

Adaptive Turbo Reduced-Rank Equalization for MIMO Channels

Yakun Sun, *Member, IEEE*, Vinayak Tripathi, *Member, IEEE*, and Michael L. Honig, *Fellow, IEEE*

Abstract—An adaptive iterative (turbo) decision-feedback equalizer (DFE) for channels with intersymbol interference (ISI) is presented. The filters are computed directly from the soft decisions and received data to minimize a least-squares (LS) cost function. Numerical results show that this method gives a substantial improvement in performance relative to a turbo DFE computed from an exact channel estimate, assuming perfect feedback. Adaptive reduced-rank estimation methods are also presented, based on the multistage Wiener filter (MSWF). The adaptive reduced-rank turbo DFE for single-input/single-output channels is extended to multiple-input/multiple-output (MIMO) channels with ISI and multiple receive antennas. Numerical results show that for MIMO channels with limited training, the reduced-rank turbo DFE can perform significantly better than the full-rank turbo DFE.

Index Terms—Adaptive filters, multiple-input/multiple-output (MIMO) channels, space-time processing, turbo equalization.

I. INTRODUCTION

ACHIEVING high spectral efficiencies over frequency-selective channels requires channel coding and mitigation of intersymbol interference (ISI). Viewing the code and channel as a serially concatenated code leads to the application of the turbo principle for joint equalization and decoding [1], [2]. Initial work on turbo equalization considered a maximum *a posteriori* (MAP) symbol-by-symbol equalizer with an MAP symbol-by-symbol decoder [2]. In order to reduce the complexity of the turbo equalizer, it has been recognized that the MAP equalizer can be replaced by a decision-feedback equalizer (DFE) with soft cancellation of ISI [3].

Previous work on turbo decision-feedback equalization has assumed that the receiver either knows or estimates the channel [4]–[8]. The channel estimate is then used to compute the DFE filter coefficients assuming perfect feedback. In practice, this approach is appropriate given limited training overhead,

Manuscript received August 16, 2003; revised April 23, 2004; accepted September 20, 2004. The editor coordinating the review of this paper and approving it for publication is L. Vandendorpe. This work was supported by the Northwestern–Motorola Center for Communications, and by the U.S. Army Research Office under Grant DAAD19-99-1-0288. This paper was presented in part at the IEEE Military Communications (MILCOM) Conference, Anaheim, CA, October, 2002.

Y. Sun was with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA. He is now with Motorola, Inc., Libertyville, IL 60048 USA (e-mail: yksun@motorola.com).

V. Tripathi was with Motorola Labs, Schaumburg, IL 60196 USA. He is now with the Economics Department, Princeton University, Princeton, NJ 08544 USA (e-mail: vtripath@princeton.edu).

M. L. Honig is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA (e-mail: mh@ece.northwestern.edu).

Digital Object Identifier 10.1109/TWC.2005.858037

since typically there are relatively few channel coefficients to estimate compared with the number of filter coefficients. However, the performance of the turbo equalizer is sensitive to channel estimation error [9], [10]. Another drawback of this approach is that even with perfect channel estimates, the resulting filters are generally suboptimal, due to the assumption of perfect feedback.

Here we introduce an adaptive turbo equalizer in which the filter coefficients are directly estimated from the received data and soft decoder outputs according to a least-squares (LS) criterion. This approach is motivated by a similar adaptive approach to turbo multiuser detection for CDMA [11]. When the number of filter coefficients is larger than the number of training symbols, it is infeasible to use the training symbols alone to estimate the DFE filters directly. Consequently, a hybrid scheme is proposed in which the channel is estimated from the training sequence for the first iteration, and in subsequent iterations, the filters are estimated directly from the soft decoder outputs. Adapting the filter coefficients directly makes the turbo DFE less sensitive to channel estimation error and more robust to error propagation. We observe, in fact, that the LS adaptive turbo DFE outperforms a nonadaptive turbo equalizer, which uses a perfect channel estimate and assumes perfect feedback.

Related work on adaptive turbo equalization for single-input/single-output channels has been presented in [12]. In that work, a stochastic-gradient algorithm is used to update the equalizer coefficients. In this paper, we show that adapting the filter coefficients according to an LS criterion is better than the stochastic gradient approach, particularly for short packet lengths.

We also extend the LS turbo DFE for the single-antenna channel to a system with an arbitrary number of receive antennas and other cochannel users. Related work on noniterative space-time equalization is presented in [13]–[15]. Space-time turbo equalization for multiple-input multiple-output (MIMO) channels is studied in [16] and [17], assuming both exact and estimated channels. Here we consider two scenarios: 1) space-time equalization with cochannel interference (CCI) and 2) a single-user MIMO channel with multiple transmit and receive antennas. At the receiver, the ISI and CCI are suppressed or cancelled with an adaptive MIMO DFE. Because channel estimation is difficult in the presence of strong CCI, in that situation, we directly estimate the DFE coefficients, which minimize the LS cost function in the initial iteration as well as in subsequent iterations.

The number of DFE coefficients increases with the number of antennas and the channel delay spread. With limited training, the LS filter estimates degrade as the filter lengths increase,

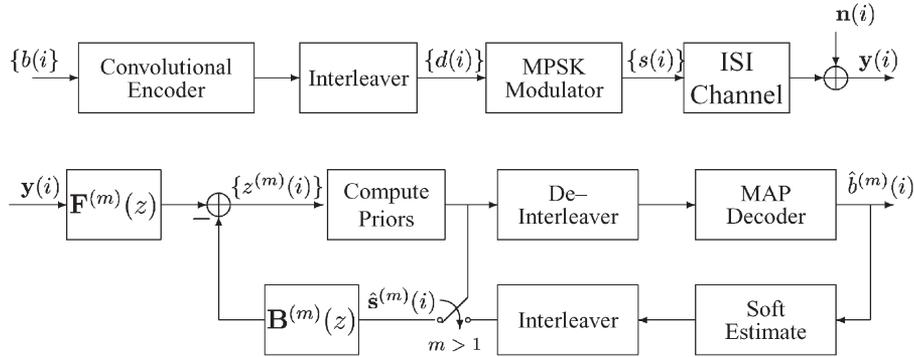


Fig. 1. Block diagram of a transmitter and turbo DFE for a single-user single-antenna system.

and the performance of the turbo equalizer is limited by the estimation error. Furthermore, the complexity of LS estimation also increases with filter length. To improve performance with limited training, we apply reduced-rank estimation based on the multistage Wiener filter (MSWF) [18], [19] (see also [20], which applies the MSWF to noniterative equalization). Numerical results are presented, which show that with limited training, the reduced-rank algorithm can perform substantially better than the full-rank (conventional) LS algorithm. In addition, the reduced-rank algorithm requires less computation than the full-rank algorithm when the number of filter coefficients is large.

The outline of the paper follows. In Section II, we present the adaptive LS turbo DFE for a single-user single-antenna system. The associated performance is compared with the turbo DFE using a channel estimate, and the stochastic-gradient DFE. The reduced-rank turbo DFE is also introduced in this section. In Section III, we generalize the LS turbo DFE to a multiuser multi-antenna system and a single-user MIMO system. Section IV concludes the paper.

II. SINGLE-USER SINGLE-ANTENNA SYSTEM

A. System Model

To begin, consider the single-user single-antenna system model shown in Fig. 1. At the transmitter, a block of information bits $\{b(i)\}$ is convolutionally encoded and interleaved. The coded interleaved bits $\{d(i)\}$ are input to a multitone phase-shift keying (MPSK) modulator (this work is easily extended to other modulation formats). The resulting T_S symbols, $\{s(i)\}$ along with a midamble of T_0 binary training symbols (motivated by the Enhanced Data rate for Global System for Mobile communications (GSM) Evolution (EDGE) system [21]), form a packet that is transmitted over an ISI channel.¹ The channel is modeled as a tapped-delay line with L symbol-spaced taps, and is represented by the vector $\mathbf{h} = [h_0 \cdots h_{L-1}]^T$, where $(\cdot)^T$ denotes transpose. Here we assume symbol-synchronous sampling for convenience. A fractionally spaced equalizer can be used with asynchronous sampling.

At the receiver, the noisy received sequence $\{y(i)\}$, corresponding to a packet, is processed N_f samples at a time. At time i , the vector of N_f received samples is given by

$$\mathbf{y}(i) = \mathbf{H}\mathbf{s}(i) + \mathbf{n}(i) \quad (1)$$

where $\mathbf{y}(i) = [y(i + \delta - 1) \cdots y(i + \delta - N_f)]^T$ and δ is a delay, which indicates the position of the cursor. The corresponding vector of coded MPSK symbols is $\mathbf{s}(i) = [s(i + \delta - 1) \cdots s(i + \delta - N_f - L + 1)]^T$, and $\mathbf{n}(i)$ is a vector of Gaussian noise samples with covariance matrix $\sigma^2 \mathbf{I}$. The $N_f \times (N_f + L - 1)$ Toeplitz channel matrix is given by

$$\mathbf{H} = \begin{bmatrix} h_0 & \cdots & h_{L-1} & 0 & \cdots & \cdots & 0 \\ 0 & h_0 & \cdots & h_{L-1} & & & \vdots \\ \vdots & & \ddots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & h_0 & \cdots & h_{L-1} \end{bmatrix}. \quad (2)$$

B. Turbo DFE

In the first iteration of the turbo-equalization algorithm, the received sequence $\{y(i)\}$ is the input to a symbol-spaced DFE that consists of a feedforward filter $\mathbf{f}^{(m)}$ of length N_f , and a feedback filter $\mathbf{b}^{(m)}$ of length N_b . The superscript m denotes the iteration number, so that for the first iteration $m = 1$. At time i , the output of the DFE for the m th iteration is

$$z^{(m)}(i) = \left(\mathbf{f}^{(m)}\right)^\dagger \mathbf{y}(i) - \left(\mathbf{b}^{(m)}\right)^\dagger \hat{\mathbf{s}}^{(m)}(i) \quad (3)$$

where “ \dagger ” denotes complex conjugate transpose. As explained in what follows, the $N_b \times 1$ vector $\hat{\mathbf{s}}^{(m)}(i) = [\hat{s}^{(m)}(i + \delta - 1) \cdots \hat{s}^{(m)}(i + \delta - N_b)]^T$, which is the input to the feedback filter, contains soft decisions of the corresponding transmitted symbols. For $m > 1$, these values are computed by the MAP decoder at iteration $m - 1$. For $m = 1$, each $\hat{s}^{(1)}(i)$ is computed directly from the corresponding DFE output $z^{(1)}(i)$, and used to cancel ISI associated with successive symbols.

For $m = 1$, the DFE is a causal filter and the first δ elements of $\mathbf{b}^{(1)}$ are set to zero. The delay δ is selected so that the channel impulse response falls within the span of the feedforward filter. The DFE outputs $\{z^{(1)}(i)\}$, are buffered and used to compute the *a priori* probabilities $\{P(z^{(1)}(i)|d_j(s(i)) = \pm 1), j = 1, \dots, \log_2 M\}$, where $\{d_j(s(i)), j = 1, \dots, \log_2 M\}$ is the

¹The sequence index i spans a different range for the information bits, coded bits, and transmitted symbols, corresponding to different sequence lengths. In what follows, we will refer to i as the time or symbol index, corresponding to the symbol sequence $\{s(i)\}$.

sequence of $\log_2 M$ bits corresponding to the MPSK symbol $s(i)$. These probabilities are deinterleaved and fed into an MAP decoder that generates the *a posteriori* probabilities $P(d_j(s(i)) = \pm 1 | \mathbf{z})$, where \mathbf{z} is the vector of DFE outputs corresponding to the entire packet. This completes the first iteration.

In subsequent iterations, the outputs of the MAP decoder are reinterleaved, and used to compute soft symbol estimates. The symbol estimates are fed into a noncausal DFE, or ISI canceller [22], whose output at time i is again given by (3), where for $m \geq 2$ only the δ th element in $\mathbf{b}^{(m)}$ is set to zero. As suggested by the notation in (3), the filters are updated in every iteration, but remain constant for the duration of a packet. In the final iteration, the MAP decoder outputs hard decisions for the information sequence.

C. Computations of Filters

In the first iteration, we wish to select filters that minimize the mean-squared-error (MSE) cost function given by

$$\mathcal{E}_{\text{MSE}} = E \left[\left| s - \left(\mathbf{f}^{(1)} \right)^\dagger \mathbf{y} + \left(\mathbf{b}^{(1)} \right)^\dagger \hat{\mathbf{s}} \right|^2 \right] \quad (4)$$

where the dependence on i is not shown for convenience. Assuming that the DFE has perfect feedback, i.e., $\hat{\mathbf{s}} = \mathbf{s}$ (also, $N_b = N_f + L - 1$), the optimum causal filters are [23]

$$\begin{aligned} \mathbf{f}^{(1)} &= \left(\mathbf{H}_{1:\delta} \mathbf{H}_{1:\delta}^\dagger + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{h}_\delta \\ \mathbf{b}^{(1)} &= [\mathbf{0}_{N_f \times \delta} \quad \mathbf{H}_{\delta+1:N_f+L-1}]^\dagger \mathbf{f}^{(1)} \end{aligned} \quad (5)$$

where \mathbf{h}_k is the k th column of \mathbf{H} and $\mathbf{H}_{j:k}$ represents the matrix comprised of the j th to k th columns of \mathbf{H} , and $\mathbf{0}_{N_f \times \delta}$ denotes an $N_f \times \delta$ matrix of zeros.

Note that the MMSE DFE shown here is a time-invariant filter, which depends only on the channel. In contrast, a time-varying MMSE filter, which depends on the MAP outputs, is proposed in [8]. In that case, the MMSE filters must be updated every symbol, as opposed to once per channel realization for the time-invariant DFE. Because of the high complexity associated with the time-varying filter, in what follows, we will use the time-invariant MMSE DFE as a benchmark for comparison with adaptive estimation schemes.

Using the midamble as a training sequence, an LS estimate of the channel vector \mathbf{h} is given by

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \sum_{i=i_0+L-1}^{i_0+T_0} \|\mathbf{y}(i) - \mathbf{H} \mathbf{s}_T(i)\|^2 \quad (6)$$

where i_0 is the index for the first training symbol in the packet, and $\mathbf{s}_T(i)$ denotes the i th training symbol. For the first iteration, the adaptive DFE is computed from (5) with the estimated channel.

For iterations $m \geq 2$, the causal DFE is replaced with a noncausal DFE, which cancels both pre- and postcursor ISI. In this paper, we consider the following three adaptive methods for recomputing the filters at each iteration.

1) *Channel-Estimated Adaptation*: In this approach, the channel is reestimated at each iteration using the soft symbol estimates from the previous iteration as additional training. This type of data-directed channel estimation has been considered in [24] and [25]. The updated channel estimate is then used to recompute the minimum MSE (MMSE) filters. For the causal DFE, these filters are given by (5), whereas for the noncausal two-sided DFE, the MMSE filters are given by [22]

$$\begin{aligned} \mathbf{f}^{(m)} &= \frac{1}{\|\mathbf{h}_\delta^{(m)}\|^2 + \sigma^2} \mathbf{h}_\delta^{(m)} \\ \mathbf{b}^{(m)} &= [\mathbf{H}_{1:\delta-1} \quad \mathbf{0}_{N_f \times 1} \quad \mathbf{H}_{\delta+1:N_f+L-1}]^\dagger \mathbf{f}^{(m)} \end{aligned} \quad (7)$$

again assuming perfect feedback. The channel-estimated DFE filter coefficients are computed from (7) with $\mathbf{h}_\delta^{(m)}$ replaced by the corresponding channel estimate $\hat{\mathbf{h}}_\delta^{(m)}$. If $\delta = L$, then the feedforward filter is simply a scaled matched filter and the feedback filter cancels both pre- and postcursor ISI.

2) *LS Direct Adaptation*: Instead of first estimating the channel taps and then computing the filters according to (7), we can use the soft symbol estimates to compute the filter coefficients directly. Our objective is to minimize the LS cost function

$$\mathcal{E}_{\text{LS}} = \sum_{i=1}^T \left| \hat{\mathbf{s}}^{(m)}(i) - \left(\mathbf{f}^{(m)} \right)^\dagger \mathbf{y}(i) + \left(\mathbf{b}^{(m)} \right)^\dagger \hat{\mathbf{s}}^{(m)}(i) \right|^2 \quad (8)$$

where $T = T_0 + T_S$, is the total length of the packet. The noncausal DFE can be expressed in terms of the concatenated feedforward and feedback filters $\begin{bmatrix} \mathbf{f}^{(m)} \\ \mathbf{b}^{(m)} \end{bmatrix}$ and input vector $\mathbf{x}(i) = \begin{bmatrix} \mathbf{y}^{(i)} \\ \check{\mathbf{s}}^{(m)}(i) \end{bmatrix}$, where $\check{\mathbf{s}}^{(m)}(i)$ is the vector $\hat{\mathbf{s}}^{(m)}(i)$ with the δ th element deleted, and $\mathbf{b}^{(m)}$ contains the corresponding elements of the feedback filter $\mathbf{b}^{(m)}$. Using this notation, the filters that minimize the LS cost function can be written as [23]

$$\begin{aligned} \begin{bmatrix} \mathbf{f}^{(m)} \\ \check{\mathbf{b}}^{(m)} \end{bmatrix} &= \mathbf{R}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{R}_{\mathbf{s}\mathbf{x}} \\ \mathbf{R}_{\mathbf{x}\mathbf{x}} &= \sum_{i=1}^T \mathbf{x}(i) \mathbf{x}^\dagger(i) \\ \mathbf{R}_{\mathbf{s}\mathbf{x}} &= \sum_{i=1}^T \mathbf{x}(i) \left(\hat{\mathbf{s}}^{(m)}(i) \right)^* \end{aligned} \quad (9)$$

The principal advantage of this approach over the preceding approach, based on channel estimation, is that it does not rely on the perfect feedback assumption.

3) *Least-Mean-Square (LMS) Adaptation*: For the sake of comparison, we also consider the stochastic gradient, or LMS adaptation of filter coefficients proposed in [12]. In this case, the filter coefficients are updated as

$$\begin{aligned} \mathbf{f}^{(m)}(i+1) &= \mathbf{f}^{(m)}(i) - \alpha \mathbf{y}^*(i) \left[\hat{\mathbf{s}}^{(m-1)}(i) - \tilde{z}(i) \right] \\ \check{\mathbf{b}}^{(m)}(i+1) &= \check{\mathbf{b}}^{(m)}(i) + \alpha \left(\check{\mathbf{s}}^{(m-1)}(i) \right)^* \left[\hat{\mathbf{s}}^{(m-1)}(i) - \tilde{z}(i) \right] \\ \tilde{z}(i) &= \left(\mathbf{f}^{(m)}(i) \right)^\dagger \mathbf{y}(i) - \left(\check{\mathbf{b}}^{(m)}(i) \right)^\dagger \check{\mathbf{s}}^{(m-1)}(i) \end{aligned} \quad (10)$$

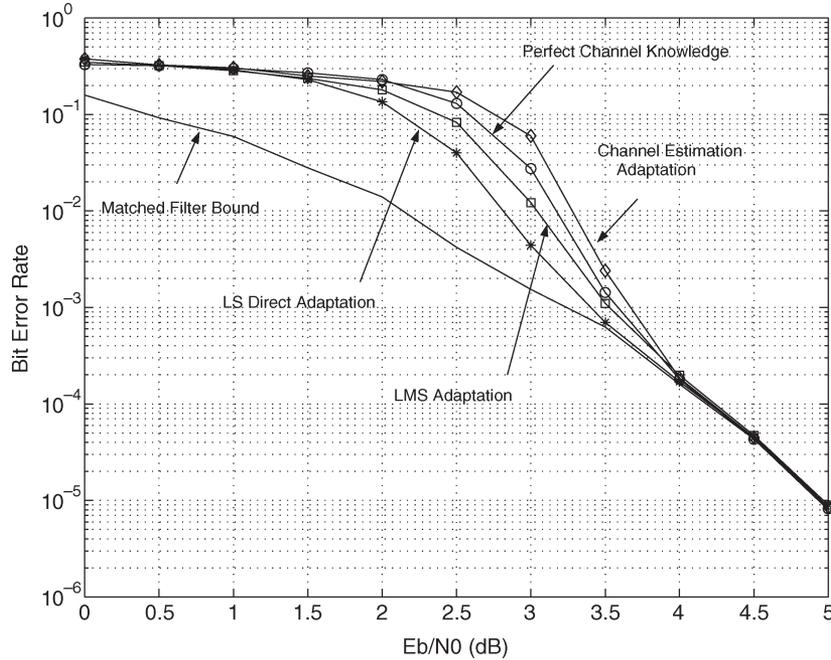


Fig. 2. BER versus E_b/N_0 (dB) for adaptive and channel-estimated turbo DFEs.

for $i = 0, \dots, T - 1$, where α is an appropriate step size, and the filters are initialized as the corresponding filters from the last iteration, i.e., $\mathbf{f}^{(m)}(0) = \mathbf{f}^{(m-1)}(T)$ and $\check{\mathbf{b}}^{(m)}(0) = \check{\mathbf{b}}^{(m-1)}(T)$. Note that once the LMS algorithm runs through the entire packet, the final DFE outputs for the m th iteration are generated using $\mathbf{f}^{(m)}(T)$ and $\check{\mathbf{b}}^{(m)}(T)$ in (3).

D. Computation of Priors and Feedback Symbols

The DFE outputs $\{z(i)\}$ are buffered and used to compute the *a priori* probabilities, which are the inputs to the MAP decoder (to simplify the notation, the iteration index is omitted in this section). To do this, the filter outputs are modeled as Gaussian random variables with mean and variance given by

$$\begin{aligned} \mu &= \frac{1}{T} \sum_{i=1}^T |\hat{\mathbf{s}}^*(i)z(i)| \\ \sigma^2 &= \frac{1}{T} \sum_{i=1}^T |\hat{\mathbf{s}}^*(i)z(i) - \mu|^2. \end{aligned} \tag{11}$$

The priors can then be expressed as

$$P(z(i)|d_j(s(i)) = \pm 1) = \mathcal{K} \sum_{s \in \Omega_j^\pm} e^{-\frac{|z(i) - \mu s|^2}{\sigma^2}} \tag{12}$$

where \mathcal{K} is a scaling factor that normalizes the probability density, and Ω_j^\pm is the set of symbols in the signal constellation for which the j th bit is ± 1 .

The MAP decoder outputs the sequence of *a posteriori* probabilities $\{P(d_j(s(i)) = \pm 1|\mathbf{z})\}$. These are used to compute the sequence of soft symbol estimates $\{\hat{s}(i)\}$. The *a posteriori*

probability distribution of an MPSK signal constellation at time instant i is given by

$$P(s(i) = e^{j\frac{2\ell\pi}{M}}|\mathbf{z}) = \prod_{j=1}^{\log_2 M} P(d_j(s(i)) = d_j(e^{j\frac{2\ell\pi}{M}})|\mathbf{z}) \tag{13}$$

where $\ell = 0, \dots, \log_2 M - 1$. The soft symbol estimate at time i is then

$$\hat{s}(i) = E[s(i)|\mathbf{z}] = \sum_{\ell=0}^{\log_2 M - 1} P(s(i) = e^{j\frac{2\ell\pi}{M}}|\mathbf{z}) e^{j\frac{2\ell\pi}{M}}. \tag{14}$$

E. Numerical Results

Fig. 2 shows plots of bit error rate (BER) versus E_b/N_0 , where E_b is the energy per information bit and $N_0 = 2\sigma^2$, for different turbo equalizers. In addition to the three adaptive schemes previously described, the figure also shows the matched filter bound, corresponding to the filter in (7) with perfect feedback (i.e., perfect ISI cancellation). The performance of the turbo equalizer based on channel estimation with perfect channel knowledge (computed from (5) and (7) using the actual channel \mathbf{h}) and soft feedback is also presented for comparison.

The simulation model is similar to the one described in [12]. The information bits are encoded with a rate-1/2 constraint-length-5 convolutional code with generator polynomials [23, 35] (octal). The coded bits are randomly interleaved and modulated using a Gray-encoded QPSK constellation before being transmitted over the static Proakis B channel, specified in [12], with channel taps $\mathbf{h} = [0.407 \ 0.815 \ 0.407]^T$. The DFE feedforward and feedback filters have lengths $N_f = 12$ and $N_b = 15$, respectively. The turbo equalizer iterates ten times,

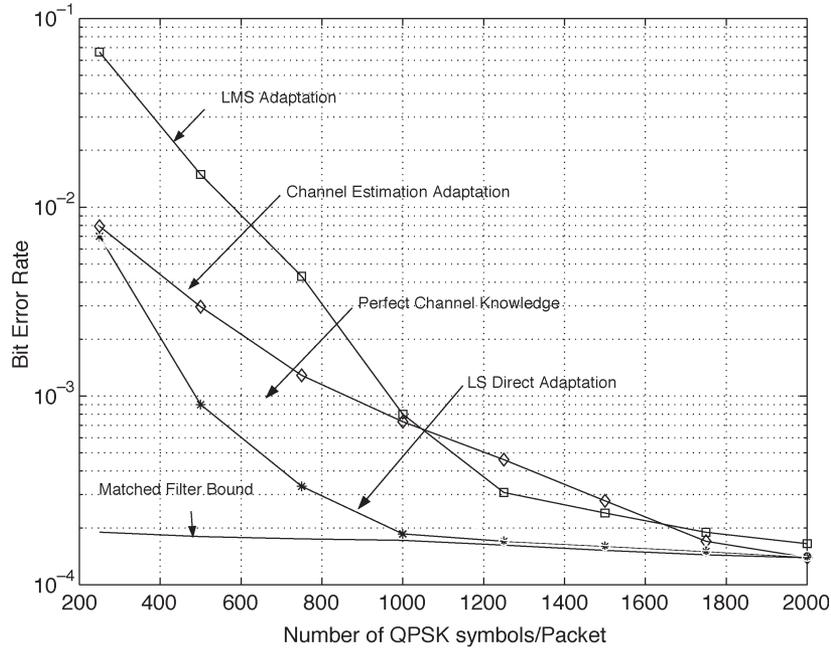


Fig. 3. BER as a function of packet length.

and the LMS-algorithm step size is $\alpha = 0.0005$. The results are averaged over packets containing 2000 coded QPSK symbols, and a midamble of 200 binary training symbols.

Fig. 2 shows that the BER for each of the three adaptation schemes converges to the matched filter bound, with the LS adaptive turbo equalizer converging at the lowest SNR. In fact, LS direct adaptation outperforms the nonadaptive turbo equalizer with perfect channel knowledge. This is because the adaptive technique does not assume perfect feedback to compute the DFE filters. As expected, the channel-estimated turbo equalizer performs somewhat worse than the turbo equalizer with perfect channel knowledge. Finally, the LMS adaptive turbo equalizer does not rely on the perfect feedback assumption, but because of its relatively slow convergence, does not perform as well as the LS turbo DFE. Still, because of the large packet size (2000 symbols), LMS adaptation also provides better performance than the channel-estimated turbo DFE.

Fig. 3 shows BER versus packet length at an SNR of 4 dB. The length of the packet is increased while keeping the length of the midamble fixed at $T_0 = 200$ symbols. As the packet length increases, the LS adaptive turbo equalizer converges much faster to the matched filter bound than the LMS and channel-estimated turbo DFEs. It also significantly outperforms the nonadaptive turbo DFE with perfect channel knowledge. The performance improvement of the nonadaptive DFE with packet length reflects the gain associated with longer interleaving depths. With short packets, the adaptive LS and channel-estimated turbo equalizers perform about the same as the turbo equalizer with perfect channel knowledge. This is because the 200-symbol-long midamble provides an accurate initialization of the equalizer filters in the first iteration, and the short packets inhibit an adaptive algorithm from providing further improvements during subsequent iterations. With LMS adaptation, we observe a significant degradation in performance

as the packet size decreases. This degradation can be mitigated to some extent by reducing the step size α .

F. Reduced-Rank Turbo DFE

A property of the LS estimation algorithms previously presented is that both the complexity and the amount of training required for acceptable performance increase with the filter length. This may present a problem with highly dispersive channels. One method for reducing the amount of training overhead is reduced-rank estimation of the feedforward and feedback filters [26, Sec. 8.4]. Here we focus on the class of adaptive reduced-rank algorithms introduced in [19], which are motivated by the MSWF [18]. These algorithms can provide a significant reduction in training when used with adaptive linear filters [19].

A reduced-rank filter projects the received signal onto a lower dimensional subspace. Both the filtering and the filter estimation take place within this subspace. Let \mathbf{S}_D be an $N_f \times D$ matrix with column vectors that are a set of basis vectors for a D -dimensional subspace, where $D < N_f$. The projected received vector corresponding to symbol i is then given by

$$\tilde{\mathbf{y}}(i) = \mathbf{S}_D^\dagger \mathbf{y}(i).$$

The sequence of projected received vectors $\{\tilde{\mathbf{y}}(i)\}$ is the input to the linear filter $\tilde{\mathbf{c}}$ ($D \times 1$ vector), which has output

$$z(i) = \tilde{\mathbf{c}}^\dagger \tilde{\mathbf{y}}(i).$$

The vector $\tilde{\mathbf{c}}(i)$, which minimizes the LS cost function $\sum_i |s(i) - \tilde{\mathbf{c}}^\dagger(i) \tilde{\mathbf{y}}(i)|^2$, is

$$\tilde{\mathbf{c}} = \left(\mathbf{S}_D^\dagger \mathbf{R}_{\mathbf{y}\mathbf{y}} \mathbf{S}_D \right)^{-1} \left(\mathbf{S}_D^\dagger \mathbf{R}_{\mathbf{s}\mathbf{y}} \right) \quad (15)$$

where

$$\mathbf{R}_{yy} = \sum_i \mathbf{y}(i)\mathbf{y}^\dagger(i) \quad (16)$$

$$\mathbf{R}_{sy} = \sum_i \hat{\mathbf{s}}^*(i)\mathbf{y}(i). \quad (17)$$

Here we focus on the reduced-rank algorithm presented in [19], for which

$$\mathbf{S}_D = [\mathbf{R}_{sy} \ \mathbf{R}_{yy}\mathbf{R}_{sy} \ \mathbf{R}_{yy}^2\mathbf{R}_{sy} \ \cdots \ \mathbf{R}_{yy}^{D-1}\mathbf{R}_{sy}]. \quad (18)$$

To apply reduced-rank filtering to the DFE, we first observe that the feedforward filter can be represented as the concatenation of a linear LS filter with an error-estimation filter [23]. The latter filter is then the feedback filter. A reduced-rank DFE uses reduced-rank approximations for both of these filters (see also [11] and [20], which present reduced-rank DFEs in other related applications).

In the first iteration, we compute the full-rank linear filter as

$$\mathbf{c} = \mathbf{R}_{yy}^{-1}\mathbf{R}_{sy} \quad (19)$$

where in (16) and (17), the sums are over the training midamble $\hat{\mathbf{s}}(i) = s(i)$ and the reduced-rank approximation of the linear filter is

$$\tilde{\mathbf{c}} = \mathbf{S}_D \tilde{\mathbf{c}} = \mathbf{S}_D \left(\mathbf{S}_D^\dagger \mathbf{R}_{yy} \mathbf{S}_D \right)^{-1} \left(\mathbf{S}_D^\dagger \mathbf{R}_{sy} \right) \quad (20)$$

where $\tilde{\mathbf{c}}$ is given in (15). In subsequent iterations, the linear filters still have these forms, but \mathbf{R}_{yy} and \mathbf{R}_{sy} are computed over the whole packet.

The output error using the soft symbol estimate is

$$\mathbf{e}^{(m)}(i) = \hat{\mathbf{s}}^{(m)}(i) - \left(\mathbf{c}^{(m)} \right)^\dagger \mathbf{y}(i). \quad (21)$$

Let $\mathbf{e}^{(m)}(i) = [e^{(m)}(i + \delta - 1) \ \cdots \ e^{(m)}(i + \delta - N_b)]^T$. The error-estimation filter forms an LS estimate of $\mathbf{e}^{(m)}(i)$ given the other elements in $\mathbf{e}^{(m)}(i)$. That is, we select $\mathbf{b}^{(m)}$ to minimize $\sum_i |\mathbf{e}^{(m)}(i) - (\mathbf{b}^{(m)})^\dagger \mathbf{e}^{(m)}(i)|^2$, where the sum is over the training midamble in the first iteration, and over the whole packet in subsequent iterations. Also, the first δ components of $\mathbf{b}^{(1)}$ are constrained to be zero, and the δ th component of $\mathbf{b}^{(m)}$ is constrained to be zero for $m \geq 2$. Let $\check{\mathbf{e}}^{(m)}$ contain the unconstrained components of $\mathbf{e}^{(m)}$, i.e., (22), shown at the bottom of the page. The full-rank error-estimation (feedback) filter is

$$\check{\mathbf{b}}^{(m)} = \check{\mathbf{R}}_e^{-1} \check{\mathbf{r}}_e \quad (23)$$

where

$$\check{\mathbf{R}}_e = \sum_i \check{\mathbf{e}}^{(m)}(i) \left(\check{\mathbf{e}}^{(m)}(i) \right)^\dagger \quad (24)$$

and

$$\check{\mathbf{r}}_e = \sum_i \left(\mathbf{e}^{(m)}(i) \right)^* \check{\mathbf{e}}^{(m)}(i). \quad (25)$$

The feedforward filter is obtained by convolving the elements of $\mathbf{c}^{(m)}$ with $\bar{\mathbf{b}}^{(m)}$, where $\bar{\mathbf{b}}^{(m)}$ is the feedback filter $\mathbf{b}^{(m)}$ with the δ th component set to 1.

Given this representation for the feedback filter, a reduced-rank approximation is given by

$$\check{\mathbf{b}}^{(m)} = \mathbf{S}_e \left(\mathbf{S}_e^\dagger \check{\mathbf{R}}_e \mathbf{S}_e \right)^{-1} \mathbf{S}_e^\dagger \check{\mathbf{r}}_e \quad (26)$$

where

$$\mathbf{S}_e = [\check{\mathbf{r}}_e \ \check{\mathbf{R}}_e \check{\mathbf{r}}_e \ \check{\mathbf{R}}_e^2 \check{\mathbf{r}}_e \ \cdots \ \check{\mathbf{R}}_e^{D_b-1} \check{\mathbf{r}}_e] \quad (27)$$

and D_b is the associated rank. The feedforward filter $\mathbf{f}^{(m)}$ is the convolution of the reduced-rank approximation of the linear filter in (20) with the reduced-rank feedback filter in (26).

The ranks D and D_b for the first (linear LS) and second (error estimation) filters, respectively, can be chosen independently. In what follows, D and D_b are chosen to minimize the decision-error metric $\mathcal{M}_{D,D_b} = \sum_{i=1}^T |\hat{z}_{D,D_b}^{(m)}(i) - z_{D,D_b}^{(m)}(i)|^2$, where $z_{D,D_b}^{(m)}(i)$ is the reduced-rank DFE output for iteration m at time i , and $\hat{z}_{D,D_b}^{(m)}(i)$ is the corresponding hard decision on the transmitted symbol, i.e., the closest symbol in the constellation (other rank-selection criteria are discussed in [19] and [27]). We therefore have to reestimate the equalizer and compute the filter outputs (but not decode) for each choice of rank(s). Since the ranks that achieve close to full-rank performance are bounded by (small) constants [19], independent of the filter length, the computational complexity of the reduced-rank equalizer with rank estimation still increases linearly with packet size. To reduce complexity, D and D_b are optimized sequentially. Namely, D is first selected to minimize the decision-error metric at the output of the linear filter, followed by optimization of the feedback filter rank D_b . Numerical experiments show that this (suboptimal) adaptive rank-selection method performs nearly as well as jointly optimal rank selection.

Fig. 4 shows block error rate (BLER) versus number of training symbols for adaptive full- and reduced-rank DFEs. For this plot, the feedforward filter has 24 taps, the feedback filter has 16 taps, $E_b/N_0 = 18$ dB, and the dispersive channel has 13 independent complex Gaussian taps with power-delay profile $[-6, -3, 0, -3, -6, -10, \dots, -10]$ dB. The channel is constant within a burst, but is independent from burst to burst. Each

$$\check{\mathbf{e}}^{(m)} = \begin{cases} [e^{(m)}(i-1) \ \cdots \ e^{(m)}(i+\delta-N_b)]^T, & m = 1 \\ [e^{(m)}(i+\delta-1) \ \cdots \ e^{(m)}(i+1) \ e^{(m)}(i-1) \ \cdots \ e^{(m)}(i+\delta-N_b)]^T, & m > 1 \end{cases} \quad (22)$$

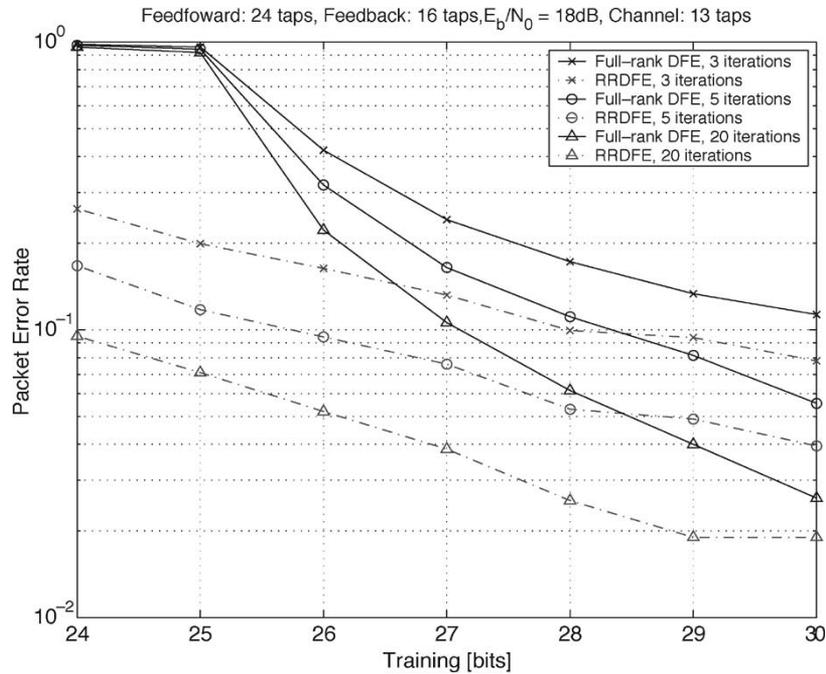


Fig. 4. BLER versus training symbols for adaptive full-rank and reduced-rank DFEs.

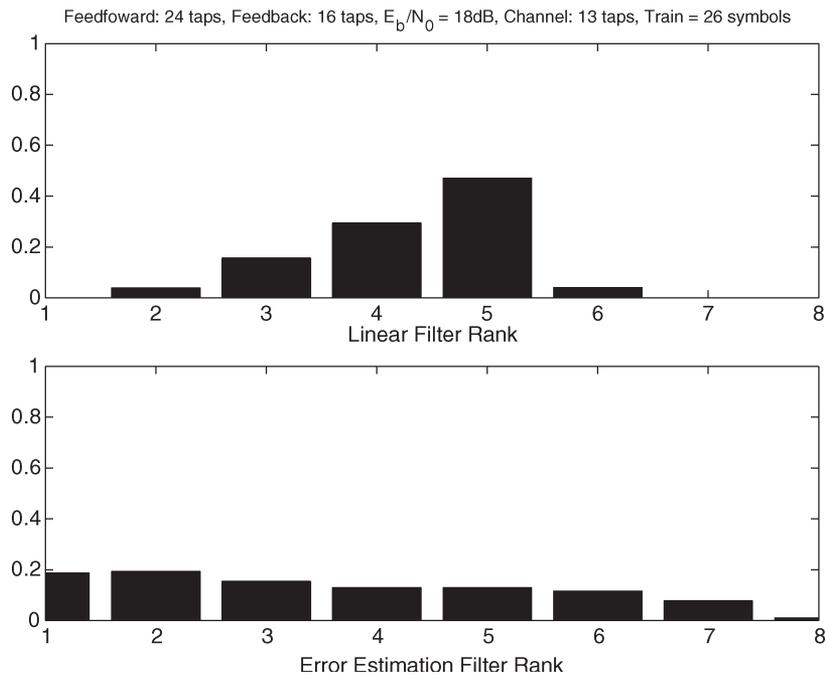


Fig. 5. Histogram of optimized ranks for DFE filters corresponding to the parameters used in Fig. 4 with 26 training symbols.

packet, or block, contains 464 Gray-encoded 8-PSK symbols. In the numerical examples that follow, the turbo DFE iterates 20 times. The reduced-rank results assume a reduced-rank channel estimate for the first iteration. These results show that the reduced-rank algorithms offer a significant performance gain relative to the full-rank algorithms for the parameters selected. In general, for a fixed training length, this gain increases with the size of the filters. For short filter lengths, e.g., $N_f < 10$ and $N_b < 10$, reduced-rank estimation does not offer a significant performance advantage. With fixed filter lengths,

the performance gain decreases as the training length increases, since both the reduced-rank (with optimal rank) and full-rank estimates converge to the MMSE estimate.

Fig. 5 shows a normalized histogram of the optimal ranks for the feedforward and feedback filters at the 20th iteration corresponding to the parameters used to generate Fig. 4 with 26 training symbols. The optimal ranks are typically quite small in comparison with the filter lengths. Depending on the packet length, estimation of a low-rank filter may require much less computation than estimating the corresponding full-rank

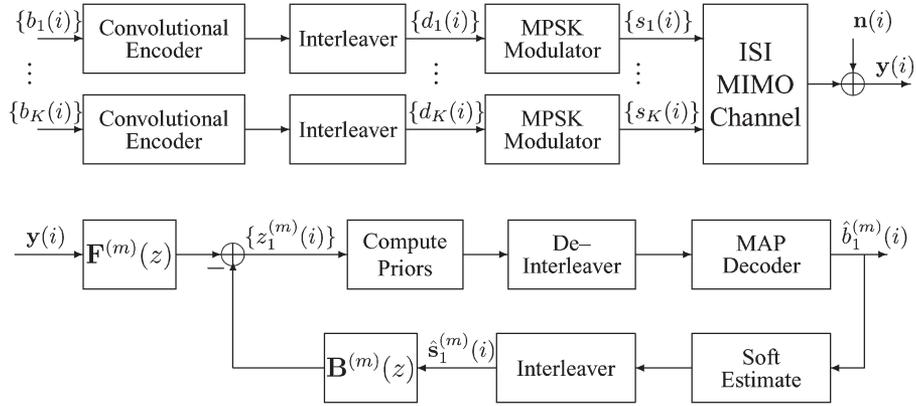


Fig. 6. Block diagram of the transmitters (top diagram) and turbo DFE (bottom diagram) for a multiuser MIMO channel.

filter. The optimal ranks are observed to change from iteration to iteration, and generally increase with the number of iterations.

III. MIMO CHANNELS

A. Multiuser Multiantenna System

In this section, we consider turbo equalization for MIMO generalizations of the previous single-antenna model. We start with the multiuser multireceive antenna model, shown in Fig. 6, in which K cochannel users each transmit with a single transmit antenna to a receiver with N antennas. The information bits for each user are coded and modulated as described previously, although the coding and modulation schemes can differ across users. Here we assume that the interfering users are in different cells, hence only the desired user's symbols are estimated and fed back for ISI cancellation.

The channel between user k and antenna n is represented by the vector $\mathbf{h}_{nk} = [h_0^{(n,k)} \cdots h_{L-1}^{(n,k)}]^T$ where L is the maximum impulse-response length across users and antennas. Letting $\mathbf{y}_n(i)$, $1 \leq n \leq N$, denote the vector of received samples at time i on antenna n , the stacked vector $\mathbf{y}(i) = [\mathbf{y}_1^T \cdots \mathbf{y}_N^T]^T$ again satisfies (1), where $\mathbf{s}^T(i) = [\mathbf{s}_1^T \cdots \mathbf{s}_N^T]$, \mathbf{s}_k being the vector of transmitted symbols from user k , and \mathbf{n} is the associated $(N_f \cdot N) \times 1$ noise vector with an identity covariance matrix. The channel matrix is

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1K} \\ \vdots & & \vdots \\ \mathbf{H}_{N1} & \cdots & \mathbf{H}_{NK} \end{bmatrix} \quad (28)$$

where \mathbf{H}_{nk} is the $N_f \times (N_f + L - 1)$ Toeplitz channel matrix formed from \mathbf{h}_{nk} , as in (2).

Suppose user 1 is the desired user. The output of the DFE is given by

$$z_1^{(m)}(i) = \left(\mathbf{f}_1^{(m)}\right)^\dagger \mathbf{y}(i) - \left(\mathbf{b}_1^{(m)}\right)^\dagger \hat{\mathbf{s}}_1^{(m)}(i) \quad (29)$$

where $\hat{\mathbf{s}}_1^{(m)}(i) = [\hat{s}_1^{(m)}(i + \delta - 1) \cdots \hat{s}_1^{(m)}(i + \delta - N_b)]^T$ is the $N_b \times 1$ vector of soft symbol estimates for user 1. In contrast to the single-user single-antenna model, here the feedforward and feedback filters are directly computed from the soft estimates according to an LS cost function in both the initial and subsequent iterations. This is because the MMSE solution for the filters depends not only on the desired user's channel, but also on the interferers' channels, which are difficult to estimate in the absence of training information from the cochannel users. Also, the total number of channel parameters in this case may be more than the number of filter coefficients.

In the first iteration, the DFE must be a causal filter, hence the first δ elements of $\hat{\mathbf{s}}_1(i)$ cannot contribute to $z_1^{(1)}(i)$, and the first δ elements of $\mathbf{b}_1^{(1)}$ are set to zero. In succeeding iterations, estimates of precursor symbols are available, and only the δ th element of $\mathbf{b}^{(m)}$ is constrained to be zero. To compute the DFE filters, we therefore define (30), shown at the bottom of the page. The full-rank LS filters are then given by (9), where $\mathbf{x}^T(i) = [\mathbf{y}^T(i) (\hat{\mathbf{s}}_1^{(m)}(i))^T]$, and the sum in (9) is over the training midamble in the first iteration, and over the whole packet in subsequent iterations.

Reduced-rank estimates of the feedforward and feedback filters are again obtained by first computing \mathbf{c}_1 from (20), where the symbol estimate $\hat{\mathbf{s}}^{(m)}$ is replaced by $\hat{\mathbf{s}}_1^{(m)}$. The corresponding error $\mathbf{e}_1^{(m)}(i) = \hat{\mathbf{s}}_1^{(m)}(i) - (\mathbf{c}_1^{(m)})^\dagger \mathbf{y}(i)$ is used in (26) to compute the error-estimation filter $\mathbf{b}^{(m)}$ [to make the dimensions of the feedback filter and corresponding input vector

$$\check{\mathbf{s}}_1^{(m)}(i) = \begin{cases} [\hat{s}_1^{(m)}(i-1) \cdots \hat{s}_1^{(m)}(i+\delta-N_b)]^T, & m=1 \\ [\hat{s}_1^{(m)}(i+\delta-1) \cdots \hat{s}_1^{(m)}(i+1) \hat{s}_1^{(m)}(i-1) \cdots \hat{s}_1^{(m)}(i+\delta-N_b)]^T, & m>1 \end{cases} \quad (30)$$

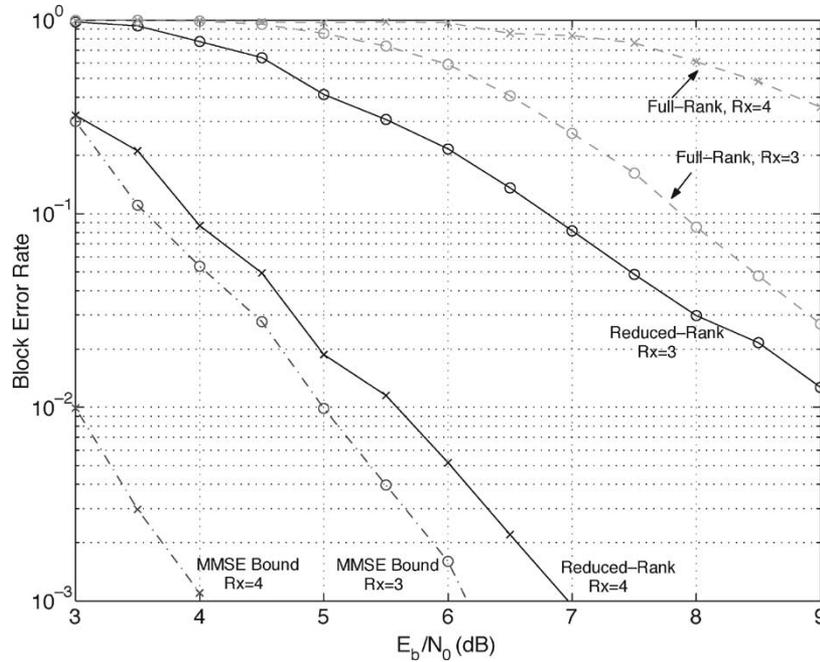


Fig. 7. BLER versus E_b/N_0 for full- and reduced-rank space-time turbo DFEs. There are two cochannel users and multiple receiver antennas.

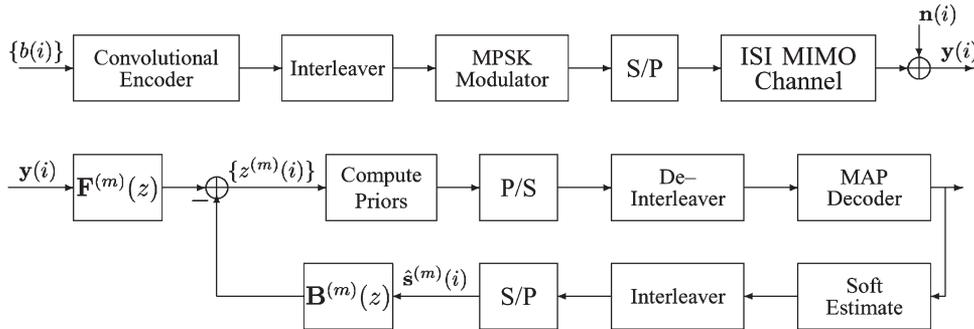


Fig. 8. Block diagram of transmitter and turbo DFE for a single-user multiantenna channel. S/P refers to the conversion of a serial symbol stream to parallel symbol streams across transmit antennas.

consistent, we must again define an error vector $\tilde{\mathbf{e}}_1^{(m)}(i)$, which is analogous to $\tilde{\mathbf{s}}_1^{(m)}$ in (30).

Fig. 7 shows BLER versus E_b/N_0 for a system with two equal-power cochannel users, each with one transmit antenna, and multiple receiver antennas. Each channel has $L = 5$ equal-power taps. The block-fading channel coefficients are complex Gaussian, and are independent across taps and users. The packet format corresponds to an EDGE system. A midamble of training symbols and three tail symbols on each side are inserted into the 8-PSK modulated symbols to form a packet. Namely, the block length (interleaving depth) is 1392 bits, which spans four bursts with 116 symbols each. A training sequence of 26 bits is added to each burst. The feedforward and feedback filters for each antenna have six taps each. Curves corresponding to both adaptive full- and reduced-rank turbo DFEs are shown. The results are averaged over channel realizations. For comparison, we also show the MMSE bound, which corresponds to perfect ISI cancellation for the desired user (since only symbols from the desired user are fed back), and a linear MMSE filter to suppress CCI.

The results show that with three receive antennas, the reduced-rank turbo DFE offers a marginal performance gain relative to the full-rank DFE. With four antennas, the performance gain is quite large. The performance of the full-rank DFE with four antennas is worse than that with three antennas. This can be explained by observing that in going from three to four antennas, the number of feedforward filter coefficients increases from 18 to 24. Since there are only 26 training bits, the estimation problem for the full-rank DFE becomes ill-conditioned with four antennas. The performance of the reduced-rank DFE does not degrade from the increase in the number of filter coefficients associated with additional receiver antennas, hence the reduced-rank algorithm can still exploit available diversity.

B. Single-User MIMO System

Consider now a single-user MIMO system with K transmit antennas. As shown in Fig. 8, we assume that the block of information bits is convolutionally coded, interleaved, modulated,

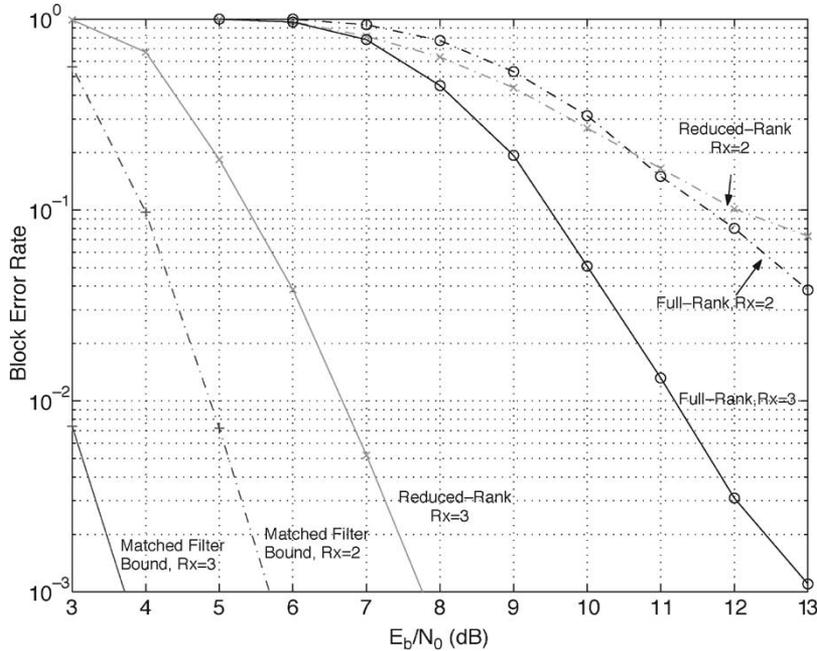


Fig. 9. BLER versus E_b/N_0 for full- and reduced-rank turbo MIMO DFEs. There are two transmitters, and curves are shown for both two and three receive antennas.

and then split into equal-length subblocks, which are assigned to each of the transmit antennas. The midamble of training symbols and tail symbols are inserted into subblocks separately to form one packet for each antenna. The transmit antennas are analogous to the K cochannel users in the preceding section. In contrast with the preceding situation, the receiver now attempts to demodulate all K transmitted bit streams, so that the MIMO feedback filter in the DFE attempts to cancel the corresponding CCI, in addition to the ISI. The system model in (1) still applies; however, we must compute feedforward and feedback filters for each transmit antenna. The K DFE outputs must be buffered and merged to reconstruct the original transmitted block. This single symbol stream is then deinterleaved and decoded to obtain the original packet.

Let $\mathbf{f}_k^{(m)}$ and $\mathbf{b}_k^{(m)}$ denote the feedforward and feedback filters corresponding to transmit antenna k at iteration m . For the first iteration, only the soft estimates $\hat{\mathbf{s}}_k^{(1)}$ are fed back through $\mathbf{b}_k^{(1)}$ to cancel postcursor ISI. That is, the DFE does not attempt to cancel CCI in the first iteration, since the symbol estimates from other antennas are error prone, and would cause substantial error propagation. In subsequent iterations, the vector $\hat{\mathbf{s}}^{(m)}(i) = [\hat{s}_1^{(m)}(i), \dots, \hat{s}_K^{(m)}(i)]$ is fed back, which contains all estimates of symbols in $s(i)$ that interfere with $\mathbf{s}_k(i)$. That is, the feedback filter attempts to cancel pre- and postcursor ISI and all CCI. Each feedback filter has KN_b coefficients. Hence, for $m = 1$, $z_k^{(1)}(i)$ is given by (29), and for $m > 1$, $z_k^{(m)}(i) = (\mathbf{f}_k^{(m)})^\dagger \mathbf{y}(i) - (\mathbf{b}_k^{(m)})^\dagger \hat{\mathbf{s}}^{(m)}(i)$.

We can again estimate a reduced-rank DFE for each transmitter antenna $k = 1, \dots, K$. The output error corresponding to transmitter k is

$$\mathbf{e}_k^{(m)}(i) = \hat{\mathbf{s}}_k^{(m)}(i) - (\mathbf{c}_k^{(m)})^\dagger \mathbf{y}(i) \quad (31)$$

where \mathbf{c}_k is the linear equalizer for transmitter k , computed according to (20). For the first iteration, the error-estimation filter for each k is computed exactly the same way as described in Section III-A. In subsequent iterations, we compute an LS estimate of the error $\mathbf{e}_k^{(m)}(i)$ given all available output errors corresponding to all transmit antennas. That is, we define

$$\mathbf{e}^{(m)}(i) = \left[(\mathbf{e}_1^{(m)}(i))^T \cdots (\mathbf{e}_K^{(m)}(i))^T \right]^T$$

and

$$\check{\mathbf{e}}_{k,K}^{(m)}(i) = \left[(\mathbf{e}_1^{(m)}(i))^T \cdots (\check{\mathbf{e}}_k^{(m)}(i))^T \cdots (\mathbf{e}_K^{(m)}(i))^T \right]^T$$

where $\check{\mathbf{e}}_k^{(m)}(i)$ is defined in the preceding section for $m \geq 2$ and $k = 1, \dots, K$. The error-estimation filter $\check{\mathbf{b}}_k^{(m)}$ estimates $\mathbf{e}_k^{(m)}(i)$ from $\check{\mathbf{e}}_{k,K}^{(m)}(i)$. Both full- and reduced-rank error-estimation filters can be computed, as previously described. For the numerical results, which follow, the ranks associated with different receive filters, corresponding to different transmit antennas, are chosen to optimize performance, and generally vary across receivers.

Fig. 9 presents BLER versus E_b/N_0 for a single-user MIMO channel with two transmit antennas. Results are shown for two and three receive antennas. Each channel, corresponding to a transmit-receive antenna pair, contains $L = 5$ complex Gaussian taps with power-delay profile $[-2, -1, 0, -1, -2]$ dB. The transmitted packets again follow the EDGE format

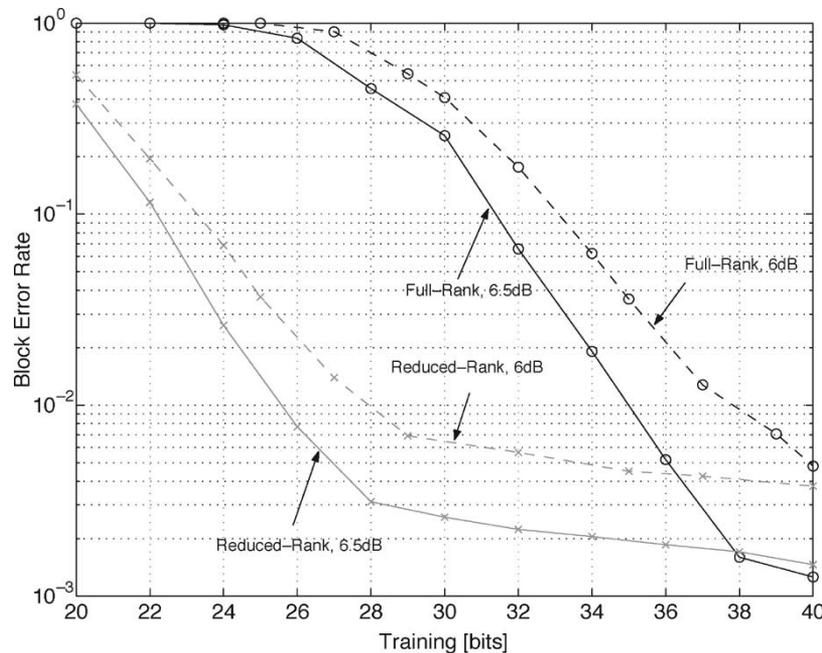


Fig. 10. BLER versus training for full- and reduced-rank turbo MIMO DFEs. There are two transmit and three receive antennas.

previously described. Each feedforward and feedback filter has seven taps. The matched filter bound in this case corresponds to perfect MIMO channel knowledge and perfect feedback for all transmit antennas. With two receive antennas, the full- and reduced-rank DFEs give similar performance, whereas with three receiver antennas, the reduced-rank DFE offers a substantial performance gain. In both cases, the performance improves as the number of receive antennas (R_x) is increased from two to three. Fig. 10 shows BLER versus training interval for the MIMO system with three receive antennas. Results are shown for $E_b/N_0 = 6$ and 6.5 dB. The transmitted packet still corresponds to an EDGE system, except that we vary the length of the midamble. For the cases shown, the reduced-rank DFE requires significantly less training than the full-rank DFE.

IV. CONCLUSION

An adaptive turbo DFE has been presented in which the filter coefficients are computed directly from the input samples and the soft decisions at the output of the MAP decoder. The filters are selected to minimize a LS cost function. Numerical results show that this method gives a substantial performance improvement relative to both a turbo DFE computed from an exact channel estimate assuming perfect feedback, and an adaptive turbo DFE, which uses the LMS algorithm [12]. Reduced-rank estimates of the DFE filters have also been presented, which require less training overhead, relative to full-rank estimates.

The adaptive turbo DFE was extended to MIMO channels. Two situations were considered: 1) the feedback filter only attempts to cancel ISI for each symbol stream and 2) the feedback filter attempts to cancel all ISI and CCI. Reduced-rank estimation algorithms were also presented in this context. As the number of antennas increases with limited training, the

performance becomes limited by estimation error. In that case, the reduced-rank DFE can perform substantially better than the full-rank DFE. Of course, this performance gap diminishes as the training interval increases. We therefore conclude that with limited training, the reduced-rank DFE can more effectively exploit available diversity.

REFERENCES

- [1] S. Kaiser and J. Hagenauer, "Multi-carrier CDMA with iterative decoding and soft-interference cancellation," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, Phoenix, AZ, Nov. 1997, vol. 1, pp. 6–10.
- [2] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Jun. 1995.
- [3] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Int. Symp. Turbo Codes*, Brest, France, Sep. 1997, pp. 96–102.
- [4] S. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," in *Proc. Int. Conf. Communications (ICC)*, New Orleans, LA, May 2000, vol. 3, pp. 1238–1242.
- [5] I. Fijalkow, A. Roumy, S. Rongier, D. Pirez, and P. Vila, "Improved interference cancellation for turbo equalization," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Istanbul, Turkey, May 2000, vol. 1, pp. 416–419.
- [6] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
- [7] A. Singer, J. Nelson, and R. Koetter, "Linear iterative turbo equalization (LITE) for dual channels," in *Proc. 33rd Asilomar Conf. Signals, Systems and Computers*, Monterey, CA, Oct. 1999, pp. 1670–1674.
- [8] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.
- [9] R. Otnes and M. Tuchler, "Soft iterative channel estimation for turbo equalization: Comparison of channel estimation algorithms," in *Proc. IEEE Int. Conf. Communication Systems*, Singapore, Nov. 2002, pp. 72–76.
- [10] G. Bauch and V. Franz, "A comparison of soft-in/soft-out algorithms for turbo-detection," in *Proc. IEEE Int. Conf. Telecommunications*, Porto-Carras, Greece, Jun. 1998, pp. 259–263.

- [11] M. L. Honig, G. Woodward, and Y. Sun, "Adaptive iterative multiuser decision feedback detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 477–485, Mar. 2004.
- [12] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: Adaptive equalization and channel decoding jointly optimized," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1744–1752, Sep. 2001.
- [13] X. Wang and H. V. Poor, "Space-time multiuser detection in multipath CDMA channels," *IEEE Trans. Signal Process.*, vol. 9, no. 47, pp. 2356–2374, Sep. 1999.
- [14] S. L. Ariyavisitakul, J. H. Winters, and N. R. Sollenberger, "Joint equalization and interference suppression for high data rate wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1214–1220, Jul. 2000.
- [15] S. L. Ariyavisitakul, J. H. Winters, and I. Lee, "Optimum space-time processors with dispersive interference: Unified analysis and required filter span," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1073–1083, Jul. 1999.
- [16] G. Bauch and N. Al-Dhahir, "Reduced-complexity space-time turbo equalization for frequency-selective MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 819–828, Oct. 2002.
- [17] T. Abe and T. Matsumoto, "Space-time turbo equalization in frequency selective MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 3, pp. 469–475, May 2003.
- [18] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.
- [19] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.
- [20] G. Died, C. Mensing, W. Utschick, J. A. Nossek, and M. D. Zoltowski, "Multi-stage MMSE decision feedback equalization for EDGE," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Hong Kong, Apr. 2003, vol. 4, pp. 509–512.
- [21] European Telecommunications Standards Institute (ETSI), *Digital Cellular Telecommunications System (phase2+): Multiplexing and Multiple Access on the Radio Path*. GSM 5.02, Version 8.5.0, Release 1999.
- [22] A. Gersho and T. L. Lim, "Adaptive cancellation of intersymbol interference for data transmission," *Bell Syst. Tech. J.*, vol. 60, no. 11, pp. 1997–2021, Nov. 1981.
- [23] G. Woodward, R. Ratasuk, M. L. Honig, and P. B. Rapajic, "Minimum mean squared error multiuser decision feedback detection for DS-SS-CDMA," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2104–2112, Dec. 2002.
- [24] C. Luschi *et al.*, "Advanced signal processing algorithms for energy efficient wireless communications," *Proc. IEEE*, vol. 88, no. 10, pp. 1633–1649, Oct. 2000.
- [25] D. Raphaeli and Y. Zarai, "Combined turbo equalization and turbo decoding," *IEEE Commun. Lett.*, vol. 2, no. 4, pp. 107–109, Apr. 1998.
- [26] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1993.
- [27] H. Qian and S. N. Batalama, "Data record-based criteria for the selection of an auxiliary vector estimator of the MMSE/MVDR filter," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1700–1708, Oct. 2003.



Yakun Sun (S'01–M'04) received the B.S. degree from Tsinghua University, Beijing, China, in 1999. He received the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 2001 and 2004, respectively.

He is currently with Motorola, Inc., Libertyville, IL, where he works on physical-layer algorithms, and wireless-system analysis and design. His research interests include multiantenna/multicarrier wireless systems, channel estimation and equalization, iterative receivers, and resource allocation.



Vinayak Tripathi (S'99–M'01) received the B.S. degree from Cornell University, Ithaca, NY, in 1997, and the M.S. degree from the University of Michigan, Ann Arbor, in 1998, both in electrical engineering. He is currently pursuing the Ph.D. degree in financial economics at Princeton University, Princeton, NJ.

He worked as a Research Engineer at the Communication Systems Research Lab, Motorola Labs, Schaumburg, IL, until August 2003.



Michael L. Honig (S'80–M'81–SM'92–F'97) received the B.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively.

He subsequently joined Bell Laboratories, Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983, he joined the Systems Principles Research Division, Bellcore, where he worked on digital subscriber lines and wireless communications. Since the fall of 1994, he has been with Northwestern University, Evanston, IL, where he is a Professor in the Electrical and Computer Engineering Department. He has held Visiting Scholar positions at the U.S. Naval Research Laboratory (San Diego), the University of California, Berkeley, the University of Sydney, and Princeton University. He was a Guest Editor for the *European Transactions on Telecommunications and Wireless Personal Communications*.

Dr. Honig has served as an Editor for the *IEEE TRANSACTIONS ON INFORMATION THEORY* (1998–2000) and the *IEEE TRANSACTIONS ON COMMUNICATIONS* (1990–1995). He has also served as a member of the Digital Signal Processing Technical Committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society (1997–2002). He is a corecipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.