Fast Adaptive Equalization/Diversity Combining for Time-Varying Dispersive Channels

Heung-No Lee and Gregory J. Pottie, *Member, IEEE*

***Abstract*— We examine adaptive equalization and diversity combining methods for fast Rayleigh-fading frequency selective channels. We assume a block adaptive receiver in which the receiver coefficients are obtained from feedforward channel estimation. For the feedforward channel estimation, we propose a novel reduced dimension channel estimation procedure, where the number of unknown parameters are reduced using *a priori* information of the transmit shaping filter's impulse response. Fewer unknown parameters require a shorter training sequence. We obtain least-squares, maximum-likelihood, and maximum *a posteriori* (MAP) estimators for the reduced dimension channel estimation problem. For symbol detection, we propose the use of a matched filtered diversity combining decision feedback equalizer (DFE) instead of a straightforward diversity combining DFE. The matched filter form has lower computational complexity and provides a well-conditioned matrix inversion. To cope with fast time-varying channels, we introduce a new DFE coefficient computation algorithm which is obtained by incorporating the channel variation during the decision delay into the minimum mean square error (MMSE) criterion. We refer to this as the non-Toeplitz DFE (NT-DFE). We also show the feasibility of a suboptimal receiver which has lower complexity than a recursive least squares adaptation, with performance close to the optimal NT-DFE.**

***Index Terms*—Adaptive equalizer, channel parameter estimation, fading dispersive channels.**

I. Introduction

WIRELESS digital communications systems such as IS-54, GSM, and personal communication systems (PCS's) suffer from many channel impairments, including signal fading, multipath delay spread, and Doppler spread. Signal fading results in a very low instantaneous channel signal-to-noise ratio (SNR), which necessitates diversity combining. The multipath delay spread and stringent transmit filtering requirements may cause severe intersymbol interference (ISI). Thus, an equalizer is mandatory due to ISI-induced irreducible error floors. These become significant once τ_o, the root-mean-square (rms) delay spread of a multipath delay profile (MPDP), exceeds about 1/10 the symbol period (T) [1], [3], [8], [32]. This multipath-induced ISI problem in slow fading can be effectively dealt with using an equalizer and explicit diversity combining.

However, the receiver design for the frequency-selective fading channels is quite demanding in fast fading. For example, in IS-54, with a carrier frequency of 900 MHz and assuming a mobile moving at a maximum highway speed of 120 km/h, the maximum normalized Doppler fading rate $f_{dm}T$ (the product of the maximum Doppler fading rate and the symbol period) reaches up to 0.0042 [8]. This implies that the minimum time between the two fading nulls is 5 ms $(1/2f_{dm})$, which is even shorter than the proposed burst length of 6.7 ms. To deal with such rapid fading and dispersive channels, we need not only a fast and robust channel tracking algorithm but also a highly optimized detection algorithm which makes full use of all of the acquired channel information.

For tracking of fast changing dispersive channels, a block adaptive decision feedback equalizer (DFE) based on feedforward channel estimation [4], [10], [20] has been shown to be more effective than the conventional symbol by symbol adaptation methods, such as least mean squares (LMS) or even recursive least squares (RLS) [5], [26]. Other block adaptive schemes, based on the feedforward channel estimation but using the maximum-likelihood sequence estimator (detector), can be found in [6], [9], and [11]. In this paper we follow the block adaptive framework of [20]. In particular, the receiver is assumed to operate on continuous transmitted frames, where each frame consists of training and unknown data segments. A "snap-shot" channel estimate is obtained from the training segment. Channel tracking during the data segments is performed by interpolating a set of the snap-shot channel estimates. With the interpolated channel estimates, the receiver filter coefficients are computed. Thus, the block adaptive strategy we discuss in this paper will be applicable to an IS-54 downlink [31] and future PCS's [35].

In this paper we extend previous results [10], [20] on the block adaptive strategy using channel estimate-based DFE. First, we propose a new channel estimation method that is robust against fast fading and also requires shorter training sequences. We also compare two possible diversity-combining DFE implementations and report a structure that is more robust and less computationally complex for the block adaptive strategy. We further propose a new DFE coefficient computation algorithm to deal with very fast fades. We illustrate the improved performance of the receiver through computer simulations. Finally, we show the feasibility of a low computational complexity but suboptimal solution for tracking fast fades.

Paper approved by S. Ariyavisitakul, the Editor for Wireless Techniques and Fading of the IEEE Communications Society. Manuscript received July 26, 1996; revised December 23, 1996, August 5, 1997, and November 10, 1997. This paper was presented in part at the 1997 IEEE 47th Vehicular Technology Conference, Phoenix, AZ, May 1997.

The authors are with the Electrical Engineering Department, University of California at Los Angeles, Los Angeles, CA 90095 (e-mail: hlee@ee.ucla.edu; pottie@icsl.ucla.edu).

Publisher Item Identifier S 0090-6778(98)06651-3.

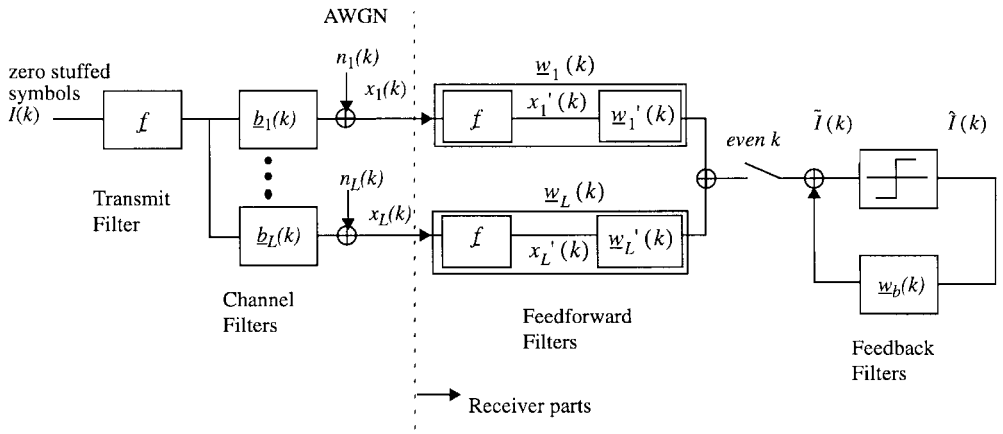


Fig. 1. The baseband L-diversity channel model and the straightforward diversity-combining DFE receiver.

The remainder of this paper is organized as follows. Section II describes the underlying system and channel model. Section III describes the new channel estimation method and derives channel estimators under different criteria. Section IV discusses the channel interpolation scheme. Section V compares the two possible diversity-combining DFE realizations and derives the new DFE coefficient computation algorithm. Section VI presents simulation results for the different receiver schemes. Performance of an RLS channel tracking DFE is presented for comparison. Finally, Section VII presents our conclusions.

We adopt the following submatrix convention throughout the paper:

- for an $[Ni \times Nj]$ matrix A, the row index ranges over $i = 0, 1, \cdots, Ni-1$, and the column index $j = 0, 1, \cdots, Nj-1$;
- $A_{(i1:i2,\, j1:j2)}$ denotes the submatrix of A, including rows of indexes from $i1$ to $i2$ and columns of indexes from $j1$ to $j2$;
- $A_{(:,\, j1:j2)}$ denotes the submatrix of A, including columns from $j1$ to $j2$ and all rows;
- an underbar, i.e., $\underline{B_l}$, denotes a column vector, and $B_{l,\, i1:i2}$ denotes the subvector of $\underline{B_l}$, including rows from $i1$ to $i2$.

II. Baseband Equivalent Channel Model

Fig. 1 includes the baseband equivalent channel model. A square root raised cosine (SRRC) transmit filter $f(t)$ with a rolloff $\beta = 0.35$ is assumed. The L-diversity received signals, corrupted by independent additive complex-valued white Gaussian noise (AWGN), are assumed to be mutually independent. Since $x_l(t)$, the received signal at each diversity branch, is bandlimited with an excess bandwidth of $(1+\beta)(1/T)$, $T/2$-spaced sampling is considered, i.e., $x_l(k) := x_l(t = kT/2)$, where k denotes the $T/2$-spaced epoch index throughout the paper. The noise is also assumed to be $T/2$-spaced sampled, and the sampled noise $n_l(k)$ has a variance σ_n^2. To represent symbol transmission at the rate of $1/T$, we use the zero-stuffed sequence $\{I(k)\}$ such that the value of $I(k)$ at every odd k is zero and $I(k)$ at even k represents the symbol transmitted at the baud rate.

For the $T/2$-spaced sampled system, the impulse responses of the transmit and the channel filters can be realized with $T/2$-spaced tapped delay line FIR filters [20]. Each filter is represented by a column vector. A $T/2$-spaced sampled version of the transmit filter $f(t)$ is represented by a unit energy $[31 \times 1]$ column vector $\underline{f}$, which corresponds to a 15-symbol truncation. The time-varying impulse response of the channel is expressed as the baseband equivalent complex-valued response and is represented by a $[N_R \times 1]$ column vector having complex-valued time-varying taps, i.e., $\underline{b_l}(k) = [b_{l,0}(k) \cdots b_{l,N_R-1}(k)]^T$, where N_R is the number of the time-varying channel taps ($N_R = 3$ in this paper). Finally, an overall channel impulse response is represented by $\underline{g_l}(k)$, i.e., $\underline{g_l}(k) := \underline{f} \otimes \underline{b_l}(k)$, where $\otimes$ denotes the convolution operation.

The mountainous terrain multipath delay profile (MT-MPDP) has the worst delay spread among the various land mobile MPDP's [1], [14], [29]. We use the MT-MPDP of [20], where the $T/2$-spaced channel coefficients have relative average powers of $\alpha_0^2 = 0$ dB, $\alpha_1^2 = -5$ dB, and $\alpha_2^2 = -15$ dB, and $\sum_{i=0}^{2} \alpha_i^2 = 1$. The rms delay spread of this model is about 1/4 symbol period (10.40μs for a symbol period T of 41.7 μs). We assume that each tap gain undergoes Rayleigh amplitude fading according to Jakes' model [15] with a time-correlation function

$$\phi(\Delta t) = J_o(2\pi f_{dm} \Delta t) \tag{1}$$

and a Doppler spreading spectrum

$$S(f_d) = \begin{cases} (1-(f_d/f_{dm})^2)^{-1/2}, & |f_d| \le f_{dm} \\ 0, & |f_d| \ge f_{dm} \end{cases} \tag{2}$$

where $J_o(\cdot)$ implies the zeroth order Bessel function of the first kind, and f_{dm} is the maximum Doppler fading rate.

Then, each channel tap coefficient $b_{l,i}(k)$ is modeled as

$$b_{l,i}(k) = \alpha_i \rho\left(t = k\frac{T}{2};\ ST_{li}\right) \tag{3}$$

where $\rho(t;\ ST_{li}) = N_D \sum_{j=0}^{8} d_j \cos(2\pi\delta_j f_{dm} t + ST_{li})$, N_D is a normalization coefficient that makes the second moment of $\rho(t;\ ST_{li})$ equal to 1.0, d_j is the complex-valued amplitude,

and δ_j is the $(j+1)$th Doppler shift. In addition, wide sense stationary uncorrelated scattering [24] of the three Rayleigh-fading tap coefficients is assumed. This is simulated by having independent random starting phases ST_{li} in (3). That is, for $l = 1, \cdots, L$ and $i = 0, 1, \cdots, N_R - 1$, each ST_{li} randomly takes a real number at the start of each sample path generation [20]. Finally, in this paper, assuming a symbol rate $(1/T)$ of 24 ksps, fast fading corresponds to $f_{dm} = 100$ Hz ($f_{dm}T = 0.0042$) and slow fading to $f_{dm} = 10$ Hz ($f_{dm}T = 0.000\,42$).

III. Snap-Shot Channel Impulse Response Estimation

The channel estimation is performed at the receiver by observing the channel response to the training segment. We make a "snap-shot" assumption that during the observation interval mT, the channel is effectively time invariant. We focus on a single channel branch to describe the channel estimation procedure since each diversity channel branch has an identical structure. Thus, we drop the indexes for epoch and diversity, such as $\underline{b}$ for a channel $\underline{b}_l(k)$ and $\underline{g}$ for an overall channel $\underline{g}_l(k)$, during $k = 0, 1, \cdots, 2m-1$. The estimators to be derived are all *consistent* estimators under the static channel assumption. However, in fast fading this assumption leads to an estimation error if m is large.

In Section III-A we describe a novel reduced-dimension channel estimation method and obtain a new channel estimation equation. In Section III-B we apply three classical optimality criteria to the new channel estimation equation and obtain three estimators under each criterion. Section III-C provides the minimum mean square estimation error analysis of the three estimators. Section III-D discusses the training sequences. In Section VI we investigate the performance of these estimators, in terms of minimum-mean-square estimation errors and bit-error rate (BER).

A. Problem Formulation for the Snap-Shot Channel Estimations

The use of a bandwidth efficient transmit filter (the square root raised cosine pulse shaping filter in Fig. 1) increases the effective span of the overall channel impulse response (CIR). For the *flat* fading channel, a receive filter matched to the SRRC filter can be used to recover an isolated source pulse with zero crossings every T seconds. In the frequency-selective channel, however, the composite pulse $\underline{f} \otimes \underline{b} \otimes \underline{f}$ is no longer Nyquist due to the multipath channel $\underline{b}$. Thus, we need an estimate of the overall channel response for optimum symbol detection.

From the system description of Section II, the overall channel filter $\underline{g}$, which is the convolution of the SRRC filter $\underline{f}$ and the channel $\underline{b}$, spans 16 symbol periods and can be represented by a $[33 \times 1]$ vector. To reduce the length of a training sequence required, a truncated channel is used in the channel estimation equation. That is, the estimation accuracy is traded off for shorter training since a longer CIR requires longer training sequences. We represent the truncated overall CIR with a $[2N_c \times 1]$ vector $\underline{\tilde{g}}$, where $N_c < 15$.

For the $T/2$-spaced system, the received signal over an observation period is described by

$$x(k) \cong \sum_{i=0}^{2N_c-1} \tilde{g}_i I(k-i) + n(k) \tag{4}$$

where $\{I(k)\}$ is the $T/2$-spaced zero-stuffed input symbol sequence, i.e., it includes known training symbols for even k and 0.0 for odd k. Since k is the $T/2$-spaced epoch index, the zero-stuffed input sequence $\{I(k)\}$ represents the input symbols at the symbol rate of $1/T$.

Note that there are $2N_c$ unknown parameters in (4). Previous channel estimation methods (e.g., see [9], [10], and [20] for $T/2$-spaced sampling and [30] and [33] for T-spaced sampling) estimate these parameters without exploiting the fact that the overall CIR is a convolution of the transmit shaping filter and the time-varying channel filter. In fact, the impulse response of the shaping filter is *a priori* known to the receiver and thus the unknown parameters are only the channel taps.

Using the *a priori* knowledge of the transmit filter, we reduce the number of unknown parameters in the channel estimation equation. The truncated CIR $\underline{\tilde{g}}$, the convolution of a truncated transmit filter and the channel filter, can be represented by $\underline{\tilde{g}} = \mathcal{F}\underline{b}$, where $\mathcal{F}$ is a $[2N_c \times N_R]$ matrix whose elements $\mathcal{F}_{(i,j)}$ can be determined for the SRRC filter (Appendix A). Then, (4) can be rewritten as

$$x(k) \cong \sum_{i=0}^{2N_c-1} \sum_{j=0}^{N_R-1} \mathcal{F}_{(i,j)} b_j I(k-i) + n(k). \tag{5}$$

Note that the number of unknown parameters in (5) is now N_R. This brings about a number of benefits. First, with fewer unknown parameters, a shorter observation window is needed. Second, with a shorter observation period, the snap-shot channel estimation performs robustly in fast fading. Since a snap-shot channel estimation problem relies on a fixed channel during the observation period, a long observation interval may become counterproductive [10], [20]. Finally, the estimates will be more accurate when there are fewer parameters to be estimated. Having obtained the estimates of the channel, the overall channel can be computed from the convolution of the estimate $\underline{b}$ and the SRRC filter.

We now continue the problem formulation for the estimation of the channel $\underline{b}$. It can be shown that $\underline{\tilde{r}} := [x(0)\, x(1)\, x(2)\, \cdots\, x(2m-1)]^T$, which is a $[2m \times 1]$ vector that includes the $T/2$-spaced sampled received signal during an observation period mT, can be divided into even and odd indexed parts, and each part produces a T-spaced channel estimation problem, i.e.,

$$\underline{r} = XF\underline{b} + \underline{n} \tag{6}$$

where $\underline{r} = [x(0)\, x(2)\, \cdots\, x(2m-1)]^T$ for the even symbol sequence and $\underline{r} = [x(1)\, x(3)\, \cdots\, x(2m-1)]^T$ for the odd symbol sequence. X is a $[m \times N_c]$ Toeplitz matrix whose elements are determined from the training sequence of length

N_t, $N_t = m + N_c - 1$ such that

$$X = \begin{bmatrix} I(0) & I(-2) & \cdots & I(-2(N_c-1)) \\ I(2) & I(0) & \cdots & I(-2(N_c-2)) \\ \cdots & \cdots & \cdots & \cdots \\ I(2(m-1)) & I(2(m-2)) & \cdots & I(2(m-N_c)) \end{bmatrix}. \tag{7}$$

F is an $[N_c \times N_R]$ SRRC matrix, an *a priori* matrix that can be determined for the even and odd parts as described in Appendix A. The $[m \times 1]$ noise vector $\underline{n}$ is a multivariate Gaussian with a zero-mean vector and a covariance matrix of $R_n = \sigma_n^2 \mathbf{I}_m$, where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix.

From the two equations representing odd and even symbols, two estimators of $\underline{b}$ can be obtained. We choose the one that yields a smaller theoretical mean square estimation error for a given training sequence.

B. Three Classical Estimation Criteria for the Snap-Shot Estimates of $\underline{b}$

Given the new estimation model of (6), least-squares estimation (LSE), maximum-likelihood estimation (MLE), and maximum *a posteriori* (MAP) estimation criteria are considered for the estimation of $\underline{b}$. In the derivations of these estimators the training matrix X and the *a priori* matrix F are assumed to be fixed in their contents and dimensions. In addition, we consider cases with $m \geq N_R$ in this paper. The inverse matrices for each estimation operator to be derived are assumed to be well-defined with an optimal or suboptimal choice of the training sequence matrix X. The optimal training sequences will be discussed in Section III-D.

If there is no *a priori* statistical knowledge about the noise and the channel, the LSE of $\underline{b}$ can be considered, i.e.,

$$\hat{\underline{b}}_{\mathrm{LSE}} := \arg\min_{\underline{b}} |\underline{r} - XF\underline{b}|^2 = (F^H X^H XF)^{-1}(XF)^H \underline{r} \tag{8}$$

where the superscript "H" implies the conjugate transpose operation of a matrix and arg denotes the argument. This results in the lowest complexity estimator among the three. The $[N_R \times m]$ matrix $(F^H X^H XF)^{-1}(XF)^H$ can be precomputed and stored, and then an estimate of $\underline{b}$ can be obtained by multiplying it with the observation vector $\underline{r}$. We note that Khayrallah *et al.* [36] have independently obtained a similar LSE expression as (8) using the idea of incorporating the knowledge of the modulation filter into a least square channel estimation operation at the receiver in a GSM-based system.

The maximum-likelihood estimator of $\underline{b}$ can be obtained as follows:

$$\begin{aligned} \hat{\underline{b}}_{\mathrm{ML}} &:= \arg\max_{\underline{b}}(p(\underline{r}|\underline{b})) \\ &= \arg\max_{\underline{b}}[-(\underline{r} - XF\underline{b})^H R_n^{-1}(\underline{r} - XF\underline{b})] \end{aligned} \tag{9}$$

where R_n is the covariance matrix of the noise. Setting the gradient of the quadratic term equal to zero, we obtain

$$\hat{\underline{b}}_{\mathrm{ML}} = (F^H X^H R_n^{-1} XF)^{-1}(X^H F^H R_n^{-1}) \cdot \underline{r}. \tag{10}$$

Thus, MLE requires the second-order statistics of the noise, such as the noise covariance matrix R_n. MLE therefore performs better than LSE provided that the noise is correlated and the autocorrelation function of the noise is known. In our estimation model of (6), however, we have assumed white noise, i.e., $R_n = \sigma_n^2 \mathbf{I}_m$, and thus LSE and MLE are identical.

Note that interpreting R_n^{-1} as a weighting matrix, MLE of $\underline{b}$ can be interpreted as optimally weighted LSE. Thus, MLE and LSE are similar in a sense that both minimize a square residual error $(\underline{r} - XF\hat{\underline{b}})$ but not the estimation error $\hat{\underline{b}} - \underline{b}$.

An estimator which directly minimizes the mean square estimation error of $\underline{b}$ requires *a priori* statistical knowledge of $\underline{b}$. The MAP estimator is in this category. In particular, the MAP estimator is by definition

$$\hat{\underline{b}}_{\mathrm{MAP}} := \arg\max_{\underline{b}}(p(\underline{b}|\underline{r})) = \arg\max_{\underline{b}}\left(\frac{p(\underline{b}, \underline{r})}{p(\underline{r})}\right). \tag{11}$$

In our case, the noise vector is a multivariate Gaussian and thus the posterior density $p(\underline{b}|\underline{r})$ is also a Gaussian distribution where the mode and the mean coincide. With some algebraic manipulations of the posterior density, we obtain

$$\begin{aligned} \hat{\underline{b}}_{\mathrm{MAP}} &= E\{\underline{b}|\underline{r}\} \\ &= R_b(F^H X^H)(XFR_b F^H X^H + R_n)^{-1}\underline{r} \end{aligned} \tag{12}$$

where $R_b := E\{\underline{b}\,\underline{b}^H\}$. This MAP estimator of $\underline{b}$ amounts to the minimum mean square estimator of $\underline{b}$ (Appendix B).

Note that MAP not only requires R_n but also R_b. Thus, in practice it can be employed only after enough information about the noise variance and multipath has been obtained. While collecting the information, we can employ LSE. In this paper, we assume that they are estimated. In particular, diagonal elements of the channel correlation matrix R_b, the average powers of each multipath, i.e., the MPDP $\{\alpha_i^2\}_{i=0,1,2}$ defined in Section II, are assumed to be estimated; off-diagonal elements are all zero valued assuming wide sense stationary uncorrelated scattering of the multipath components.

C. The Mean Square Channel Estimation Errors

It is useful to compare the estimators in terms of their theoretical mean square channel estimation error (MSCEE) performance. We first derive the mean square estimation error matrix for each criterion. Then, the MSCEE is obtained from the *trace* of the mean square estimation error matrix. These theoretical results will be compared with simulations in Section VI.

MLE is an unbiased estimator. This can be easily verified by taking the *expectation* of the following equation, which is obtained from substituting (6) into (10), i.e.,

$$\hat{\underline{b}}_{\mathrm{ML}} = \underline{b} + (F^H X^H R_n^{-1} XF)^{-1} XF)^{-1}(F^H X^H R_n^{-1})\underline{n} \tag{13}$$

and by using $E\{\underline{n}\} = \underline{0}_m$, where $\underline{0}_m$defines an $[m \times 1]$ vector of element of zeros. Similar steps can be taken to show that LSE is also unbiased. Thus, for MLE and LSE, the error covariance matrix of the estimator is equal to the mean square channel estimation error matrix. The mean square estimation error matrix of MLE is

$$\begin{aligned} \Theta_{\mathrm{ML}} &:= E\{(\hat{\underline{b}}_{\mathrm{ML}} - E(\hat{\underline{b}}_{\mathrm{ML}}))(\hat{\underline{b}}_{\mathrm{ML}} - E(\hat{\underline{b}}_{\mathrm{ML}}))^H\} \\ &= (F^H X^H R_n^{-1} XF)^{-1}. \end{aligned} \tag{14}$$

The error covariance matrix of (14) attains the Cramer–Rao lower bound (unbiased class), as shown in Appendix C. Thus, $\hat{\underline{b}}_{\mathrm{ML}}$ is the best linear unbiased estimator for the estimation problem of (6).

In our problem, however, $R_n = \sigma_n^2 \mathbf{I}_m$ is assumed; thus, LSE and MLE produce identical results, i.e.,

$$\Theta_{\mathrm{ML}} = \sigma_n^2 (F^H X^H X F)^{-1} = \Theta_{\mathrm{LSE}} \tag{15}$$

and $\hat{\underline{b}}_{\mathrm{ML}} = \hat{\underline{b}}_{\mathrm{LSE}}$.

Since the MAP estimator is a biased estimator, we directly obtain the mean square error matrix. By defining $\hat{\underline{b}}_{\mathrm{MAP}} := B^+ \underline{r}$, where $B^+ := R_b (F^H X^H)(X F R_b F^H X^H + R_n)^{-1}$, we have

$$\begin{aligned}\Theta_{\mathrm{MAP}} :=& E\{(\hat{\underline{b}}_{\mathrm{MAP}} - \underline{b})(\hat{\underline{b}}_{\mathrm{MAP}} - \underline{b})^H\} \\ =& B^+ E\{\underline{r}\,\underline{r}^H\} B^{+H} - E\{\underline{b}\,\underline{r}^H\} B^{+H} \\ & - B^+ E\{\underline{r}\,\underline{b}^H\} + E\{\underline{b}\,\underline{b}^H\}. \end{aligned} \tag{16}$$

Then, the MSCEE matrix is

$$\begin{aligned}\Theta_{\mathrm{MAP}} =& E\{\underline{b}\,\underline{b}^H\} - B^+ E\{\underline{r}\,\underline{b}^H\} \\ =& (R_b - B^+ X F R_b). \end{aligned} \tag{17}$$

As defined in (6) there are two *a priori* matrices F for the even and the odd observation vectors, and thus two estimators of $\underline{b}$ can be obtained for each estimation criterion. From the two, we select the one that produces a smaller mean square estimation error.

D. Binary Training Sequences for LSE of $\underline{b}$

Crozier [4] tabulates binary training sequences (BTS's; binary sequences of 1 and -1) for different channel lengths N_c and observation sequence lengths m. The sequence is designed to minimize the *trace* of the error covariance matrix of an LSE of $\underline{\tilde{g}}$. They are found either from exhaustive computer search or using the "m-sequences." The same design concept can be applied to the least squares estimator of $\underline{b}$, and new training sequences which minimize the *trace* of the error covariance matrix of the LSE of $\underline{b}$ can be obtained by the procedure of Appendix D. However, for two reasons, Crozier's BTS will be used for our system simulations. First, the improvement, afforded by the new sequence compared to the BTS, is typically less than 1 dB in SNR, while the exact SNR saving depends on the value of N_c and m. For a short training sequence, the difference narrows. Thus, the BTS is nearly optimal. Second, the elements of the new sequence are real valued and not usually members of a digital signal constellation.

IV. Channel Tracking by Interpolation

The periodic training provides a snap-shot channel estimate at the end of each training interval. Then, during the unknown data segments, the estimates of the time-varying channels are obtained from interpolation on a set of estimates of $\underline{b}$. From the interpolated estimates, optimum receiver coefficients such as the matched filter and DFE filter coefficients are obtained. We follow the interpolation framework of [20]. The main difference here is that we interpolate the channel estimate $\underline{b}$, instead of the overall channel estimate $\underline{g}$.

Two parameters are of importance for interpolation, relating the performance of interpolation to the issues of training overhead and interpolation delay. One parameter is the frequency of periodic channel estimation, i.e., the length of a frame B, where a frame consists of a training block of length N_t and a data block of length N_d, i.e., $B = N_t + N_d$. According to the sampling theorem, B should satisfy $B \leq 1/2 f_{dm} T$. For instance, if $f_{dm} T = 0.0042$, the shortest null to null distance of a fading tap is about 240 symbols. Thus, B should be less than 120 symbols. The other parameter is Q, the number of estimates of $\underline{b}$ used in each interpolation. In this paper only $Q = 4$ will be considered. Thus, an interpolation over four consecutive channel estimates, two past and two future, is performed to obtain an interpolate of $\underline{b}$ at an epoch during the middle data segment. The maximum interpolation delay for $Q = 4$ is $3B$ symbol periods. We use a sinc function (sin x/x) interpolator [20].

The snap-shot estimation leads to an estimation lag since the channel is, in fact, continuously varying during the training observation interval. As the result, an estimate $\underline{b}$, which is obtained at the end of a training interval, say at an epoch $k = K$, does not best evaluate the channel $\underline{b}(K)$. To be used in the interpolation steps, the epoch of the estimate $\underline{b}$ is chosen to be $k = K - m$, such that it is the median epoch of the observation interval $k = K - 2m + 1, \cdots, K - 1, K$. This choice is shown to minimize the mean square estimation errors from our simulation.

At the expense of the interpolation delay, channel tracking by interpolation provides immunity to the decision delay problem that is inherent in recursive channel tracking methods which rely on the detected symbols to obtain a channel state update. This advantage, however, has not been fully utilized in symbol detection in previous research; for example, see [10] and [20]. In Section V a new DFE computation algorithm that fully exploits interpolation will be developed.

V. Diversity-Combining DFE

In this section we discuss how to obtain diversity-combining DFE coefficients using the interpolated channel estimates. In Section V-A we apply the MMSE criterion to the receiver structure depicted in Fig. 1, and obtain the "straightforward" solution. This reformulation of the DFE algorithm contributes in two respects. First, we incorporate the channel variation during the decision delay into the MMSE criterion, and thus the solution will fully exploit the channel interpolation scheme of Section IV. That is, the channel variations during the decision delay are available from interpolation, and all should be utilized in finding the optimum receiver coefficients. Second, from the straightforward solution we derive the "matched filtered" solution, providing insights into how the two are related.

For the channel-estimate based receiver, the coefficients of diversity combining DFE are computed from the channel estimates, which involves some form of matrix inversion. The straightforward diversity combining DFE turns out to

be problematic due to the severe eigenvalue spread in the correlation matrix of the Wiener–Hopf equation. Advancing a few critical steps from the straightforward solution, however, we arrive at the matched filtered form of diversity combining DFE which solves the problem of eigenvalue spreads, while still taking into account the channel variation over the decision delay.

A. The Minimum Mean Square Error Diversity Combining DFE

The receiver side of Fig. 1 depicts the straightforward diversity combining DFE. We apply the minimum mean square error (MMSE) criterion to this structure and obtain the straightforward solution. Before the derivation, we now review notations and assumptions. k still denotes the $T/2$-spaced epoch index and, thus, "even k" corresponds to symbol rate sampling. $n_l(k)$ denotes the complex-valued AWGN at the lth branch, with zero mean and variance of σ_n^2. We assume that the additive noise sources $n_l(k)$ for each diversity branch are mutually independent and also independent of the input symbol. We also assume that the input symbol sequence $\{I(k)\}$ for even k, i.e., the sequence without zero stuffing, consists of complex-valued zero-mean unit variance and mutually uncorrelated random variables. Each $T/2$-spaced interpolated overall channel is assumed perfectly estimated and is denoted by a $[\tilde{N}_g \times 1]$ vector $\underline{g_l}(k)$. We assume that $\tilde{N}_g$ is even, and $N_g = \tilde{N}/2$. For the receiver, $\underline{w_l}(k)$ denotes the $T/2$-spaced feedforward filter at lth diversity branch. A $[N_b \times 1]$ vector $\underline{w_b}(k)$ denotes the T-spaced feedback filter. We choose the received signal $x_l(k)$ as the input signal to the adaptive filters $\underline{w_l}(k)$ and $\underline{w_b}(k)$. For the straightforward solution in Section V-B, we use $x_l'(k)$ as the input signal to the adaptive filters $\underline{w_l'}$ and $\underline{w_b}(k)$ for reasons to be discussed.

Now denoting $\underline{w}(k) := [\underline{w_l}^T(k) \cdots \underline{w_L}^T(k)\ \underline{w_b}^T(k)]^T$, we want to find the optimal vector $\underline{w_o}(k)$ that minimizes the mean-square error at each decision instant ($k = 0, 2, 4, \cdots$), i.e.,

$$\underline{w_o}(k) = \arg\min_{\underline{w}(k)} E\{|\tilde{I}(k) - I(k-\tilde{\Delta})|^2 | \underline{g_l}(q),\ q = k, k-1, \cdots, k-\tilde{\Delta},\ l = 1, \cdots, L\} \quad (18)$$

where $\tilde{\Delta}$ is the required decision delay in units of $T/2$. We also assume that $\tilde{\Delta}$ is even, and $\Delta = \tilde{\Delta}/2$. The predecision value $\tilde{I}(k)$ is now described by

$$\tilde{I}(k) = \sum_{l=1}^{L} \underline{x_l}^T(k)\underline{w_l}(k) + \underline{I_b}^T \underline{w_b}(k) \quad (19)$$

where $\underline{I_b}(k) = [I(k-2-\tilde{\Delta}) \cdots I(k-2N_b-\tilde{\Delta})]^T$ (assuming the past decisions were correct) and $\underline{x_l}(k) := [x_l(k)\ x_l(k-1) \cdots x_l(k-\tilde{\Delta})]^T$. By defining $\underline{s}(k) := [\underline{x_1}^T(k) \cdots \underline{x_L}^T(k)\ \underline{I_b}^T(k)]^T$, $\tilde{I}(k)$ can be written compactly as $\tilde{I}(k) = \underline{s}^T(k)\underline{w}(k)$.

Note that the length of $\underline{x_l}(k)$ is $\tilde{\Delta}+1$ and thus the length of the feedforward filter $\underline{w_l}(k)$ is also $\tilde{\Delta}+1$. At this point, we are simply assuming a very large feedforward filter length for the derivation of the matched filtered solution. The desired relationship between the feedforward filter of a finite length and the decision delay can be established later for each solution.

Now each $\underline{x_l}(k)$, the input vector to the feedforward filter $\underline{w_l}(k)$, can be written as

$$\underline{x_l}(k) = H_l(k)\underline{I}(k) + \underline{n_l}(k) \quad (20)$$

where $H_l(k)$ is given in (20a), shown at the bottom of the page, $\underline{I}(k) := [I(k)\ I(k-2) \cdots I(k-2(\Delta+N_g-1))]^T$ and $\underline{n_l}(k) := [n_l(k)\ n_l(k-1) \cdots n_l(k-\tilde{\Delta})]^T$.

In what follows, assuming a decision at $k=0$, we omit the notation of epoch (the parenthesis) from the matrices for brevity and retrieve it after the solution is derived. We also omit the notation for the mathematical *conditioning* operation in (18) but it is understood that the mathematical *expectation* is meant to apply only to the noise and the input sequences.

Then the mean square measure of (18) can be compactly written as $E\{|\underline{s}^T\underline{w} - I(-\tilde{\Delta})|^2\}$. Invocation of the orthogonality principle gives $E\{\underline{s}(\underline{s}^H\underline{w}^* - I^*(-\tilde{\Delta}))\} = 0$, which results in the Wiener–Hopf normal equation

$$E\{\underline{s}\,\underline{s}^H\}\underline{w}^* = E\{\underline{s}I^*(-\tilde{\Delta})\}. \quad (21)$$

Now denoting $\underline{x} := [\underline{x_1}^T \cdots \underline{x_1}^T]T$, $R_{xx} := E\{\underline{x}\,\underline{x}^H\}$, $R_{xI} := E\{\underline{x}\,\underline{I_b}^H\}$, $\underline{w_f} := [\underline{w_1}^T \cdots \underline{w_L}^T]^T$, and $\underline{c} := E\{\underline{x}I^*(-\Delta)\}$, (21) can be written as

$$\begin{aligned} &\begin{bmatrix} R_{xx} & R_{xI} \\ R_{xI}^H & E(\underline{I_b}\underline{I_b}^H) \end{bmatrix} \begin{bmatrix} \underline{w_f} \\ \underline{w_b} \end{bmatrix}^* \\ &= \begin{bmatrix} E(\underline{x}\,\underline{x}^H) & E(\underline{x}\,\underline{I_b}^H) \\ E(\underline{I_b}\underline{x}^H) & \mathbf{I}_{N_b} \end{bmatrix} \begin{bmatrix} \underline{w_f} \\ \underline{w_b} \end{bmatrix}^* \\ &= \begin{bmatrix} \underline{c} \\ \underline{0}_{N_b} \end{bmatrix} \end{aligned} \quad (22)$$

and in a detailed form as

$$\begin{bmatrix} E(\underline{x_1}\underline{x_1}^H) & E(\underline{x_1}\underline{x_2}^H) & \cdots & E(\underline{x_1}\underline{I_b}^H) \\ E(\underline{x_2}\underline{x_1}^H) & E(\underline{x_2}\underline{x_2}^H) & \cdots & E(\underline{x_2}\underline{I_b}^H) \\ \cdots & \cdots & \cdots & \cdots \\ E(\underline{I_b}\underline{x_1}^H) & E(\underline{I_b}\underline{x_2}^H) & \cdots & \mathbf{I}_{N_b} \end{bmatrix} \begin{bmatrix} \underline{w_1} \\ \underline{w_2} \\ \cdots \\ \underline{w_b} \end{bmatrix}^* = \begin{bmatrix} \underline{c_1} \\ \underline{c_2} \\ \cdots \\ \underline{0}_{N_b} \end{bmatrix} \quad (23)$$

$$H_l(k) := \begin{bmatrix} g_{l,0}(k) & g_{l,2}(k) & \cdots & g_{l,\tilde{N}_g-2}(k) & 0 & 0 & \cdots \\ 0 & g_{l,1}(k-1) & g_{l,3}(k-1) & \cdots & g_{l,\tilde{N}_g-1}(k-1) & 0 & \cdots \\ 0 & g_{l,0}(k-2) & g_{l,2}(k-2) & \cdots & g_{l,\tilde{N}_g-2}(k-2) & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdots & 0 & g_{l,1}(k-\tilde{\Delta}-1) & g_{l,3}(k-\tilde{\Delta}-1) & \cdots & g_{l,\tilde{N}_g-1}(k-\tilde{\Delta}-1) \\ \cdots & \cdots & 0 & g_{l,0}(k-\tilde{\Delta}) & g_{l,2}(k-\tilde{\Delta}) & \cdots & g_{l,\tilde{N}_g-2}(k-\tilde{\Delta}) \end{bmatrix} \quad (20a)$$

where we have used $E\{\underline{I}_b I^*(-\tilde{\Delta})\} = \underline{0}_{N_b}$ and $E(\underline{I}_b \underline{I}_b^H) = \mathbf{I}_{N_b}$. Recall our assumption that the noise sources at each diversity branch are mutually independent and also independent of the input. Next, each $[(\tilde{\Delta}+1)\times 1]$ cross-correlation vector can be shown to be the $(\Delta+1)$th column of H_l, i.e.,

$$\begin{aligned}\underline{c}_l &= E\{H_l \underline{I} + \underline{n}_l) I^*(-\Delta)\} \\ &= H_{l,(:,\Delta)}, \qquad \text{for } l = 1, \cdots, L. \end{aligned} \tag{24}$$

The individual submatrices of (23) can be identified for i, $j = 1, \cdots, L$ as

$$\begin{aligned} E\{\underline{x}_i \underline{x}_j^H\} &= E\{(H_i \underline{I} + \underline{n}_i)(H_j \underline{I} + \underline{n}_j)^H\} \\ &= H_i H_j^H + \sigma_n^2 \delta(i-j)\Phi \end{aligned} \tag{25}$$

where $\delta(\bullet)$ is the Kronecker delta function and Φ, a $[(\tilde{\Delta}+1)\times(\tilde{\Delta}+1)]$ noise autocorrelation matrix, is equal to $\mathbf{I}_{\tilde{\Delta}+1}$. The other $[(\tilde{\Delta}+1)\times N_b]$ submatrices of (23) are

$$\begin{aligned} E\{\underline{x}_i \underline{I}_b^H\} &= E\{(H_i \underline{I} + \underline{n}_i)\underline{I}_b^H\} \\ &= H_i E\{\underline{I}\,\underline{I}_b^H\} \\ &= H_{i,(:,\Delta+1:\Delta+N_b)}, \qquad \text{for } i = 1, \cdots, L \end{aligned} \tag{26}$$

and for $j = 1, \cdots, L$, $E\{\underline{I}_b \underline{x}_j^H\} = E\{\underline{x}_j \underline{I}_b^H\}^H$.

In Appendix E the derivation for the matched filtered solution is continued from the results obtained so far. We term the optimal solution $\underline{w}$ obtained from (23) the "straightforward" solution.

B. A Straightforward Solution

The main purpose of this subsection is to illustrate the shortcomings of the straightforward solution of (23). In doing so, we use $x'_l(k)$, the received signal after the receive SRRC filter as depicted in Fig. 1, to be the input signal to the adaptive filters $\underline{w}'_l(k)$ and $\underline{w}_b(k)$. The benefit of this approach is that the required filter length $\tilde{N}'_f$ of the $T/2$-spaced feedforward filter $\underline{w}'_l(k)$ can be shorter than that of $\underline{w}_l(k)$. The solution is isomorphic to (23) with some minor modifications to (20)–(26), which are given as follows:

- $x'_l(k)$ now replaces $x_l(k)$;
- thus, $\underline{g}_l(k)$ represents the composite pulse $(\underline{f} \otimes \underline{f} \otimes \underline{b}_l(k))_{30-N_g+1:30+N_g}$, truncated with length $\tilde{N}_g$;
- the i, jth element of the noise autocorrelation matrix in (25) is now $\Phi_{(i,j)} = (\underline{f} \otimes \underline{f})_{30+i-j}$; note that from the definition of the SRRC filter $\underline{f}$ in Section II, the composite pulse $\underline{f} \otimes \underline{f}$, approximates the $T/2$-spaced sampled raised cosine pulse with rolloff $\beta = 0.35$, and $(\underline{f} \otimes \underline{f})_{30}$ is the main tap of value 1.0;
- a $[\tilde{N}'_f \times 1]$ subvector $\underline{x}'_l(k) := [x'_l(k-\tilde{\Delta}-\tilde{N}'_f+1) \cdots x'_l(k-\tilde{\Delta})]^T$ replaces the full $[(\tilde{\Delta}+1)\times 1]$ vector $\underline{x}_l(k)$ in (20);
- thus, for the rest of the equations, H_l represents the $[\tilde{N}'_f \times (N_g+\Delta)]$ submatrix $H_{l,(\tilde{\Delta}-\tilde{N}_f+2:\tilde{\Delta}+1,:)}$;
- the cross-correlation vectors $\underline{c}_l$, each individual submatrix of $E\{\underline{x}_i \underline{x}_j^H\}$, and that of $E\{\underline{x}_i \underline{I}_b^H\}$ are truncated to have the correct dimension of $[\tilde{N}_f \times 1]$, $[\tilde{N}'_f \times \tilde{N}'_f]$, and $[\tilde{N}'_f \times N_b]$, respectively.

The decision delay of this solution takes the form $\tilde{\Delta} \le \tilde{N}'_f - 1 + \Delta_g$ (equality with a sufficient number of feedback filter taps), where Δ_g is the main tap location of the channel $\underline{g}(k)$. We note that when the channel variation over $\tilde{N}'_f$ is ignored, the straightforward solution of $\underline{w}'_l$ and $\underline{w}_b(k)$ reduces to the one of [20].

In fact, the $L-[N'_f \times 1]$ $T/2$-spaced feedforward filter $\underline{w}_f$ can be obtained from solving

$$[R_{xx} - R_{xI} R_{xI}^H]\underline{w}_f^* = \underline{c} \tag{27}$$

and the $[N_b \times 1]$ T-spaced feedback filter from

$$\underline{w}_b^* = -R_{xI}^H \underline{w}_f^*. \tag{28}$$

We now illustrate two major drawbacks of the straightforward methods. First, the computational complexity increases exponentially with diversity order. The complexity is order $(LN'_f)^3$, provided we use a Cholesky factorization to solve (27). Second, a more serious problem is that the matrix $[R_{xx} - R_{xI} R_{xI}^H]$ becomes extremely unstable for a diversity order $L > 1$, i.e., a huge condition number (the ratio of the largest and the smallest eigenvalue) occurs.

A large eigenvalue spread occurs when the cross-correlation submatrices of R_{xx} of (23) have large values. Recalling that since the *expectation* does not apply to the channels $\underline{g}_l(k)$ from (18), the cross correlation between any two diversity channels at any given time is not zero valued in general. These nonzero off-diagonal matrices in R_{xx} are the main cause of the large eigenvalue spreads of the matrix R_{xx}. For a simple illustration, consider a two-by-two correlation matrix, i.e.,

$$\begin{aligned} A &:= \begin{bmatrix} E(X_1X_1) & E(X_1X_2) \\ E(X_2X_1) & E(X_2X_2) \end{bmatrix} \\ &= \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \qquad \text{for which a diagonalization} \\ &\qquad \text{reduces to } \begin{bmatrix} a-\xi & 0 \\ 0 & b+\xi \end{bmatrix} \end{aligned} \tag{29}$$

where $c^2 \le a \cdot b$ from the Schwartz inequality and where ξ is denoted as $\xi := \frac{1}{2}(\sqrt{(a-b)^2+4c^2} - |a-b|)$. It is evident from this example that there will be only one significant eigenvalue when the cross correlation c tends to its maximum. Likewise, the condition number of R_{xx} may become very large whenever the cross-correlation submatrices have large values. In our simulation, at high SNR (more than 15 dB), the order of the condition number of matrix R_{xx} for a relatively small N'_f (=4) often reaches up to 10^5 for $L=2$ or 10^8 for $L=4$. Therefore, without a regularization technique to relieve the eigenvalue spreads, the DFE coefficients obtained at high SNR often become unreliable due to magnification of the channel estimation errors.

C. A Matched Filtered DFE Solution

We now discuss the matched filtered diversity-combining DFE, depicted in Fig. 2. The receive signals are matched filtered by $\underline{M}_l(k)$. The matched filtered signals are then combined and sampled at the symbol rate without loss of information. The T-spaced sampled combined signal is then

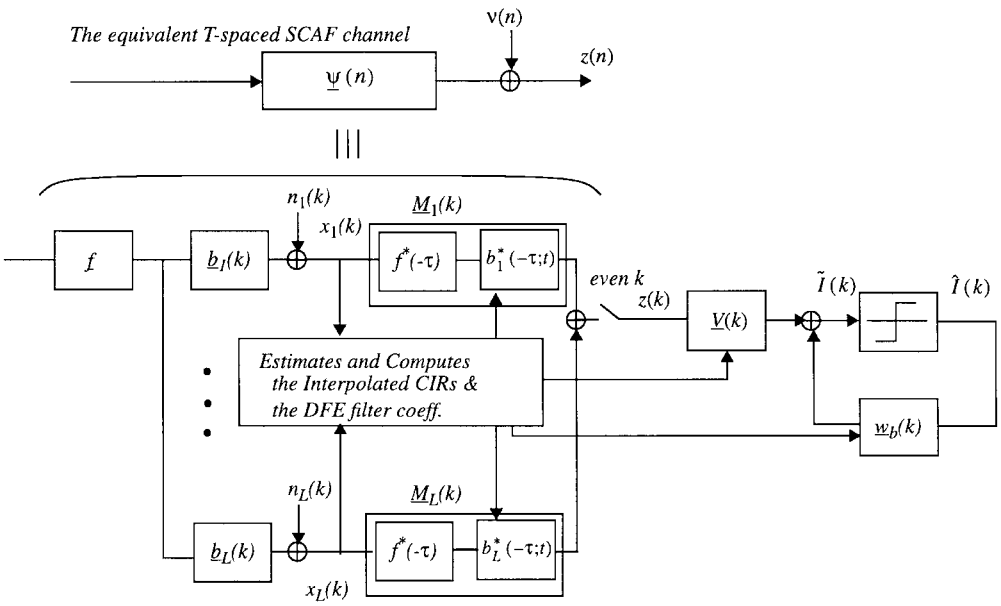


Fig. 2. The matched filtered diversity-combining DFE and the equivalent T-spaced SCAF channel.

fed to the T-spaced feedforward filter $\underline{V}(k)$. The correlation matrix of this structure does not suffer from large eigenvalue spreads.

From the presentation in [2] it can be observed that, although not explicitly stated, the L-diversity-combining DFE problem can be transformed into an equivalent symbol-spaced sampled single-channel DFE problem when a matched filter is used for each diversity branch. This implies that for a finite-length diversity-combining DFE, the dimension of the correlation matrix becomes independent of diversity order L. Thus, there will be no cross-correlation submatrices to spread the eigenvalues. This motivated us to consider the matched filtered DFE solution.

The derivations in [2], however, are performed using a quasistatic channel assumption and focus on obtaining mean square errors of an infinite-order DFE from which the Chernoff upper bounds on bit-error probability can be related. What we need, however, is a solution (with finite filter lengths) that takes into account the rapid channel variation over a decision delay. This can be accomplished relatively easily with the matrix representation of signals and filters as we have developed in Section V-A. In Appendix E the derivation for the matched filtered solution is continued from the straightforward solution (23). In this section we summarize the result.

For the matched filtered diversity-combining DFE, the decision delay is the summation of the matched filter and feedforward filter lengths. It is defined in units of $T/2$-spaced epochs as

$$\tilde{\Delta} = \tilde{N}_g + 2(N_f - 1), \text{ or in units of } T\text{-spaced epochs as } \Delta = N_g + N_f - 1 \quad (30)$$

where N_f is the length of T-spaced feedforward filter $\underline{V}(k)$.

We again assume that $k = 0$ is the current epoch, for the description of the matched and DFE filters. The $[\tilde{N}_g \times 1]$ matched filter can be identified as

$$\underline{M}_l = [g^*_{l,\tilde{N}_g-1}(-1)\ g^*_{l,\tilde{N}_g-2}(-2)\ \cdots\ g^*_{l,0}(-\tilde{N}_g)]^T. \quad (31)$$

Note the decreasing epoch index of the vector elements. Thus, each matched filter at an epoch k needs $\tilde{N}_g$ previous snap-shot channel estimates.

To describe the DFE filters, it is convenient to first define a T-spaced sampled summed channel autocorrelation function (SCAF) at the epoch $k - r$, i.e.,

$$\psi_a(r) := \sum_{l=1}^{L} \psi_{l,a}(r), \qquad \text{for } r = 0, -1, \cdots, -N_f + 1 \quad (32)$$

where the lth diversity channel autocorrelation function $\psi_{l,a}(r)$ is defined as

$$\psi_{l,a}(R) = \sum_{q=|a|}^{\tilde{N}_g-1-|a|} g^*_{l,q-a}(q - \tilde{N}_g - a + 2r) \cdot g_{l,q+a}(q - \tilde{N}_g - a + 2r)$$

for $a = -N_g + 1, \cdots, 0, \cdots, N_g - 1$, and $\psi_{l,a}(r) = 0$ for $a \geq |N_g|$. Note that the phases of the main terms $\psi_{l,0}(\bullet)$ are equalized. Thus, the main term $\psi_0(\bullet)$ is the result of L constructive additions while the rest $\{\psi_a(\bullet)\}_{a\neq 0}$ are not.

Now, the i, jth element of an $[N_f \times N_f]$ correlation matrix $\mathcal{R}$ can be described as

$$\mathcal{R}_{(i,j)} = \sum_{q=-N_g+1}^{N_f-1-i} \psi_q(-i) \cdot \psi^*_{q+i-j}(-j) + \sigma_n^2 \psi_{j-i}(-j), \quad \text{for } i, j = 0, 1, \cdots, N_f - 1. \quad (33)$$

An $[N_f \times 1]$ cross-correlation vector $\underline{P}$ can be identified as, for $i = 0, 1, \cdots, N_f - 1$

$$P_i = \psi_{N_f-1-i}(-i). \quad (34)$$

Finally, an $[N_b \times N_f]$ matrix $\mathcal{B}$ is, for $i = 0, 1, \cdots, N_b - 1$, and $j = 0, 1, \cdots, N_f - 1$

$$\mathcal{B}_{(i,j)} = \psi^*_{N_f+i-j}(-j). \tag{35}$$

Then, the T-spaced feedforward filter $\underline{V}$ can be obtained from solving $\mathcal{R}\underline{V}^* = \underline{P}$, and the feedback filter $\underline{w}_b$ from $\underline{w}_b^* = -\mathcal{B}\underline{V}^*$. We refer to this as the matched filtered non-Toeplitz DFE (NT-DFE). Note that this solution utilizes all of the channel state information during the last $\tilde{\Delta}(T/2)$ period. When time invariance of the channel over the decision delay is assumed, the channel matrix $H_l(k)$ of (20) becomes block Toeplitz, and all of the epoch terms inside the parenthesis of (33)–(35) are ignored. We refer to this as a matched filtered Toeplitz DFE (T-DFE).

From inspection of (33)–(35), it is worthwhile to note that the L-diversity-combining DFE problem has been transformed into an equivalent single-channel T-spaced DFE problem, as depicted in Fig. 2. Denoting n as the symbol-spaced epoch index, the equivalent T-spaced channel is the $[(2N_g - 1) \times 1]$ vector $\underline{\psi}(n)$ whose elements are the $\{\psi_a(n)\}_{a<|N_g|}$ of (32), i.e., $\psi_{-N_g+1}(n)$ is the first element of the vector $\underline{\psi}(n)$. The equivalent T-spaced noise is denoted by $v(n)$ and the noise autocorrelation function is $\sigma_n^2 \cdot \psi_a(n)$. Then, the T-spaced sampled $z(n)$ can be described by $z(n) = \underline{\psi}(n)^T \underline{I}(n) + v(n)$, where $\underline{I}(n) := [I(n)\, I(n-1) \cdots I(n-2N_g+1)]^T$. Now, for an $[N_f \times 1]$ input vector $\underline{z}(n) = [z(n) \cdots z(n-N_f+1)]^T$, the predecision value $\tilde{I}(n)$ analogous to (19) can be defined as $\tilde{I}(n) = \underline{z}^T(n)\underline{V}(n) + \underline{I}_b^T(n)\underline{w}_b(n)$. Then, following the standard procedure analogous to (20)–(26), the matched filtered solution of (33)–(35) can be reproduced.

In summary, we have derived the matched filtered diversity-combining DFE, which provides a lower complexity and more stable solution than the straightforward approach does. The dimension of the correlation matrix $\mathcal{R}$ $[N_f \times N_f]$ stays the same for any diversity order. Note that there are no cross-correlation submatrices to spread the eigenvalues; instead, the eigenvalue spread of the correlation matrix is now fully determined by the SCAF vector $\underline{\psi}(n)$. Also note that SCAF values $\{\psi_a(n)\}_{a\neq 0}$ in (32) correspond to ISI terms relative to the phase-equalized main term $\psi_0(n)$, and that their energies relative to the main term decrease for increasing L. Thus, the correlation matrix tends to be more stable for increasing L. It is well known that explicit diversity reduces ISI. In the matched filtered diversity-combining DFE, it also helps to stabilize the DFE computation algorithm. Thus, the DFE coefficients obtained from the channel estimates become less susceptible to channel estimation noise enhancement.

Comparing NT-DFE and T-DFE of the matched filtered diversity-combining DFE, the NT-DFE is optimal because it uses all of the channel state information during the decision delay which is provided by channel interpolation. The T-DFE uses only partial information and is thus suboptimal, but it has lower complexity than the NT-DFE. The NT-DFE provides a performance advantage over the T-DFE only when the channel is in fast fading and tracked with a reasonable accuracy. In Section VI the above comparisons will be made through computer simulations.

VI. System Simulations and Discussion of Results

The performance of the complete system was investigated through complex baseband computer simulations. A Monte Carlo method with 2000–50 000 independent trials was used. To evaluate the adaptation on continuously transmitted frames, each trial consisted of 5–16 frames, where a frame is a block of B symbols including the N_t training symbols. During a trial, each fading coefficient of the L-diversity channels was continuously varied at a given fading rate according to (3). From trial to trial, an independent set of channels was generated by selecting a new set of random starting phases ST_{li} in (3) at the start of each trial. The $T/2$-spaced complex-valued additive noise samples were independently generated. The modulation schemes used were quadrature phase-shift keying (QPSK) and differentially encoded QPSK (DQPSK) to compare to the existing literature. The SNR in this section implies the long-term average SNR of the three-path fading channel.

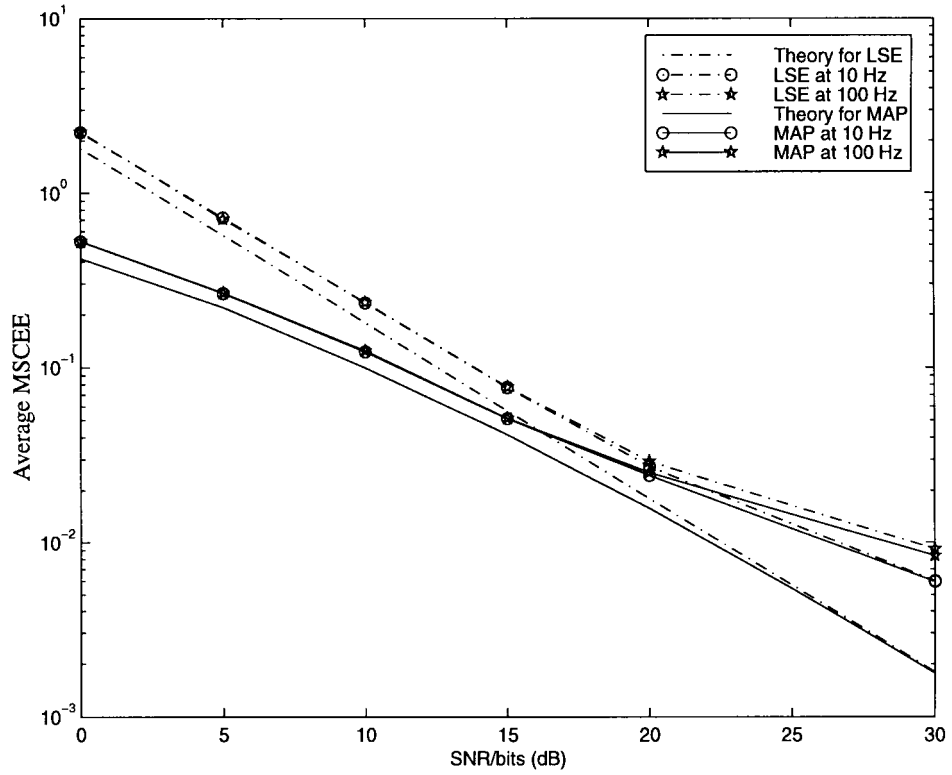
A. Mean Square Error Performance of the Channel Estimators

In Fig. 3 the performance of two channel estimators, LSE and MAP, in terms of the MSCEE are assessed both in theory and in simulation. For the simulation results, the channel estimate, which is obtained at the end of the training observation window, is compared with $\underline{b}(k)$, where k is the median epoch of the window.

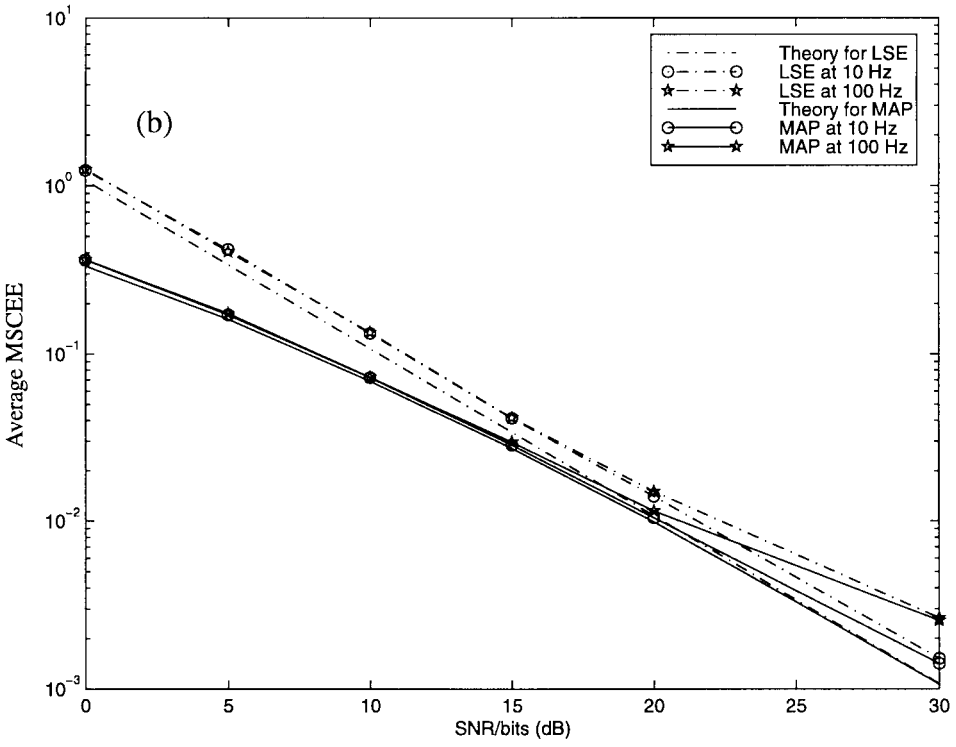
Fig. 3(a) shows the MSCEE for (the training length N_t, the truncation length N_c) $= (7, 4)$ and Fig. 3(b) for (11, 6). We first note the effect of truncation at high SNR. Recall the channel estimation of (5) where we truncate the length of the overall channel to be N_c symbol intervals. The slow fading curves for $(N_t, N_c) = (11, 6)$ stay very close to theory out up to 30 dB, whereas the slow fading curves for $(N_t, N_c) = (7, 4)$ deviate from theory at high SNR due to the truncation errors. This suggests that truncation at $N_c = 6$ is sufficient for the purpose of channel estimation. Second, in both figures we observe that the fast-fading curves show some deviations from the slow-fading curves at 30 dB. These degradations are due to the snap-shot assumption that during the observation periods $m(= N_t - N_c + 1)$, the channel is fixed. Thus, we do observe some MSCEE degradations from the theory due to the truncation error and snap-shot assumption; however, they are very much suppressed and do not form an irreducible MSCEE floor. Finally, we note the marked advantage of MAP estimation over LSE at low SNR's. For the BER simulations in the following sections we use (N_t, N_c) =(11,6) and a frame length $B = 80$.

B. BER Performance in Slow Fading

In Fig. 4 the BER performances of two receivers are compared with theoretical matched filter bounds (MFB). The MFB is the lowest attainable bound since it is obtained assuming the transmitted pulses are far enough apart so that no ISI occurs. A similar approach reported in [19] has been used to obtain the MFB's for our three-path channel model. *Flat* fading indicates the matched filter bound for the single-path Rayleigh-fading channel.



(a)



(b)

Fig. 3. Theoretical and simulation mean square channel estimation errors (MSCEE's). (a) For $(N_t,\ N_c) = (7,\ 4)$. (b) For $(N_t,\ N_c) = (11,\ 6)$.

In Fig. 4 "LSE and interpolation NT-DFE" refers to the use of least-squares channel estimation, channel tracking by interpolation, and NT-DFE. "RLS channel tracking T-DFE" refers to the use of a recursive least-squares algorithm to track the time-varying channel (i.e., without channel interpolation) and the T-DFE for symbol detection. The T-DFE is used since the channel states during the decision delay are not available with the recursive adaptation. With regard to the use of the RLS algorithm, the channel-estimate approach, instead of a conventional direct adaptation on the DFE coefficients

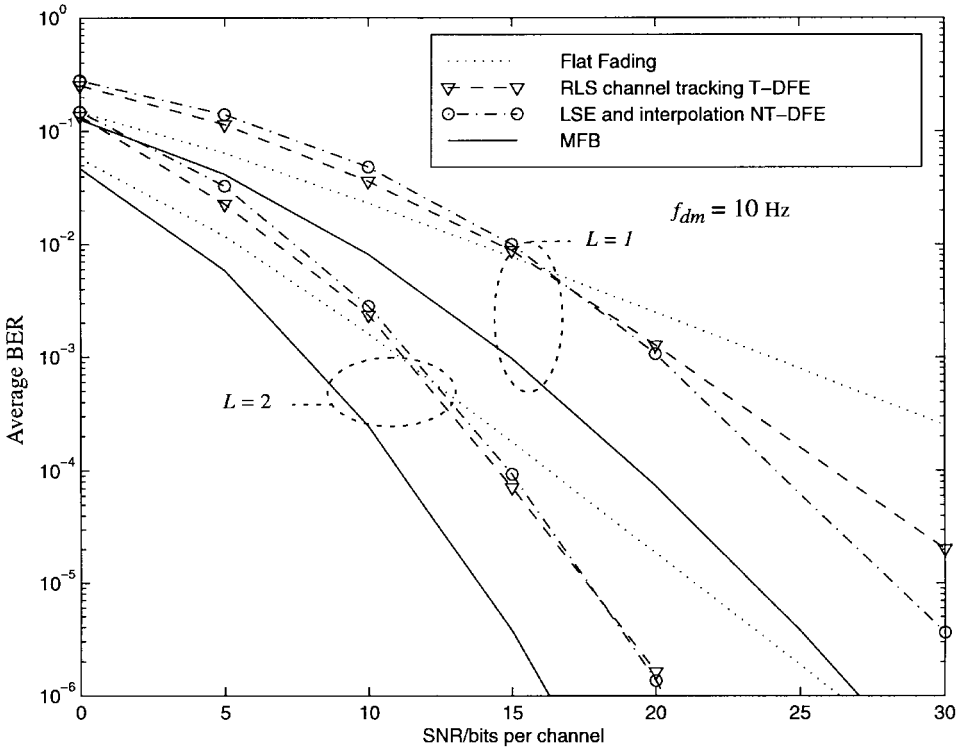


Fig. 4. QPSK average BER as a function of SNR in slow fading—RLS channel tracking T-DFE, LSE NT-DFE, and theoretical matched filter bounds are compared.

without channel estimation, is selected since the channel-estimate based DFE (without diversity) has been shown to be more effective than the direct DFE adaptation [5], [26]. Specifically, we use the exponential windowing RLS algorithm from [26]. To be a fair comparison, the same known training blocks are inserted in the data stream. Thus, during the training segment, the RLS algorithm and DFE filters are refreshed at the same rate. Furthermore, the exponential weighting factor w of the RLS algorithm is optimized at various SNR's, fade rates, and channel lengths. For this, the following equation is adopted from [26]:

$$\mathrm{SNR} = \frac{(N_c+1)}{(2f_{dm}\pi T)^2}\,\frac{(1-\omega)^3}{(1+\omega)^2}. \tag{36}$$

The filter orders used in the simulation are $((\tilde{N}_g, N_f, N_b) = (20,5,5)$. The channel is slow fading with $f_{dm} = 10$ Hz ($f_{dm}T = 0.000\,42$). We note that the slopes of the BER curves for both methods (e.g., about 10^{-2} per 10 dB SNR for $L = 1$) are close to those of their MFB's and steeper than those (e.g., 10^{-1} per 10 dB for $L = 1$) of the theoretical flat-fading channel. This indicates that both receivers take advantage of the implicit diversity gain, which is inherent in the frequency-selective channel. The RLS T-DFE and LSE NT-DFE show comparable performance in slow fading.

C. BER Performance in Fast Fading

In Fig. 5 the BER performance of the RLS T-DFE with DQPSK signaling is evaluated for $f_{dm} = \{10, 50, 100\}$. Since the T-DFE ignores the channel variation over the receive filter lengths (i.e., the matched and feedforward filter), longer filters might become counterproductive in fast fading (this behavior is also observed in [10] without diversity). Thus, at $f_{dm} = 100$ Hz we use shorter filters of $(\tilde{N}_g, N_f, N_b) = (12,4,4)$, which have been determined to be optimal from our simulations. We observe that at $f_{dm} = 100$ Hz, the irreducible BER's are too high (0.1 for $L = 1$ and 0.01 for $L = 2$) to be of any practical use. Therefore, we confirm that RLS actually fails to track the three-tap fast Rayleigh-fading channel.

In Fig. 6 the BER performances of LSE and MAP NT-DFE receivers with DQPSK signaling are evaluated at $f_{dm} = 100$ Hz. "MAP NT-DFE" refers to the use of maximum *a posteriori* channel estimation, channel tracking by interpolation, and NT-DFE. We note that both receivers show a superior and robust BER performance against fast fading. LSE and MAP NT-DFE curves are not even flat out up to 30 dB. We also show results for the NT-DFE, which is denoted by "ideal CIR NT-DFE" (the use of perfect channel at all epochs). The NT-DFE exhibits no sign of irreducible error floors, in contrast to the T-DFE, which will display irreducible error floors even with the ideal CIR supplied. An example can be found in [10], where a DFE receiver, using T-DFE without diversity, shows relatively high irreducible BER floors even using perfect channel estimates at all epochs.

We observe that LSE and MAP NT-DFE curves show less than 1 dB difference below a BER of 10^{-3}. This suggests that the use of LSE is a reasonable design choice, at least for an uncoded system. In addition, we note that the throughput rate at this BER is $(B - N_t)/B = (80 - 11)/80 = 0.8625$.

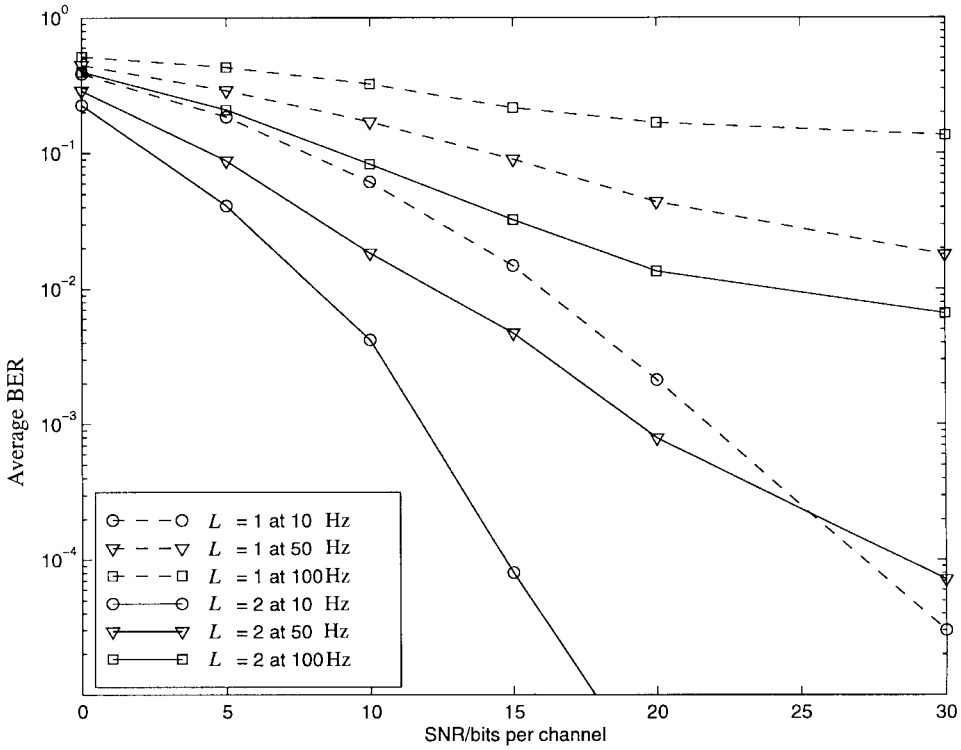


Fig. 5. DQPSK average BER as a function of SNR for RLS channel tracking Toeplitz DFE receiver.

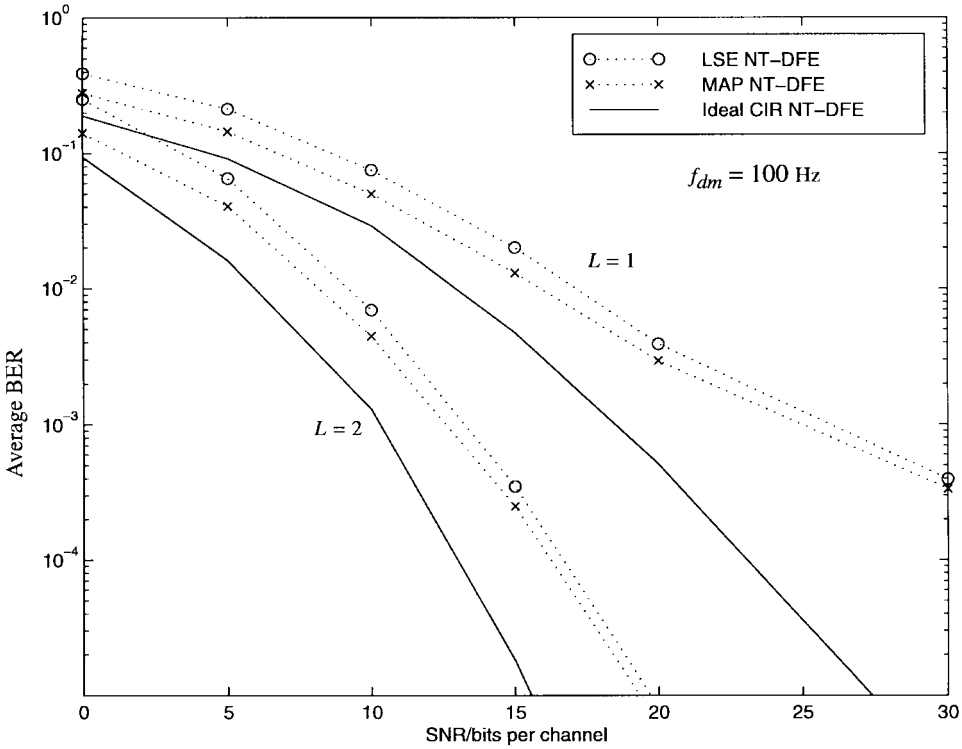


Fig. 6. DQPSK average BER as a function of SNR—LSE and MAP NT-DFE, with channel tracking by interpolation, are compared. Ideal CIR NT-DFE implies NT-DFE with perfect channel information.

D. The Sources of the BER Degradation in Fast Fading

In Fig. 7 the NT-DFE and T-DFE receivers are compared in fast fading. For each receiver, we also compare three different modes of obtaining channel impulse responses. They are: 1) interpolation on the maximum *a posteriori* channel estimates, 2) interpolation on perfect channel estimates, and 3) use of the correct channel values at all epochs (ideal channel reference). Comparison of these curves should identify the main cause

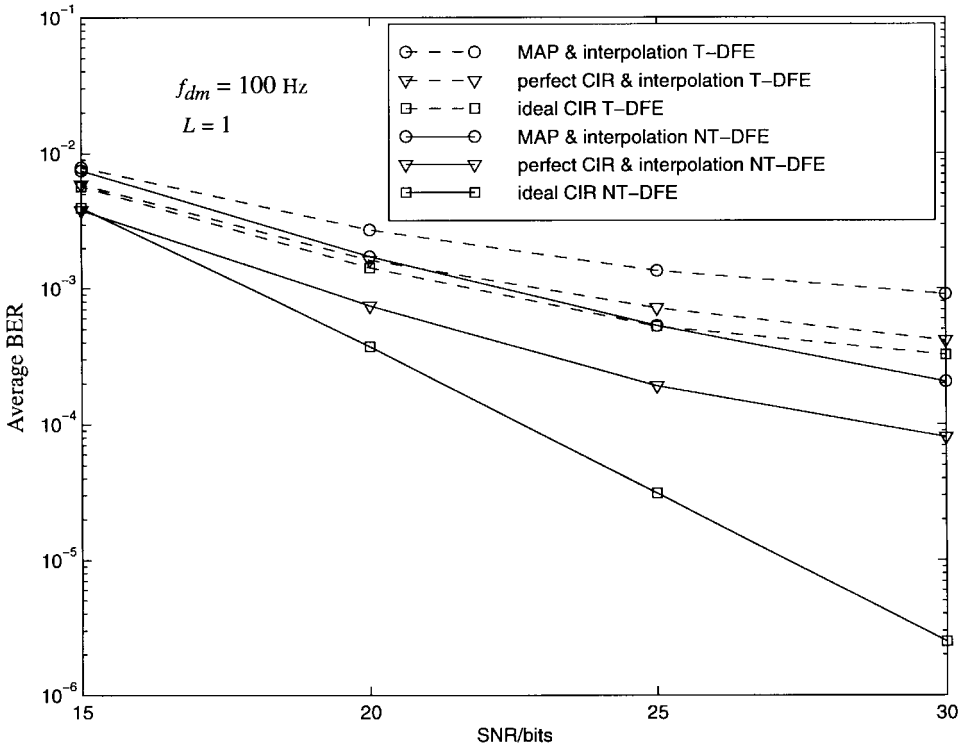


Fig. 7. QPSK average BER simulations to determine the source of error floors.

of BER degradations in fast fading. The optimal filter orders for the T-DFE are again $(\tilde{N}_g, N_f, N_b) = (12, 4, 4)$. The filter orders used for the NT-DFE are (20, 5, 5).

First, note that the T-DFE curves show higher BER floors. Even the ideal CIR T-DFE produces a higher BER floor than the MAP NT-DFE does. This illustrates the detrimental consequence of ignoring the channel variation during the decision delay of a T-DFE receiver. Second, by comparing the NT-DFE curves, it can be seen that the BER degradations are mainly due to the interpolation errors. The interpolator performs poorly in the middle of the data segment, and the decision errors occur predominantly in the middle of the data segment. This problem persists even at $B = 40$, for which the BER at 30 dB is about 3×10^{-5} (not shown in the figures).

E. The Suboptimal T-DFE and the DFE Update Periods

In Fig. 8 we investigate the impact on the BER of increasing the DFE update periods μ, where μ is the number of symbol periods between any two updates of DFE filters. Again, the BER performances of the T-DFE and NT-DFE are compared at $f_{dm} = 100$ Hz. We use the optimal filter orders $(\tilde{N}_g, N_f, N_b) = (12, 4, 4)$ for the Toeplitz case. For the non-Toeplitz case, (20, 5, 5) are used for $\mu = 1$, while shorter filter orders (16, 4, 4) are used for other values of μ. The MAP channel estimator is used for both. First, note that the performance difference of the two deepens for a higher diversity order and for a higher SNR. Second, the NT-DFE receiver maintains its superiority to T-DFE only for $\mu = 1$, as the BER gain quickly disappears for $\mu > 1$. This suggests that if μ larger than one is selected for a lower computational complexity, the use of a T-DFE receiver is suitable.

F. Computational Complexity

In Table I we summarize the number of complex multiplications and divisions required for the RLS T-DFE, LSE T-DFE, and LSE NT-DFE. The first and second rows indicate the required number of operations for the channel tracking techniques. For channel tracking by interpolation, we assume that the $T/2$-sampled sinc function is stored. Then, the interpolated channel $\underline{b}$ can be obtained from $N_R Q$ complex multiplications and the convolution of $\underline{b}$ and $\underline{f}$ requires another $N_R \tilde{N}_g$. The matched filter coefficient vector $\underline{M_l}$ can be obtained without any computation since it is a pure mapping from the interpolated overall channel. The third row indicates the number of operations required to form the summed channel autocorrelation function, the correlation matrix, the cross-correlation vector, and the feedback filter matrix. The fourth row gives the required number of computations to solve for the feedforward filter with length N_f provided the Cholesky factorization [24] is used. The shaded region implies that the numbers of multiplication can be divided by μ, where μ is the DFE filter update period in units of symbol period.

The last row gives sample calculations for $L = 1, 2$, $\mu = 1$ with a typical set of filter lengths and channel estimation parameters $(\tilde{N}_g, N_f, N_b, B, N_t, N_c) = (12, 6, 4, 4, 80, 11, 6)$. This gives $N_g = \tilde{N}_g/2$ and $m = N_t - N_c + 1 = 6$. For T-DFE receivers, the required number of operations for a DFE update period of $\mu = 5$ is also calculated and presented inside parentheses. The sample calculation, together with the BER results in Fig. 8, shows the feasibility of the T-DFE receiver in a practical application. That is, at a per-symbol complexity of less than 100 complex multiplications and divisions, we

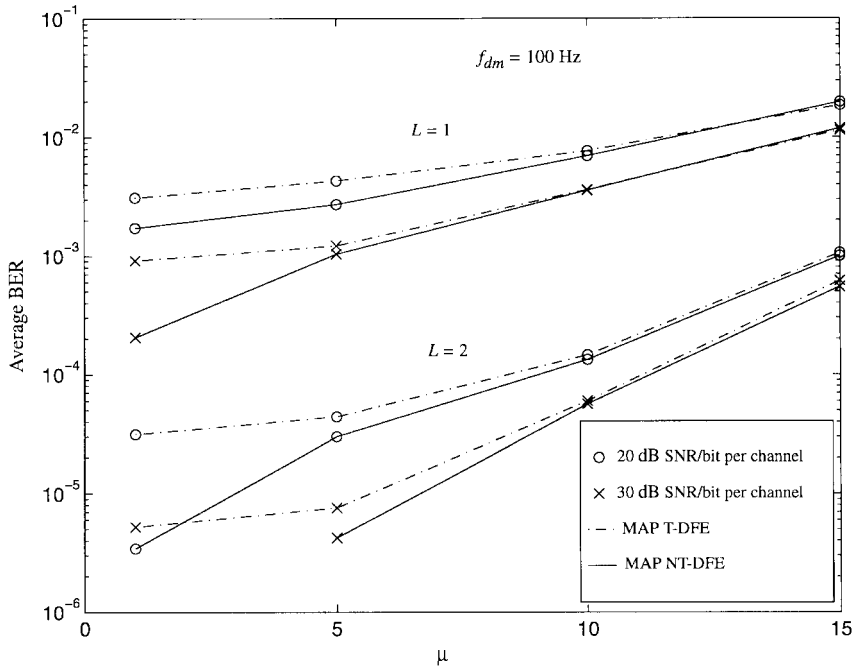


Fig. 8. QPSK average BER as a function of the DFE filter update period (μ).

TABLE I
THE NUMBER OF COMPLEX MULTIPLICATIONS AND DIVISIONS REQUIRED

	RLS channel tracking T-DFE	LSE T-DFE	LSE NT-DFE
Channel estimation	$9L\tilde{N}_g$ + 6	$\frac{(LmN_R)}{(B-Nt)}$	$\frac{(LmN_R)}{(B-Nt)}$
Channel tracking by the interpolation	N/A	$2LN_R(Q+\tilde{N}_g)$	$2LN_R(Q+\tilde{N}_g)$
Forming the SCAF, the corr. matrix, the cross-corr. vector, and the feedback matrix	$\frac{L}{2}(\tilde{N}_g^2+\tilde{N}_g)$ + $N_f(N_g+N_f-1)$	$\frac{L}{2}(\tilde{N}_g^2+\tilde{N}_g)$ + $N_f(N_g+N_f-1)$	$L(\tilde{N}_g^2+\tilde{N}_g-1)$ + $N_f^2(N_g+N_f)$ - $\frac{1}{6}(4N_f^3+9N_f^2-N_f)$
Feedforward filters	$\frac{1}{6}(N_f^3+9N_f^2-4N_f)$ + N_f	$\frac{1}{6}(N_f^3+9N_f^2-4N_f)$ + N_f	$\frac{1}{6}(N_f^3+9N_f^2-4N_f)$ + N_f
Feedback filter	N_fN_b	N_fN_b	N_fN_b
L = 1, 2 and μ = 1 (μ = 5)	279 (147) 465 (271)	263 (53) 436 (88)	464 714

achieve BER of about 10^{-3} for $L=1$ or 10^{-5} for $L=2$, at the reasonable SNR of 20–30 dB.

VII. CONCLUDING REMARKS

We have presented robust channel estimation methods which require little training overhead over the fast Rayleigh-fading dispersive channel. It has been shown through simulations that channel tracking by interpolation along with our proposed channel estimation method is significantly better than the RLS channel tracking method and previously published feedforward channel estimation methods in terms of both the throughput and the BER performance.

For the block adaptive diversity-combining DFE scheme, we have proposed the matched filtered approach because of its stable performance in the presence of channel estimation errors. The matched filtered DFE simplifies the L-diversity combining decision feedback equalizer into an equivalent single-channel DFE problem. This provides a reduced computational burden in tracking the optimum coefficients of the receiver and leads to a well-conditioned correlation matrix.

We have derived a matched filtered diversity-combining NT-DFE which takes into account the channel variation over the decision delay. This NT-DFE can obtain the full benefit of the channel interpolation and thus provides a benchmark for performance. While optimal, the NT-DFE incurs relatively high computational complexity, and thus for a suboptimal but low-complexity solution, we propose the use of the T-DFE, which still provides better performance than the RLS algorithm.

For higher SNR's, however, the BER curves of the NT-DFE still tend to flatten out, due to interpolation errors. Design of an interpolation strategy to overcome this problem would be an interesting research exercise.

Appendix A
The SRRC Filter Matrix F

An example with $N_c = 6$ and $N_R = 3$ is sufficient to describe the procedure of obtaining the SRRC filter matrices in Section III. $\underline{f}$ denotes the 31-tap $T/2$-spaced SRRC filter, i.e., $\underline{f} := [f0\ f1\ \cdots\ f_{30}]^T$, where f_{15} is the main tap of the SRRC vector. Then, a [12 × 1] truncated overall CIR $\underline{\tilde{g}}$ can be described by a matrix and a vector multiplication as $\underline{\tilde{g}} = \mathcal{F}\underline{b}$, where $\mathcal{F}$ is a [12 × 3] Toeplitz matrix which can be described by the first row and the first column.

The first row is $[f_{10}\ f_9\ f_8]^T$ and the first column is $f_{10}\ f_{11}\ \cdots\ f_{21}]^T$. Thus, $\mathcal{F}$ is

$$\mathcal{F} := \begin{bmatrix} f_{10} & f_9 & f_8 \\ f_{11} & f_{10} & f_9 \\ \cdots & \cdots & \cdots \\ f_{21} & f_{20} & f_{19} \end{bmatrix}. \tag{A.1}$$

Finally, a [6 × 3] matrix F in (6) for even $\underline{r}$ is obtained from taking the six even rows of $\mathcal{F}$. Similarly, taking the six odd rows constitutes F for odd $\underline{r}$.

Appendix B
The MMSE Estimator of $\underline{b}$

The MAP estimator coincides with the MMSE estimator of $\underline{b}$. Define the linear estimator that achieves the MMSE as $\hat{\underline{b}} = B^+\underline{r}$, where B^+ is the linear estimation operator. Then, $\hat{\underline{b}}$ will satisfy the following equality:

$$\hat{\underline{b}} = \arg\min_{\hat{\underline{b}}}(E\{(\hat{\underline{b}} - \underline{b})^H(\hat{\underline{b}} - \underline{b})\}). \tag{B.1}$$

The orthogonality relation $E\{(\hat{\underline{b}} - \underline{b})\underline{r}^H\} = \underline{0}$ leads to $B^+E\{\underline{r}\,\underline{r}^H\} = E\{\underline{b}\,\underline{r}^H\}$ and finally

$$B^+ = E\{\underline{b}\,\underline{r}^H\}E\{\underline{r}\,\underline{r}^H\}^{-1} \tag{B.2}$$

where $E\{\underline{r}\,\underline{r}^H\} = E\{(XF\underline{b} + \underline{n})(XF\underline{b} + \underline{n})^H\} = (XFR_bF^HX^H + R_n)$ and $E\{\underline{b}\,\underline{r}^H\} = E\{\underline{b}(XF\underline{b} + \underline{n})^H\} = R_bF^HX^H$. We note that B^+ is identical to the MAP operator in (12).

Appendix C
Cramer–Rao Lower Bound

The maximum-likelihood estimator derived in Section III-B achieves the Cramer–Rao lower bound for unbiased estimators. The Cramer–Rao lower bound for unbiased estimators is

$$\mathrm{cov}(\hat{\underline{b}}) \geq M^{-1} \tag{C.1}$$

where M is the Fisher information matrix defined as

$$M := E\left[\left(\frac{\partial}{\partial \underline{b}}\ln(p(\underline{r}|\underline{b}))\right)\left(\frac{\partial}{\partial \underline{b}}\ln(p(\underline{r}|\underline{b}))\right)^H\right]. \tag{C.2}$$

We will prove the equality, i.e., $\mathrm{cov}(\hat{\underline{b}}_{\mathrm{ML}}) = M^{-1}$.

Note

$$\begin{aligned}\frac{\partial}{\partial \underline{b}}&\ln(p(\underline{r}|\underline{b})) \\ &= -\tfrac{1}{2}(-2(XF)^HR_n^{-1}\underline{r} + 2((XF)^HR_n^{-1}(XF))\underline{b}) \\ &= ((XF)^HR_n^{-1}(XF))(\hat{\underline{b}}_{\mathrm{ML}} - \underline{b})\end{aligned} \tag{C.3}$$

then

$$\begin{aligned}M &= [(XF)^HR_n^{-1}(XF)]\,\mathrm{cov}(\hat{\underline{b}}_{\mathrm{ML}})[(XF)^HR_n^{-1}(XF)] \\ &= [(XF)^HR_n^{-1}(XF)].\end{aligned} \tag{C.4}$$

Since we already know $\mathrm{cov}(\hat{\underline{b}}_{\mathrm{ML}}) = [(XF)^HR_n^{-1}(XF)]^{-1}$ from (14), the equality in (C.1) is proved.

Appendix D
Optimum Training Sequences

LSE applied to an $[m \times 1]$ observation vector $\underline{r} = X\underline{g} + \underline{n}$ produces the least-square estimate of $\underline{g}$, i.e.,

$$\hat{\underline{g}} = X^\dagger\underline{r} \tag{D.1}$$

where $X^\dagger = (X^HX)^{-1}X^H$ and $\underline{g}$ is an $[N_c \times 1]$ vector. Then, the covariance matrix of the estimator is

$$\Theta := E(\hat{\underline{g}} - E(\hat{\underline{g}}))(\hat{\underline{g}} - E(\hat{\underline{g}}))^H = \sigma_n^2(X^HX)^{-1}. \tag{D.2}$$

Therefore, the optimum sequences (stored in matrix X) are the sequences that satisfy

$$X = \arg\min_X\{\mathrm{tr}(\Theta)\}. \tag{D.3}$$

Given N_c and m, the optimum binary training sequence (OBTS) satisfies (D.3) and makes the matrix X^HX as close as possible to diagonal.

From (15), the covariance matrix of $\hat{\underline{b}}_{\mathrm{LSE}}$ is

$$\begin{aligned}\Theta_{\mathrm{LSE}} &:= E\{(\hat{\underline{b}}_{\mathrm{LSE}} - E(\hat{\underline{b}}_{\mathrm{LSE}}))(\hat{\underline{b}}_{\mathrm{LSE}} - E(\hat{\underline{b}}_{\mathrm{LSE}}))^H\} \\ &= \sigma_n^2(F^HX^HXF)^{-1}.\end{aligned} \tag{D.4}$$

Now, the optimal training sequence is the sequence that makes (F^HX^HXF) to be as close as possible to diagonal. This optimization problem can be approached by representing the convolution operation XF in a matrix and vector multiplication form as $\mathsf{F}\underline{x} = \underline{c}$, where $\underline{x} = [I_0\ I_1\ I_2\ \cdots\ I_{N_t-1}]$ is an $[N_t \times 1]$ vector, F is now an $[N_c \times N_t]$ matrix, and $\underline{c}$ is the

$[N_c \times 1]$ OBTS vector. Then, the new optimal sequence $\underline{x}$ that minimizes the trace$\{\Theta_{\mathrm{LSE}}\}$ can be obtained as

$$\underline{x} = \mathsf{F}^{\dagger}\underline{c} \tag{D.5}$$

where $\mathsf{F}^{\dagger}$ is the generalized inverse of F [34]. The vector $\underline{x}$ should be scaled so that the energy of the scaled vector $\underline{x}$ is N_t.

Appendix E
Derivation of the Matched Filtered NT-DFE Solution

In this section we continue the derivation of the NT-DFE solution from the straightforward solution (23). The L simultaneous matrix equations of (23) can now be rewritten as

$$\begin{aligned}(H_1H_1^H+\sigma_n^2\mathbf{I})\underline{w}_1^*+H_1H_2^H\underline{w}_2^*+\cdots+H_1E\{\underline{I}\,\underline{I}_b^H\}\underline{w}_b^* &= \underline{c}_1, \qquad \text{for } l=1\\ H_2H_1^H\underline{w}_1^*+(H_2H_2^H+\sigma_n^2\mathbf{I})\underline{w}_2^*+\cdots+H_2E\{\underline{I}\,\underline{I}_b^H\}\underline{w}_b^* &= \underline{c}_2, \qquad \text{for } l=2\end{aligned} \tag{E.1}$$

and similarly for up to $l=L$, and the last matrix equation for the feedback part is

$$E\{\underline{I}_b\underline{I}^H\}H_1^H\underline{w}_1^*+E\{\underline{I}_b\underline{I}^H\}H_2^H\underline{w}_2^*+\cdots+\underline{w}_b^*=\underline{0}_{N_b}. \tag{E.2}$$

Substituting (E.2), arranged in terms of $\underline{w}_b^*$, into (E.1), each of the L matrix equations of (E.1) becomes

$$H_l\sum_{r=1}^{L}H_r^H\underline{w}_r^*+\sigma_n^2\underline{w}_l^*-H_lE\{\underline{I}\,\underline{I}_b^H\}E\{\underline{I}_b\underline{I}^H\}\sum_{r=1}^{L}H_r^H\underline{w}_r^* = \underline{c}_l. \tag{E.3}$$

Now, by defining $\dot{M}_l := H_l - H_lE\{\underline{I}\,\underline{I}_b^H\}E\{\underline{I}_b\underline{I}^H\}$ and $\underline{\tilde{V}}^* := \sum_{r=1}^{L}H_r^H\underline{w}_r^*$, and rearranging with respect to $\underline{w}_l^*$, (E.3) produces

$$\underline{w}_l^* = \frac{1}{\sigma_n^2}\left(-\dot{M}_l\cdot\underline{\tilde{V}}^*+\underline{c}_l\right)=M_l\underline{V}^*, \qquad \text{for each } l \tag{E.4}$$

where the $[(\tilde{\Delta}+1)\times(\Delta+1)]$ matrix M_l is equal to the submatrix of H_l

$$M_l = H_{l,(:,\,0:\Delta)} \tag{E.5}$$

and the elements of the $[(\Delta+1)\times 1]$ vector $\underline{V}$ are defined as

$$V_i := \begin{cases}(1-\tilde{V}_i)/\sigma_n^2, & i=\Delta\\ (-\tilde{V}_i)/\sigma_n^2, & i=0,\cdots,\Delta-1.\end{cases} \tag{E.6}$$

We note from (E.4) that each feedforward filter $\underline{w}_l$ can be decomposed into a matched filter at each diversity branch and a T-spaced feedforward filter which is common to all the diversity branches.

Next, premultiplying M_l^H and substituting $\underline{w}_l^*$ of (E.4) into the corresponding lth equation of (E.3), and then summing over all the L equations, produces

$$\left[\sum_{l=1}^{L}M_l^HM_l\cdot\sum_{r=1}^{L}M_r^HM_r+\sigma_n^2\sum_{l=1}^{L}M_l^HM_l\right]\underline{V}^*=\sum_{l=1}^{L}M_l^H\underline{c}_l. \tag{E.7}$$

Now define a $[(\Delta+1)\times(\Delta+1)]$ matrix $\psi_l := M_l^HM_l$, and note $\psi_l=M_l^HM_l=\psi_l^H$, and similarly for $\psi := \sum_{l=1}^{L}\psi_l=\psi^H$. Also note that $\sum_{l=1}^{L}M_l^H\underline{c}_l$ of (E.7) is just the $(\Delta+1)$th column of ψ, i.e., $\sum_{l=1}^{L}M_l^H\underline{c}_l=\psi_{(:,\,\Delta)}$. Finally, substituting (E.4) into (E.2), we have the feedback coefficients

$$\begin{aligned}\underline{w}_b^* &= -\left(\sum_{l=1}^{L}E\{\underline{I}_b\underline{I}^H\}H_l^HM_l\right)\underline{V}^*\\ &= \left(\sum_{l=1}^{L}(H_{l,\,(\Delta+1:\,\Delta+N_g,\,:)})^HM_l\right)\underline{V}^*.\end{aligned} \tag{E.8}$$

From inspection of (E.4) and (E.7), all of the necessary information on the matched filtered diversity-combining structure can be obtained. Specifically, the matched filter coefficients of (31) can be determined from M_l^H, the combined signal of the diversity matched filter outputs is the input signal to the T-spaced feedforward filter $\underline{V}$, and the decision delay should be $\tilde{\Delta}=\tilde{N}_g+2(N_f-1)$ for a T-spaced feedforward filter length with N_f.

In particular, considering $\underline{V}_{\Delta-N_f+1:\,\Delta}$ for a feedforward filter with length N_f, (E.7) can be reduced to

$$\mathcal{R}\underline{V}^*_{\Delta-N_f+1:\,\Delta}=\psi_{(\Delta-N_f+1:\,\Delta,\,\Delta)} \tag{E.9}$$

where $\mathcal{R} := (\psi\psi^H+\sigma_n^2\psi)_{(\Delta-N_f+1:\,\Delta,\,\Delta-N_f+1:\,\Delta)}$. The $[N_b\times 1]$ feedback filter can be obtained from

$$\underline{w}_b^*=-\mathcal{B}\underline{V}^*_{\Delta-N_f+1:\,\Delta} \tag{E.10}$$

where the $[N_b\times N_f]$ matrix $\mathcal{B} := (\sum_{l=1}^{L}(H_{l,\,(\Delta+1:\,\Delta+N_g,\,:)})^HM_l)_{(0:\,N_b-1,\,\Delta-N_f+1:\,\Delta)}$.

References

[1] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Commun. Mag.*, vol. 33, pp. 42–49, Jan. 1995.

[2] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio—Part I: Theoretical considerations and Part II: Numerical results," *IEEE Trans. Commun.*, vol. 40, pp. 885–907, May 1992.

[3] J. C. Chuang, "The effects of time delay spread on portable radio communications channels with digital modulation," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 879–889, June 1987.

[4] S. N. Crozier, "Short-block, data detection techniques employing channel estimation for fading, time-dispersive channels," Ph.D. dissertation, Dept. Syst. Comput. Eng., Carleton Univ., Ottawa, Ont., Canada, May 1990.

[5] G. W. Davidson, D. D. Falconer, and A. U. H. Sheikh, "An investigation of block-adaptive decision feedback equalization for frequency selective fading channels," *Can. J. Elect. Comp. Eng.*, vol. 13, nos. 3/4, pp. 106–116, 1988.

[6] G. Deng, J. Cavers, and P. Ho, "A reduced dimensionality propagation model for frequency selective Rayleigh fading channels," in *Proc. IEEE ICC'95*, Seattle, WA, 1995, pp. 1158–1162.

[7] E. Eleftheriou and D. D. Falconer, "Adaptive equalization tech. for HF channels," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 238–247, Feb. 1987.

[8] D. D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Commun. Mag.*, vol. 33, pp. 50–57, Jan. 1995.

[9] S. Fechtel and H. Meyr, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. Commun.*, vol. 42, pp. 1639–1650, Feb./Mar./Apr. 1994.

[10] S. Fechtel and H. Meyr, "An investigation of channel estimation and equalization techniques for moderately rapid fading HF-channels," in *Proc. IEEE ICC'91*, vol. 2, Denver, CO, 1991, pp. 768–772.

[11] ——, "Optimal feedforward estimation of frequency-selective fading radio channels using statistical channel information," in *Proc. ICC'92*, vol. 2, Chicago, IL, 1992, pp. 677–681.
[12] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.
[13] B. Glance and L. J. Greenstein, "Frequency-selective fading effects in digital mobile radio with diversity combining," *IEEE Trans. Commun.*, vol. COM-31, pp. 1085–1094, Sept. 1983.
[14] H. Hashemi," Simulation of the urban radio propagation channel," *IEEE Trans. Veh. Technol.*, vol. VT-28, pp. 213–225, Aug. 1979.
[15] W. C. Jakes, Ed., *Microwave Mobile Communications.* New York: Wiley, 1974.
[16] G. K. Kaleh, "Channel estimation for block transmission systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 110–121, Jan. 1995.
[17] P. Kam, "Bit error rate probabilities of MDPSK over the nonselective Rayleigh fading channel with diversity reception," *IEEE Trans. Commun.*, vol. 39, pp. 220–224, Feb. 1991.
[18] J. Lin, J. G. Proakis, F. Ling, and H. Lev-Ari, "Optimal tracking of time-varying channels: A frequency domain approach for known and new algorithms," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 141–153, Jan. 1995.
[19] F. Ling, "Matched filter-bound for time-discrete multipath Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 710–713, Feb./Mar./Apr. 1995.
[20] N. 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.
[21] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, vol. 17, pp. 56–64, Jan. 1971.
[22] ——, "Theoretical and measured performance of a DFE modem on a fading multipath channel," *IEEE Trans. Veh. Technol.*, vol. COM-25, pp. 1144–1153, Oct. 1977.
[23] ——, "Digital transmission performance on fading dispersive diversity channels," *IEEE Trans. Commun.*, vol. COM-21, pp. 33–39, Jan. 1973.
[24] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1989.
[25] J. Salz, "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, no. 8, pp. 1341–1373, Oct. 1973.
[26] P. K. Shukla and L. F. Turner, "Channel-estimation-based adaptive DFE forfading multipath radio channels," *Proc. Inst. Elect. Eng.* vol. 138, pt. I, pp. 525–543, Dec. 1991.
[27] P. K. Shukla and L. F. Turner, "Examination of an adaptive DFE and MLSE/near-MLSE for fading multipath radio channels," *Proc. Inst. Elect. Eng.*, vol. 139, pt. I, pp. 418–428, Aug. 1992.
[28] M. Stojanovic, J. G. Proakis, and J. Catipovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Trans. Commun.*, vol. 43, pp. 877–886, Feb./Mar./Apr. 1995.
[29] R. Steel, *Mobile Radio Communications.* New York: IEEE Press, 1992.
[30] R. A. Ziegler and J. M. Cioffi, "Estimation of time-varying digital radiochannels," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 134–151, May 1992.
[31] *Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard*, Telecommunications Industry Association TIA/EIA IS-54, Dec. 1991.
[32] A. A. Abu-Dayya and N. C. Beaulieu, "Micro- and macrodiversity MDPSK on shadowed frequency-selective channels," *IEEE Trans. Commun.*, vol. 43, pp. 2334–2343, Aug. 1995.
[33] K. A. Hamied, M. Rahman, and M. S. El-Hennawey, "A new channel estimator for fast start-up equalization," *IEEE Trans. Commun.*, vol. 39, pp. 177–181, Feb. 1991.
[34] C. Rao, *Linear Statistical Inference and Its Applications.* New York: Wiley, 1973.
[35] H. Meyr and R. Subramanian, "Advanced digital receiver principles and technologies for PCS," *IEEE Commun. Mag.*, vol. 33, pp. 68–78, Jan. 1995.
[36] A. S. Khayrallah, R. Ramésh, G. Bottomley, and D. Koilpillai, "Improved channel estimation with side information," in *IEEE 47th Vehicular Technology Conf. Proc.*, vol. 2, Phoenix, AZ, May 1997, pp. 1049–1053.
[37] H.-N. Lee and G. J. Pottie, "Channel estimation based adaptive equalization/diversity combining for time-varying dispersive channels," in *Proc. IEEE 47th Vehicular Technology Conf.*, Phoenix, AZ, vol. 2, May 1997, pp. 884–888.

Heung-No Lee was born in Choong-Nam, Korea. He received the B.S. degree in 1993 and the M.S. degree in 1994, both in electrical engineering, from the University of California, Los Angeles, where he is currently working towards the Ph.D. degree in electrical engineering.

He has worked in the areas of adaptive channel estimation, equalization, and channel coding for multipath ISI fading channels. His current research interests include estimation and detection techniques and system design for personal communication transceivers.

Gregory J. Pottie (S'85–M'88) was born in Wilmington, DE. He received the B.Sc. degree in engineering physics from Queen's University, Kingston, Ont., Canada, in 1984, and the M.Eng. and Ph.D. degrees from McMaster University, Hamilton, Ont., Canada, in 1985 and 1988, respectively. From 1989 until 1991 he worked in the transmission research department of Codex/Motorola, Mansfield, MA, with projects including high-speed digital subscriber lines, and coding and equalization schemes for voice-band modems. He is presently an Associate Professor with the Electrical Engineering Department, University of California, Los Angeles. His research interests include systems design for wireless distributed sensor networks and personal communication transceivers, as well as channel coding and digital subscriber lines.