respectively, which is known to have orthogonal columns resulting in a diagonal Gram matrix.

Although we have shown results only for binary training sequences, the search technique presented in this paper applies to the more general case of nonbinary sequences, e.g., QPSK, 8-PSK, and 16-QAM. However, experimentation with QPSK sequences showed that while the search space is larger, the auto- and crosscorrelation properties of the resulting sequences are no better than for the binary sequences listed in Table I.

V. DESIGN ISSUES AND PERFORMANCE

In this section, several design issues are treated, namely the choice of interpolator order, choice of frame length, and efficiency. The performance of the joint channel estimation scheme is then investigated, using the channel estimation quality measures $\sigma_{e_m}^2(k)$ and $\overline{\rho_m}(k)$ defined in (22) and (23), respectively. Unless otherwise specified, the optimal interpolator $\mathbf{U}_{opt}(k) = \mathbf{P}(k)\mathbf{R}_z^\#$ is used so that $\mathbf{R}_e(k)$ in (20) is equal to $\mathbf{R}_e^{opt}(k)$.

Frequency-selective fading is considered using various values of Doppler spread f_{D_m} and rms delay spread $\tau_{rms,m}$. The power delay profile is assumed to be exponential, and the relative delay τ_m is assumed to be distributed uniformly on the interval $[-T/2, T/2]$. Also, $\sigma_{g_m}^2$ is set to 1/2 resulting in the m th user's SNR being $\Gamma_m = P_m/N_o$. Unless otherwise specified, the root-raised cosine transmit pulse $u(t)$ has 50% excess bandwidth. Furthermore, according to Fig. 2, the parameter L_c is set to 4 with $L_1 = -2$ and $L_2 = 2$. The training sequences listed in Table I are used in all cases. Recall that these sequences are of the minimum length $N_t = M(L_c + 1) + L_c$.

A. Interpolator Order

Fig. 4 shows a plot of channel estimation error versus interpolator order, defined as $2Q + 1$. As can be seen, the use of more than about nine training blocks ($Q = 4$) to form the estimate of $\mathbf{h}(k)$ does not significantly decrease the channel estimation error. This behavior was found to be representative of a large variety of fading and SNR conditions. It is also consistent with that observed in [7] for the case of a single user and flat fading. Note that user 2 experiences slightly better performance than user 1 does since the training sequence for user 2 happens to have slightly better autocorrelation properties.

B. Frame Length

Fig. 5 shows a plot of channel estimation error versus frame length N . For a fixed Doppler spread, as N is increased beyond a critical value, the channel estimation error increases sharply, due to the fact that the fading channels are not sampled often enough to allow proper interpolation. Clearly, as the Doppler spread increases, the fading channels must be sampled at a

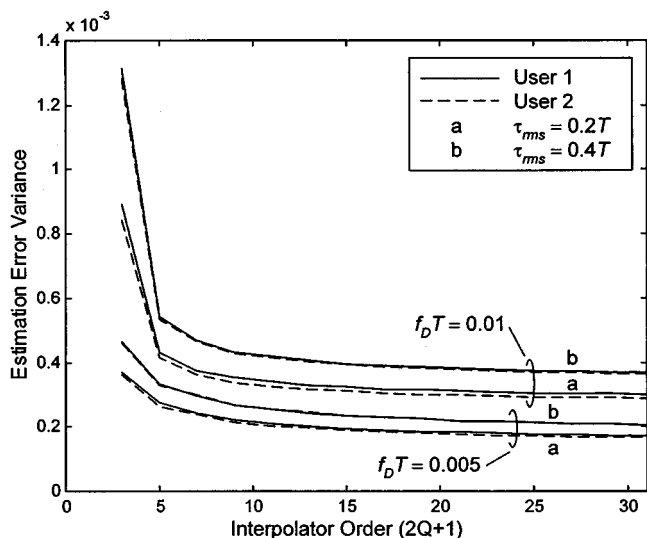


Fig. 4. Channel estimation error versus interpolator order for two equipower users at an SNR of 30 dB. The frame length is $N = 2N_t = 28$ symbols.

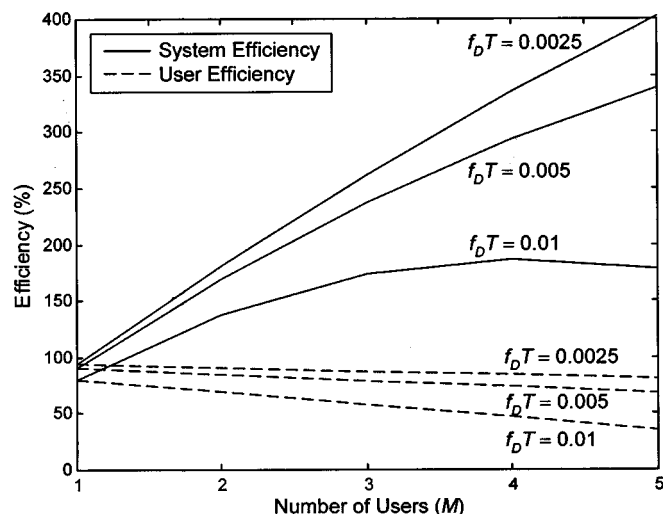


Fig. 6. User and system efficiency for the critical frame lengths found in the previous figure.

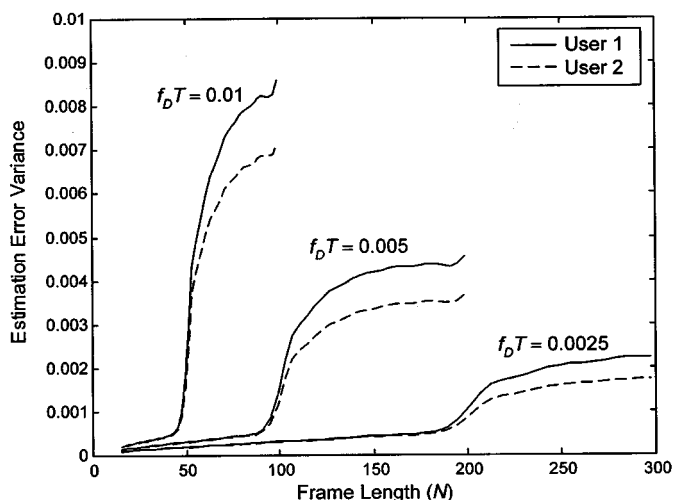


Fig. 5. Channel estimation error versus frame length for two equipower users at an SNR of 30 dB with interpolator order nine. The fading is frequency-selective with $\tau_{rms}/T = 0.2$ for each user.

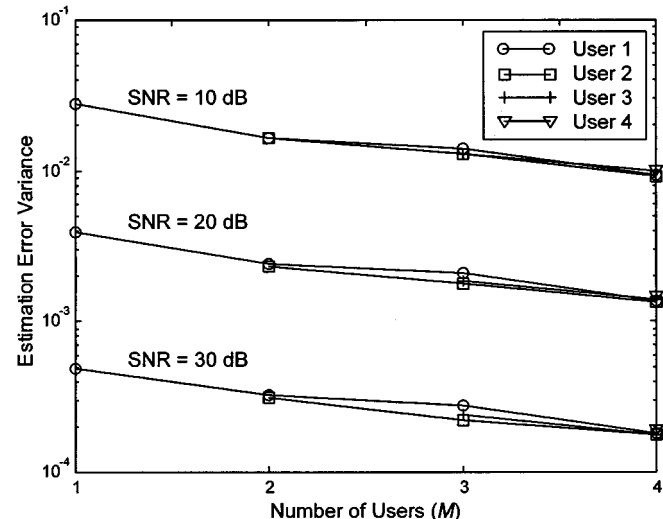


Fig. 7. Channel estimation error versus the number of equipower users with frame length $N = 50$ and interpolator order nine. The fading is frequency-selective with $\tau_{rms}/T = 0.2$ and $f_{D_m}T = 0.005$ for each user.

higher rate (shorter frame length): for $f_{D_m}T = 0.0025, 0.005,$ and $0.01,$ the critical frame lengths are approximately 180, 90, and 45 symbols, respectively. These values correspond closely to the inverse of the Nyquist rate $2f_{D_m}T$. Again, this behavior is consistent with that observed in [7].

C. Efficiency

The transmission efficiency—or throughput—experienced by any user is given by the ratio of the number of data symbols per frame ($N - N_t$) to the frame length N . As the number of users increases, so does the required length of the training sequence, causing the user efficiency to drop. Using the minimum training sequence length found earlier, the user efficiency is $\eta_u = (N - M(L_c + 1) - L_c)/N$. Fig. 6 shows a plot of user efficiency versus the number of users for the critical values of N found above. This plot illustrates significantly reduced efficiency for short frame lengths and a large number of users. In the extreme of fast fading ($f_{D_m}T = 0.01$) with four users

and a frame length of $N = 45,$ the user efficiency drops from its value of 80% corresponding to a single user to a value near 50%. Remember, though, that in the case of frequency reuse within a cell, system capacity may be enhanced through joint detection and/or antenna arrays [3], [6] by allowing four users to share the same frequency/time slot which offsets this reduction in user efficiency. Therefore, we define system efficiency as $\eta_s = M\eta_u$ and plot it on Fig. 6, where an optimal value of M can be seen. This optimal value and the corresponding optimal η_s both increase for slower fading where the frame length can be much greater than for fast fading.

D. Performance

Fig. 7 shows a plot of channel estimation error versus number of users. The estimation error actually decreases with each additional user, which is due to longer training sequences as each additional user is added. Again, each user has a slightly different

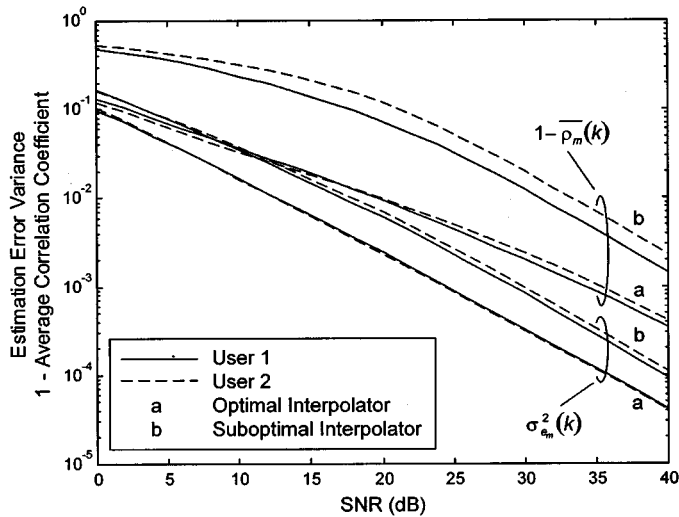


Fig. 8. Channel estimation error and average correlation coefficient for optimal and suboptimal interpolators with two equipower users, frame length $N = 50$, and interpolator order nine. The fading is frequency-selective with $\tau_{\text{rmsm}}/T = 0.2$ and $f_{Dm}T = 0.005$ for each user.

estimation error variance due to the fact that users' training sequences have slightly different autocorrelation properties.

In addition to the consideration of multiple users, our analysis provides an extension to the single-user results of [8] by accounting for coloration of the sampled noise sequence and correlation between the channel taps. Fig. 8 shows the estimation gain achieved by considering these two effects. This graph plots the estimation error variance $\sigma_{e_m}^2(k)$ and one minus the average correlation coefficient $(1 - \bar{\rho}_m(k))$ for both the optimal interpolator $\mathbf{U}_{\text{opt}}(k) = \mathbf{P}(k)\mathbf{R}_z^\#$ and the suboptimal interpolator $\mathbf{U}(k) = \mathbf{P}'(k)\mathbf{R}_z^\#$. The matrices $\mathbf{P}'(k)$ and $\mathbf{R}_z^\#$ are obtained by modifying $\mathbf{P}(k)$ and \mathbf{R}_z as follows: the off-diagonal elements of $\mathbf{R}_h(\cdot)$ in (28) and (30) are set to zero, and the autocorrelation function of the sampled noise sequence in (30) is redefined so that $\phi_n(j) = N_o$ for $j = 0$ and zero otherwise. Observing the estimation error curves, one can see that, for moderate SNR, up to a 6-dB gain in estimation error may be achieved by considering the noise coloration and the inter-tap correlations. An even a larger improvement is observed in the correlation coefficient. It should be noted that this benefit comes at little or no cost in computational load when performing the channel estimation.

This plot also demonstrates that the asymptotic channel estimation error varies inversely with SNR, as expected. Similarly, the channel estimate and the true channel become increasingly correlated with increasing SNR. For example, at an SNR of 40 dB, use of the optimal interpolator results in an average correlation coefficient of approximately $\bar{\rho}_m(k) = 0.9997$ —a high degree of correlation. This is extremely desirable, since, as observed in [7] for the case of a single user and flat fading, the behavior of $\bar{\rho}_m(k)$ with SNR results in the elimination of the bit error rate (BER) floor that is commonly observed in systems employing differential detection.

The above graphs have all illustrated the performance for equipower users. In normal system operation, though, we can expect power differences among the users. Even in this case, the *absolute* estimation error variance, given by $\text{trace}[\mathbf{R}_{e_m}(k)]$

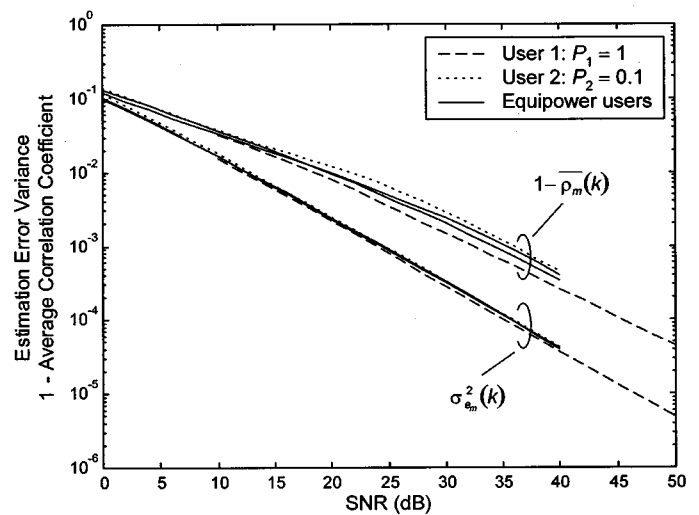


Fig. 9. Channel estimation error and average correlation coefficient for two nonequipower users with frame length $N = 50$ and interpolator order nine. The fading is frequency-selective with $\tau_{\text{rmsm}}/T = 0.2$ and $f_{Dm}T = 0.005$ for each user.

[see (21)], is approximately equal for all users, since the estimation error is produced primarily by the receiver noise $n(t)$. On the other hand, the *relative* estimation error variance $\sigma_{e_m}^2(k)$, defined in (22), is normalized by the sum of channel tap variances, which, in turn, is proportional to $A_m^2 = 2P_m$ [see (5)]. Therefore, at a given noise level, the relative error variance is close to inversely proportional to signal power, giving stronger users better relative error results than weaker users.

Fig. 9 shows the relative estimation error variance and average correlation coefficient for the case of two users with a power difference of 10 dB. For reference, the performance of the equipower case ($P_1 = P_2 = 1$) is shown on the same graph. Note that each user's curves are plotted against the user's *own* SNR, defined in (8). As a result, the curves are closely spaced which illustrates the approximate inverse dependence on signal power, as discussed above. In a typical operating scenario, all users are detected at a common receiver noise level N_o . Consequently, the weak and strong users operate at SNR values that are 10 dB apart, and the horizontal axis must be interpreted in this light. For example, if the strong user is at 30 dB SNR, then the weak user is at 20 dB with a relative estimation error variance roughly 10 dB greater than that of the strong user. Similarly, the average correlation coefficient for the stronger user is significantly better (closer to unity) than for the weaker user.

In all of the results presented here, it is assumed that the interpolator is designed assuming perfect knowledge of the Doppler fade rate f_{Dm} , the rms delay spread τ_{rmsm} , and the SNR Γ_m for each user. In a practical situation, these parameters may not be known at design time and may, in fact, be different for each user. Furthermore, they may change as the scattering environment changes. As mentioned previously, this may be handled by adopting a worst-case design methodology. This approach was investigated here by designing an interpolator using worst-case values of f_{Dm} and τ_{rmsm} (equal for all users) and a typical operating SNR Γ_m (again, equal for all users). The performance of this fixed interpolator was then investigated in an environment with different channel and SNR conditions.

Not surprisingly, it was found that better performance may be obtained by optimizing the interpolator to match the actual fading and SNR conditions. More importantly, however, it was found that the performance is neither degraded nor improved if the actual channel conditions are better than those designed for (lower f_{D_m} and τ_{rms_m} than the design values) and if the actual SNR's are different from the design value. If the actual channel conditions are worse than those designed for, significant degradation in performance occurs.

VI. CONCLUSIONS

In this paper, we have developed a pilot-based MMSE technique for jointly estimating the channels of multiple cochannel users in a TDMA system that is useful for a variety of multiuser detection and interference cancellation schemes. The paper makes two key contributions: first, it accounts for time variation of the channels both within and between training periods. The former is essential in a multiuser environment where the training sequences are necessarily longer than in a single-user environment, resulting in significant variation during training. Second, a simple strategy for the selection of appropriate training sequences for the multiple cochannel users is developed. The selection strategy is demonstrated for the special case of BPSK sequences.

Several design issues are considered including the choice of interpolator order, the choice of frame length, and efficiency. Results show that the user throughput decreases with each additional user since the minimum length of the training sequences grows linearly with the number of users. However, system efficiency may increase, since, through the use of joint detection and/or antenna arrays, multiple users may be allowed to share the same frequency/time slot.

Performance results are presented, and it is shown that the channel estimation error decreases with each additional user due to increasing training sequence lengths. Furthermore, it is shown that for nonequipower users, the absolute channel estimation error variance for all users is approximately equal, irrespective of power differences, but the degree of correlation between the channel estimate and the true channel is higher for the stronger users than the weaker users.

REFERENCES

- [1] Z. Zvonar and D. Brady, "Multiuser detection in single-path fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 1729–1739, Feb./Mar./Apr. 1994.
- [2] L. K. Rasmussen *et al.*, "Impact of estimation error on multiuser detection in CDMA," in *Proc. IEEE VTC'98*, Ottawa, Canada, May 18–21, 1998, pp. 1844–1848.
- [3] S. J. Grant and J. K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Commun.*, vol. 46, pp. 1038–1049, Aug. 1998.
- [4] J. Joung and G. L. Stüber, "Performance of truncated co-channel interference canceling MLSE for TDMA systems," in *Proc. IEEE VTC'98*, Ottawa, Canada, May 18–21, 1998, pp. 1710–1714.
- [5] K. Giridhar *et al.*, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Trans. Commun.*, vol. 45, pp. 473–484, Apr. 1997.
- [6] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 528–539, July 1984.

- [7] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.
- [8] S. A. Fechtel and H. Meyer, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. Commun.*, vol. 42, pp. 1639–1650, Feb./Mar./Apr. 1994.
- [9] W. K. Lo, D. D. Falconer, and A. U. H. Sheikh, "Adaptive equalization and diversity combining for mobile radio using interpolated channel estimates," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 636–645, Aug. 1991.
- [10] B. C. Ng, M. Cedervall, and A. Paulraj, "A structured channel estimator for maximum-likelihood sequence detection," *IEEE Commun. Lett.*, vol. 1, pp. 52–55, Mar. 1997.
- [11] M. L. Moher and J. H. Lodge, "TCMP—A modulation and coding strategy for Rician fading channels," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1347–1355, Dec. 1989.
- [12] F. Ling, "Coherent detection with reference symbol based channel estimation for direct sequence CDMA uplink communications," in *Proc. IEEE VTC'93*, Secaucus, NJ, May 1993, pp. 335–344.
- [13] O. Nesper and P. Ho, "A pilot symbol assisted interference cancellation scheme for an asynchronous DS/CDMA system," in *Proc. IEEE Globecom'96*, London, U.K., November 1996, pp. 1447–1451.
- [14] P. A. Ranta, A. Hottinen, and Z. Honkasalo, "Co-channel interference cancelling receiver for TDMA mobile systems," in *Proc. IEEE ICC'95*, Seattle, WA, June 18–22, 1995, pp. 17–21.
- [15] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [16] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *Proc. Inst. Elect. Eng.*, pt. F, vol. 138, pp. 371–378, Aug. 1991.
- [17] C. L. Ng, K. B. Letaief, and R. D. Murch, "Complex optimal sequences with constant magnitude for fast channel estimation initialization," *IEEE Trans. Commun.*, vol. 46, pp. 305–308, Mar. 1998.
- [18] D. V. Sarwate and M. B. Pursley, "Crosscorrelation properties of pseudo-random and related sequences," *Proc. IEEE*, vol. 68, pp. 593–619, May 1990.



Stephen J. Grant (S'95–A'00) was born in Vancouver, BC, Canada, in 1969. He received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada in 1993 and the M.A.Sc. and Ph.D. degrees in electrical engineering from Simon Fraser University, Burnaby, BC, Canada, in 1996 and 2000, respectively. His M.A.Sc. thesis involved the analysis, design, and implementation of a DSP-controlled adaptive feedforward linearizer for RF power amplifiers. His Ph.D. work concerned increasing the capacity of TDMA cellular systems through joint detection with diversity antenna arrays.

He is now with the Advanced Development and Research Group of Ericsson, Inc., Research Triangle Park, NC. His current research interests include multiuser detection and channel estimation, adaptive antenna arrays, transmit diversity, and space-time coding.



James K. Cavers (M'90–SM'99) was born in Port Alice, BC, Canada in 1944. He received the B.A.Sc. degree in engineering physics and the Ph.D. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1966 and 1970, respectively.

From 1970 to 1979, he was an Assistant, then Associate, Professor in the Department of Systems Engineering at Carleton University in Ottawa. He spent 1979 to 1982 as a Program Manager at MacDonald Dettwiler and Associates, in Vancouver, followed by a year as Senior Engineer at Glenayre Electronics, also in Vancouver. In 1983, he joined the School of Engineering Science at Simon Fraser University, Burnaby, BC, Canada, where he holds the rank of Professor. From 1990 to 1994, he was Director of the School. His research interests include modulation and detection for mobile communications, and integrated RF/DSP design.

Dr. Cavers is the 1995 recipient of the Gold Medal in Engineering and Applied Science from the Science Council of B.C. and the 1998 recipient of the Principal Innovation Award of the E.C. Manning Foundation.