Multiuser Channel Estimation for Detection of Cochannel Signals

Stephen J. Grant and James K. Cavers
School of Engineering Science, Simon Fraser University
Burnaby, B.C., Canada V5A 1S6
http://www.ensc.sfu.ca/people/grad/grantq

Abstract— **Estimation of the channel impulse responses of multiple cochannel users is a key requirement of all multiuser detection and interference cancellation techniques, though little attention has been paid to subject in the context of TDMA systems. This paper addresses a pilot-based technique for multiuser channel estimation in a TDMA system, and makes several new contributions: it allows for time variation of the channels within and between training sequences, it accounts for colouration of the sampled noise sequence as well as correlation between the channel taps, and and it considers users to be asynchronous resulting in a technique whereby explicit timing recovery is unnecessary. In addition, it addresses selection of appropriate training sequences.**

I. Introduction

Multiuser detection and interference cancellation techniques for both CDMA and TDMA systems [1]-[3] have received much attention recently due to their potential for increasing system capacity. One aspect common to all of the multiuser detection and interference cancellation techniques, though, 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., [4],[5]. For the case of multiuser systems, pilot-based channel estimation has been studied extensively only for CDMA, e.g., [6], where processing gain suppresses interference in the channel estimator.

This paper addresses pilot-based techniques for estimating the channel impulse responses of multiple cochannel users in a TDMA system. In contrast to [7], it allows time variation of the channels within and between the training sequences—an essential feature even at moderate fading rates—and addresses the selection of appropriate training sequences. Furthermore, it accounts for colouration of the sampled noise sequence, as well as correlation between the channel taps, thereby rendering the channel estimates truly optimal. These effects have previously been ignored, even in the most comprehensive study of channel estimation for single-user systems [5].

II. Signal and Channel Models

Fig. 1 shows a diagram of the transmission of M cochannel signals through independently fading, dispersive channels, each represented by the time-variant channel impulse response (CIR) $g_m(\tau;t)$, where τ denotes the memory of the impulse response, and t denotes the time variation. The mth user's transmitted signal is given by

$$s_m(t) = \sqrt{2P_m}\sum_n c_m(n)\,u(t-nT-\tau_m) \qquad (1)$$

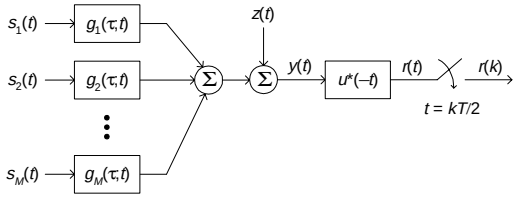

Fig. 1. Cochannel signal model.

where $c_m(n)$ is a data or training symbol, T is the symbol period, P_m is the power in the bandpass equivalent of $s_m(t)$, and $u(t)$ is the transmit pulse with deterministic autocorrelation function $x(\alpha) = \int u(t)\,u^*(t-\alpha)\,dt$. The relative delay τ_m appears in (1) since, in general, the signals from the various users arrive asynchronously due to differing propagation delays.

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. The output of the matched filter $u^*(-t)$, using (1), is

$$r(t) = \sum_{m=1}^{M}\sum_n c_m(n)\,h_m(t-nT;t) + n(t) \qquad (2)$$

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

$$h_m(\tau;t) = \sqrt{2P_m}g_m(\tau-\tau_m;t) \circledast x(t) \qquad (3)$$

and $\circledast$ denotes convolution. The autocorrelation function of the filtered noise process $n(t)$ is $\phi_n(\alpha) = N_o x(\alpha)$. Notice that the relative delay τ_m is considered part of the channel impulse response.

MMSE estimation of the users' channels requires knowledge of the second order statistics of $h_m(\tau;t)$ summarized by the following correlation function:

$$\begin{aligned} R_{h_m}(\tau_1,\tau_2,\alpha) &= \frac{1}{2}E\left[h_m(\tau_1;t)\,h_m^*(\tau_2;t-\alpha)\right] \\ &= 2P_m R_{g_m}(\alpha)\cdot \\ &\quad \int P_m(\tau)\,x(\tau_1-\tau)\,x^*(\tau_2-\tau)\,d\tau \end{aligned} \qquad (4)$$

where $P_m(\tau) = \frac{1}{2}E\left[|g_m(\tau-\tau_m;0)|^2\right]$ and $R_{g_m}(\alpha)$ is the normalized temporal autocorrelation function of the channel.

For isotropic scattering, $R_{g_m}(\alpha) = J_0(2\pi f_{D_m}\alpha)$ where f_{D_m} is the maximum Doppler shift for user m. The latter equality in (4) is a consequence of assuming a wide sense stationary uncorrelated scattering (WSSUS) channel as well as a separable scattering function.

Denoting $f(\tau_m)$ as the probability density function (pdf) of the relative delay τ_m and $P_{g_m}(\tau)$ as the power delay profile of the channel (e.g. exponential with area $\sigma^2_{g_m}$ and RMS delay spread τ_{rms_m}), the function $P_m(\tau)$ is given by $P_m(\tau) = \int P_{g_m}(\tau - \tau_m) f(\tau_m)\, d\tau_m$, i.e. the convolution of $P_{g_m}(\tau)$ and $f(\tau_m)$. As can be seen, the calculation of $R_{h_m}(\tau_1, \tau_2, \alpha)$ does not depend on τ_m itself—only on its pdf. In other words, explicit timing recovery is unnecessary; the relative delay τ_m is simply estimated as part of the channel.

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

$$r(k) = \sum_{m=1}^{M} \mathbf{c}_m^T(k)\, \mathbf{h}_m(k) + n(k). \tag{5}$$

The vector $\mathbf{h}_m(k)$ consists of samples of $h_m(\tau; t)$ at $T/2$-spaced delays evaluated at time $t = kT/2$. It is assumed here that $h_m(\tau; t)$ is non-causal and of finite duration spanning the range of delays $\tau \in \left[L_1 T, \left(L_2 + \frac{1}{2}\right) T\right]$ where $L_1 \leq 0$ and $L_2 \geq 0$. Consequently, both $\mathbf{h}_m(k)$ and the symbol vector $\mathbf{c}_m^T(k)$ are of finite length $2(L_c + 1)$ where $L_c = L_2 - L_1$. The vector $\mathbf{c}_m(k)$ contains the symbols $c_m(\lfloor k/2 \rfloor - n)$ for $n \in \{L_1, \ldots, L_2\}$ interspersed with zeros. For k even, the zeros appear in the 1st, 3rd, 5th, ... positions of $\mathbf{c}_m(k)$, and for k odd appear in the 0th, 2nd, 4th, ... positions.

The autocorrelation matrix of $\mathbf{h}_m(k)$ is given by $\mathbf{R}_{\mathbf{h}_m}(j) = \frac{1}{2}\mathrm{E}\left[\mathbf{h}_m(k)\, \mathbf{h}_m^\dagger(k-j)\right]$. Using (4), its u, vth element is

$$\{\mathbf{R}_{\mathbf{h}_m}(j)\}_{u,v} = R_{h_m}\left(\frac{uT}{2}, \frac{vT}{2}, \frac{jT}{2}\right) \tag{6}$$

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

III. Joint Channel Estimation

MMSE estimation of the M users' channels relies upon the periodic insertion of a unique training sequence into each user's data sequence, the choice of which will be discussed in Section IV. It is assumed that all users' training sequences are inserted at the same time, although different propagation delays make their arrivals asynchronous. 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 variation of the channels are tracked. The frame structure, along with symbol and frame indexing conventions used throughout this paper, is shown in Fig. 2. As can be seen, the length of each frame is N symbols, and the length

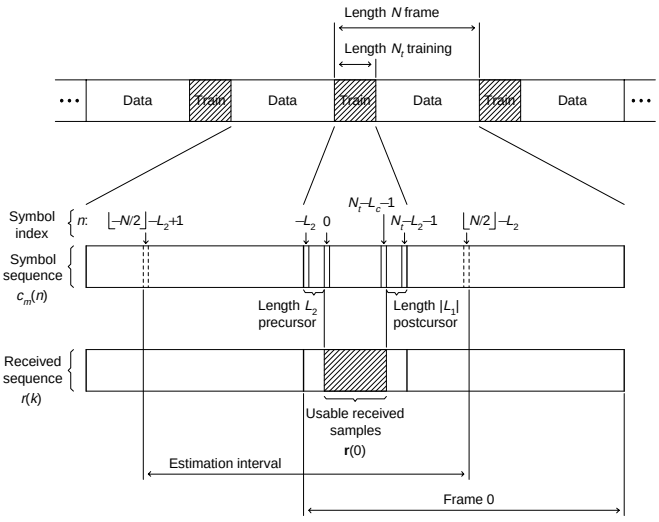

Fig. 2. Frame structure and indexing conventions.

of each training sequence is N_t symbols. Note that n indexes symbols, and k samples, so that $n = \lfloor k/2 \rfloor$.

Since the users' channels are to be estimated jointly, we define the vector $\mathbf{h}(k)$ as the concatenation of the M users' channel vectors:

$$\mathbf{h}(k) = \left[\begin{array}{cccc} \mathbf{h}_1^T(k) & \mathbf{h}_2^T(k) & \cdots & \mathbf{h}_M^T(k) \end{array}\right]^T. \tag{7}$$

Since the users' channels fade independently, the autocorrelation matrix of $\mathbf{h}(k)$, denoted $\mathbf{R}_{\mathbf{h}}(j)$, is block diagonal, with the mth block given by $\mathbf{R}_{\mathbf{h}_m}(j)$.

Consider the MMSE estimation of $\mathbf{h}(k)$ in the interval $\lfloor -N/2 \rfloor - L_2 < n \leq \lfloor N/2 \rfloor - L_2$ that is, mid-frame to mid-frame either side of the training period in frame-0. 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

$$\mathbf{z} = \left[\begin{array}{cccc} \mathbf{r}^T(-Q) & \cdots & \mathbf{r}^T(0) & \cdots & \mathbf{r}^T(Q) \end{array}\right]^T \tag{8}$$

where the vector $\mathbf{r}(q)$ contains a subset of the received samples during the qth training block (called the usable samples in Fig. 2). 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 $\mathbf{R}_{\mathbf{z}} = \frac{1}{2}\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]$, 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 $\mathbf{z}$ and avoid explicit inversion of $\mathbf{R}_{\mathbf{z}}$. Accordingly, we use eigendecomposition to write the covariance matrix of $\mathbf{z}$ as $\mathbf{R}_{\mathbf{z}} = \mathbf{M}\mathbf{\Lambda}\mathbf{M}^\dagger$, where $\mathbf{M} = [\mathbf{M}_1\ \mathbf{M}_2]$ and $\mathbf{\Lambda}$ is a block diagonal matrix with diagonal blocks $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$. $\mathbf{\Lambda}_1$ contains the

dominant eigenvalues of $\mathbf{R_z}$, and $\mathbf{\Lambda}_2$ contains those eigenvalues that fall below some very small threshold. The non-square matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ contain the eigenvectors (as columns) corresponding to the eigenvalues in $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ respectively. Now, we 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 $\mathbf{v}(k) = \mathrm{E}[\mathbf{h}(k) \mid \mathbf{w}]$. Since $\mathbf{h}(k)$ and $\mathbf{w}$ are jointly Gaussian, the conditional mean is linear in $\mathbf{w}$, and is given by

$$\mathbf{v}(k) = \frac{1}{2}E\left[\mathbf{h}(k)\,\mathbf{w}^{\dagger}\right](\mathbf{R_w})^{-1}\mathbf{w} = \mathbf{P}(k)\,\mathbf{R_z^{\#}}\mathbf{z} \quad (9)$$

where $\mathbf{P}(k) = \frac{1}{2}\mathrm{E}\left[\mathbf{h}(k)\,\mathbf{z}^{\dagger}\right]$ and $\mathbf{R_z^{\#}} = \mathbf{M}_1\mathbf{\Lambda}_1^{-1}\mathbf{M}_1^{\dagger}$. The latter quantity is recognized as the generalized pseudoinverse, or Moore-Penrose generalized inverse, of $\mathbf{R_z}$ [8]. Note that for short training sequences, $\mathbf{R_z}$ may not be ill-conditioned; in this case, $\mathbf{R_z^{\#}} = \mathbf{R_z^{-1}}$.

For the optimal channel estimate $\mathbf{v}(k)$, the channel estimation error at each position in the frame is denoted $\mathbf{e}(k)$ such that

$$\mathbf{h}(k) = \mathbf{v}(k) + \mathbf{e}(k) \quad (10)$$

where $\mathbf{v}(k)$ and $\mathbf{e}(k)$ are uncorrelated. The estimation error covariance matrix is, in turn, given by

$$\mathbf{R_e}(k) = \mathbf{R_h}(0) - \mathbf{P}(k)\,\mathbf{R_z^{\#}}\mathbf{P}^{\dagger}(k) \quad (11)$$

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

$$\sigma_{e_m}^2(k) = \frac{\mathrm{tr}\left[\mathbf{R}_{\mathbf{e}_m}(k)\right]}{\mathrm{tr}\left[\mathbf{R}_{\mathbf{h}_m}(0)\right]} \quad (12)$$

where $\mathbf{R}_{\mathbf{e}_m}(k)$ is the mth block of the main diagonal of $\mathbf{R_e}(k)$. Although this error measure depends directly on k, we found very little variation across the frame.

We now examine the optimal estimator in (9) in more detail to obtain the required matrices. It is convenient to first introduce the following data matrix:

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_M \end{array}\right] \quad (13)$$

where jth row of $\mathbf{A}_m$ is simply the symbol vector $\mathbf{c}_m^T(j)$ (defined in relation to (5)), where $j \in \{0, 1, \ldots, 2(N_t - L_c) - 1\}$. Due to the precursor and postcursor in each training sequence, $\mathbf{A}_m$ consists only of symbols from the mth user's training sequence, and not on adjacent data symbols. Using (5), (7), (13), the jth component of $\mathbf{r}(q)$ can be written as

$$r(2qN + j) = \mathbf{a}_j\mathbf{h}(2qN + j) + n(2qN + j) \quad (14)$$

where $\mathbf{a}_j$ is the jth row of the data matrix $\mathbf{A}$. With this expression in hand, the elements of the matrices $\mathbf{P}(k)$ and $\mathbf{R_z}$ may be easily determined.

Using (8), the qth submatrix of $\mathbf{P}(k)$ is $\mathbf{P}_q(k) = \frac{1}{2}\mathrm{E}\left[\mathbf{h}(k)\,\mathbf{r}^{\dagger}(q)\right]$, where $q \in \{-Q, \ldots, Q\}$. Now using (14), and assuming the noise and channel fading process are uncorrelated, the jth column of $\mathbf{P}_q(k)$ is

$$\left\{\mathbf{P}_q(k)\right\}_j = \mathbf{R_h}(k - 2qN - j)\,\mathbf{a}_j^{\dagger}. \quad (15)$$

Using (8) again, the q,pth submatrix of $\mathbf{R_z}$ is $\mathbf{R}_{\mathbf{z}_{q,p}} = \frac{1}{2}\mathrm{E}\left[\mathbf{r}(q)\,\mathbf{r}^{\dagger}(p)\right]$. Using (14), the i,jth element of $\mathbf{R}_{\mathbf{z}_{q,p}}$ is

$$\begin{aligned}\left\{\mathbf{R}_{\mathbf{z}_{q,p}}\right\}_{i,j} &= \mathbf{a}_i\mathbf{R_h}(2(q-p)N + i - j)\,\mathbf{a}_j^{\dagger} + \\ &\quad \phi_n(2(q-p)N + i - j). \end{aligned} \quad (16)$$

Here, $\phi_n(j) = N_0 x\left(\frac{jT}{2}\right)$ is the autocorrelation function of the (coloured) noise sequence. Observing (15) and (16), one can see that the interpolation matrix $\mathbf{P}(k)\,\mathbf{R_z^{\#}}$ depends only on the channel autocorrelation matrix $\mathbf{R_h}(j)$, the data matrix $\mathbf{A}$, and the noise autocorrelation function, which are all known at design time.

IV. Choice of Training Sequences

Optimal choice 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 $\sigma_{e_m}^2(k)$ defined in (12). 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, it is assumed that the users' channels vary slowly enough that they may be considered constant over the duration of each training period; second, the matrix $\mathbf{R_z}$ is assumed to be non-singular so that $\mathbf{R_z^{\#}} = \mathbf{R_z^{-1}}$; and third, the noise sequence $n(k)$ is assumed to be white. It must be emphasized, though, that these assumptions are made for the purposes of training sequence selection only. The resulting sequences are then used to calculate the optimal channel estimate $\mathbf{v}(k)$ as in (9).

Using the above assumptions, it is shown in [9] that $\mathbf{v}(k)$ may be calculated by first forming the least squares (LS) channel estimate $\widehat{\mathbf{h}}_{LS}(q) = \left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{r}(q)$ in each of the $2Q + 1$ training periods centered about frame-0, and then interpolating the acquired LS estimates between training periods. The estimation error covariance matrix for the acquired LS estimates is simply $\mathbf{R_e^{LS}} = N_o\mathbf{G}^{-1}$ where the Gram matrix $\mathbf{G} = \mathbf{A}^{\dagger}\mathbf{A}$.

This immediately suggests that the training sequences have a minimum required length. In order to form the LS estimates, the matrix $\mathbf{G}$ must be non-singular. 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 $N_t = M(L_c + 1) + L_c$.

The LS formulation also suggests a simplified criterion for choosing good training sequences. Rather than choosing the sequences to minimize $\sigma^2_{\hat{e}_m}(k)$ for each user (the optimal criterion), in this paper the sequences are chosen to minimize the trace of $\mathbf{G}$ — an easier task since $\mathbf{G}$ depends only on the set of training sequences. 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.

As is shown in [10] for the case of a single user, trace$\left[\mathbf{G}^{-1}\right]$ is minimized by choosing a single training sequence such that $\mathbf{G}$ is diagonal. In the multiuser case, this implies that the M different training sequences should be chosen such that they have both perfect autocorrelation properties and perfect crosscorrelation properties. This is generally very difficult to achieve for arbitrary M and L_c. Consequently, in this paper, BPSK training sequences are chosen such that the off-diagonal elements of $\mathbf{G}$ 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. Because the number of combinations of M different sequences can be extremely large, a sequential search is used, rather than an exhaustive one, to build up a set of M training sequences one-by-one. Due to space constraints, the resulting sequences are not listed here, but are contained in [9].

V. Design Issues and Performance

In this section, several design issues as well as the performance of the joint channel estimation scheme are investigated. Fig. 3 shows the effect of interpolator order, defined as $2Q+1$, on the estimation error. As can be seen, the use of more than about 9 training blocks ($Q = 4$) does not significantly decrease the channel estimation error variance. This behaviour was found to be representative of a large variety of fading and SNR conditions. Furthermore, it is consistent with that observed in [4] for the case of a single user and flat fading.

Fig. 4 shows the effect of frame length N on the estimation error. For a fixed Doppler spread, as N is increased beyond a critical value, the channel estimation error variance 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 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 $2f_{D_m}T$.

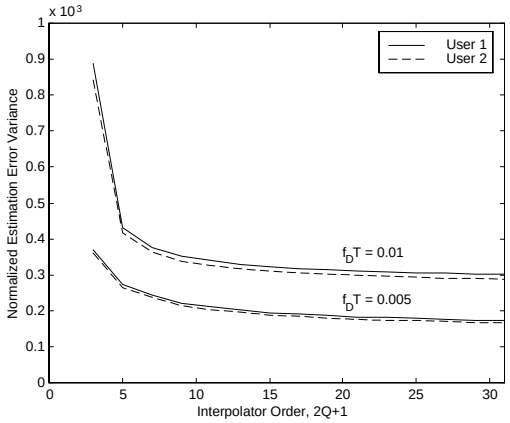

Fig. 3. Effect of interpolator order: 2 equipower users, $\tau_{rms_m}/T = 0.2$, $N = 2N_t = 28$.

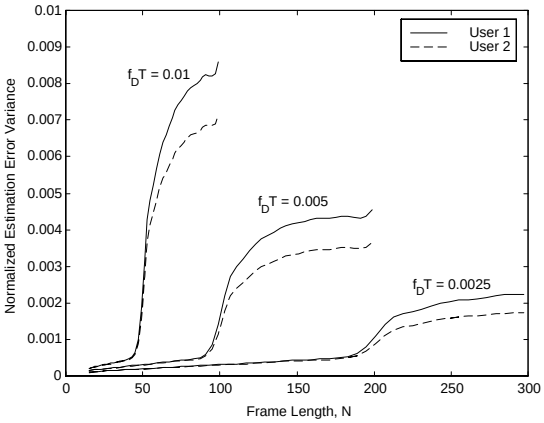

Fig. 4. Effect of frame length: 2 equipower users, $\tau_{rms_m}/T = 0.2$, $Q = 4$.

Again, this behaviour is consistent with that observed in [4].

The transmission efficiency, or throughput, experienced by any user is given by the ratio of the number of data symbols per frame to the frame length. As the number of users increases, so does the required length of 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. 5 shows a plot of user efficiency vs. number of users for the critical frame lengths found in Fig. 4. 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 4 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 cell, system capacity is enhanced by allowing 4 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 in on Fig. 5, where an optimal value of

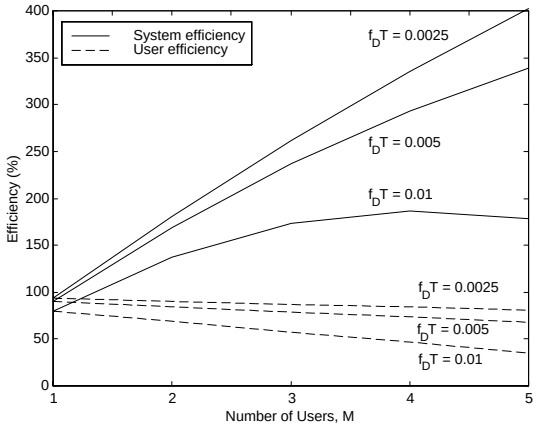

Fig. 5. User and system efficiency for $L_c = 4$ and the critical frame lengths found in Fig. 4.

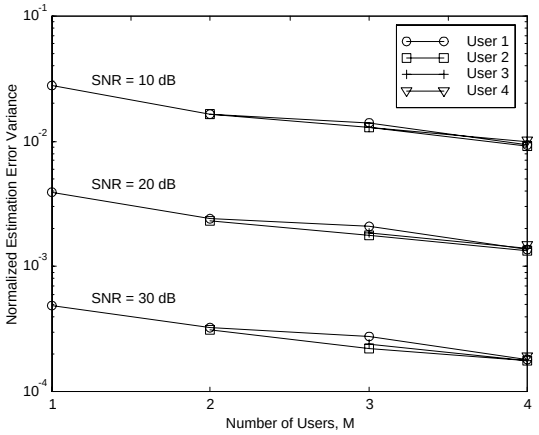

Fig. 6. Effect of number of users: $\tau_{rms_m}/T = 0.2$, $f_{D_m} = 0.005$, $N = 50$, $Q = 4$.

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.

Fig. 6 shows the effect of the number of users on estimation error. As can be seen, the estimation error variance actually decreases with each additional user, which is due to longer training sequences as each additional user is added. This plot also shows that each user has a slightly different estimation error variance which is due to the fact that the users' training sequences have different autocorrelation properties and are not mutually orthogonal. Furthermore, Fig. 6 illustrates that, as expected, estimation error variance decreases as the inverse of SNR.

VI. Conclusions

In this paper, we have developed a pilot-based 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. Several design issues are considered, including the selection of multiple training sequences, choice of interpolator order, the choice of frame length, and efficiency. Results show that, although the user throughput decreases with each additional user due to increased training sequence length, the system efficiency increases, since multiple users are allowed to share the same frequency/time slot. Furthermore, it is shown that the channel estimation error per user is inversely related to SNR, and actually decreases with each additional user.

References

[1] Z. Zvonar and D. Brady, "Multiuser detection in single-path fading channels," *IEEE Trans. Comm.*, vol. 42, pp. 1729-1739, Feb./March/April 1994.

[2] S.J. Grant and J.K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Comm.*, vol. 46, pp. 1038-1049, Aug. 1998.

[3] J. Joung and G.L. Stüber, "Performance of truncated co-channel interference canceling MLSE for TDMA systems," *Proceedings of IEEE VTC'98*, Ottawa, Canada, May 18-21, 1998, pp. 1710-1714.

[4] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 40, 686-693, Nov. 1991.

[5] S.A. Fechtel and H. Meyer, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. Comm.*, vol. 42, pp. 1639-1650, Feb./March/April 1994.

[6] O. Nesper and P. Ho, "A pilot symbol assisted interference cancellation scheme for an asynchronous DS/CDMA system," *IEEE Globecom'96*, London, UK, Nov., 1996, pp. 1447-1451.

[7] P.A. Ranta, A. Hottinen, and Z. Honkasalo, "Co-channel interference cancelling receiver for TDMA mobile systems," *IEEE ICC'95*, Seattle, WA, June 18-22, 1995, pp. 17-21.

[8] S. Haykin, *Adaptive Filter Theory 3rd Edition*, Upper Saddle River, NJ: Prentice-Hall, 1996.

[9] S.J. Grant and J.K. Cavers, "Multiuser channel estimation for detection of cochannel signals," submitted to *IEEE Trans. Comm.*, Oct. 1998.

[10] S.N. Crozier, D.D. Falconer, and S.A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proc.-F*, vol. 138, pp. 371-378, Aug. 1991.