A Class of Block-Iterative Equalizers For Intersymbol Interference Channels

Albert M. Chan and Gregory W. Wornell

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract—**A new and efficient class of nonlinear equalizers is introduced 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. Asymptotically they achieve the performance of maximum-likelihood sequence detection (MLSD), but only have a computational complexity on the order of a linear equalizer (LE). And because their structure allows cancellation of both pre- and post-cursor ISI, iterated-decision equalizers outperform the minimum mean-square error decision-feedback equalizer (DFE) by 2.5 dB on severe ISI channels even with uncoded systems. Even more importantly, unlike the DFE, iterated-decision equalizers can be readily used in conjunction with error-control coding, making them attractive for a wealth of applications.**

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 often suffer from excessive noise enhancement. Maximum-likelihood sequence detection (MLSD) 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) [1] 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.

The key limitations of the DFE arise out of the sequential way in which this equalizer 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, 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.

This work has been supported in part by ONR under Grant No. N00014-96-1-0930, NSF under Grant No. CCR-9979363, and Qualcomm, Inc.

In this paper, we introduce a remarkably efficient multipass equalizer that is a particularly attractive alternative to all these classical equalizers. In particular, this new equalizer achieves the performance of MLSD, retains the low complexity of the LE, and is fully compatible with the use of coding.

II. Discrete-Time 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 coded or uncoded symbols $x[n]$, each with energy $\mathcal{E}_s$ (typical trellis codes used with random bit streams generally produce white symbol streams [2], as do random codes). The symbols $x[n]$ are corrupted by a convolution with the channel impulse response, $a[n]$, and by additive noise, $w[n]$, to produce the received symbols[1] $r[n] = a[n] * x[n] + w[n]$. 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 $x[n]$. (To simplify the exposition, we focus on symbol-spaced equalization. However, the fractionally spaced generalizations required in practice follow in a straightforward manner, as developed in [3].)

In Section III, we focus on the case in which the receiver has exact knowledge of $a[n]$ in order to develop the basic theory and fundamental limits. In Section IV, we consider the case of practical interest in which the receiver does not have *a priori* knowledge of the channel, and develop a corresponding adaptive equalizer. 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.

Also, 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

[1] The symbol $*$ denotes the convolution operation.

0-7803-6283-7/00/$10.00 © 2000 IEEE

data $x[n]$ and the noise $w[n]$. Note 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].

III. The Iterated-Decision Equalizer

The iterated-decision equalizer we now develop, which can be viewed as an analog to the multipass multiuser detector developed in [5], processes the received data in a block-iterative fashion. Specifically, during each iteration or "pass," a linear filter is applied to all the 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 $\tilde{r}^l[n] = b^l[n] * r[n]$. 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., $\tilde{x}^l[n] = \tilde{r}^l[n] - \hat{z}^l[n]$, where $\hat{z}^l[n] = d^l[n] * \hat{x}^{l-1}[n]$. Since $\hat{z}^l[n]$ is intended to be some kind of ISI estimate, we impose the constraint that the zero-delay tap of $d^l[n]$ be zero or, equivalently, that $\frac{1}{2\pi}\int_{-\pi}^{\pi} D^l(\omega)d\omega = 0$. The slicer then generates the hard decisions $\hat{x}^l[n]$ from $\tilde{x}^l[n]$ using a minimum-distance rule. (Note that if error-control coding is employed, it suffices to replace the slicer with, e.g., the appropriate soft-decision decoder for the code.)

Analogous to counterparts in [5] and [4], we have that when $x[n]$ and $\hat{x}^{l-1}[n]$ are sequences of zero-mean uncorrelated symbols, each with energy $\mathcal{E}_s$, such that their normalized correlation is of the form $E[x^*[n] \cdot \hat{x}^{l-1}[k]]/\mathcal{E}_s \approx \rho_x^{l-1}\delta[n-k]$, then the slicer input at each iteration can be expressed as

$$\tilde{x}^l[n] \approx E[AB^l]x[n] + v^l[n] \tag{1}$$

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 whose variance is a function of $B^l(\omega)$ and $D^l(\omega)$, and where the accuracy of the approximation in (1) increases with the length L of the impulse response $a[n]$.

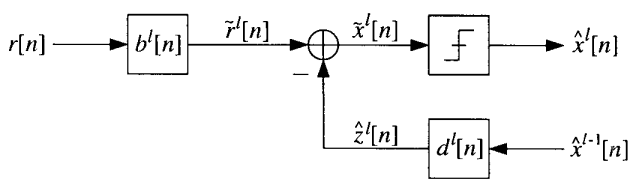

Fig. 1. Iterated-decision equalizer structure.

The second-order model (1) turns out to be a useful one for analyzing and optimizing the performance of the iterated-decision equalizer. In particular, the signal-to-interference+noise ratio (SINR) at the slicer input during each pass, defined as $\gamma^l = \mathcal{E}_s|E[AB^l]|^2/\mathrm{var}\, v^l[n]$, achieves a maximum value of

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

when

$$B^l(\omega) \propto \frac{A^*(\omega)}{\mathcal{N}_0 + \mathcal{E}_s(1-(\rho_x^{l-1})^2)|A(\omega)|^2} \tag{3}$$

$$D^l(\omega) = \rho_x^{l-1}\left(A(\omega)B^l(\omega) - E[AB^l]\right). \tag{4}$$

The optimal $D^l(\omega)$ 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 though, the correlation coefficient ρ_x^{l-1} describes our confidence in the quality of the estimate $\hat{x}^{l-1}[n]$. Thus, while the strictly causal feedback filter of the DFE subtracts out only postcursor ISI, the noncausal nature of the feedback filter allows the iterated-decision equalizer to cancel both precursor and postcursor ISI. Note that the center tap of $d^l[n]$ is indeed zero, as stipulated earlier. Note also that during the first ($l = 1$) 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.

Next, the properties of $v^l[n]$ imply 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 [6]

$$\Pr(\epsilon^l) = 2\mathcal{Q}\left(\sin\left(\frac{\pi}{M}\right)\sqrt{2\gamma^l}\right), \tag{5}$$

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

By substituting (3) and (4) into (2) and exploiting the fact that arbitrarily close samples of the random process $A(\omega)$ are asymptotically independent in the limit as $L \to \infty$, it can be shown that

$$\gamma^l = \left(\frac{1}{\xi^l e^{\xi^l} E_1(\xi^l)} - 1\right) \cdot \frac{1}{1-(\rho_x^{l-1})^2} \tag{6}$$

where $E_1(s) = \int_s^\infty e^{-t}/t\, dt$ and $\xi^l = \zeta/(1-(\rho_x^{l-1})^2)$, with $1/\zeta = \mathcal{E}_s L\sigma_a^2/\mathcal{N}_0$ being the expected SNR at which the transmission is received.

As the above development suggests, successful implementation of the iterated-decision equalizer requires that the normalized correlation between the symbols $x[n]$ and $\hat{x}^{l-1}[n]$ at the lth pass be computable. The iterative algorithm for computing the set of correlation coefficients

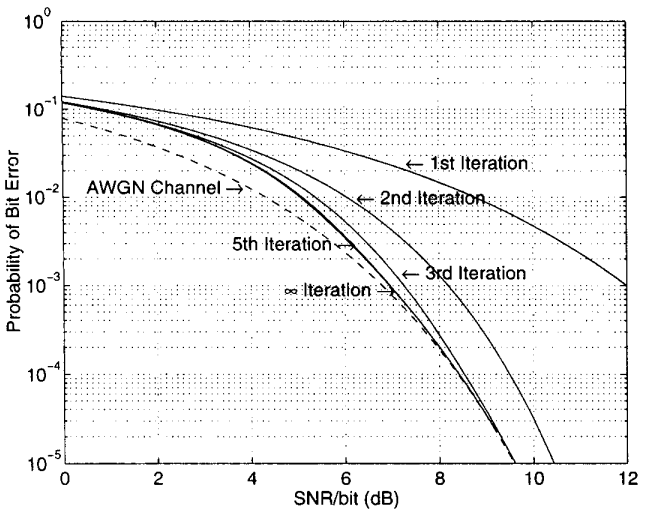

Fig. 2. Theoretical QPSK bit-error rate for the iterated-decision equalizer as a function of SNR per bit and the number of decoding iterations.

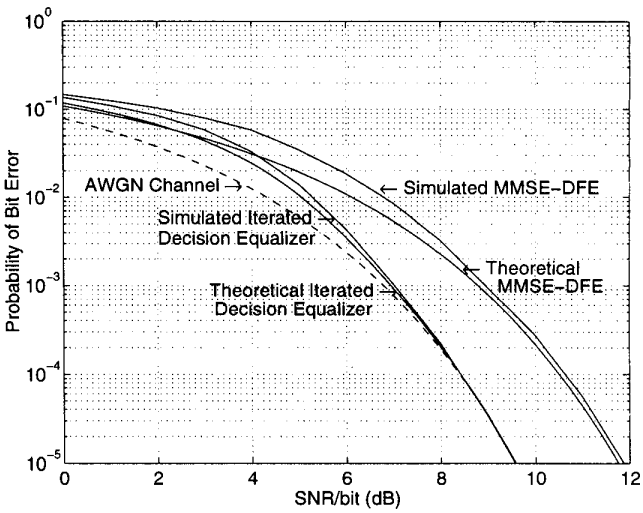

Fig. 3. Theoretical and experimentally observed QPSK bit-error rate for the iterated-decision equalizer and the MMSE-DFE as a function of SNR per bit.

ρ_x^l, and in turn predicting the sequence of symbol error probabilities is as follows.

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 (6). [It is worth pointing out that for shorter ISI channels, we can alternatively (and in some cases more accurately) compute γ^l from ρ_x^{l-1} via (2), 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 (5).
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 [5]

$$\rho_x^l = 1 - 2\sin^2\left(\frac{\pi}{M}\right)\Pr(\epsilon^l). \tag{7}$$

5. Increment l and go to step 2.

A straightforward analysis of the algorithm reveals that the sequence of error probabilities $\Pr(\epsilon^1), \Pr(\epsilon^2), \ldots$ is monotonically decreasing, suggesting that additional iterations always improve performance. However, the error rate performance for a given SNR of $1/\zeta$ eventually converges to a steady-state value of $\Pr(\epsilon^\infty)$.

In Fig. 2, bit error rate is plotted as a function of SNR 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 the multipass equalizer is comparable to that of the DFE or the LE. It is also significant to note that as the SNR increases ($\zeta \to 0$), the slicer input SINR $\gamma \to 1/\zeta$ which, when substituted into (5), gives the performance of the classical AWGN channel. Thus, perfect ISI cancellation is approached at high SNR.

We plot in Fig. 3 the theoretical performance of the multipass equalizer and the ideal minimum mean-square error decision-feedback equalizer (MMSE-DFE) as $L \to \infty$ [3], as well as the associated experimental performance when $L = 256$. We can readily see that at moderate to high SNR, the iterated-decision equalizer requires significantly less transmit power than the MMSE-DFE to achieve the same error rate. In fact, at high SNR, the iterated-decision equalizer theoretically requires $10\Gamma_0 \log e \approx 2.507$ dB less transmit power [3] to achieve the same probability of error as the ideal MMSE-DFE, where $\Gamma_0 = 0.57721\cdots$ denotes Euler's constant.

IV. The Adaptive Iterated-Decision Equalizer

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.

The multipass 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 $n = 0, 1, \ldots, N-1$ is

transmitted; some of the symbols (not necessarily at the head of the packet) are for training, while the rest are data symbols.

The structure of the adaptive multipass equalizer is similar to that of the multipass equalizer described in Section III. (Multichannel generalizations follow in a straightforward manner, as developed in [3].) The difference is that the filters $b^l[n]$ and $d^l[n]$ for the lth iteration are now 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 lth pass of the equalizer where $l = 1, 2, 3, \ldots$, the slicer input $\tilde{x}^l[n]$ can be expressed as[2]

$$\tilde{x}^l[n] = \mathbf{c}^{l\dagger}\mathbf{q}^l[n] \tag{8}$$

where

$$\mathbf{c}^l = \begin{bmatrix} \mathbf{b}^l \\ -\mathbf{d}^l \end{bmatrix} \qquad \mathbf{q}^l[n] = \begin{bmatrix} \mathbf{r}[n] \\ \hat{\mathbf{x}}^{l-1}[n] \end{bmatrix} \tag{9}$$

and[3]

$$\mathbf{b}^l = \begin{bmatrix} b^l[-J_1] & \cdots & b^l[0] & \cdots & b^l[J_2] \end{bmatrix}^\dagger \tag{10}$$

$$\mathbf{d}^l = \begin{bmatrix} d^l[-K_1] & \cdots & d^l[-1] & d^l[1] & \cdots & d^l[K_2] \end{bmatrix}^\dagger \tag{11}$$

$$\mathbf{r}[n] = \begin{bmatrix} r[n+J_1] & \cdots & r[n] & \cdots & r[n-J_2] \end{bmatrix}^T \tag{12}$$

$$\hat{\mathbf{x}}^{l-1}[n] = \begin{bmatrix} \hat{x}^{l-1}[n+K_1] & \cdots & \hat{x}^{l-1}[n+1] & \hat{x}^{l-1}[n-1] & \cdots & \hat{x}^{l-1}[n-K_2] \end{bmatrix}^T \tag{13}$$

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

In the lth iteration, there are two sets of data available to the receiver: $r[n]$ and $\hat{x}^{l-1}[n]$, $n = 0, 1, \ldots, N-1$. 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:

$$\mathcal{E}(\mathbf{c}^l) = \sum_{n=-\infty}^{\infty} |\hat{x}^{l-1}[n] - \mathbf{c}^{l\dagger}\mathbf{q}^l[n]|^2. \tag{14}$$

Since this is a linear least-squares estimation problem, the optimum $\mathbf{c}^l$ is [7]

$$\mathbf{c}^l_{\text{opt}} = [\mathbf{\Phi}^l]^{-1}\mathbf{u}^l, \tag{15}$$

where

$$\mathbf{\Phi}^l = \sum_{n=-\infty}^{\infty} \mathbf{q}^l[n]\mathbf{q}^{l\dagger}[n] \tag{16}$$

and

$$\mathbf{u}^l = \sum_{n=-\infty}^{\infty} \hat{x}^{l-1^*}[n]\mathbf{q}^l[n]. \tag{17}$$

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 [8].

We now turn to a couple of implementation issues. First, we would ideally like our finite-length adaptive filters to approximate (3) and (4), which are infinite length. The optimal $b^l[n]$ in (3) includes a filter matched to $a[n]$, and the optimal $d^l[n]$ in (4) 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 multipass equalizer as a function of the number of iterations, for varying amounts of training data. As expected, the curves are monotonically decreasing, suggesting that additional iterations always improve performance. 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.

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 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 [7]. 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 in Section III. 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

[2]The superscript † denotes the conjugate-transpose operation.
[3]The superscript ^{T} denotes transposition.

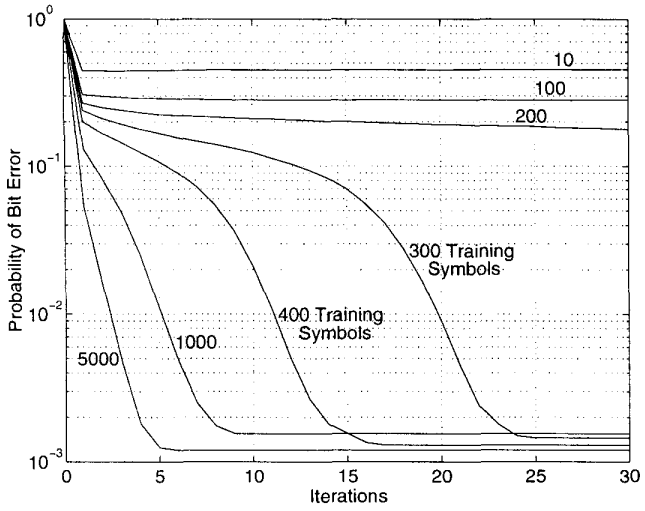

Fig. 4. Experimentally observed QPSK bit-error rate for the adaptive iterated-decision equalizer as a function of the number of decoding iterations and the number of training symbols transmitted with each block of 10000 data symbols at an SNR per bit of 7 dB. The 100-tap channels were equalized using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

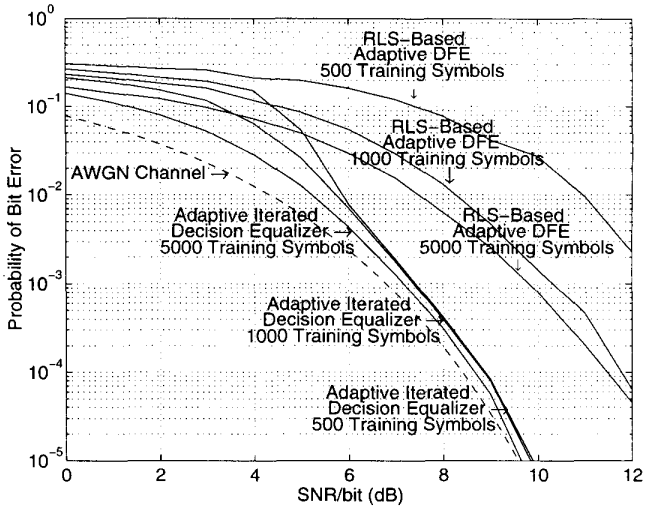

Fig. 5. Experimentally observed QPSK bit-error rate for the adaptive iterated-decision equalizer and the RLS-based adaptive DFE (with forgetting factor $\lambda = 1$) as a function of SNR per bit. Blocks of 10000 data symbols were transmitted through 128-tap channels, which were equalized using 257 feedforward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

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 multipass equalizer essentially requires the computation of one large matrix inverse per iteration for *all* the symbols in the packet, with the number of required iterations being typically small.

V. Coded Systems

For bandlimited channels with strong frequency-dependent distortion, coding must be combined with equalization techniques so that capacity is approached. In typical implementations, the DFE cancels postcursor ISI by using delay-free symbol decisions, which in a coded system are often highly unreliable compared to maximum likelihood (ML) decisions, so performance is often poor as a result. From this perspective, the iterated-decision equalizer, which avoids this problem, is a compelling alternative to the DFE in coded systems.

The structure of a communication system that combines the iterated-decision equalizer with coding is shown in Fig. 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 $d^l[n]$.

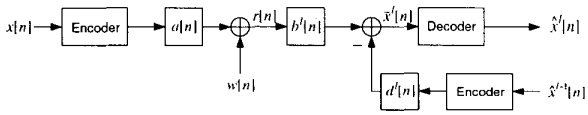

Fig. 6. Structure of a communication system that combines iterated-decision equalization with channel coding.

References

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

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

[3] A. M. Chan, "A class of batch-iterative methods for the equalization of intersymbol interference channels," dissertation, M.I.T., Aug. 1999.

[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, pp. 191–205, Jan. 1997.

[5] 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, pp. 403–18, Sep.–Oct. 1998.

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

[7] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1996.

[8] H. Lütkepohl, *Handbook of Matrices*. Chichester, England: Wiley, 1996.