



A New Class of Efficient Block-Iterative Interference Cancellation Techniques for Digital Communication Receivers*

ALBERT M. CHAN AND GREGORY W. WORNELL

*Department of Electrical Engineering and Computer Science, and the Research Laboratory of Electronics,
Massachusetts Institute of Technology, MIT, Rm. 36-677, Cambridge, MA 02139, USA*

Received November 13, 2000; Revised August 24, 2001

Abstract. A new and efficient class of nonlinear receivers is introduced for digital communication systems. These “iterated-decision” receivers use optimized multipass algorithms to successively cancel interference from a block of received data and generate symbol decisions whose reliability increases monotonically with each iteration. Two variants of such receivers are discussed: the iterated-decision equalizer and the iterated-decision multiuser detector. Iterated-decision equalizers, designed to equalize intersymbol interference (ISI) channels, asymptotically achieve the performance of maximum-likelihood sequence detection (MLSD), but only have a computational complexity on the order of a linear equalizer (LE). Even more importantly, unlike the decision-feedback equalizer (DFE), iterated-decision equalizers can be readily used in conjunction with error-control coding. Similarly, iterated-decision multiuser detectors, designed to cancel multiple-access interference (MAI) in typical wireless environments, approach the performance of the optimum multiuser detector in uncoded systems with a computational complexity comparable to a decorrelating detector or a linear minimum mean-square error (MMSE) multiuser detector.

Keywords: equalization, multiuser detection, decision-feedback equalizer, multipass receivers, multistage detectors, iterative decoding, stripping, interference cancellation

1. Introduction

Over the last several decades, a variety of equalization techniques have been proposed for use on intersymbol interference (ISI) channels. Linear equalizers (LE) are attractive from a complexity perspective, but often suffer from excessive noise enhancement. Maximum-likelihood sequence detection (MLSD) [1] is an asymptotically optimum receiver in terms of bit-error rate performance, but its high complexity has invariably precluded its use in practice. Decision-feedback equalizers (DFE) [2] are a widely used compromise, retaining a complexity comparable to the LE, but incurring much less noise enhancement. However, DFEs still

have some serious shortcomings. First, decisions made at the slicer can only be fed back to improve future decisions due to the sequential way in which the receiver processes data. Thus, only postcursor ISI can be subtracted, so even if ideal postcursor ISI cancellation is assumed, the performance of the DFE is still limited by possible residual precursor ISI and noise enhancement. Second, and even more importantly, the sequential structure of the DFE makes it essentially incompatible for use in conjunction with error-control coding (on channels not known at the transmitter, as is the case of interest in this paper). As a result, use of the DFE has been largely restricted to uncoded systems.

In parallel to these developments, a variety of multiuser detectors have been proposed for code-division multiple-access (CDMA) channels over the last decade and a half as solutions to the problem of mitigating multiple-access interference (MAI) [3]. Given the close

*This work has been supported in part by Qualcomm, Inc., the Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0002, and Sanders, a Lockheed-Martin Company.

coupling between the problems of suppressing ISI and MAI, there are, not surprisingly, close relationships between corresponding solutions to these problems. For example, decorrelating detectors and linear minimum mean-square error (MMSE) multiuser detectors—the counterparts of zero-forcing and MMSE linear equalizers—are attractive from a complexity perspective, but suffer from noise enhancement. Optimum maximum-likelihood (ML) multiuser detection, while superior in performance, is not a practical option because of its high complexity.

In this paper, we introduce a class of promising multipass receivers that is a particularly attractive alternative to all these conventional equalizers and detectors. Specifically, in Section 2 we describe the iterated-decision equalizer, and in Section 3 we describe the corresponding iterated-decision multiuser detector. In particular, we show that these new receivers achieve asymptotically optimum performance while requiring surprisingly low complexity.

2. The Iterated-Decision Equalizer

In the discrete-time baseband model of the pulse amplitude modulation (PAM) communication system we consider, the transmitted data is a white M -ary phase-shift keying (PSK) stream of coded or uncoded symbols $x[n]$, each with energy \mathcal{E}_s . The symbols $x[n]$ are corrupted by a convolution with the impulse response of the channel, $a[n]$, and by additive noise, $w[n]$, to produce the received symbols

$$r[n] = \sum_k a[k]x[n-k] + w[n]. \quad (1)$$

The noise $w[n]$ is a zero-mean, complex-valued, circularly symmetric, stationary, white Gaussian noise sequence with variance \mathcal{N}_0 that is independent of $x[n]$. The associated channel frequency response is denoted by

$$A(\omega) = \sum_n a[n]e^{-j\omega n}. \quad (2)$$

As increasingly aggressive data rates are pursued in wideband systems to meet escalating traffic requirements, ISI becomes increasingly severe. Accordingly, in this paper we pay special attention to the performance and properties of the equalizers in this regime. For the purposes of analysis, a convenient severe-ISI channel model we will exploit is one in which

$a[n]$ is a finite impulse response (FIR) filter of length L , where L is large and the taps are mutually independent, zero-mean, complex-valued, circularly symmetric Gaussian random variables with variance σ_a^2 . The channel taps $a[n]$ are also independent of the data $x[n]$ and the noise $w[n]$. It is also worth pointing out that this is also a good channel model for many wireless systems employing transmitter antenna diversity in the form of linear space-time coding [4].

In this section and Section 2.1, we summarize the results of [5], which focuses on the basic theory and fundamental limits of the iterated-decision equalizer when the receiver has accurate knowledge of $a[n]$. In Section 2.2, we develop and analyze adaptive implementations in which the channel coefficients $a[n]$ are not known *a priori*. Examining the fixed and adaptive scenarios separately and comparing their results allows system designers to isolate channel tracking effects from overall equalizer behavior. We emphasize that in both cases, we restrict our attention to transmitters that have no knowledge of the channel, which is the usual case for reasonably rapidly time-varying channels.

The iterated-decision equalizer we now develop processes the received data in a block-iterative fashion. Specifically, during each iteration or “pass,” a linear filter is applied to a block of received data, and tentative decisions made in the previous iteration are then used to construct and subtract out an estimate of the ISI. The resulting ISI-reduced data is then passed on to a slicer, which makes a new set of tentative decisions. With each successive iteration, increasingly refined hard decisions are generated using this strategy.

The detailed structure of the iterated-decision equalizer is depicted in Fig. 1. The parameters of all systems and signals associated with the l th pass are denoted using the superscript l . On the l th pass of the equalizer where $l = 1, 2, 3, \dots$, the received data $r[n]$ is first processed by a linear filter $b^l[n]$, producing the sequence

$$\tilde{r}^l[n] = \sum_k b^l[k]r[n-k]. \quad (3)$$

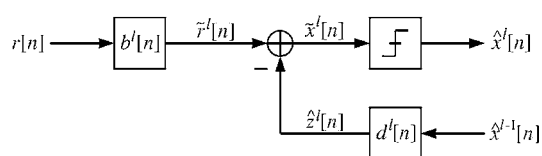


Figure 1. Iterated-decision equalizer structure.

Next, an appropriately constructed estimate $\hat{z}^l[n]$ of the ISI is subtracted from $\tilde{r}^l[n]$ to produce $\tilde{x}^l[n]$, i.e.,

$$\tilde{x}^l[n] = \tilde{r}^l[n] - \hat{z}^l[n] \quad (4)$$

where

$$\hat{z}^l[n] = \sum_k d^l[k] \hat{x}^{l-1}[n-k]. \quad (5)$$

(In subsequent analysis, we will show that $\hat{x}^0[n]$ is never required for the first iteration, so the sequence may remain undefined.) Since $\hat{z}^l[n]$ is intended to be

$a[n]$. A formal statement of this result and its associated proof is developed in [5].

The second-order model (8) turns out to be a useful one for analyzing and optimizing the performance of the iterated-decision equalizer. In particular, it can be used to obtain a surprisingly accurate estimate of the symbol error rate for M -ary PSK even though we ignore the higher-order statistical dependencies.

The first step in developing these results is to observe that (8) implies that the signal-to-interference+noise ratio (SINR) at the slicer input during each pass can be written, using (9), as

$$\gamma^l(B^l, D^l) = \frac{\mathcal{E}_s |E[AB^l]|^2}{\mathcal{N}_0 E[|B^l|^2] + \mathcal{E}_s (1 - (\rho^{l-1})^2) \text{var}[AB^l] + \mathcal{E}_s E[|D^l - \rho^{l-1}(AB^l - E[AB^l])|^2]} \quad (10)$$

some kind of ISI estimate, we restrict attention to the case in which

$$d^l[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} D^l(\omega) d\omega = 0. \quad (6)$$

Finally, the slicer then generates the hard decisions $\hat{x}^l[n]$ from $\tilde{x}^l[n]$ using a minimum-distance rule.

The composite system consisting of the channel in cascade with l iterations of the multipass equalizer can be conveniently characterized. In particular, when $x[n]$ and $\hat{x}^{l-1}[n]$ are sequences of zero-mean uncorrelated symbols with energy \mathcal{E}_s , such that their normalized correlation is of the form

$$\frac{E[x^*[n] \cdot \hat{x}^{l-1}[k]]}{\mathcal{E}_s} \approx \rho^{l-1} \delta[n-k], \quad (7)$$

the slicer input after l iterations can be expressed as¹

$$\tilde{x}^l[n] \approx E[AB^l]x[n] + v^l[n], \quad (8)$$

where $A(\omega)$ and $B^l(\omega)$ are the frequency responses of $a[n]$ and $b^l[n]$ respectively, where $v^l[n]$ is a complex-valued, marginally Gaussian, zero-mean white noise sequence, uncorrelated with the input symbol stream $x[n]$, and having variance

$$\begin{aligned} \text{var } v^l[n] &= \mathcal{N}_0 E[|B^l|^2] + \mathcal{E}_s (1 - (\rho^{l-1})^2) \text{var}[AB^l] \\ &\quad + \mathcal{E}_s E[|D^l - \rho^{l-1}(AB^l - E[AB^l])|^2], \end{aligned} \quad (9)$$

and where the accuracy of the approximation in (8) increases with the length L of the impulse response

and that the probability of symbol error at the l th iteration may be approximated by the high signal-to-noise ratio (SNR) formula for the M -ary PSK symbol error rate of a symbol-by-symbol threshold detector for additive white Gaussian noise (AWGN) channels, given by [6]

$$\Pr(\epsilon^l) = 2\mathcal{Q}\left(\sin\left(\frac{\pi}{M}\right)\sqrt{2\gamma^l}\right), \quad (11)$$

where

$$\mathcal{Q}(v) = \frac{1}{\sqrt{2\pi}} \int_v^{\infty} e^{-t^2/2} dt. \quad (12)$$

For the special case of QPSK ($M=4$), the extension of (11) to arbitrary SNRs is given by [6]

$$\Pr(\epsilon^l) = \mathcal{Q}(\sqrt{\gamma^l})[2 - \mathcal{Q}(\sqrt{\gamma^l})]. \quad (13)$$

Note that this equivalent channel model effectively suggests that, in the absence of coding, we replace the computationally expensive Viterbi-algorithm-based MLSD with a simple symbol-by-symbol detector, as if the channel were an AWGN channel.²

Since the probability of error given by (11) or (13) is a monotonically decreasing function of SINR, a natural equalizer design strategy involves maximizing the SINR over all $B^l(\omega)$ and $D^l(\omega)$. Thus, the optimal filters are [5]

$$B^l(\omega) \propto \frac{A^*(\omega)}{\mathcal{N}_0 + \mathcal{E}_s (1 - (\rho^{l-1})^2) |A(\omega)|^2} \quad (14)$$

$$D^l(\omega) = \rho^{l-1} (A(\omega)B^l(\omega) - E[AB^l]). \quad (15)$$

The result for $D^l(\omega)$ is intuitively satisfying. If $\hat{x}^{l-1}[n] = x[n]$ so that $\rho^{l-1} = 1$, then the output of $D^l(\omega)$ exactly reproduces the ISI component of $\tilde{r}^l[n]$. More generally, ρ^{l-1} describes our confidence in the quality of the estimate $\hat{x}^{l-1}[n]$. If $\hat{x}^{l-1}[n]$ is a poor estimate of $x[n]$, then ρ^{l-1} will in turn be low, and consequently a smaller weighting is applied to the ISI estimate that is to be subtracted from $\tilde{r}^l[n]$. On the other hand, if $\hat{x}^{l-1}[n]$ is an excellent estimate of $x[n]$, then $\rho^{l-1} \approx 1$, and nearly all of the ISI is subtracted from $\tilde{r}^l[n]$. Thus, while the strictly causal feedback filter of the DFE subtracts out only postcursor ISI, the noncausal nature of the filter $d^l[n]$ allows the iterated-decision equalizer to cancel both precursor and postcursor ISI. Note also that the center tap of $d^l[n]$ is indeed asymptotically zero, as stipulated by (6).

Some comments can also be made about the special case when $l = 1$. During the first pass, the feedback branch is not used because $\rho^0 = 0$, so the sequence $\hat{x}^0[n]$ does not need to be defined. Moreover, the filter $B^1(\omega)$ takes the form

$$B^1(\omega) \propto \frac{A^*(\omega)}{\mathcal{N}_0 + \mathcal{E}_s |A(\omega)|^2}, \quad (16)$$

which is the minimum mean-square error linear equalizer (MMSE-LE). Thus the performance of the iterated-decision equalizer, after just one iteration, is identical to the performance of the MMSE-LE. In Section 2.1, we show that the equalizer, when using multiple iterations, performs significantly better than both the MMSE-LE and the minimum mean-square error decision-feedback equalizer (MMSE-DFE).

We now proceed to simplify the SINR expression that characterizes the resulting performance. With the optimum $B^l(\omega)$ and $D^l(\omega)$, the SINR from (10) becomes [5]

$$\gamma^l = \left(\frac{1}{E\left[\frac{1}{1+\alpha^l}\right]} - 1 \right) \cdot \frac{1}{1 - (\rho^{l-1})^2}, \quad (17)$$

where

$$\alpha^l(\omega) = \frac{\mathcal{E}_s (1 - (\rho^{l-1})^2) |A(\omega)|^2}{\mathcal{N}_0}. \quad (18)$$

Now since our channel model implies that $A(\omega)$ is a complex-valued, circularly symmetric Gaussian random variable with zero mean and variance $L\sigma_a^2$, it follows that $\alpha^l(\omega)$ is exponentially distributed with mean

$$\frac{1}{\xi^l} = \frac{(1 - (\rho^{l-1})^2)}{\zeta}, \quad (19)$$

where

$$\frac{1}{\zeta} = \frac{\mathcal{E}_s L \sigma_a^2}{\mathcal{N}_0} \quad (20)$$

is the expected SNR at which the transmission is received. Evaluating the expectation in (17), our simplified SINR expression is [5]

$$\gamma^l = \left(\frac{1}{\xi^l e^{\xi^l} E_1(\xi^l)} - 1 \right) \cdot \frac{1}{1 - (\rho^{l-1})^2}, \quad (21)$$

where

$$E_1(s) = \int_s^\infty \frac{e^{-t}}{t} dt \quad (22)$$

is the exponential integral.

Equation (21) can, in turn, be used in the following convenient iterative algorithm for determining the set of correlation coefficients ρ^l to be used at each iteration, and simultaneously predicting the associated sequence of symbol error probabilities:

1. Set $\rho^0 = 0$ and let $l = 1$.
2. Compute the SINR γ^l at the slicer input on the l th decoding pass from ρ^{l-1} via (21), (19), and (20). [It is worth pointing out that for shorter ISI channels, we can alternatively (and in some cases more accurately) compute γ^l from ρ^{l-1} via (17) and (18), where the expectation is replaced by a frequency average.]
3. Compute the symbol error probability $\Pr(\epsilon^l)$ at the slicer output from γ^l via (11).
4. Compute the normalized correlation coefficient ρ^l between the symbols $x[n]$ and the decisions $\hat{x}^l[n]$ generated at the slicer via the approximation [8]

$$\rho^l \approx 1 - 2 \sin^2 \left(\frac{\pi}{M} \right) \Pr(\epsilon^l). \quad (23)$$

5. Increment l and go to step 2.

In the special case of QPSK, it can be shown that the algorithm can be streamlined by eliminating Step 3 and replacing the approximation (23) with the exact formula

$$\rho^l = 1 - 2Q(\sqrt{\gamma^l}). \quad (24)$$

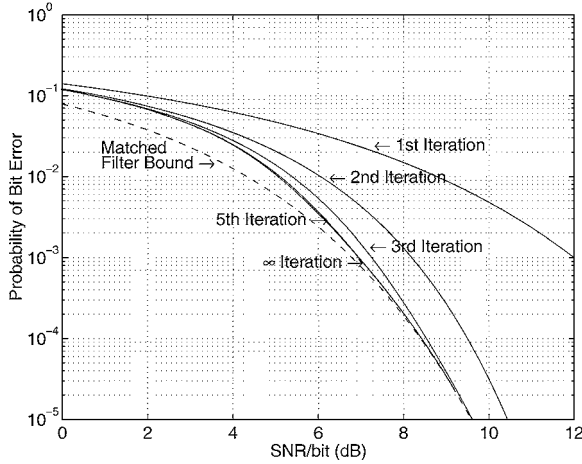


Figure 2. Theoretical iterated-decision equalizer performance as a function of SNR per bit. The successively lower solid curves depict the QPSK bit-error rate as a function of SNR per bit for 1, 2, 3, 5, and ∞ decoding iterations.

2.1. Performance

In Fig. 2, bit-error rate is plotted as a function of SNR per bit for 1, 2, 3, 5, and an infinite number of iterations. We observe that steady-state performance is approximately achieved with comparatively few iterations, after which additional iterations provide only negligibly small gains in performance. It is significant that few passes are required to converge to typical target bit-error rates, since the amount of computation is directly proportional to the number of passes required; we emphasize that the complexity of a single pass of the iterated-decision equalizer is comparable to that of the DFE or the LE.

Figure 3 compares the theoretical performance of the iterated-decision equalizer when the number of channel taps $L \rightarrow \infty$ with experimentally obtained results when $L = 256$. The experimental results are indeed consistent with theoretical predictions, especially at high SNR ($\zeta \rightarrow 0$) where it has been theoretically shown [5] that the equalizer achieves the matched filter bound, i.e., $\gamma \rightarrow 1/\zeta$.

For comparison, in Fig. 3 we also plot the theoretical error rates of the ideal MMSE-DFE, the MMSE-LE, and the zero-forcing linear equalizer (ZF-LE), based on their asymptotic SINRs in the large ISI limit [5]:

$$\gamma_{\text{MMSE-DFE}} = \exp\{e^\zeta E_1(\zeta)\} - 1 \quad (25)$$

$$\gamma_{\text{MMSE-LE}} = \frac{1}{\zeta e^\zeta E_1(\zeta)} - 1 \quad (26)$$

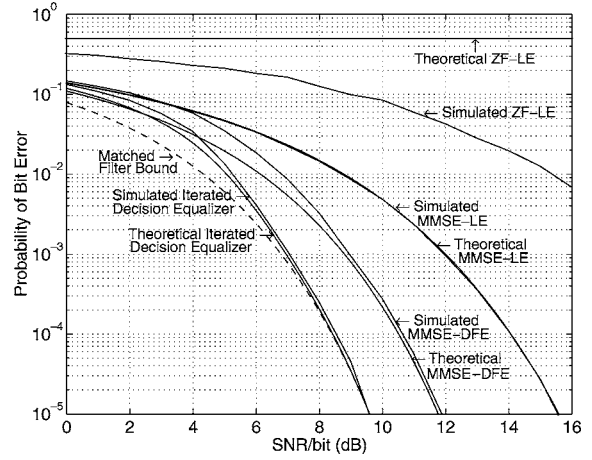


Figure 3. Theoretical ($L \rightarrow \infty$) and experimentally observed ($L = 256$) performance for various equalizers. The solid curves depict QPSK bit-error rates for the iterated-decision equalizer, MMSE-DFE, MMSE-LE, and ZF-LE as a function of SNR per bit.

$$\gamma_{\text{ZF-LE}} = 0. \quad (27)$$

We can readily see that at moderate to high SNR, the iterated-decision equalizer requires significantly less transmit power than any of the other equalizers to achieve the same probability of error. Specifically, at high SNR ($\zeta \rightarrow 0$), we have from [5] that $\gamma_{\text{MMSE-DFE}} \rightarrow 1/\zeta e^{\Gamma_0}$ and $\gamma_{\text{MMSE-LE}} \rightarrow 1/[\zeta(-\Gamma_0 - \ln \zeta)] - 1$, where $\Gamma_0 = 0.57721 \dots$ denotes Euler's constant. Thus, the MMSE-DFE theoretically requires e^{Γ_0} times or $10\Gamma_0 \log e \approx 2.507$ dB more transmit power to achieve the same probability of error as the iterated-decision equalizer. Moreover, as $\zeta \rightarrow 0$, the MMSE-LE requires increasingly more transmit power than the iterated-decision equalizer to achieve the same probability of error. The ZF-LE is even worse: $\gamma_{\text{ZF-LE}} = 0$ for all ζ , which is expected since the zeros of the random channel converge uniformly on the unit circle in the long ISI limit [9]. These results emphasize the strong suboptimality of conventional equalizers.

The performance of the iterated-decision equalizer for channels whose taps are few in number, non-Gaussian, and/or correlated is discussed in [5, 10].

2.2. Adaptive Implementations

We now develop an adaptive implementation of the iterated-decision equalizer, in which optimal FIR filter coefficients are selected automatically (from the received data) without explicit knowledge of the channel

characteristics. We focus on the single channel case; multichannel generalizations follow in a straightforward manner, as developed in [10].

The iterated-decision equalizer is designed to process received data in a block-iterative fashion, so it is ideally suited for packet communication in which the packet size is chosen small enough that the channel encountered by each packet appears linear time-invariant. As is typically the case with other adaptive equalizers, the adaptive iterated-decision equalizer makes use of training symbols sent along in the packet with the data symbols. Suppose that a block of white M -ary PSK symbols $x[n]$ for $n=0, 1, \dots, N-1$ is transmitted; some of the symbols (not necessarily at the head of the packet) are for training, while the rest are data symbols.

In the adaptive implementation of the iterated-decision equalizer, the filters $b^l[n]$ and $d^l[n]$ for the l th iteration are finite-length filters. Specifically, $b^l[n]$ has J_1 strictly anticausal taps and J_2 strictly causal taps plus a center tap, while $d^l[n]$ has K_1 strictly anticausal taps and K_2 strictly causal taps with no center tap.

Before the first pass ($l=1$), we need to initialize the hard decisions $\hat{x}^0[n]$. Since the locations and values of the training symbols in $x[n]$ are known at the receiver, we set $\hat{x}^0[n] = x[n]$ for the n corresponding to those locations. For all the other n between 0 and $N-1$ inclusive, we set $\hat{x}^0[n]$ to be a “neutral” value—for white PSK symbols, this value should be zero.

On the l th pass of the equalizer where $l=1, 2, 3, \dots$, the slicer input $\tilde{x}^l[n]$ can be expressed as³

$$\tilde{x}^l[n] = \mathbf{c}^l \mathbf{q}^l[n] \quad (28)$$

where

$$\mathbf{c}^l = \begin{bmatrix} (b^l[-J_1])^\dagger \\ \vdots \\ (b^l[J_2])^\dagger \\ (-d^l[-K_1])^\dagger \\ \vdots \\ (-d^l[-1])^\dagger \\ (-d^l[1])^\dagger \\ \vdots \\ (-d^l[K_2])^\dagger \end{bmatrix} \quad \mathbf{q}^l[n] = \begin{bmatrix} r[n+J_1] \\ \vdots \\ r[n-J_2] \\ \hat{x}^{l-1}[n+K_1] \\ \vdots \\ \hat{x}^{l-1}[n+1] \\ \hat{x}^{l-1}[n-1] \\ \vdots \\ \hat{x}^{l-1}[n-K_2] \end{bmatrix} \quad (29)$$

Using a minimum-distance rule, the slicer then generates the hard decisions $\hat{x}^l[n]$ from $\tilde{x}^l[n]$ for all n between 0 and $N-1$ inclusive, except for those n corres-

ponding to the locations of training symbols in $x[n]$. For those n , we set $\hat{x}^l[n] = x[n]$.

In the l th iteration, there are two sets of data available to the receiver: $r[n]$ and $\hat{x}^{l-1}[n]$, $n=0, 1, \dots, N-1$. If we assume that $x[n] \approx \hat{x}^{l-1}[n]$ for the purposes of determining the optimal filters (as is similarly done in the adaptive DFE in decision-directed mode), then it is reasonable to choose $b^l[n]$ and $d^l[n]$ so as to minimize the sum of error squares:

$$\mathcal{E}(\mathbf{c}^l) = \sum_{n=-\infty}^{\infty} |\hat{x}^{l-1}[n] - \mathbf{c}^l \mathbf{q}^l[n]|^2. \quad (30)$$

Since this is a linear least-squares estimation problem, the optimum \mathbf{c}^l is [11]

$$\mathbf{c}_{\text{opt}}^l = [\mathbf{\Phi}^l]^{-1} \mathbf{u}^l, \quad (31)$$

where $\mathbf{\Phi}^l = \sum_{n=-\infty}^{\infty} \mathbf{q}^l[n] \mathbf{q}^{l\dagger}[n]$ and $\mathbf{u}^l = \sum_{n=-\infty}^{\infty} \hat{x}^{l-1*}[n] \mathbf{q}^l[n]$. The resulting equalizer lends itself readily to practical implementation, even for large filter lengths. In particular, the matrix $\mathbf{\Phi}^l$ can be efficiently computed using correlation functions involving $r[n]$ and $\hat{x}^{l-1}[n]$, and $[\mathbf{\Phi}^l]^{-1}$ can be efficiently computed using formulas for the inversion of a partitioned matrix [12].

We now turn to a couple of implementation issues. First, we would ideally like our finite-length adaptive filters to approximate (14) and (15), which are infinite length. The optimal $b^l[n]$ in (14) includes a filter matched to $a[n]$, and the optimal $d^l[n]$ in (15) includes a cascade of $a[n]$ and the corresponding matched filter, suggesting that a reasonable rule of thumb is to select $J_1 = J_2 = K_1 = K_2 = L$. Second, the block-iterative nature of the equalizer allows the training symbols to be located anywhere in the packet. Since—in contrast to the DFE—the locations do not appear to affect equalizer performance, we arbitrarily choose to uniformly space the training symbols within each packet.

In Fig. 4, we plot the bit-error rate of the adaptive iterated-decision equalizer as a function of the number of iterations, for varying amounts of training data. The graph strongly suggests that there is a threshold for the number of training symbols, below which the adaptive equalizer performs poorly and above which the bit-error rate consistently converges to approximately the same steady-state value regardless of the exact number of training symbols. The excess training data is still important though, since the bit-error rate converges quicker with more training data.

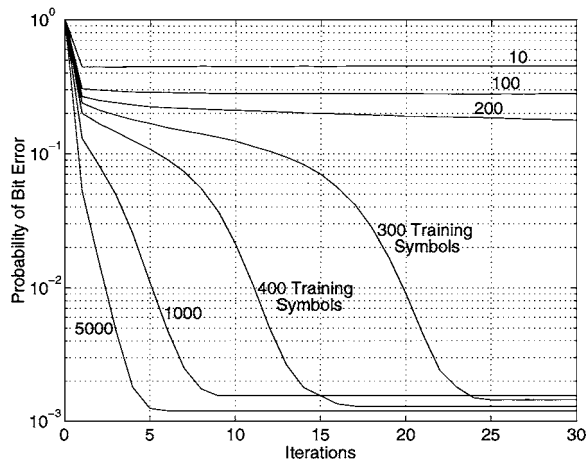


Figure 4. Experimentally observed QPSK bit-error rate for the adaptive iterated-decision equalizer as a function of the number of decoding iterations and the number of training symbols transmitted with each block of 10000 data symbols at an SNR per bit of 7 dB. The 100-tap channels were equalized using 201 feedforward taps and 200 feedback taps.

We next examine the probability of bit error as a function of SNR for varying amounts of training data. From Fig. 5 we see that, as expected, performance improves as the amount of training data is increased. Moreover, only a modest amount of training symbols is required at high SNR for the adaptive equalizer to

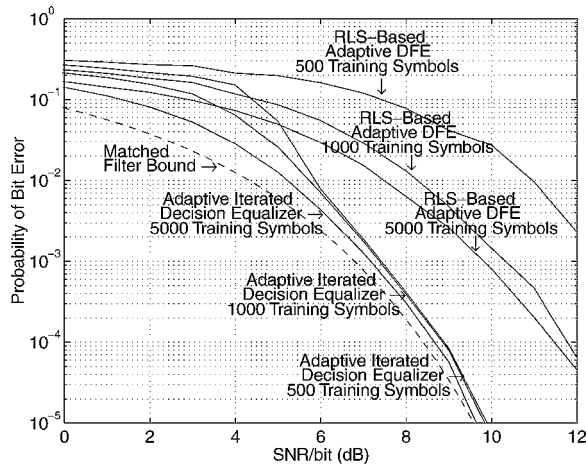


Figure 5. Experimentally observed QPSK bit-error rate for the adaptive iterated-decision equalizer and the RLS-based adaptive DFE (with forgetting factor $\lambda = 1$) as a function of SNR per bit. Blocks of 10000 data symbols were transmitted through 128-tap channels, which were equalized using 257 feedforward taps and 256 noncausal feedback taps in the case of the iterated-decision equalizer, and using 257 feedforward taps and 128 strictly causal feedback taps in the case of the DFE.

perform as if the channel were exactly known at the receiver.

For comparison purposes, we also plot in Fig. 5 the performance of the recursive least squares (RLS) based implementation of the adaptive DFE [11]. The DFE performs significantly worse than the iterated-decision equalizer for comparable amounts of training data. Indeed, the high SNR gap is even larger than the 2.507 dB determined for the nonadaptive case. This is because, as Figs. 3 and 5 show, the performance of the adaptive DFE is not accurately predicted by the nonadaptive MMSE-DFE, even in the long ISI limit. It is also worth stressing that the RLS-based adaptive DFE is much more computationally expensive than the adaptive iterated-decision equalizer because the RLS-based DFE requires the multiplication of large matrices for *each* transmitted symbol, whereas the iterated-decision equalizer essentially requires the computation of one large matrix inverse per iteration for *all* the symbols in the packets, with the number of required iterations being typically small.

2.3. Coded Implementations

For ideal bandlimited AWGN channels, powerful coding schemes such as trellis-coded modulation with maximum likelihood (ML) decoding can improve the performance over uncoded PAM so that channel capacity is approached.

On the other hand, for bandlimited channels with strong frequency-dependent distortion, coding must be combined with equalization techniques. While the MMSE-DFE has certain attractive characteristics in the context of coded systems [13, 14], in many practical settings it is difficult to use effectively. In particular, in typical implementations the MMSE-DFE cancels postcursor ISI by using delay-free symbol decisions, which in a coded system are often highly unreliable compared to ML decisions, and performance is often poor as a result. From this perspective, the iterated-decision equalizer, which avoids this problem, is a compelling alternative to the MMSE-DFE in coded systems.

The structure of a communication system that combines the iterated-decision equalizer with coding is shown in Fig. 6. Although the sequence $x[n]$ is first encoded before it is transmitted, the approximation in (8) is still valid because typical trellis codes and random codes generally produce white symbol streams [7]. What makes the iterated-decision equalizer an

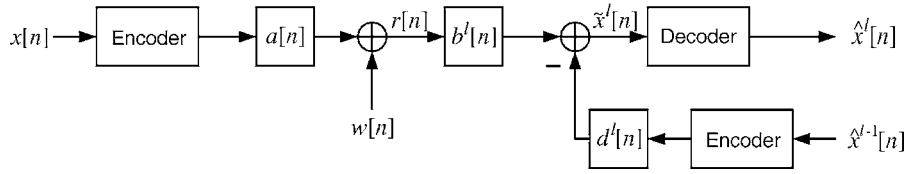


Figure 6. Structure of a communication system that combines iterated-decision equalization with channel coding.

attractive choice when coding schemes are involved is that the structure of the equalizer allows equalization and coding to be largely separable issues. One of the main differences now in the iterated-decision equalizer is that the symbol-by-symbol slicer has been replaced by a soft-decision ML decoder; the other is that the batch of decisions must be re-encoded before being processed by the filter $d^l[n]$. For shorter ISI channels, performance of the system may be improved by inserting an interleaver after each encoder to reduce correlation between adjacent symbols, and by inserting a corresponding deinterleaver before the decoder to reduce the correlation of the residual ISI and noise. Among a variety of interesting issues to explore is the relationship between the structure and performance of such coded systems and those developed in [15].

3. The Iterated-Decision Multiuser Detector

We now develop the counterpart of the iterated-decision equalizer for the multiuser detection problem. As we will see, the resulting detectors are structurally similar to multistage detectors [16] in that they both generate tentative decisions for all users at each iteration and subsequently use these to cancel MAI at the next iteration. However, unlike those original multistage detectors, the new detectors developed in this section explicitly take into account the reliability of tentative

decisions and are optimized to maximize the signal-to-interference + noise (SINR) ratio at each iteration.

For the purposes of illustration (and to simplify exposition), we consider a P -user discrete-time synchronous channel model, where the i th user modulates an M -ary PSK symbol x_i onto a randomly generated signature sequence $\mathbf{h}_i = [h_i[1], h_i[2], \dots, h_i[Q]]^T$ of length Q assigned to that user, where the taps of the sequence are mutually independent, zero-mean, complex-valued, circularly symmetric Gaussian random variables with variance $1/Q$. The received signal is

$$\mathbf{r} = \mathbf{H}\mathbf{A}\mathbf{x} + \mathbf{w}, \quad (32)$$

where $\mathbf{H} = [\mathbf{h}_1 | \dots | \mathbf{h}_P]$ is the $Q \times P$ matrix of signatures, $\mathbf{A} = \text{diag}\{A_1, \dots, A_P\}$ is the $P \times P$ diagonal matrix of received amplitudes, $\mathbf{x} = [x_1, x_2, \dots, x_P]^T$ is the $P \times 1$ vector of data symbols, and \mathbf{w} is a Q -dimensional Gaussian vector with independent zero-mean, complex-valued, circularly symmetric components of variance \mathcal{N}_0 .

The structure of the iterated-decision multiuser detector is depicted in Fig. 7. The parameters of all systems and signals associated with the l th pass are denoted using the superscript l . On the l th pass of the equalizer where $l = 1, 2, 3, \dots$, the received vector \mathbf{r} is first premultiplied by a $P \times Q$ matrix $\mathbf{B}^{l\dagger} = [\mathbf{b}_1^l | \dots | \mathbf{b}_P^l]^\dagger$, producing the $P \times 1$ vector

$$\tilde{\mathbf{r}}^l = \mathbf{B}^{l\dagger} \mathbf{r}. \quad (33)$$

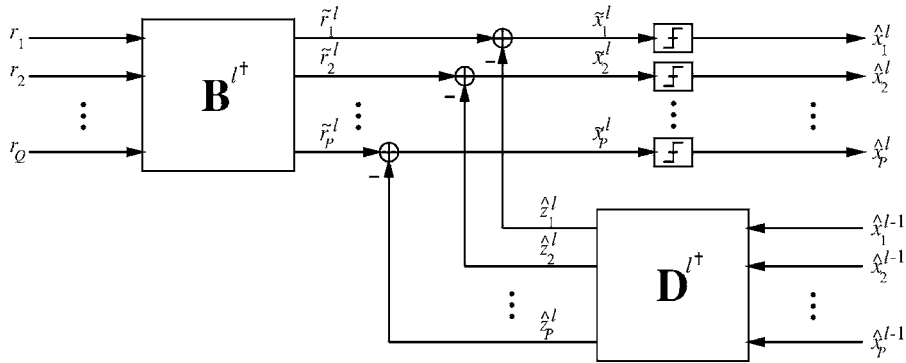


Figure 7. Iterated-decision multiuser detector structure.

Next, an appropriately constructed estimate $\hat{\mathbf{z}}^l$ of the MAI is subtracted from $\tilde{\mathbf{r}}^l$ to produce $\tilde{\mathbf{x}}^l$, i.e.,

$$\tilde{\mathbf{x}}^l = \tilde{\mathbf{r}}^l - \hat{\mathbf{z}}^l \quad (34)$$

where

$$\hat{\mathbf{z}}^l = \mathbf{D}^{l\ddagger} \hat{\mathbf{x}}^{l-1} \quad (35)$$

with $\mathbf{D}^l = [\mathbf{d}_1^l | \dots | \mathbf{d}_P^l]$, a $P \times P$ matrix. (In subsequent analysis, we will show that $\hat{\mathbf{x}}^0$ is never required for the first iteration, so the vector may remain undefined.) Since $\hat{\mathbf{z}}^l$ is intended to be some kind of MAI estimate, we restrict attention to the case in which

$$(\mathbf{D}^l)_{11} = (\mathbf{D}^l)_{22} = \dots = (\mathbf{D}^l)_{PP} = 0. \quad (36)$$

Finally, a bank of slicers then generates the $P \times 1$ vector of hard decisions $\hat{\mathbf{x}}^l$ from $\tilde{\mathbf{x}}^l$ using a minimum-distance rule.

Let us now characterize the composite system consisting of the signatures, the channel, and the multipass multiuser detector.

Let \mathbf{x} and $\hat{\mathbf{x}}^{l-1}$ be vectors of zero-mean uncorrelated symbols, each with energy \mathcal{E}_s ; and let the normalized correlation matrix of the two vectors be expressed in the form

$$\frac{E[\mathbf{x} \cdot \hat{\mathbf{x}}^{l-1\ddagger}]}{\mathcal{E}_s} = \rho^{l-1} \triangleq \text{diag}\{\rho_1^{l-1}, \rho_2^{l-1}, \dots, \rho_P^{l-1}\}, \quad (37)$$

where ρ_i^{l-1} can be interpreted as a measure of the reliability of \hat{x}_i^{l-1} . Moreover, let \mathbf{D}^l satisfy the natural requirement (37). Then, the slicer input \tilde{x}_i^l defined via (35) with (34), (36), and (33) satisfies, for $i = 1, 2, \dots, P$,

$$\tilde{x}_i^l = \mathbf{b}_i^{l\ddagger} \mathbf{h}_i A_i x_i + v_i^l \quad (38)$$

$$\begin{aligned} & + \mathcal{E}_s (\mathbf{d}_i^l - \rho^{l-1\ddagger} (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii}))^\ddagger \\ & \times (\mathbf{d}_i^l - \rho^{l-1\ddagger} (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii}))^\ddagger \end{aligned} \quad (39)$$

where \mathbf{I} is the identity matrix and \mathbf{E}_{ii} is the $P \times P$ matrix with a 1 in the i th row and column as its only nonzero entry.

Equation (39) implies that the SINR at the i th slicer input during each pass can be written as

$$\gamma_i^l(\mathbf{b}_i^l, \mathbf{d}_i^l) = \frac{\mathcal{E}_s |\mathbf{b}_i^{l\ddagger} \mathbf{h}_i A_i|^2}{\text{var } v_i^l} \quad (40)$$

and that the probability of symbol error for the i th user at the l th iteration can be approximated by (11) for the general M -ary PSK case or (13) for the QPSK case. Since the probability of error given by (11) or (13) is a monotonically decreasing function of SINR, a natural detector design strategy involves maximizing the SINR of the i th user over all \mathbf{b}_i^l and \mathbf{d}_i^l .

For a given filter \mathbf{b}_i^l , it is straightforward to find the optimal filter \mathbf{d}_i^l . In particular, note that \mathbf{d}_i^l appears only in a non-negative denominator term of the SINR expression given by (41) and (40), and that term can be made exactly zero by setting

$$\mathbf{d}_i^l = \rho^{l-1\ddagger} (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii})^\ddagger \quad \text{for } i = 1, 2, \dots, P \quad (41)$$

or, equivalently,

$$\begin{aligned} \mathbf{D}^l &= \rho^{l-1\ddagger} [\mathbf{B}^{l\ddagger} \mathbf{H} \mathbf{A} \\ & - \text{diag}\{(\mathbf{B}^{l\ddagger} \mathbf{H} \mathbf{A})_{11}, \dots, (\mathbf{B}^{l\ddagger} \mathbf{H} \mathbf{A})_{PP}\}]^\ddagger. \end{aligned} \quad (42)$$

Using (42) to eliminate \mathbf{d}_i^l , the SINR expression in (41) now simplifies to

$$\gamma_i^l(\mathbf{b}_i^l) = \frac{\mathcal{E}_s |\mathbf{b}_i^{l\ddagger} \mathbf{h}_i A_i|^2}{\mathcal{N}_0 \mathbf{b}_i^{l\ddagger} \mathbf{b}_i^l + \mathcal{E}_s (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii}) (\mathbf{I} - \rho^{l-1} \rho^{l-1\ddagger}) (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii})^\ddagger}. \quad (43)$$

where v_i^l is complex-valued, marginally Gaussian, zero-mean, and uncorrelated with x_i , having variance

$$\begin{aligned} \text{var } v_i^l &= \mathcal{N}_0 \mathbf{b}_i^{l\ddagger} \mathbf{b}_i^l + \mathcal{E}_s (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii}) \\ & \times (\mathbf{I} - \rho^{l-1} \rho^{l-1\ddagger}) (\mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\ddagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii})^\ddagger \end{aligned}$$

This result for \mathbf{d}_i^l is intuitively satisfying. If $\hat{x}_i^{l-1} = x_i$ so that $\rho_i^{l-1} = 1$, then the inner product $\mathbf{d}_i^{l\ddagger} \hat{\mathbf{x}}^{l-1}$ exactly reproduces the MAI component of \tilde{r}_i^l . More generally, ρ_i^{l-1} describes our confidence in the quality of the estimate \hat{x}_i^{l-1} . If \hat{x}_i^{l-1} is a poor estimate of x_i , then ρ_i^{l-1} will in turn be low, and consequently a smaller

weighting is applied to the MAI estimate that is to be subtracted from \tilde{r}_i^l . On the other hand, if \hat{x}_i^{l-1} is an excellent estimate of x_i , then $\rho_i^{l-1} \approx 1$, and nearly all of the MAI is subtracted from \tilde{r}_i^l . Note that the diagonal of \mathbf{D}^l is indeed asymptotically zero, as stipulated by (37).

Next, we optimize the vector \mathbf{b}_i^l . The identity

$$\begin{aligned} & (\mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii}) (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \\ & \quad \times (\mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A} - \mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A} \mathbf{E}_{ii})^\dagger \\ & = (\mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A}) (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) (\mathbf{b}_i^{l\dagger} \mathbf{H} \mathbf{A})^\dagger \\ & \quad - (1 - (\rho_i^{l-1})^2) |\mathbf{b}_i^{l\dagger} \mathbf{h}_i A_i|^2 \end{aligned} \quad (44)$$

can be used to rewrite (44) as

$$\gamma_i^l(\mathbf{b}_i^l) = \frac{1}{\frac{1}{\phi_i^l(\mathbf{b}_i^l)} - (1 - (\rho_i^{l-1})^2)}, \quad (45)$$

where

$$\phi_i^l(\mathbf{b}_i^l) = \frac{\mathcal{E}_s |\mathbf{b}_i^{l\dagger} \mathbf{h}_i A_i|^2}{\mathbf{b}_i^{l\dagger} [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger] \mathbf{b}_i^l}. \quad (46)$$

Using the Schwarz inequality, we have⁴

$$\begin{aligned} & |\mathbf{b}_i^{l\dagger} \mathbf{h}_i A_i|^2 \\ & = |\mathbf{b}_i^{l\dagger} [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{1/2} \\ & \quad \times [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1/2} \mathbf{h}_i A_i|^2 \\ & \leq \mathbf{b}_i^{l\dagger} [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger] \mathbf{b}_i^l \\ & \quad \times A_i^* \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1} \mathbf{h}_i A_i \end{aligned} \quad (47)$$

with equality if and only if

$$\mathbf{b}_i^l \propto [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1} \mathbf{h}_i A_i \quad \text{for } i = 1, 2, \dots, P. \quad (48)$$

Substituting (48) into (47), we see that (49) maximizes (47) and, in turn, (44). When we choose the proportionality constant to be the same for $i = 1, 2, \dots, P$, we may write⁵

$$\mathbf{B}^l \propto [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1} \mathbf{H} \mathbf{A}. \quad (49)$$

Some comments can be made about the special case when $l = 1$. During the first pass, feedback is not used

because $\rho^0 = \mathbf{0}$, so the vector $\hat{\mathbf{x}}^0$ does not need to be defined. Moreover, the filter \mathbf{B}^1 takes the form

$$\mathbf{B}^1 \propto [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1} \mathbf{H} \mathbf{A}, \quad (50)$$

which is an expression for the linear MMSE multiuser detector. Thus the performance of the iterated-decision multiuser detector, after just one iteration, is identical to the performance of the linear MMSE multiuser detector. In Section 3.1, we show that the iterated-decision multiuser detector, when using multiple iterations, performs significantly better than the linear MMSE multiuser detector.

The iterated-decision multiuser detector also has an interesting relationship with another multiuser detector. If ρ^{l-1} is set to \mathbf{I} , then the matrices (43) and (50) for the iterated-decision multiuser detector become the matrices used for the multistage detector [16]. In other words, the iterated-decision multiuser detector explicitly takes into account the reliability of tentative decisions, while the multistage detector assumes that all tentative decisions are correct. As we will see in Section 3.1, this difference is the reason that the decisions of the former asymptotically converge to the optimum ones, while the decisions of the latter often diverge.

We now proceed to simplify the SINR expression that characterizes the resulting performance for the i th user. With the optimum \mathbf{b}_i^l and \mathbf{d}_i^l , we have, substituting (48) into (47),

$$\begin{aligned} \phi_i^l & = \mathcal{E}_s A_i^* \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s \mathbf{H} \mathbf{A} \\ & \quad \times (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger]^{-1} \mathbf{h}_i A_i. \end{aligned} \quad (51)$$

After some algebraic manipulation, the SINR from (46), with (52), then becomes

$$\gamma_i^l = \left(\frac{1}{([\mathbf{I} + \boldsymbol{\alpha}^l]^{-1})_{ii}} - 1 \right) \cdot \frac{1}{1 - (\rho_i^{l-1})^2} \quad (52)$$

where

$$\boldsymbol{\alpha}^l = \frac{\mathcal{E}_s (\mathbf{I} - \rho^{l-1} \rho^{l-1\dagger}) \mathbf{A}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{A}}{\mathcal{N}_0}. \quad (53)$$

For the case of accurate power control, i.e., $\mathbf{A} = \mathbf{A} \mathbf{I}$ so $\rho^{l-1} = \rho^{l-1} \mathbf{I}$, it is shown in Appendix A that in the large system limit ($P \rightarrow \infty$ with $\beta \triangleq P/Q$ held constant), the SINR in (53) for each user converges in the mean-square sense to

$$\gamma^l = \left(\frac{1}{1 - \frac{\xi^l}{4\beta} \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right)} - 1 \right) \cdot \frac{1}{1 - (\rho^{l-1})^2} \quad (54)$$

where

$$\mathcal{F}(y, z) \triangleq \left(\sqrt{y(1 + \sqrt{z})^2 + 1} - \sqrt{y(1 - \sqrt{z})^2 + 1} \right)^2 \quad (55)$$

and

$$\frac{1}{\xi^l} = \frac{(1 - (\rho^{l-1})^2)}{\zeta}, \quad (56)$$

with

$$\frac{1}{\zeta} = \frac{\mathcal{E}_s |A|^2}{\mathcal{N}_0}, \quad (57)$$

the received SNR.

The iterative algorithm for computing the set of correlation coefficients ρ^l , and in turn predicting the sequence of symbol error probabilities is as follows.

1. Set $\rho^0 = 0$ and let $l = 1$.
2. Compute the SINR γ^l from ρ^{l-1} via (55), (57), and (58). [For smaller systems, we can alternatively (and in some cases more accurately) compute γ^l from ρ^{l-1} by averaging (53) over all users.]
3. Compute the symbol error probability $\Pr(\epsilon^l)$ from γ^l via (11).
4. Compute ρ^l via (23).
5. Increment l and go to step 2.

In the special case of QPSK, it can be shown that the algorithm can be streamlined by eliminating Step 3 and replacing the approximation (23) with the exact formula in (24).

3.1. Performance

From Steps 2 and 3 of the algorithm, we see that $\Pr(\epsilon^l)$ can be expressed as

$$\Pr(\epsilon^l) = \mathcal{G}(\zeta, \beta, \rho^{l-1}), \quad (58)$$

where $\mathcal{G}(\cdot, \cdot, \cdot)$ is a monotonically decreasing function in both SNR $1/\zeta$ and correlation ρ^{l-1} , but a monotonically increasing function in β . The monotonicity of $\mathcal{G}(\cdot, \cdot, \cdot)$ is illustrated in Fig. 8 where the successively lower solid curves plot $\mathcal{G}(\zeta, \beta, \rho)$ as a function

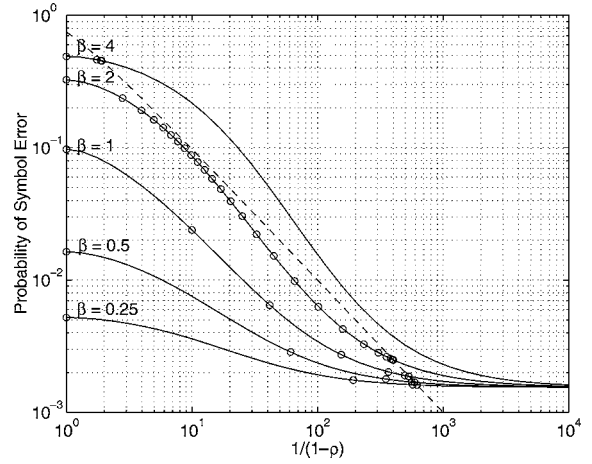


Figure 8. Iterated-decision multiuser detector performance, with power control. The successively higher solid curves plot QPSK symbol error rate as a function of the correlation coefficient ρ for $\beta = P/Q$ values of 0.25, 0.5, 1, 2, and 4, with an SNR per bit of 7 dB. Along each curve, \circ 's identify the theoretically predicted decreasing error rates achieved with $l = 1, 2, \dots$ decoding passes, and the intersections with the dashed line are the steady-state values ($l \rightarrow \infty$).

of $1/(1 - \rho)$ for various values of β , with an SNR per bit of 7 dB and power control. Meanwhile, from Step 4 of the algorithm, we see that we can also express $\Pr(\epsilon^l)$ as

$$\Pr(\epsilon^l) = \mathcal{H}(\rho^l), \quad (59)$$

where $\mathcal{H}(\cdot)$ is a monotonically decreasing function of ρ^l . The dashed line in Fig. 8 plots $\mathcal{H}(\rho)$ as a function of $1/(1 - \rho)$.

At a given $1/\zeta$ and β , the sequence of error probabilities $\Pr(\epsilon^l)$ and correlation coefficients ρ^l can be obtained by starting at the left end of the solid curve (corresponding to $\rho^0 = 0$) and then successively moving horizontally to the right from the solid curve to the dashed line, and then moving downward from the dashed line to the solid curve. Each "step" of the resulting descending staircase corresponds to one pass of the multiuser detector. In Fig. 8, the sequence of operating points is indicated on the solid curves with the \circ symbols.

That the sequence of error probabilities $\Pr(\epsilon^1), \Pr(\epsilon^2), \dots$ obtained by the recursive algorithm is monotonically decreasing suggests that additional iterations always improve performance. The error rate performance for a given SNR of $1/\zeta$ and a given β eventually converges to a steady-state value of $\Pr(\epsilon^\infty)$, which is the unique solution to the equation

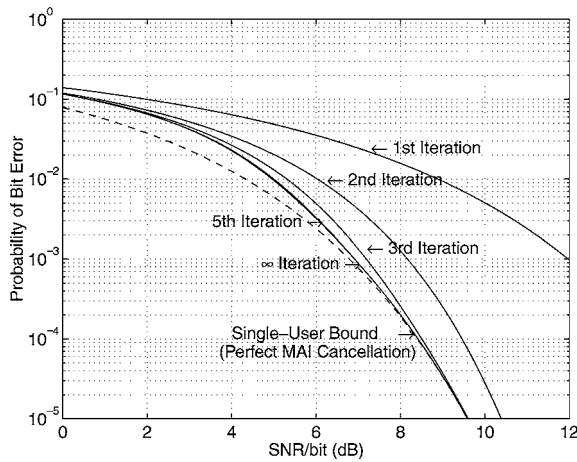


Figure 9. Theoretical iterated-decision multiuser detector performance with power control, as a function of SNR per bit. The successively lower solid curves depict the QPSK bit-error rate with $\beta = P/Q = 0.77$ as a function of SNR per bit for 1, 2, 3, 5, and ∞ decoding iterations.

$$\Pr(\epsilon^\infty) = \mathcal{G}(\zeta, \beta, \mathcal{H}^{-1}(\Pr(\epsilon^\infty))), \quad (60)$$

corresponding to the intersection of the dashed line and the appropriate solid curve in Fig. 8.

If β is relatively small, Fig. 8 suggests that steady-state performance is approximately achieved with comparatively few iterations, after which additional iterations provide only negligibly small gains in performance. This observation can also be readily made from Fig. 9, where bit-error rate is plotted as a function of SNR per bit for 1, 2, 3, 5, and an infinite number of iterations, with $\beta = 0.77$. It is significant that, for small β , few passes are required to converge to typical target bit-error rates, since the amount of computation is directly proportional to the number of passes required; we emphasize that the complexity of a single pass of the iterated-decision multiuser detector is comparable to that of the decorrelating detector or the linear MMSE multiuser detector.

As β increases, Fig. 8 shows that the gap between the solid curve and the dashed curve decreases. Thus the “steps” of the descending staircase get smaller, and there is a significant increase in the number of iterations required to approximately achieve steady-state performance. Moreover, the probability of error at steady-state becomes slightly larger.

When β is greater than some SNR-dependent threshold, not only can (61) have multiple solutions, but one of the solutions occurs at a high probability of error, as illustrated by the curve in Fig. 8 corresponding

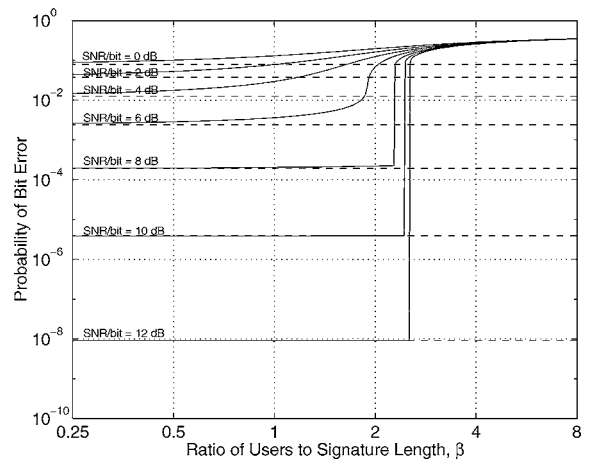


Figure 10. Theoretical iterated-decision multiuser detector performance with power control, as a function of $\beta = P/Q$. The solid curves depict the QPSK bit-error rate as a function of β for various values of SNR per bit, while the corresponding dashed curves depict the single-user bound.

to $\beta = 4$. The dependence of the threshold on SNR is shown in Fig. 10. As the SNR increases, the β threshold increases, and the bit-error rate curve becomes much sharper at the threshold. Our experiments show that in the high SNR regime the threshold is near $\beta \approx e$.

In Fig. 11, we compare the theoretical ($Q \rightarrow \infty$) and simulated ($Q = 128$) bit-error rates of the iterated-decision multiuser detector with the bit-error rates

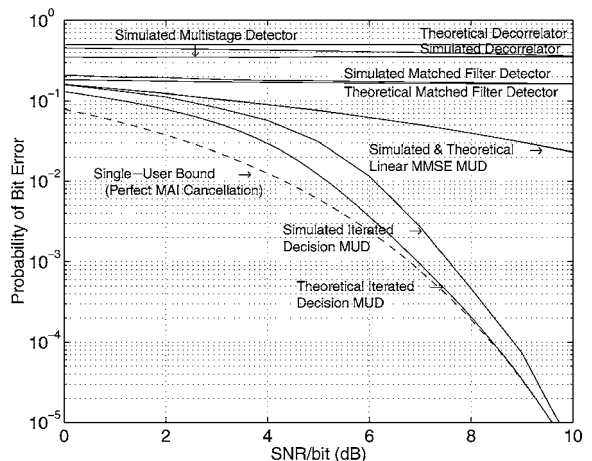


Figure 11. Theoretical ($Q \rightarrow \infty$) and experimentally observed ($Q = 128$) performance for various multiuser detectors, with power control. The solid curves depict QPSK bit-error rates with $\beta = P/Q = 1$ for the iterated-decision multiuser detector, multistage detector, linear MMSE multiuser detector, decorrelating detector, and matched-filter multiuser detector as a function of SNR per bit.

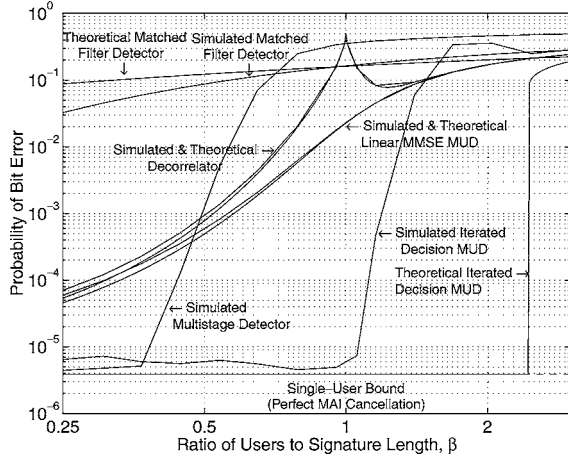


Figure 12. Theoretical ($Q \rightarrow \infty$) and experimentally observed ($Q = 128$) performance for various multiuser detectors, with power control. The solid curves depict QPSK bit-error rates at an SNR per bit of 10 dB for the iterated-decision multiuser detector, multistage detector, linear MMSE multiuser detector, decorrelating detector, and matched-filter multiuser detector as a function of $\beta = P/Q$.

of various other multiuser detectors as a function of SNR, with $\beta = 1$ and power control. The iterated-decision multiuser detector significantly outperforms the other detectors at moderate to high SNR, and asymptotically approaches the single-user bound. Thus, perfect MAI cancellation is approached at high SNR.

Next, in Fig. 12, we compare the effect of β on the simulated bit-error rates of the various multiuser detectors⁶ when decoding $Q = 128$ simultaneous users at an SNR per bit of 10 dB with power control. The iterated-decision multiuser detector has clearly superior performance when $\beta \lesssim 1.5$. Figure 12 also shows the corresponding theoretical curves for $Q \rightarrow \infty$.

3.2. Adaptive Implementations

In Section 3, we derived the optimal matrices \mathbf{B}^l and \mathbf{D}^l for known values of the channel and the user signatures. We now develop an adaptive implementation of the iterated-decision multiuser detector, in which optimal matrices are selected automatically (from the received data) without explicit knowledge of the channel or the signatures. Furthermore, we assume that the packet size is chosen small enough such that the channel encountered by each user's packet appears fixed.

We consider a P -user discrete-time synchronous channel model, where the i th user modulates an M -ary

PSK symbol sequence $x_i[n]$ for $n = 0, 1, \dots, N - 1$ onto a signature sequence \mathbf{h}_i of length Q assigned to that user; some of these symbols (not necessarily at the head of the packet) are for training, while the rest are data symbols. The received vector sequence is

$$\mathbf{r}[n] = \mathbf{H}\mathbf{A}\mathbf{x}[n] + \mathbf{w}[n], \quad (61)$$

where $\mathbf{H} = [\mathbf{h}_1 | \dots | \mathbf{h}_P]$ is the $Q \times P$ matrix of signatures, $\mathbf{A} = \text{diag}\{A_1, \dots, A_P\}$ is the $P \times P$ diagonal matrix of received amplitudes, $\mathbf{x}[n] = [x_1[n], x_2[n], \dots, x_P[n]]^T$ is the $P \times 1$ vector sequence of data symbols, and $\mathbf{w}[n]$ is a noise vector sequence.

Before the first pass ($l = 1$) of the adaptive iterated-decision multiuser detector, we need to initialize the hard decisions $\hat{x}_i^0[n]$ for each user's packet. Since the locations and values of the training symbols in each packet are known at the receiver, we set $\hat{x}_i^0[n] = x_i[n]$ for the i and n corresponding to those locations. For all other locations in the packets, we set $\hat{x}_i^0[n]$ to be a "neutral" value—for white PSK symbols, this value should be zero.

On the l th pass of the detector where $l = 1, 2, 3, \dots$, each received vector $\mathbf{r}[n]$ for $n = 0, 1, \dots, N - 1$ is first premultiplied by a $P \times Q$ matrix $\mathbf{B}^l = [\mathbf{b}_1^l | \dots | \mathbf{b}_P^l]^\dagger$, producing the $P \times 1$ vector

$$\tilde{\mathbf{r}}^l[n] = \mathbf{B}^{l\dagger} \mathbf{r}[n]. \quad (62)$$

Next, an appropriately constructed estimate $\hat{\mathbf{z}}^l[n]$ of the MAI in that symbol period is subtracted from $\tilde{\mathbf{r}}^l[n]$ to produce $\tilde{\mathbf{x}}^l[n]$, i.e.,

$$\tilde{\mathbf{x}}^l[n] = \tilde{\mathbf{r}}^l[n] - \hat{\mathbf{z}}^l[n] \quad (63)$$

where

$$\hat{\mathbf{z}}^l[n] = \mathbf{D}^{l\dagger} \hat{\mathbf{x}}^{l-1}[n] \quad (64)$$

with $\mathbf{D}^l = [\mathbf{d}_1^l | \dots | \mathbf{d}_P^l]$, a $P \times P$ matrix. Since $\hat{\mathbf{z}}^l[n]$ is intended to be some kind of MAI estimate, we restrict attention to the case in which

$$(\mathbf{D}^l)_{11} = (\mathbf{D}^l)_{22} = \dots = (\mathbf{D}^l)_{PP} = 0. \quad (65)$$

Thus, the i th component of the slicer input $\tilde{\mathbf{x}}^l[n]$ can be expressed as

$$\tilde{x}_i^l[n] = \mathbf{c}_i^{l\dagger} \mathbf{q}_i^l[n] \quad (66)$$

where

$$\mathbf{c}_i^l = \begin{bmatrix} b_{1,i}^l \\ \vdots \\ b_{Q,i}^l \\ -d_{1,i}^l \\ \vdots \\ -d_{i-1,i}^l \\ -d_{i+1,i}^l \\ \vdots \\ -d_{P,i}^l \end{bmatrix} \quad \mathbf{q}_i^l[n] = \begin{bmatrix} r_1[n] \\ \vdots \\ r_Q[n] \\ \hat{x}_1^{l-1}[n] \\ \vdots \\ \hat{x}_{i-1}^{l-1}[n] \\ \hat{x}_{i+1}^{l-1}[n] \\ \vdots \\ \hat{x}_P^{l-1}[n] \end{bmatrix} \quad (67)$$

with $b_{j,k}^l$ and $d_{j,k}^l$ being the jk th elements of \mathbf{B}^l and \mathbf{D}^l respectively. The slicer then generates the hard decisions $\hat{x}_i^l[n]$ from $\tilde{x}_i^l[n]$ for all i and n , except for those values corresponding to the locations of training symbols in $x_i[n]$. For those n , we set $\hat{x}_i^l[n] = x_i[n]$.

If we assume that $x_i[n] \approx \hat{x}_i^{l-1}[n]$ for all i and all n for the purposes of determining the optimal matrices, then it is reasonable to choose \mathbf{b}_i and \mathbf{d}_i so as to minimize the sum of error squares:

$$\mathcal{E}(\mathbf{c}_i^l) = \sum_{n=-\infty}^{\infty} |\hat{x}_i^{l-1}[n] - \mathbf{c}_i^{l*} \mathbf{q}_i^l[n]|^2. \quad (68)$$

Since this is a linear least-squares estimation problem, the optimum \mathbf{c}_i^l is [11]

$$\mathbf{c}_{i,\text{opt}}^l = [\Phi_i^l]^{-1} \mathbf{u}_i^l, \quad (69)$$

where $\Phi_i^l = \sum_{n=-\infty}^{\infty} \mathbf{q}_i^l[n] \mathbf{q}_i^{l*}[n]$ and $\mathbf{u}_i^l = \sum_{n=-\infty}^{\infty} \hat{x}_i^{l-1*}[n] \mathbf{q}_i^l[n]$. The matrices Φ_i^l can be efficiently obtained by eliminating the $(Q+i)$ th row and column of $\Phi^l = \sum_{n=-\infty}^{\infty} \mathbf{q}^l[n] \mathbf{q}^{l*}[n]$ where $(\mathbf{q}^l[n])^T = [(\mathbf{r}[n])^T (\hat{\mathbf{x}}^{l-1}[n])^T]$, and $[\Phi_i^l]^{-1}$ can be efficiently computed using formulas for the inversion of a partitioned matrix [12].

The block-iterative nature of the multiuser detector allows the training symbols to be located anywhere in the users' packets. Since the locations do not appear to affect performance, we arbitrarily choose to uniformly space the training symbols within each user's packet.

In Fig. 13, we plot the probability of bit error as a function of SNR for varying amounts of training data. We see that, as expected, performance improves as the amount of training data is increased. Moreover, only a modest amount of training symbols is required at high

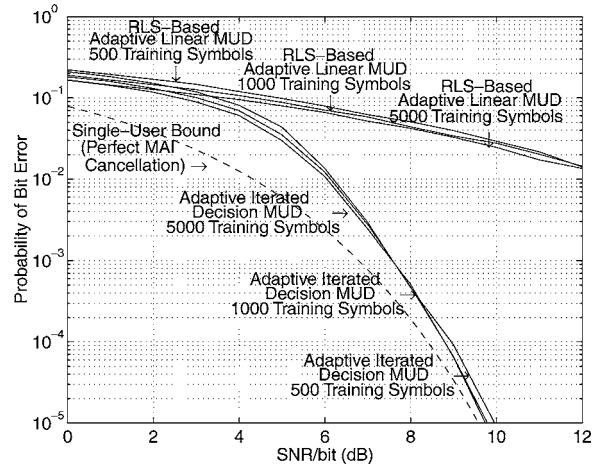


Figure 13. Experimentally observed ($Q = 128$) QPSK bit-error rates for the adaptive iterated-decision multiuser detector and the RLS-based adaptive linear multiuser detector (forgetting factor $\lambda = 1$), with $\beta = P/Q = 1$ and power control. Each user transmitted packets consisting of 10000 data symbols and either 500, 1000, or 5000 training symbols.

SNR for the adaptive multiuser detector to perform as if the channel and the signatures were exactly known at the receiver.

For comparison purposes, we also plot in Fig. 13 the performance of the RLS-based implementation of the adaptive linear multiuser detector [19]. The linear multiuser detector performs significantly worse than the iterated-decision multiuser detector for comparable amounts of training data.

3.3. Coded Implementations

For coded systems, an iterated-decision multiuser decoder is readily obtained, and takes a form analogous to the iterated-decision equalizer-decoder structure described in Section 2.3.

A communication system that combines iterated-decision multiuser detection with coding is depicted in Fig. 14. The data streams $x_i[n]$, $i = 1, 2, \dots, P$ of the P users are encoded using separate encoders, and the corresponding streams of coded symbols are $\bar{x}_i[n]$ for $i = 1, 2, \dots, P$ and $n = 0, 1, \dots, N - 1$. The received vector sequence is thus

$$\mathbf{r}[n] = \mathbf{H}\mathbf{A}[n]\bar{\mathbf{x}}[n] + \mathbf{w}[n], \quad (70)$$

where $\bar{\mathbf{x}}[n] = [\bar{x}_1[n], \bar{x}_2[n], \dots, \bar{x}_P[n]]^T$. As in Section 3.2, on the l th pass of the multiuser detector, each received vector $\mathbf{r}[n]$ for $n = 0, 1, \dots, N - 1$ is processed independently to produce a corresponding

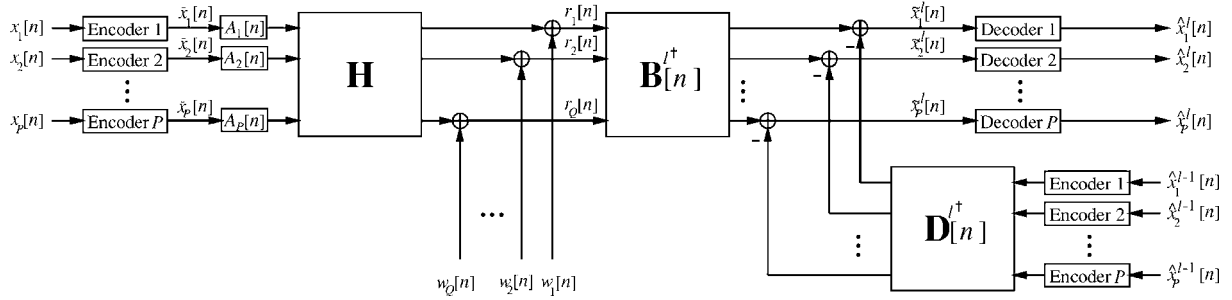


Figure 14. Structure of a communication system that combines iterated-decision multiuser detection with channel coding.

vector $\tilde{\mathbf{x}}^l[n]$. The sequences $\tilde{x}_i^l[n]$ for $i = 1, 2, \dots, P$ are then input to a bank of soft-decision ML decoders, thereby producing $\hat{x}_i^l[n]$ for $i = 1, 2, \dots, P$, the tentative decisions for $x_i[n]$. These tentative decisions must be re-encoded before being processed by the matrix $\mathbf{D}^l[n]$. Performance may be improved by using an interleaver after each encoder and a deinterleaver before each decoder.

This multiuser decoder structure for coded systems can be compared to those developed in [20, 21], which have a significantly different but similarly intriguing receiver structure.

Appendix A: Derivation of SINR Expression (55)

With accurate power control, (52) becomes

$$\phi_i^l = \mathcal{E}_s |A|^2 \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{H} \mathbf{H}^\dagger]^{-1} \mathbf{h}_i \quad (71)$$

Substituting (73) into (46), we get

$$\begin{aligned} \gamma_i^l &= \frac{1}{\frac{1}{\mathcal{E}_s |A|^2 \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{H} \mathbf{H}^\dagger]^{-1} \mathbf{h}_i} - (1 - (\rho^{l-1})^2)} \\ &= \frac{\mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{H} \mathbf{H}^\dagger]^{-1} \mathbf{h}_i}{1 - \mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{h}_i^\dagger [\mathcal{N}_0 \mathbf{I} + \mathcal{E}_s (1 - (\rho^{l-1})^2) |A|^2 \mathbf{H} \mathbf{H}^\dagger]^{-1} \mathbf{h}_i} \cdot \frac{1}{1 - (\rho^{l-1})^2} \\ &= \left(\frac{1}{\chi_i^l} - 1 \right) \cdot \frac{1}{1 - (\rho^{l-1})^2} \end{aligned} \quad (72)$$

where

$$\chi_i^l = 1 - \frac{1}{\xi^l} \mathbf{h}_i^\dagger \left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H} \mathbf{H}^\dagger \right]^{-1} \mathbf{h}_i \quad (73)$$

with ξ^l defined by (57).

The derivation of (55) requires the following two lemmas.

Lemma 1. *In the limit as $P \rightarrow \infty$ with $\beta \triangleq P/Q$ held constant, the expected value of χ_i^l is*

$$E[\chi_i^l] = \bar{\chi}_i^l \triangleq 1 - \frac{\xi^l}{4\beta} \cdot \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right) \quad (74)$$

where

$$\mathcal{F}(y, z) \triangleq (\sqrt{y(1 + \sqrt{z})^2 + 1} - \sqrt{y(1 - \sqrt{z})^2 + 1})^2. \quad (75)$$

Proof: From (75), the expected value of χ_i^l is

$$\begin{aligned}
&= E \left[\left(\left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H}^\dagger \mathbf{H} \right]^{-1} \right)_{ii} \right] \\
&= \frac{1}{P} E \left[\text{trace} \left(\left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H}^\dagger \mathbf{H} \right]^{-1} \right) \right] \\
&= \frac{1}{P} E \left[\sum_{j=1}^P \lambda_j \left(\left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H}^\dagger \mathbf{H} \right]^{-1} \right) \right] \\
&= E \left[\lambda \left(\left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H}^\dagger \mathbf{H} \right]^{-1} \right) \right] \\
&= E \left[\frac{1}{1 + \frac{1}{\xi^l} \lambda(\mathbf{H}^\dagger \mathbf{H})} \right], \tag{76}
\end{aligned}$$

where we have used the identity⁷

$$\mathbf{Y}(\mathbf{I} + \mathbf{X}\mathbf{Y})^{-1} = (\mathbf{I} + \mathbf{Y}\mathbf{X})^{-1}\mathbf{Y} \tag{77}$$

and the fact that the trace of a square matrix is equal to the sum of its eigenvalues (denoted by λ_j).

If the ratio of the number of users to the signature length is, or converges to a constant:

$$\lim_{P \rightarrow \infty} \frac{P}{Q} = \beta \in (0, +\infty), \tag{78}$$

then the percentage of the P eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$ that lie below x converges to the cumulative distribution function of the probability density function [22]

$$f_\beta(x) = [1 - \beta^{-1}]^+ \delta(x) + \frac{\sqrt{[x-s]^+ [t-x]^+}}{2\pi\beta x} \tag{79}$$

where

$$\begin{aligned}
s &= (1 - \sqrt{\beta})^2 \\
t &= (1 + \sqrt{\beta})^2
\end{aligned} \tag{80}$$

and the operator $[\cdot]^+$ is defined according to

$$[u]^+ \triangleq \max\{0, u\}. \tag{81}$$

Thus, we can compute the limit as $P \rightarrow \infty$ of (78) as [3]

$$\begin{aligned}
\lim_{P \rightarrow \infty} E[\chi_i^l] &= \lim_{P \rightarrow \infty} E \left[\frac{1}{1 + \frac{1}{\xi^l} \lambda(\mathbf{H}^\dagger \mathbf{H})} \right] \\
&= \int_0^\infty \frac{1}{1 + \frac{1}{\xi^l} x} f_\beta(x) dx \\
&= 1 - \frac{\xi^l}{4\beta} \cdot \mathcal{F} \left(\frac{1}{\xi^l}, \beta \right). \quad \square \tag{82}
\end{aligned}$$

We can use Lemma 1 to prove an even stronger result.

Lemma 2. *In the limit as $P \rightarrow \infty$ with β held constant,*

$$\chi_i^l \xrightarrow{\text{m.s.}} 1 - \frac{\xi^l}{4\beta} \cdot \mathcal{F} \left(\frac{1}{\xi^l}, \beta \right). \tag{83}$$

Proof: Consider the normalized variance of χ_i^l :

$$\begin{aligned}
\frac{\text{var}(\chi_i^l)}{(E[\chi_i^l])^2} &\leq E \left[\frac{(\chi_i^l)^2 + (E[\chi_i^l])^2 - 2\chi_i^l E[\chi_i^l]}{\chi_i^l (E[\chi_i^l])^2} \right] \\
&= E \left[\frac{1}{\chi_i^l} \right] - \frac{1}{E[\chi_i^l]}, \tag{84}
\end{aligned}$$

where the upper bound comes from the fact that $0 \leq \chi_i^l \leq 1$. Thus, to show the mean-square convergence result in (85), we need to show that

$$\lim_{P \rightarrow \infty} E \left[\frac{1}{\chi_i^l} \right] = \lim_{P \rightarrow \infty} \frac{1}{E[\chi_i^l]}. \tag{85}$$

To this end, we develop a useful expression for χ_i^l . Let

$$\boldsymbol{\Sigma}_i^l = \mathbf{I} + \frac{1}{\xi^l} \sum_{j \neq i} \mathbf{h}_j \mathbf{h}_j^\dagger. \tag{86}$$

Then

$$\begin{aligned}
\chi_i^l &= 1 - \frac{1}{\xi^l} \mathbf{h}_i^\dagger \left[\mathbf{I} + \frac{1}{\xi^l} \mathbf{H}\mathbf{H}^\dagger \right]^{-1} \mathbf{h}_i \\
&= 1 - \frac{1}{\xi^l} \mathbf{h}_i^\dagger \left[\boldsymbol{\Sigma}_i^l + \frac{1}{\xi^l} \mathbf{h}_i \mathbf{h}_i^\dagger \right]^{-1} \mathbf{h}_i \\
&= 1 - \frac{1}{\xi^l} \mathbf{h}_i^\dagger \left[(\boldsymbol{\Sigma}_i^l)^{-1} - \frac{(\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1}}{\xi^l + \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i} \right] \mathbf{h}_i \\
&= 1 - \frac{1}{\xi^l} \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i \left[1 - \frac{\mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i}{\xi^l + \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i} \right] \\
&= 1 - \frac{\mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i}{\xi^l + \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i} \\
&= \frac{1}{1 + \frac{1}{\xi^l} \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i}, \tag{87}
\end{aligned}$$

where the third equality results from the matrix inversion lemma:

$$\begin{aligned}
(\mathbf{W} + \mathbf{X}\mathbf{Z}\mathbf{Y})^{-1} &= \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{X} \\
&\quad \times (\mathbf{Z}^{-1} + \mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{W}^{-1} \tag{88}
\end{aligned}$$

where \mathbf{W} and \mathbf{Z} are invertible square matrices.

Thus,

$$\begin{aligned}
E\left[\frac{1}{\chi_i^l}\right] &= E\left[1 + \frac{1}{\xi^l} \mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i\right] \\
&= 1 + \frac{1}{\xi^l} E[\mathbf{h}_i^\dagger (\boldsymbol{\Sigma}_i^l)^{-1} \mathbf{h}_i] \\
&= 1 + \frac{1}{\xi^l} E\left[\frac{1}{Q} \sum_{j=1}^Q ([\boldsymbol{\Sigma}_i^l]^{-1})_{jj}\right] \\
&= 1 + \frac{1}{\xi^l Q} E[\text{trace}([\boldsymbol{\Sigma}_i^l]^{-1})] \\
&= 1 + \frac{1}{\xi^l Q} E\left[\sum_{j=1}^Q \lambda_j([\boldsymbol{\Sigma}_i^l]^{-1})\right] \\
&= 1 + \frac{1}{\xi^l} E[\lambda([\boldsymbol{\Sigma}_i^l]^{-1})] \\
&= 1 + \frac{1}{\xi^l} E\left[\frac{1}{1 + \frac{\beta}{\xi^l} \lambda(\frac{1}{\beta} \sum_{j \neq i} \mathbf{h}_j \mathbf{h}_j^\dagger)}\right], \quad (89)
\end{aligned}$$

where the third equality comes from the fact that the components of \mathbf{h}_i are independent of $\boldsymbol{\Sigma}_i^l$ and have zero mean and variance equal to $1/Q$. If the ratio of the number of users to the signature length is, or converges to a constant:

$$\lim_{P \rightarrow \infty} \frac{P}{Q} = \beta \in (0, +\infty), \quad (90)$$

then we can use (81) to compute the limit of (91) as $P \rightarrow \infty$ [3]:

$$\begin{aligned}
\lim_{P \rightarrow \infty} E\left[\frac{1}{\chi_i^l}\right] &= 1 + \frac{1}{\xi^l} \lim_{P \rightarrow \infty} E\left[\frac{1}{1 + \frac{\beta}{\xi^l} \lambda(\frac{1}{\beta} \sum_{j \neq i} \mathbf{h}_j \mathbf{h}_j^\dagger)}\right] \\
&= 1 + \frac{1}{\xi^l} \int_0^\infty \frac{1}{1 + \frac{\beta}{\xi^l} x} f_{1/\beta}(x) dx \\
&= 1 + \frac{1}{\xi^l} \left[1 - \frac{\beta}{4} \cdot \frac{\xi^l}{\beta} \cdot \mathcal{F}\left(\frac{\beta}{\xi^l}, \frac{1}{\beta}\right)\right] \\
&= 1 + \frac{1}{\xi^l} \left[1 - \frac{\xi^l}{4} \cdot \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right)\right] \\
&= 1 + \frac{1}{\xi^l} - \frac{1}{4} \cdot \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right). \quad (91)
\end{aligned}$$

To show (87), Lemma 1 tells us we need to check that

$$\left[1 - \frac{\xi^l}{4\beta} \cdot \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right)\right]^{-1} = 1 + \frac{1}{\xi^l} - \frac{1}{4} \cdot \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right). \quad (92)$$

But this is equivalent to checking that the equation

$$x^2 - 4\left(1 + \frac{1}{\xi^l}(1 + \beta)\right)x + \frac{16\beta}{(\xi^l)^2} = 0 \quad (93)$$

has a solution at

$$x = \mathcal{F}\left(\frac{1}{\xi^l}, \beta\right), \quad (94)$$

which can be verified by substituting (96) into (95), so we have proved (85). \square

We now proceed to show (55). With $\bar{\chi}_i^l$ as defined in (76), and with γ_i^l in (74) bounded according to $\gamma_i^l \leq 1/\zeta$, we have that as $P \rightarrow \infty$ with β held constant,

$$\begin{aligned}
&E\left[\left(\left(\frac{1}{\chi_i^l} - 1\right) \cdot \frac{1}{1 - (\rho^{l-1})^2} - \left(\frac{1}{\bar{\chi}_i^l} - 1\right) \cdot \frac{1}{1 - (\rho^{l-1})^2}\right)^2\right] \\
&= E\left[\left(\frac{1}{\bar{\chi}_i^l}\right)^2 \left(\frac{1}{\chi_i^l(1 - (\rho^{l-1})^2)}\right)^2 (\chi_i^l - \bar{\chi}_i^l)^2\right] \\
&\leq E\left[\left(\frac{1}{\bar{\chi}_i^l}\right)^2 \left(\frac{1}{\zeta} + \frac{1}{1 - (\rho^{l-1})^2}\right)^2 (\chi_i^l - \bar{\chi}_i^l)^2\right] \\
&= \left(\frac{1}{\bar{\chi}_i^l}\right)^2 \left(\frac{1}{\zeta} + \frac{1}{1 - (\rho^{l-1})^2}\right)^2 E[(\chi_i^l - \bar{\chi}_i^l)^2] \\
&\xrightarrow{\text{m.s.}} 0, \quad (95)
\end{aligned}$$

where the final limit follows from Lemma 2. So the SINR for each user (74) converges in the mean-square sense to (55).

Notes

1. Throughout the paper, our expectations involving functions of frequency ω do not depend on ω , so we omit this dependence in our notation to emphasize this.
2. When $x[n]$ is a sequence coded for the Gaussian channel, the approximation in (8) is still valid—typical trellis codes used with random bit streams generally produce white symbol streams [7], as do random codes. More will be said about coding in Section 2.3.
3. The superscripts T and \dagger denote the transpose and conjugate-transpose operations, respectively.
4. $\mathbf{F}^{1/2}$, a square root matrix of the positive semidefinite matrix \mathbf{F} , satisfies $\mathbf{F} = \mathbf{F}^{1/2\dagger} \mathbf{F}^{1/2}$.
5. Using the matrix identity (79), we may alternatively write

$$\mathbf{B}^l \propto \mathbf{H}\mathbf{A}[\mathcal{N}_0\mathbf{I} + \mathcal{E}_s(\mathbf{I} - \rho^{l-1}\rho^{l-1\dagger})\mathbf{A}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{A}]^{-1},$$

which may be easier to evaluate depending on the relative sizes of P and Q .

6. The theoretical large system performance of the decorrelator for the case $\beta > 1$ is derived in [17], where the decorrelator is defined as the Moore–Penrose generalized inverse [18] of \mathbf{H} .
7. The identity is a special case of the matrix inversion lemma (90).

References

1. G.D. Forney, Jr., “Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference,” *IEEE Trans. Inform. Theory*, vol. IT-18, 1972, pp. 363–378.
2. C.A. Belfiore and J.H. Park, Jr., “Decision-Feedback Equalization,” in *Proc. IEEE*, 1979, vol. 67, pp. 1143–1156.
3. S. Verdú, *Multiuser Detection*, Cambridge, U.K.: Cambridge University Press, 1998.
4. G.W. Wornell and M.D. Trott, “Efficient Signal Processing Techniques for Exploiting Transmit Antenna Diversity on Fading Channels,” *IEEE Trans. Signal Processing*, vol. 45, 1997, pp. 191–205.
5. A.M. Chan and G.W. Wornell, “A Class of Block-Iterative Equalizers for Intersymbol Interference Channels: Fixed Channel Results,” *IEEE Trans. Commun.*, vol. 49, Nov. 2001.
6. J.G. Proakis, *Digital Communications*, 3rd edn. New York: McGraw-Hill, 1995.
7. E. Biglieri, “Ungerboeck Codes Do Not Shape the Signal Power Spectrum,” *IEEE Trans. Inform. Theory*, vol. IT-32, 1986, pp. 595–596.
8. S. Beheshti, S.H. Isabelle, and G.W. Wornell, “Joint Intersymbol and Multiple-Access Interference Suppression Algorithms for CDMA Systems,” *European Trans. Telecomm. & Related Technol.*, vol. 9, 1998, pp. 403–418.
9. A.T. Bharucha-Reid and M. Sambandham, *Random Polynomials*, Orlando, FL: Academic Press, 1986.
10. A.M. Chan, “A Class of Batch-Iterative Methods for the Equalization of Intersymbol Interference Channels,” S.M. Dissertation, M.I.T., Aug. 1999.
11. S. Haykin, *Adaptive Filter Theory*, 3rd edn., Englewood Cliffs, NJ: Prentice Hall, 1996.
12. H. Lütkepohl, *Handbook of Matrices*, Chichester, England: Wiley, 1996.
13. J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, Jr., “MMSE Decision-Feedback Equalizers and Coding—Part I: Equalization Results,” *IEEE Trans. Commun.*, vol. 43, 1995, pp. 2582–2594.
14. J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, Jr., “MMSE Decision-Feedback Equalizers and Coding—Part II: Coding Results,” *IEEE Trans. Commun.*, vol. 43, 1995, pp. 2595–2604.
15. M. Tüchler, R. Kötter, and A. Singer, “Turbo Equalization’: Principles and New Results,” *IEEE Trans. Commun.*, submitted.
16. M.K. Varanasi and B. Aazhang, “Near-Optimum Detection in Synchronous Code-Division Multiple-Access Systems,” *IEEE Trans. Commun.*, vol. 39, May 1991, pp. 725–736.
17. Y.C. Eldar and A.M. Chan, “On Wishart Matrix Eigenvalues and Eigenvectors and the Asymptotic Performance of the Decorrelator,” *IEEE Trans. Inform. Theory*, submitted.
18. G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd edn., Baltimore, MD: Johns Hopkins University Press, 1996.
19. M.L. Honig and H.V. Poor, “Adaptive Interference Suppression” in *Wireless Communications: Signal Processing Perspectives*, H.V. Poor and G.W. Wornell (Eds.), Upper Saddle River, NJ: Prentice-Hall, 1998.
20. X. Wang and H.V. Poor, “Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA,” *IEEE Trans. Commun.*, vol. 47, 1999, pp. 1047–1061.
21. J. Boutros and G. Caire, “Iterative Multiuser Joint Decoding: Unified Framework and Asymptotic Analysis,” *IEEE Trans. Inform. Theory*, submitted.
22. Z.D. Bai and Y.Q. Yin, “Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix,” *Annals of Probability*, vol. 21, 1993, pp. 1275–1294.



Albert M. Chan received the B.A.Sc. degree from the University of Toronto, Canada, in 1997 and the S.M. degree from the Massachusetts Institute of Technology (MIT) in 1999, both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at MIT. His research interests include signal processing and communications.

Mr. Chan has served as a Teaching Assistant for probability, digital signal processing, and signals and systems courses in the MIT Department of Electrical Engineering and Computer Science. He is the recipient of the MIT Frederick C. Hennie III Award for Teaching Excellence (2000).
chanal@mit.edu



Gregory W. Wornell received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991 he has been on the faculty of the Department of Electrical Engineering and Computer Science at MIT, where he is currently an Associate Professor. He has spent leaves at the University

of California, Berkeley, CA, in 1999-2000 and at AT&T Bell Laboratories, Murray Hill, NJ, in 1992-3.

His research interests span the areas of signal processing, communication systems, and information theory, and include algorithms and architectures for wireless networks, broadband systems, and multimedia environments. He is author of a number of papers in these areas, as well as the Prentice-Hall monograph *Signal Processing with Fractals: A Wavelet-Based Approach*, and is co-editor (with H. V. Poor) of the Prentice-Hall collection *Wireless Communications: Signal Processing Perspectives*. Within the IEEE he has served as Associate Editor for Communications for *IEEE Signal Processing Letters*, and serves on the Communications Technical Committee of

the Signal Processing Society. He is also active in industry and an inventor on numerous issued and pending patents.

Among the awards he has received for teaching and research are the MIT Goodwin Medal for "conspicuously effective teaching" (1991), the ITT Career Development Chair at MIT (1993), an NSF Faculty Early Career Development Award (1995), an ONR Young Investigator Award (1996), the MIT Junior Bose Award for Excellence in Teaching (1996), the Cecil and Ida Green Career Development Chair at MIT (1996), and an MIT Graduate Student Council Teaching Award (1998). Dr. Wornell is also a member of Tau Beta Pi and Sigma Xi, and a Senior Member of the IEEE.
gww@mit.edu