A Class of Block-Iterative Equalizers for Intersymbol Interference Channels: Fixed Channel Results

Albert M. Chan, *Student Member, IEEE,* and Gregory W. Wornell, *Senior Member, IEEE*

***Abstract*—A new and efficient class of nonlinear equalizers is developed for intersymbol interference (ISI) channels. These "iterated-decision equalizers" use an optimized multipass algorithm to successively cancel ISI from a block of received data and generate symbol decisions whose reliability increases monotonically with each iteration. These equalizers have an effective complexity comparable to the decision-feedback equalizer (DFE), yet asymptotically achieve the performance of maximum-likelihood sequence detection (MLSD). We show that, because their structure allows cancellation of both precursor and postcursor ISI, iterated-decision equalizers outperform the minimum mean-square error DFE by 2.507 dB on severe ISI channels even with uncoded systems. Moreover, unlike the DFE, iterated-decision equalizers can be readily used in conjunction with error-control coding, making them attractive for a wealth of applications.**

***Index Terms*—Decision-feedback equalizer, equalization, interference cancellation, iterative decoding, multipass receivers, stripping.**

I. 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 can suffer from excessive noise enhancement. Maximum-likelihood sequence detection (MLSD) [13] is an asymptotically optimum receiver in terms of bit error rate (BER) performance, but its high complexity generally precludes its use in practice.

Decision-feedback equalizers (DFE) [3] are a widely used compromise, retaining a complexity comparable to linear equalization, but incurring much less noise enhancement. However, some significant shortcomings arise out of the sequential way in which the DFE processes the received data. First, decisions made at the slicer can only be fed back to improve future decisions. 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, the sequential structure of the DFE makes it somewhat difficult to use in conjunction with coding (on channels not known at the transmitter, as is the case of interest in this paper).

In this paper, we introduce a multipass equalizer whose structure overcomes these principal limitations of the DFE, achieving the performance of MLSD and being fully compatible with coding, while retaining the low complexity of the DFE and the LE.[1]

II. Channel Model

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 (possibly coded) 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

$$r[n] = \sum_k a[k]x[n-k] + w[n]. \tag{1}$$

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

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

Throughout this paper, we focus on the fixed channel case in which the receiver has perfect knowledge of $a[n]$, in order to develop the basic theory and fundamental limits. In the companion paper [8], 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 papers, we restrict our attention to transmitters that have no knowledge of the channel, as is often the case in, for example, wireless applications.

In an increasing number of practical wireless applications, such as underwater acoustic communications [15] and future spectrally efficient implementations of emerging ultrahigh rate wideband radio-frequency (RF) communications [19], ISI can be particularly severe. This paper devotes particular attention to this scenario, examining the performance and developing the properties of the various equalizers in this regime. In fact, we shall optimize the iterated-decision equalizer for the severe-ISI

Paper approved by C.-L. Wang, the Editor for Wireless Spread Spectum of the IEEE Communications Society. Manuscript received February 1, 2000; revised December 26, 2000. This work was supported in part by a Grant from Qualcomm, Inc., a Grant from Sanders, Lockheed-Martin Company, and by the Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0002. This paper was presented in part at the IEEE International Conference on Communications, New Orleans, LA, June 2000.

The authors are with the Department of Electrical Engineering and Computer Science, and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: chanal@mit.edu; gww@mit.edu).

Publisher Item Identifier S 0090-6778(01)10171-6.

[1]A corresponding multipass multiuser detector with analogous properties for iterative cancellation of multiple-access interference is described in [9] and [22].

case in Section III in order to develop the basic theory and fundamental limits. However, the resulting equalizers are also effective on a broad range of shorter-ISI channels as well, as we discuss at the end of Section III.

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 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 [21] with arrays having a large number of elements.

III. The Iterated-Decision Equalizer

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 lth pass are denoted using the superscript l. On the lth pass of the equalizer where $l = 1, 2, 3, \ldots$, the received data $r[n]$ is first processed by a linear filter $b^l[n]$, producing the sequence

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

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.,

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

where

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

(In subsequent analysis, we will show that $\hat{x}^0[n]$ is never required for the first iteration, so its value may remain undefined.) Note that since $\hat{z}^l[n]$ is intended to be some kind of ISI estimate, we are primarily interested in the case in which

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

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

Some important properties of the composite system, consisting of the channel and the iterated-decision equalizer, can be derived when mild restrictions are imposed on the filters

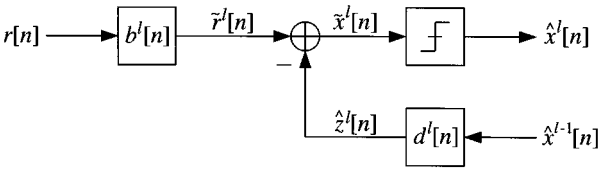

Fig. 1. Iterated-decision equalizer structure.

$b^l[n]$ and $d^l[n]$. To develop these restrictions, we make use of the following definition.

Definition 1: Let $g[n]$ be the impulse response of an LTI filter with frequency response $G(\omega)$. Then $g[n]$ is an admissible filter if $G(\omega)$ can be written in the form

$$G(\omega) = f(A(\omega)) \tag{7}$$

where $A(\omega)$ is the channel frequency response (2), and where $f(\cdot)$ is a complex-valued function that satisfies the admissibility conditions:

1) $f(\cdot)$ is bounded within any finite region of the complex plane, i.e., for every R, there exists a positive real constant $V(R) < \infty$ such that

$$|f(z)| < V(R), \qquad \text{for all } |z| < R. \tag{8}$$

2) $f(\cdot)$ grows more slowly than a quadratic exponential, i.e.,[2]

$$f(z) \sim o\left(e^{\epsilon |z|^2}\right), \qquad \text{for every } \epsilon > 0. \tag{9}$$

With our channel model, $A(\omega)$ is a complex-valued, Gaussian random variable where $E[A(\omega)] = 0$ and $\operatorname{var} A(\omega) = L\sigma_a^2$. Thus, expectations involving only admissible filters at a single frequency are functions only of $L\sigma_a^2$ and not of ω. Consequently, we shall omit the dependence on ω in our notation when dealing with such expectations.

The following useful theorem, which is analogous to ones in [21] and [2],[3] characterizes the composite system consisting of the channel in cascade with the multipass equalizer after lth iterations; a proof is provided in Appendix A.

Theorem 1: Let $x[n]$ and $\hat{x}^{l-1}[n]$ be sequences of zero-mean uncorrelated symbols, each with energy $\mathcal{E}_s$; and let the normalized correlation between the two sequences be expressed in the form[4]

$$\frac{E[x^*[n] \cdot \hat{x}^{l-1}[k]]}{\mathcal{E}_s} = \rho_x^{l-1} \delta[n-k]. \tag{10}$$

[2]The order notation $o(\cdot)$ is to be interpreted in the usual sense. If $p(z) \sim o(q(z))$, then

$$\lim_{|z| \to \infty} \frac{p(z)}{q(z)} = 0.$$

[3]While they focus on fundamentally different digital communication problems, both [21] and [2] exploit system analysis and optimization tools related to those used in this paper, which reflects the breadth of applicability of these tools.

[4]Our empirical analysis confirms that this model for the correlation between the transmitted symbols and tentative decisions is a very good one in practice. Note, too, that ρ_x^l will in general be a function of the equalizer parameters; we suppress this dependency to simplify our notation.

Moreover, let $a[n]$ be a channel impulse response of length L, where[5] $a[0], a[1], \ldots, a[L-1]$ are mutually independent, zero-mean, complex-valued, circularly symmetric Gaussian random variables with variance σ_a^2. Finally, suppose $b^l[n]$ and $d^l[n]$ are the impulse responses of admissible filters in the sense of Definition 1, and that in addition $d^l[n]$ satisfies the natural requirement (6). Then, as $L \to \infty$, we have that the slicer input $\tilde{x}^l[n]$ defined via (4) with (3), (5), and (1) satisfies, for each n,[6]

$$\tilde{x}^l[n] \xrightarrow{\text{m.s.}} E[AB^l]x[n] + v^l[n] \tag{11}$$

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

$$\begin{aligned}\operatorname{var} v^l[n] = \mathcal{N}_0 E[|B^l|^2] + \mathcal{E}_s(1-(\rho_x^{l-1})^2)\operatorname{var}[AB^l] \\ + \mathcal{E}_s E[|D^l - \rho_x^{l-1}(AB^l - E[AB^l])|^2].\end{aligned} \tag{12}$$

The second-order model (11) turns out to be a useful one for analyzing and optimizing the performance of the iterated-decision equalizer at each iteration. In particular, it can be used to obtain a surprisingly accurate estimate at each iteration of the symbol error rate for M-ary PSK even though we ignore the higher-order statistical dependencies. The evolution in performance of the equalizer with iteration is, in turn, analyzed in Section III-A.

The first step in developing these results is to observe that (11) implies that the signal-to-interference + noise ratio (SINR) at the slicer input during each pass can be written, using (12), as shown in (13) at the bottom of the page, and that the probability of symbol error at the lth iteration is well-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 [17]

$$\Pr(\epsilon^l) = 2\mathcal{Q}\left(\sin\left(\frac{\pi}{M}\right)\sqrt{2\gamma^l}\right) \tag{14}$$

where

$$\mathcal{Q}(v) = \frac{1}{\sqrt{2\pi}}\int_v^\infty e^{-t^2/2}\,dt. \tag{15}$$

For the special case of QPSK ($M = 4$), (14) can be replaced with [17]

$$\Pr(\epsilon^l) = \mathcal{Q}\left(\sqrt{\gamma^l}\right)\left[2 - \mathcal{Q}\left(\sqrt{\gamma^l}\right)\right]. \tag{16}$$

Note that this equivalent channel model effectively suggests that, in the absence of coding, the computationally expensive Viterbi-algorithm-based MLSD can be avoided—that a simple symbol-by-symbol detector may suffice, as if the channel were an AWGN channel.[7] In this case, since the probability of error for M-ary PSK given by (14) is a monotonically decreasing function of SINR, a natural equalizer design strategy involves maximizing the SINR over all $B^l(\omega)$ and $D^l(\omega)$.

For a given filter $B^l(\omega)$, it is straightforward to find the optimal filter $D^l(\omega)$. In particular, note that the SINR expression given in (13) contains $D^l(\omega)$ only once, and the nonnegative denominator term in which it appears can be made exactly zero by setting

$$D^l(\omega) = \rho_x^{l-1}(A(\omega)B^l(\omega) - E[AB^l]). \tag{17}$$

Using (17) to eliminate $D^l(\omega)$, the SINR expression in (13) now simplifies to

$$\gamma^l(B^l) = \frac{\mathcal{E}_s|E[AB^l]|^2}{\mathcal{N}_0 E[|B^l|^2] + \mathcal{E}_s(1-(\rho_x^{l-1})^2)\operatorname{var}[AB^l]}. \tag{18}$$

This result for $D^l(\omega)$ is intuitively satisfying. If $\hat{x}^{l-1}[n] = x[n]$ so that $\rho_x^{l-1} = 1$, then the output of $D^l(\omega)$ exactly reproduces the ISI component of $\tilde{r}^l[n]$. More generally, ρ_x^{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_x^{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 $\rho_x^{l-1} \approx 1$, 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 zero, as stipulated by (6).

The filter $B^l(\omega)$ can be optimized using a method analogous to one used in [20]. Specifically, we can use the identity

$$\operatorname{var}[AB^l] = E[|AB^l|^2] - |E[AB^l]|^2 \tag{19}$$

to rewrite (18) as

$$\gamma^l(B^l) = \frac{\mathcal{E}_s}{\dfrac{1}{\phi^l(B^l)} - \mathcal{E}_s(1-(\rho_x^{l-1})^2)} \tag{20}$$

where

$$\phi^l(B^l) = \frac{|E[AB^l]|^2}{E[(\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2)|B^l|^2]}. \tag{21}$$

[5]Since infinite-complexity equalizers can compensate for arbitrary time delays, we let the nonzero taps of $a[n]$ correspond to $n = 0, 1, \ldots, L-1$ without loss of generality.

[6]We use the notation $\xrightarrow{\text{m.s.}}$ to denote convergence in the mean-square sense.

[7]When $x[n]$ is a sequence coded for the Gaussian channel, Theorem 1 is still valid—typical trellis codes used with random bit streams generally produce white symbol streams [6], as do random codes. More will be said about coding in Section IV.

$$\gamma^l(B^l, D^l) = \frac{\mathcal{E}_s|E[AB^l]|^2}{\mathcal{N}_0 E[|B^l|^2] + \mathcal{E}_s(1-(\rho_x^{l-1})^2)\operatorname{var}[AB^l] + \mathcal{E}_s E[|D^l - \rho_x^{l-1}(AB^l - E[AB^l])|^2]}. \tag{13}$$

Using the Schwarz inequality, we have

$$|E[AB^l]|^2 = \left| E\left[\frac{A}{\sqrt{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2}} \cdot B^l \sqrt{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2} \right] \right|^2$$

$$\leq E\left[\frac{|A|^2}{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2} \right] \cdot E[(\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2)|B^l|^2] \quad (22)$$

with equality if and only if

$$B^l(\omega) \propto \frac{A^*(\omega)}{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A(\omega)|^2}. \quad (23)$$

So substituting (22) into (21), we see that (23) maximizes (21) and, in turn, (18).

Note that if the arbitrary complex constant of proportionality is set to unity, the optimal $B^l(\omega)$ can be implemented as the cascade of a matched filter $b_1[n] = a^*[-n]$ that realizes the numerator of (23) and a filter $b_2^l[n]$ that realizes the denominator of (23). Note that $b_2^l[n]$ changes with each iteration because of its dependence on ρ_x^{l-1}. Because the denominator is purely real, $b_2^l[n]$ is a conjugate-symmetric function in time that is generally infinite in extent. However, when the length L of $a[n]$ is finite, both sides of $b_2^l[n]$ quickly decay. Thus, $b_2^l[n]$ can be truncated and effectively approximated by a conjugate-symmetric FIR filter.

The optimal $D^l(\omega)$ given by (17) also warrants some discussion. Since the cascade of $A(\omega)$ with $B^l(\omega)$ is purely real, the frequency response $D^l(\omega)$ is purely real. Thus, the corresponding impulse response $d^l[n]$ is generally a two-sided, conjugate-symmetric sequence. Though infinite in extent, $d^l[n]$ can be truncated and effectively approximated by a conjugate-symmetric FIR filter, using the same reasoning as for $b_2^l[n]$. While the strictly causal feedback filter of the decision-feedback equalizer subtracts out only postcursor ISI, the noncausal nature of $d^l[n]$ allows the iterated-decision equalizer to cancel both precursor and postcursor ISI.

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

$$B^1(\omega) \propto \frac{A^*(\omega)}{\mathcal{N}_0 + \mathcal{E}_s|A(\omega)|^2} \quad (24)$$

which is the minimum mean-square error linear equalizer (MMSE-LE). Thus the performance of the iterated-decision equalizer, when using just one iteration, is identical to the performance of the MMSE-LE. In Section III-A, 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 $D^l(\omega)$, we have, substituting (22) into (21),

$$\phi^l = E\left[\frac{|A|^2}{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A|^2} \right]. \quad (25)$$

After some algebraic manipulation, the SINR from (20) then becomes

$$\gamma^l = \left(\frac{1}{E\left[\frac{1}{1+\alpha^l}\right]} - 1 \right) \cdot \frac{1}{1-(\rho_x^{l-1})^2} \quad (26)$$

where

$$\alpha^l(\omega) = \frac{\mathcal{E}_s(1-(\rho_x^{l-1})^2)|A(\omega)|^2}{\mathcal{N}_0}. \quad (27)$$

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

$$\frac{1}{\xi^l} = \frac{(1-(\rho_x^{l-1})^2)}{\zeta} \quad (28)$$

where

$$\frac{1}{\zeta} = \frac{\mathcal{E}_s L \sigma_a^2}{\mathcal{N}_0} \quad (29)$$

is the expected SNR at which the transmission is received. Thus, using the identity [1]

$$\int_0^\infty \frac{e^{-st}}{1+t}\, dt = e^s E_1(s) \quad (30)$$

where

$$E_1(s) = \int_s^\infty \frac{e^{-t}}{t}\, dt \quad (31)$$

is the exponential integral, it is straightforward to show that

$$E\left[\frac{1}{1+\alpha^l}\right] = \int_0^\infty \left(\frac{1}{1+\alpha}\right) \xi^l e^{-\xi^l \alpha}\, d\alpha = \xi^l e^{\xi^l} E_1(\xi^l). \quad (32)$$

Substituting (32) back into (26), our simplified SINR expression is

$$\gamma^l = \left(\frac{1}{\xi^l e^{\xi^l} E_1(\xi^l)} - 1 \right) \cdot \frac{1}{1-(\rho_x^{l-1})^2}. \quad (33)$$

Equation (33) can, in turn, be used in the following convenient iterative algorithm for determining the set of correlation coefficients ρ_x^l to be used at each iteration, and simultaneously predicting the associated sequence of symbol error probabilities.

1) Set $\rho_x^0 = 0$ and let $l = 1$.
2) Compute the SINR γ^l at the slicer input on the lth decoding pass from ρ_x^{l-1} via (33), (28), and (29). [It is worth pointing out that for shorter ISI channels, we can alternatively (and in some cases more accurately) compute γ^l

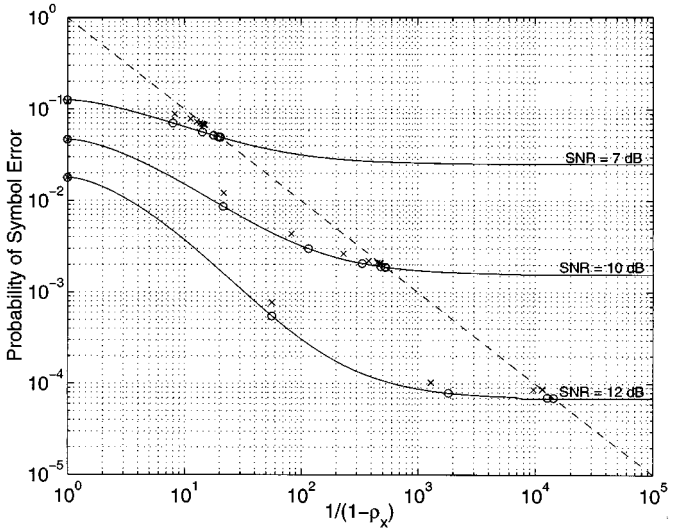

Fig. 2. Iterated-decision equalizer performance. The successively lower solid curves plot QPSK symbol error rate as a function of the correlation coefficient ρ_x for SNRs of 7, 10, and 12 dB. Along each curve, $\circ$s identify the theoretically predicted decreasing error rates achieved with $l = 1, 2, \ldots$ decoding passes, and the intersections with the dashed line are the steady-state values ($l \to \infty$). The associated experimentally obtained values for $L = 256$ are depicted using $\times$'s.

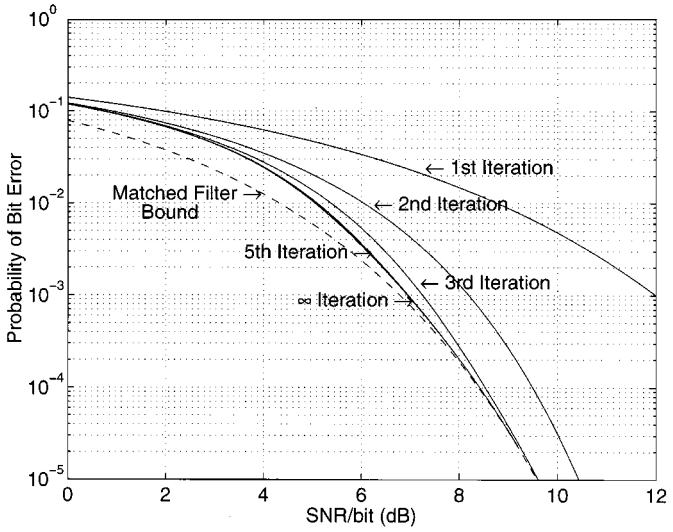

Fig. 3. Theoretical iterated-decision equalizer performance as a function of SNR per bit. The successively lower solid curves depict the QPSK BER as a function of SNR per bit for 1, 2, 3, 5, and ∞ decoding iterations. The dashed curve is the matched filter bound.

from ρ_x^{l-1} via (18) and (23), where the expectations are replaced by frequency averages.]

3) Compute the symbol error probability $\Pr(\epsilon^l)$ at the slicer output from γ^l via (14).
4) Compute the normalized correlation coefficient ρ_x^l between the symbols $x[n]$ and the decisions $\hat{x}^l[n]$ generated at the slicer via the approximation [2]

$$\rho_x^l \approx 1 - 2\sin^2\left(\frac{\pi}{M}\right)\Pr(\epsilon^l). \tag{34}$$

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 (34) with

$$\rho_x^l = 1 - 2\mathcal{Q}\left(\sqrt{\gamma^l}\right). \tag{35}$$

A. Performance

Though the iterative nature of the equalizer complicates an analysis of its performance, useful insights and approximations can nevertheless be obtained. From Steps 2 and 3 of the algorithm, we see that $\Pr(\epsilon^l)$ can be expressed as

$$\Pr(\epsilon^l) = \mathcal{F}(\zeta, \rho_x^{l-1}), \tag{36}$$

where $\mathcal{F}(\cdot, \cdot)$ is a monotonically decreasing function in both SNR $1/\zeta$ and correlation ρ_x^{l-1}. The monotonicity of $\mathcal{F}(\cdot, \cdot)$ is illustrated in Fig. 2 where the successively lower solid curves plot $\mathcal{F}(\zeta, \rho_x)$ as a function of $1/(1-\rho_x)$ for various values of $1/\zeta$. Meanwhile, from Step 4 of the algorithm, we see that we can also express $\Pr(\epsilon^l)$ as

$$\Pr(\epsilon^l) = \mathcal{G}(\rho_x^l) \tag{37}$$

where $\mathcal{G}(\cdot)$ is a monotonically decreasing function of ρ_x^l. The dashed line in Fig. 2 plots $\mathcal{G}(\rho_x)$ as a function of $1/(1-\rho_x)$.

At a given SNR $1/\zeta$, the sequence of error probabilities $\Pr(\epsilon^l)$ and correlation coefficients ρ_x^l can be obtained by starting at the left end of the solid curve (corresponding to $\rho_x^0 = 0$) 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 equalizer. In Fig. 2, the sequence of operating points is indicated on the solid curves with the $\circ$ symbols. The set of operating points obtained from Monte Carlo simulations is also indicated in Fig. 2 by the $\times$ symbols. These results suggest that the theoretical predictions are quite accurate.

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

$$\Pr(\epsilon^\infty) = \mathcal{F}(\zeta, \mathcal{G}^{-1}(\Pr(\epsilon^\infty))) \tag{38}$$

corresponding to the intersection of the dashed line and the appropriate solid curve in Fig. 2.

As Fig. 2 suggests, steady-state performance is effectively 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. 3, where BER is plotted as a function of SNR per bit for 1, 2, 3, 5, and an infinite number of iterations. It is significant that few passes are required to converge to typical target BERs, 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.

We now discuss the high-SNR ($\zeta \to 0$) performance of the equalizer. We first note from (28) that if $\rho_x \to 1$, then $\xi \to \infty$. Using the asymptotic series expansion [1]

$$E_1(t) = e^{-t}\sum_{k=0}^{\infty}(-1)^k\frac{k!}{t^{k+1}} \qquad \text{for large } t \tag{39}$$

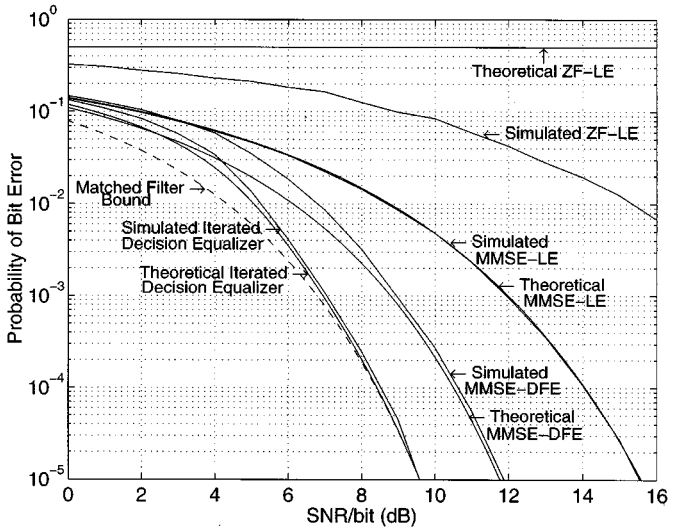

Fig. 4. Theoretical ($L \to \infty$) and experimentally observed ($L = 256$) performance for various equalizers. The solid curves depict QPSK BERs for the iterated-decision equalizer, MMSE-DFE, MMSE-LE, and ZF-LE as a function of SNR per bit. The dashed curve is the matched filter bound.

in (33) and retaining only the first two terms of the series, our SINR expression as $\rho_x \to 1$ becomes

$$\gamma_{\text{IDE}} \to \left(\frac{1}{\xi}\right) \cdot \frac{1}{1-(\rho_x)^2} = \frac{1}{\zeta} \tag{40}$$

where the equality follows from (28). When (40) is substituted into (14) or (16), we obtain the matched filter bound. Now, the asymptotic performance of the equalizer can be inferred from Fig. 2. Specifically, since the successively lower solid curves in Fig. 2 achieve increasingly higher values of ρ_x^∞ at the intersections with the dashed line, perfect ISI cancellation is thus approached at high SNR.

In Fig. 4, we compare the theoretical performance of the iterated-decision equalizer when the number of channel taps $L \to \infty$ with experimentally obtained results when $L = 256$. The experimental results are indeed consistent with theoretical predictions, especially at high SNR.

For comparison, in Fig. 4 we also plot the theoretical error rates of the ideal MMSE-DFE, the MMSE-LE, and the zero-forcing linear equalizer (ZF-LE), which are developed in Appendix B. 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 ($\zeta \to 0$), we have from Appendix B that $\gamma_{\text{MMSE-DFE}} \to 1/\zeta e^{\Gamma_0}$ and $\gamma_{\text{MMSE-LE}} \to 1/[\zeta(-\Gamma_0 - \ln\zeta)] - 1$, where $\Gamma_0 = 0.577\,21\cdots$ 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 $\zeta \to 0$, the MMSE-LE requires increasingly more transmit power than the iterated-decision equalizer to achieve the same probability of error.

Since maximum-likelihood sequence detection is optimal in terms of BER [13], the BER curve for maximum-likelihood sequence detection (not shown in Fig. 4) must lie below the BER curve for the iterated-decision equalizer but above the matched filter bound. But the theoretical ($L \to \infty$) and simulated ($L = 256$) BER curves in Fig. 4 for the iterated-decision equalizer approach the matched filter bound at high SNR, so iterated-decision equalizers asymptotically achieve the performance of maximum-likelihood sequence detection.

Some further comments on the results of Fig. 4 are worth making. First, similar experimental curves to those depicted in the figure are obtained even when the channel coefficients are not Gaussian and/or not independent. What is important is that the effective number of degrees of freedom describing the taps be large for behavior to converge to the predicted values. Thus the stronger the correlation between taps, the larger the number of them required to converge to the theoretical predictions.

It should also be stressed that when the effective number of taps (or degrees of freedom) is small, the performance of any equalizer obviously becomes realization-dependent, i.e., a function of the particular realized channel coefficients, and thus in this regime equalizer performance is less well-represented by the asymptotic expressions developed in this section of the paper. For example, simulations involving the iterated-decision equalizer with $L = 64$ channel taps show that performance close to the $L \to \infty$ case is achieved for most channels. However, about one out of every thousand 64-tap channels at an SNR per bit of 10 dB leads to a significantly higher probability of error, and about one out of every hundred 64-tap channels at an SNR per bit of 9 dB leads to a significantly higher probability of error. As another example, the simulated performance of the iterated-decision equalizer for 3-tap and 5-tap channel impulse responses was found to be strongly dependent on the particular channel realization [7]. As a useful rule of thumb, our experience suggests that the iterated-decision equalizer as we have developed it in this paper performs well (i.e., performs similarly to the MLSD and better than the MMSE-DFE) on channels whose normalized, unbiased deterministic autocorrelation function of $a[n]$ satisfies

$$R_a[n] \triangleq \frac{1}{L-|n|} \sum_k \frac{a^*[k]a[k-n]}{\left(\frac{1}{L}\sum_j |a[j]|^2\right)} \approx \delta[n]. \tag{41}$$

Essentially, this condition ensures that the channel has sufficient variation in its taps, and the larger the effective number of taps a channel has, the greater the likelihood its autocorrelation function meets this condition; see [7] for further details.

IV. Coded Systems

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 [10], [11], in a variety of practical settings it can be difficult to use effectively. In particular, in typical implementations the MMSE-DFE cancels postcursor ISI by using delay-free symbol

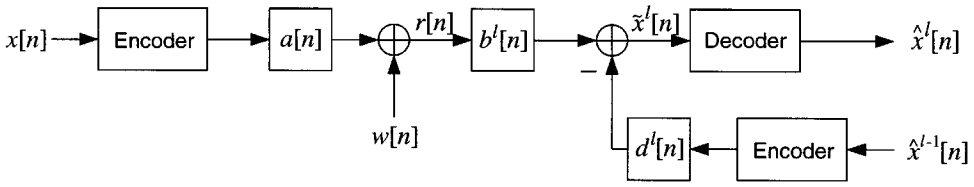

Fig. 5. Structure of a communication system that combines iterated-decision equalization with channel coding.

decisions, which in a coded system can be insufficiently reliable to yield good performance. 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. 5. Although the sequence $x[n]$ is first encoded before it is transmitted, Theorem 1 is still valid because typical trellis codes and random codes generally produce white symbol streams [6]. What makes the iterated-decision equalizer an 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.

V. Concluding Remarks

In this paper, we have developed an effective class of low-complexity nonlinear channel equalizers that rely on an iterative decoding technique to cancel ISI. These equalizers achieve a performance that approaches MLSD on severe ISI channels, yet only require complexity on the order of that of a DFE or LE. Moreover, the structure of the iterative equalizers makes them readily compatible with error-control coding, unlike DFE. Complementary results on adaptive implementations for the case in which the channel must be learned by the equalizer are developed in the companion paper [8]; an important result is that surprisingly little training data is required for the performance of adaptive implementations to approach that of the fixed channel implementations developed in this paper.

A variety of issues remain to be explored in future work. For example, the development of equalizer design and analysis techniques specifically optimized for the small-ISI scenario would also be a valuable resource for system designers to complement the large-ISI results presented here, as would tools that take into account an explicit FIR constraint on equalizer filters.

More generally, some of the richest directions for future research involve developing techniques for analyzing the behavior of systems that use coding in conjunction with iterated-decision equalization, together with useful design rules for the selection of appropriate codes, and investigating to what degree such systems can approach channel capacity. It is conceivable, for example, that counterparts to Price's result [16] or its generalizations [10], [11] could be derived for such equalizers. Finally, the iterative nature of the decoders for turbo codes [4] suggests that such codes or their variants may be particularly well matched for use with iterated-decision equalization. Thus, it may be possible to combine a turbo-style decoder and the iterated-decision equalizer into a single efficient equalizer-decoder structure comparable to the class of "turbo equalization" receivers described in [18].

Appendix A
Proof of Theorem 1

The proof requires the following pair of lemmas.

Lemma 1: Let $A(\omega)$ be a complex-valued 2π-periodic zero-mean Gaussian random process with variance σ_A^2, and let $C^l(\omega) = A(\omega)B^l(\omega)$. If $B^l(\omega)$ and $D^l(\omega)$ are admissible in the sense of Definition 1 with $B^l(\omega) = f_1(A(\omega))$ and $D^l(\omega) = f_2(A(\omega))$, then the admissibility conditions on $f_1(\cdot)$ and $f_2(\cdot)$ imply that the zero-mean random processes

$$\tilde{S}_0^l(\omega) = C^l(\omega) - E[C^l] \tag{42}$$

$$\tilde{S}_1^l(\omega) = |B^l(\omega)|^2 - E[|B^l|^2] \tag{43}$$

$$\begin{aligned}\tilde{S}_2^l(\omega) = {} & |D^l(\omega) - \rho_x^{l-1}\tilde{S}_0^l(\omega)|^2 + (1-(\rho_x^{l-1})^2)|\tilde{S}_0^l(\omega)|^2 \\ & - E[|D^l - \rho_x^{l-1}\tilde{S}_0^l|^2] - (1-(\rho_x^{l-1})^2)E[|\tilde{S}_0^l|^2]\end{aligned} \tag{44}$$

$$\tilde{S}_3^l(\omega) = D^l(\omega) \tag{45}$$

satisfy

$$\operatorname{var}\tilde{S}_i^l < \infty, \qquad \text{for } i = 0, 1, 2, 3 \tag{46}$$

$$E[|\tilde{S}_i^l|] < \infty, \qquad \text{for } i = 0, 1, 2, 3. \tag{47}$$

Proof: The proof is straightforward and follows from observing that

$$\tilde{S}_i^l(\omega) = h_i(A(\omega)), \qquad \text{for } i = 0, 1, 2, 3 \tag{48}$$

where

$$h_0(z) = zf_1(z) - \kappa_0 \tag{49}$$

$$h_1(z) = |f_1(z)|^2 - \kappa_1 \tag{50}$$

$$\begin{aligned}h_2(z) = {} & |f_2(z) - \rho_x^{l-1}(zf_1(z) - \kappa_0)|^2 \\ & + (1-(\rho_x^{l-1})^2)|zf_1(z) - \kappa_0|^2 - \kappa_2\end{aligned} \tag{51}$$

$$h_3(z) = f_2(z) \tag{52}$$

and where the κ_i are finite constants. From (49)–(52), we see that the $h_i(\cdot)$ satisfy the admissibility conditions of Definition 1

whenever $f_1(\cdot)$ and $f_2(\cdot)$ do. Using the conditions (8) and (9) in the expressions

$$\operatorname{var} \tilde{S}_i^l = \frac{1}{2\pi\sigma_A^2} \int_{-\infty}^{\infty} h_i^2(z) e^{-|z|^2/(2\sigma_A^2)}\, dz \tag{53}$$

$$E[|\tilde{S}_i^l|] = \frac{1}{2\pi\sigma_A^2} \int_{-\infty}^{\infty} |h_i(z)| e^{-|z|^2/(2\sigma_A^2)}\, dz \tag{54}$$

for $i = 0, 1, 2, 3$, we obtain our desired results. □

Lemma 2: Let $A(\omega)$ be a complex-valued 2π-periodic zero-mean Gaussian random process with variance σ_A^2 and normalized correlation function denoted by

$$\rho_A(\omega, \nu) = \frac{E[A(\omega)A^*(\nu)]}{\sigma_A^2}. \tag{55}$$

Furthermore, let the random process $\tilde{S}(\omega)$ be defined via

$$\tilde{S}(\omega) = h(A(\omega)) \tag{56}$$

for some function $h(\cdot)$ such that

$$E[\tilde{S}(\omega)] = 0 \tag{57}$$

$$E[|\tilde{S}(\omega)|] < \infty \tag{58}$$

$$\operatorname{var} \tilde{S}(\omega) < \infty. \tag{59}$$

Then, if[8] $\rho_A(\omega, \nu) \xrightarrow{\text{p.w.a.e.}} 0$, we have that

$$\mathcal{I} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |E[\tilde{S}(\omega)\tilde{S}^*(\nu)]|\, d\omega\, d\nu \to 0. \tag{60}$$

The proof of Lemma 2 is analogous to the proof of Lemma 1 in [21].

We now proceed to a proof of our main result.

First, we write

$$\tilde{x}^l[n] = v_1^l[n] + z^l[n] \tag{61}$$

where

$$v_1^l[n] = b^l[n] * w[n] \tag{62}$$

and

$$z^l[n] = c^l[n] * x[n] - d^l[n] * \hat{x}^{l-1}[n] \tag{63}$$

with

$$c^l[n] = b^l[n] * a[n]. \tag{64}$$

Let us consider $v_1^l[n]$ as defined in (62) first.

We obtain the mean and covariance, respectively, of $v_1^l[n]$ given a channel impulse response $a[\cdot]$ as

$$E[v_1^l[n]|a] = 0 \tag{65}$$

and

$$R_{v_1^l v_1^l|a}[n] = \frac{\mathcal{N}_0}{2\pi} \int_{-\pi}^{\pi} |B^l(\omega)|^2 e^{j\omega n}\, d\omega. \tag{66}$$

[8]We use $\xrightarrow{\text{p.w.a.e.}}$ to denote pointwise convergence almost everywhere.

In turn, averaging over the possible channel responses $a[\cdot]$, we obtain

$$\begin{aligned} E[R_{v_1^l v_1^l|a}[n]] &= E[|B^l|^2] \frac{\mathcal{N}_0}{2\pi} \int_{-\pi}^{\pi} e^{j\omega n}\, d\omega \\ &= \mathcal{N}_0 E[|B^l|^2] \delta[n] \end{aligned} \tag{67}$$

where we have used the fact that since the mean and variance of $A(\omega)$ are independent of ω, so are those of $B^l(\omega)$. Next, we define

$$\begin{aligned} \tilde{R}_{v_1^l v_1^l|a}[n] &\triangleq R_{v_1^l v_1^l|a}[n] - E[R_{v_1^l v_1^l|a}[n]] \\ &= \frac{\mathcal{N}_0}{2\pi} \int_{-\pi}^{\pi} \tilde{S}_1^l(\omega) e^{j\omega n}\, d\omega \end{aligned} \tag{68}$$

where $\tilde{S}_1^l(\omega)$ is given in (43). Then,

$$\begin{aligned} &\operatorname{var} R_{v_1^l v_1^l|a}[n] \\ &= E[|\tilde{R}_{v_1^l v_1^l|a}[n]|^2] \\ &= \left(\frac{\mathcal{N}_0}{2\pi}\right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E[\tilde{S}_1^l(\omega)(\tilde{S}_1^l(\nu))^*] e^{j(\omega-\nu)n} d\omega\, d\nu. \end{aligned} \tag{69}$$

Now, the channel frequency response $A(\omega)$ as defined in (2) has the following property:

$$\rho_A(\omega, \nu) = \frac{\operatorname{cov}(A(\omega), A(\nu))}{\sqrt{\operatorname{var} A(\omega)}\sqrt{\operatorname{var} A(\nu)}} \xrightarrow{\text{p.w.a.e.}} 0 \tag{70}$$

as $L \to \infty$ when $\omega \neq \nu$. Applying Lemma 1 followed by Lemma 2 with (70) to (69), we then obtain, for each n

$$\operatorname{var} R_{v_1^l v_1^l|a}[n] \to 0. \tag{71}$$

Hence, combining (67) with (71), we have, for any particular channel response $a[\cdot]$

$$R_{v_1^l v_1^l|a}[n] \xrightarrow{\text{m.s.}} \mathcal{N}_0 E[|B^l|^2] \delta[n] \tag{72}$$

for each n.

Looking next at $z^l[n]$ as defined in (63), we express $z^l[n]$ in the form

$$z^l[n] = v_2^l[n] + E[C^l]x[n]. \tag{73}$$

Then

$$v_2^l[n] = \tilde{c}^l[n] * x[n] - d^l[n] * \hat{x}^{l-1}[n] \tag{74}$$

where

$$\tilde{c}^l[n] = c^l[n] - E[C^l]\delta[n] = c^l[n] - E[c^l[n]]. \tag{75}$$

Therefore, $\tilde{c}^l[n]$ is a zero-mean sequence.

Again, for a fixed realization of $a[\cdot]$ (and hence $c^l[n]$ and $d^l[n]$), we have

$$E[v_2^l[n]|a] = \tilde{c}^l[n] * E[x[n]] - d^l[n] * E[\hat{x}^{l-1}[n]] = 0 \tag{76}$$

and

$$\begin{aligned} &R_{v_2^l v_2^l|a}[n] \\ &\quad = \frac{\mathcal{E}_s}{2\pi}\int_{-\pi}^{\pi}\{|D^l(\omega)|^2 - \rho_x^{l-1}D^l(\omega)(\tilde{C}^l(\omega))^* \\ &\qquad\qquad - \rho_x^{l-1}(D^l(\omega))^*\tilde{C}^l(\omega) + |\tilde{C}^l(\omega)|^2\}e^{j\omega n}\,d\omega \\ &\quad = \frac{\mathcal{E}_s}{2\pi}\int_{-\pi}^{\pi}\{|D^l(\omega) - \rho_x^{l-1}\tilde{C}^l(\omega)|^2 \\ &\qquad\qquad + (1-(\rho_x^{l-1})^2)|\tilde{C}^l(\omega)|^2\}e^{j\omega n}\,d\omega. \end{aligned} \tag{77}$$

However, (77) is asymptotically independent of $a[\cdot]$. To see this, first note that

$$\begin{aligned} &E[R_{v_2^l v_2^l|a}[n]] \\ &\quad = \{E[|D^l - \rho_x^{l-1}\tilde{C}^l|^2] + (1-(\rho_x^{l-1})^2)E[|\tilde{C}^l|^2]\} \\ &\qquad \cdot \frac{\mathcal{E}_s}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n}\,d\omega \\ &\quad = \mathcal{E}_s\{E[|D^l - \rho_x^{l-1}\tilde{C}^l|^2] + (1-(\rho_x^{l-1})^2)E[|\tilde{C}^l|^2]\}\delta[n] \end{aligned} \tag{78}$$

where we have used the fact that the bracketed expression on the right-hand side of (78) is the expectation of an admissible filter and thus independent of ω. Then, since

$$\begin{aligned} \tilde{R}_{v_2^l v_2^l|a}[n] &\triangleq R_{v_2^l v_2^l|a}[n] - E[R_{v_2^l v_2^l|a}[n]] \\ &= \frac{\mathcal{E}_s}{2\pi}\int_{-\pi}^{\pi}\tilde{S}_2^l(\omega)e^{j\omega n}\,d\omega \end{aligned} \tag{79}$$

where $\tilde{S}_2^l(\omega)$ is as defined in (44), we have

$$\begin{aligned} &\operatorname{var} R_{v_2^l v_2^l|a}[n] \\ &\quad = E[|\tilde{R}_{v_2^l v_2^l|a}[n]|^2] \\ &\quad = \left(\frac{\mathcal{E}_s}{2\pi}\right)^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}E[\tilde{S}_2^l(\omega)(\tilde{S}_2^l(\nu))^*]e^{j(\omega-\nu)n}\,d\omega\,d\nu. \end{aligned} \tag{80}$$

Hence, again using (70) and applying, in turn, Lemmas 1 and 2 to (80), we then obtain, for each n,

$$\operatorname{var} R_{v_2^l v_2^l|a}[n] \to 0. \tag{81}$$

Hence, combining (78) with (81), we have, for any particular channel response $a[\cdot]$,

$$\begin{aligned} R_{v_2^l v_2^l|a}[n] \xrightarrow{\text{m.s.}} \mathcal{E}_s\{&E[|D^l - \rho_x^{l-1}\tilde{C}^l|^2] \\ &+(1-(\rho_x^{l-1})^2)E[|\tilde{C}^l|^2]\}\delta[n] \end{aligned} \tag{82}$$

for each n.

Thus, we can write

$$\tilde{x}^l[n] = E[C^l]x[n] + v^l[n] \tag{83}$$

where

$$v^l[n] = v_1^l[n] + v_2^l[n] \tag{84}$$

is the equivalent noise. However, since $w[n]$ is statistically independent of $x[n]$ and can be assumed to be independent of $\hat{x}^{l-1}[n]$,

$$E[v_1^l[n](v_2^l[k])^*|a] = 0, \qquad \text{for all } n \text{ and } k \tag{85}$$

and hence

$$R_{v^l v^l|a}[n] = R_{v_1^l v_1^l|a}[n] + R_{v_2^l v_2^l|a}[n] \tag{86}$$

which, using (72) and (82) with (75) and (64), yields (12).

Finally, we need to show that for a given realization of the channel response $a[n]$ that $x[n]$ and $v^l[n]$ are asymptotically uncorrelated. Due to (84), it suffices to show that $x[n]$ is asymptotically uncorrelated with $v_1^l[n]$ and $v_2^l[n]$ individually.

First, using (62), we have

$$E[v_1^l[n]x^*[k]] = \sum_m E[w[m]x^*[k]]b^l[n-m] = 0 \tag{87}$$

where the last equality follows from the fact that the processes $w[n]$ and $x[n]$ are statistically independent.

Next, using (74), we have

$$\begin{aligned} E[v_2^l[n]x^*[k]] &= \sum_m E[x[m]x^*[k]]\tilde{c}^l[n-m] \\ &\quad - \sum_m E[\hat{x}^{l-1}[m]x^*[k]]d^l[n-m] \\ &= \mathcal{E}_s\tilde{c}^l[n-k] - \rho_x^{l-1}\mathcal{E}_s d^l[n-k] \end{aligned} \tag{88}$$

where the last equality follows from (10) and the fact that the symbol stream $x[n]$ is white. Thus, it remains only to show that $\tilde{c}^l[n] \xrightarrow{\text{m.s.}} 0$ and $d^l[n] \xrightarrow{\text{m.s.}} 0$ for all n.

To see this, we first note that

$$E[\tilde{c}^l[n]] = 0 \tag{89}$$

$$E[d^l[n]] = 0 \tag{90}$$

where (90) follows from the fact that $E[d^l[n]] = E[D^l]\delta[n]$ and the constraint (6). Next

$$\begin{aligned} &\operatorname{var}\tilde{c}^l[n] \\ &\quad = \left(\frac{1}{2\pi}\right)^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}E[\tilde{C}^l(\omega)(\tilde{C}^l(\nu))^*]e^{j(\omega-\nu)n}\,d\omega\,d\nu \end{aligned} \tag{91}$$

$$\begin{aligned} &\operatorname{var} d^l[n] \\ &\quad = \left(\frac{1}{2\pi}\right)^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}E[D^l(\omega)(D^l(\nu))^*]e^{j(\omega-\nu)n}\,d\omega\,d\nu. \end{aligned} \tag{92}$$

Hence, again using (70) and applying, in turn, Lemmas 1 and 2 to each of (91) and (92), we then obtain, for each n,

$$\operatorname{var} \tilde{c}^l[n] \to 0 \tag{93}$$

$$\operatorname{var} d^l[n] \to 0. \tag{94}$$

Hence, combining (89) and (90) with (93) and (94), we obtain the desired results.

Appendix B
Performance Predictions for Classical Equalizers

In this Appendix, we develop, for comparison with our results on the iterated-decision equalizer, performance predictions for the ideal MMSE-DFE, ideal ZF-DFE, MMSE-LE, and ZF-LE for the severe-ISI channel model described in Section II.

The input $\tilde{x}[n]$ to any equalizer slicer can be expressed as

$$\tilde{x}[n] = \mu x[n] + v[n] \tag{95}$$

where μ is a constant and $v[n]$ is the complex-valued interference + noise component. When $v[n]$ is uncorrelated with $x[n]$, using the slicer input SINR γ in conjunction with (14) or (16) leads to a convenient M-ary PSK performance approximation.

This condition is frequently met. For example, for the ideal ZF-DFE and ZF-LE, the error component $v[n]$ at the slicer is, in fact, a Gaussian process that is uncorrelated with $x[n]$ since ISI is completely removed. Moreover, $v[n]$ is a white process for the ZF-DFE. For the MMSE-LE, since it is equivalent to the first iteration of the multipass equalizer, we know $v[n]$ is similarly uncorrelated and white. And similar arguments suggest that the MMSE-DFE shares this property.

To obtain $\gamma = |\mu|^2 \mathcal{E}_s / \operatorname{var} v[n]$ for these equalizers then, we first evaluate the mean-square slicer error. For a given channel realization, the mean-square slicer error for each of the equalizers is [14]

$$\varepsilon_{\text{MMSE-DFE}} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\frac{\mathcal{E}_s}{\alpha(\omega)+1}\right) d\omega\right\} \tag{96}$$

$$\varepsilon_{\text{ZF-DFE}} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\frac{\mathcal{E}_s}{\alpha(\omega)}\right) d\omega\right\} \tag{97}$$

$$\varepsilon_{\text{MMSE-LE}} = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\mathcal{E}_s}{\alpha(\omega)+1}\right) d\omega\right\} \tag{98}$$

$$\varepsilon_{\text{ZF-LE}} = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\mathcal{E}_s}{\alpha(\omega)}\right) d\omega\right\} \tag{99}$$

where

$$\alpha(\omega) = \frac{\mathcal{E}_s |A(\omega)|^2}{\mathcal{N}_0}. \tag{100}$$

Next, to develop their characteristics in the large-ISI limit, we make use of (70) together with the fact that $A(\omega)$ is a Gaussian random process, which implies that arbitrarily close samples of $A(\omega)$, and in turn $\alpha(\omega)$, are effectively independent and identically distributed (iid). Thus, it is convenient to interpret the bracketed expressions in (96)–(99) effectively as averages of iid random variables with means [7]

$$E\left[\ln\left(\frac{\mathcal{E}_s}{\alpha+1}\right)\right] = \ln \mathcal{E}_s - e^{\zeta} E_1(\zeta) \tag{101}$$

$$E\left[\ln\left(\frac{\mathcal{E}_s}{\alpha}\right)\right] = \ln \mathcal{E}_s + \ln \zeta + \Gamma_0 \tag{102}$$

$$E\left[\frac{\mathcal{E}_s}{\alpha+1}\right] = \mathcal{E}_s \zeta e^{\zeta} E_1(\zeta) \tag{103}$$

$$E\left[\frac{\mathcal{E}_s}{\alpha}\right] \to \infty, \tag{104}$$

where ζ is defined in (29) and $\Gamma_0 = 0.577\,21\cdots$ denotes Euler's constant. An application of the strong law of large numbers [12] in each case would then suggest, substituting (101)–(104) for the respective bracketed expressions in (96)–(99), that[9]

$$\varepsilon_{\text{MMSE-DFE}} \xrightarrow{\text{a.e.}} \mathcal{E}_s \exp\{-e^{\zeta} E_1(\zeta)\} \tag{105}$$

$$\varepsilon_{\text{ZF-DFE}} \xrightarrow{\text{a.e.}} \mathcal{E}_s \zeta e^{\Gamma_0} \tag{106}$$

$$\varepsilon_{\text{MMSE-LE}} \xrightarrow{\text{a.e.}} \mathcal{E}_s \zeta e^{\zeta} E_1(\zeta) \tag{107}$$

$$\varepsilon_{\text{ZF-LE}} \xrightarrow{\text{a.e.}} \infty. \tag{108}$$

Finally, since the SINRs at the slicer input[10] are related to the mean-square slicer error according to [10]

$$\gamma_{\text{MMSE-DFE}} = \frac{\mathcal{E}_s}{\varepsilon_{\text{MMSE-DFE}}} - 1 \tag{109}$$

$$\gamma_{\text{ZF-DFE}} = \frac{\mathcal{E}_s}{\varepsilon_{\text{ZF-DFE}}} \tag{110}$$

$$\gamma_{\text{MMSE-LE}} = \frac{\mathcal{E}_s}{\varepsilon_{\text{MMSE-LE}}} - 1 \tag{111}$$

$$\gamma_{\text{ZF-LE}} = \frac{\mathcal{E}_s}{\varepsilon_{\text{ZF-LE}}}, \tag{112}$$

we substitute (105)–(108) into the corresponding (109)–(112) and conclude that

$$\gamma_{\text{MMSE-DFE}} \xrightarrow{\text{a.e.}} \exp\{e^{\zeta} E_1(\zeta)\} - 1 \tag{113}$$

$$\gamma_{\text{ZF-DFE}} \xrightarrow{\text{a.e.}} \frac{1}{\zeta e^{\Gamma_0}} \tag{114}$$

$$\gamma_{\text{MMSE-LE}} \xrightarrow{\text{a.e.}} \frac{1}{\zeta e^{\zeta} E_1(\zeta)} - 1 \tag{115}$$

$$\gamma_{\text{ZF-LE}} \to 0. \tag{116}$$

In the high-SNR regime ($\zeta \to 0$), we can use the series expansion [1]

$$E_1(t) = -\Gamma_0 - \ln t - \sum_{k=1}^{\infty} \frac{(-1)^k t^k}{t\, t!} \tag{117}$$

to show that (113) and (115), respectively, satisfy

$$\gamma_{\text{MMSE-DFE}} \to \exp(-\Gamma_0 - \ln \zeta) = \frac{1}{\zeta e^{\Gamma_0}} \tag{118}$$

$$\gamma_{\text{MMSE-LE}} \to \frac{1}{\zeta(-\Gamma_0 - \ln \zeta)} - 1. \tag{119}$$

[9] We use $\xrightarrow{\text{a.e.}}$ to denote convergence almost everywhere.

[10] The SINR for the MMSE-DFE is not the same as what Cioffi *et al.* [10] call the $\text{SNR}_{\text{MMSE-DFE}}$, which is defined as

$$\text{SNR}_{\text{MMSE-DFE}} = \frac{\mathcal{E}_s}{\varepsilon_{\text{MMSE-DFE}}}.$$

However, the SINR for the MMSE-DFE is equivalent to $\text{SNR}_{\text{MMSE-DFE,U}}$, which is the SINR for the unbiased MMSE-DFE [10]. Indeed, the SINR's for the standard MMSE-DFE and the unbiased MMSE-DFE are identical, since the unbiased MMSE-DFE merely scales both the signal and the associated interference + noise process by the same factor at the slicer input.

We conclude with some brief comments arising out the relationships between these expressions, and the quality of these approximations, as depicted in Fig. 4. First, note that (114) and (118) coincide in the high SNR limit, as one might expect since the MMSE-DFE can completely eliminate ISI in the absence of noise, which is the constraint imposed by the ZF-DFE.

The actual performance of the MMSE-DFE matches its prediction reasonably well in general, though there is some small discrepancy due to error propagation. However, the actual ZF-DFE performance deviates dramatically from its prediction (not shown in Fig. 4), due in part to numerical instabilities and error propagation in the implementation of the ZF-DFE for severe-ISI channels.

Next, a comparison of (119) to $\gamma = 1/\zeta$ for the matched filter bound reveals that as the SNR gets larger, the SNR gap between the performance curve for the MMSE-LE and the matched filter bound gets arbitrarily large. Moreover, the ZF-LE is 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 [5]. These results emphasize the strong suboptimality of linear equalizers.

Finally, note that the actual MMSE-LE performance does track its prediction quite closely, while the ZF-LE prediction is somewhat overly pessimistic: short of the asymptotic limit, at least some zeros lie quite far away from the unit circle.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1965.

[2] S. Beheshti, S. H. Isabelle, and G. W. Wornell, "Joint intersymbol and multiple-access interference suppression algorithms for CDMA systems," *Euro. Trans. Telecomm. Related Technol.*, vol. 9, pp. 403–418, Sept.–Oct. 1998.

[3] C. A. Belfiore and J. H. Park, Jr., "Decision-feedback equalization," *Proc. IEEE*, vol. 67, pp. 1143–1156, Aug. 1979.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. 1993 Int. Conf. Commun.*, vol. 2, May 1993, pp. 1064–1070.

[5] A. T. Bharucha-Reid and M. Sambandham, *Random Polynomials*. Orlando, FL: Academic, 1986.

[6] E. Biglieri, "Ungerboeck codes do not shape the signal power spectrum," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 595–596, July 1986.

[7] A. M. Chan, "A class of batch-iterative methods for the equalization of intersymbol interference channels," S.M. Thesis, M.I.T., Cambridge, MA, Aug. 1999.

[8] A. M. Chan and G. W. Wornell, "A class of block-iterative equalizers for intersymbol interference channels: Adaptive Implementations," IEEE Trans. Commun., to be published.

[9] ——, "A class of asymptotically optimum iterated-decision multiuser detectors," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Salt Lake, UT, May 2001.

[10] 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, pp. 2582–2594, Oct. 1995.

[11] ——, "MMSE decision-feedback equalizers and coding—Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2595–2604, Oct. 1995.

[12] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1968, vol. I.

[13] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[14] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Boston, MA: Kluwer, 1994.

[15] M. J. Lopez, A. C. Singer, S. L. Whitney, and G. S. Edelson, "A DFE coefficient placement algorithm for underwater digital acoustic communications," in *OCEANS '99 MTS/IEEE*, vol. 2, 1999, pp. 996–1001.

[16] R. Price, "Nonlinearly feedback-equalized PAM versus capacity for noisy filter channels," in *Proc. 1972 IEEE Int. Conf. Commun.*, June 1972, pp. 22-12–22-17.

[17] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[18] M. Tüchler, R. Kötter, and A. Singer, "'Turbo equalization': Principles and new results," *IEEE Trans. Commun.*, submitted for publication.

[19] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, pp. 36–38, Feb. 1998.

[20] G. W. Wornell, "Spread-response precoding for communication over fading channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 488–501, Mar. 1996.

[21] 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, pp. 191–205, Jan. 1997.

[22] A. M. Chan and G. W. Wornell, "A new class of efficient block-iterative interference cancellation techniques for digital communication receivers," *J. VLSI Signal Processing*, to be published.

Albert M. Chan (S'96) was born in Toronto, Canada, in 1975. He 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), Cambridge, 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. He has served as a Teaching Assistant for courses on probability, digital signal processing, and signals and systems with the MIT Department of Electrical Engineering and Computer Science.

Mr. Chan is the recipient of the MIT Frederick C. Hennie III Award for Teaching Excellence (2000).

Gregory W. Wornell (S'83–M'91–SM'00) 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 (MIT), Cambridge, all in electrical engineering and computer science, in 1985, 1987, and 1991, respectively.

Since 1991 he has been with the Department of Electrical Engineering and Computer Science at MIT, where he is currently an Associate Professor. He has spent leave of absences with the University of California, Berkeley, during 1999–2000 and with AT&T Bell Laboratories, Murray Hill, NJ, during 1992–1993. His research interests include signal processing, communication systems, and information theory, and algorithms and architectures for wireless networks, broadband systems, and multimedia environments. He is also active in industry, most recently, cofounding Chinook Communications Inc., Lexington, MA, and is an inventor on numerous issued and pending patents. He is an author of a number of papers in these areas, as well as the monograph *Signal Processing With Fractals: A Wavelet-Based Approach* (Englewood Cliffs, NJ: Prentice-Hall), and is coeditor (with H. V. Poor) of the collection *Wireless Communications: Signal Processing Perspectives* (Englewood Cliffs, NJ: Prentice-Hall.

Dr. Wornell has served as Associate Editor for Communications for the IEEE SIGNAL PROCESSING LETTERS, and serves on the Communications Technical Committee of the Signal Processing Society. 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). He is also a member of Tau Beta Pi and Sigma Xi.