Iterative Equalization with Adaptive Soft Feedback

Wolfgang H. Gerstacker, Ralf R. Müller, and Johannes B. Huber

***Abstract*—In this letter, a novel equalization algorithm applying soft-decision feedback and designed for binary transmission is introduced. In contrast to conventional decision-feedback equalization (DFE), iterations are necessary, because a simple matched filter serves as feedforward filter, which collects signal energy, but creates noncausal intersymbol interference. The rule for generating soft decisions is adapted continuously to the current state of the algorithm. In most cases, standard DFE methods are clearly outperformed. For a class of certain channel impulse responses, performance of maximum-likelihood sequence estimation is attained, in principle. The high performance of the scheme is explained using results from neural network theory.**

***Index Terms*—Equalization, iterative sequence estimation, soft-decision feedback.**

I. Introduction

THE optimum algorithm for equalization of dispersive channels producing intersymbol interference (ISI) is maximum-likelihood sequence estimation (MLSE), which can be performed by the Viterbi algorithm (VA) [1]. However, the complexity of the VA becomes prohibitive for long channel impulse responses (CIRs). Thus, suboptimum schemes have to be applied in this case, e.g., decision-feedback equalization (DFE) or reduced-state sequence estimation (RSSE), preferably in the form of delayed decision-feedback sequence estimation (DFSE) [2]. Because a minimum-phase impulse response is essential for both schemes, in general, a prefilter is necessary, which transforms the CIR into its minimum-phase equivalent. In many cases, the real-time calculation of this filter poses serious problems, e.g., in mobile communications. But even with optimized prefilter, a very high number of states might be necessary for DFSE in order to obtain high performance.

In this letter, a powerful novel iterative algorithm of low complexity is introduced, requiring no minimum-phase response and performing even better than optimized DFSE with high number of states.

II. System Model

A packet transmission according to Fig. 1 with binary pulse amplitude modulation is considered; the extension of the algorithm to continuous transmission is straightforward, cf. Section VII. All signals and systems are assumed to be real-valued.

Paper approved by S. B. Gelfand, the Editor for Transmission Systems of the IEEE Communications Society. Manuscript received October 19, 1998; revised November 15, 1999.

W. H. Gerstacker is with the Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand (e-mail: gerstack@elec.canterbury.ac.nz).

R. R. Müller is with the Forschungszentrum Telekommunikation Wien, 1040 Vienna, Austria (e-mail: mueller@ftw.at).

J. B. Huber is with the Telecommunications Laboratory, University Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: huber@LNT.de).

Publisher Item Identifier S 0090-6778(00)07525-5.

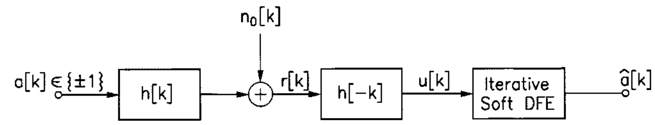

Fig. 1. Model of transmission.

The discrete-time received signal is given by $r[k] = h[k]*a[k]+n_0[k]$, where $a[k] \in \{\pm 1\}$ is the transmitted amplitude coefficient at discrete time k, $h[k]$, $k \in \{0, 1, \cdots, q_h\}$, denotes the CIR of order q_h comprising transmit filter, channel, and continuous-time receiver input filter, and $n_0[k]$ additive white Gaussian noise. The CIR, not of minimum phase in general, is assumed to be time-invariant during each burst.

For channel estimation and in order to ensure a definite inital and final state for equalization, a known training sequence of sufficient length is transmitted prior to data for each burst. Until Section VI, error-free channel estimates are assumed. In order to simplify notation, discrete time $k = 0$ is assigned to the first of L data symbols of the current burst, while $a[-q_h], \cdots, a[-1]$ and $a[L], \cdots, a[L+q_h-1]$ denote the last q_h training symbols of the current burst and the first q_h training symbols of the next burst, respectively.

In front of equalization, a matched filter $h[-k]$, whose coefficients result immediately from channel estimation, is applied to the received signal, cf. Fig. 1. Its output signal $u[k]$ can be written as

$$u[k] = \sum_{\kappa=-q_h}^{q_h} \rho_{hh}[\kappa]\, a[k-\kappa] + n[k] \tag{1}$$

with the filter autocorrelation sequence of the CIR $\rho_{hh}[\kappa] = h[\kappa]*h[-\kappa]$ and $n[k] = h[-k]*n_0[k]$. Application of a matched filter as prefilter concentrates the energy of the overall impulse response in time instant $\kappa = 0$; however, due to noncausal ISI,[1] conventional (noniterative) DFE cannot be employed.

III. Description of the Algorithm

In each iteration of the equalization algorithm, soft-decision feedback is performed sequentially starting from $k = 0$ up to $k = L - 1$ according to

$$y_\mu[k] = u[k] - \sum_{\kappa=1}^{q_h} \rho_{hh}[\kappa] \left(\hat{a}^{\mathrm{s}}_\mu[k-\kappa] + \hat{a}^{\mathrm{s}}_{\mu-1}[k+\kappa]\right). \tag{2}$$

Here, $\mu > 0$ is the number of the current iteration, and $\hat{a}^{\mathrm{s}}_\mu[k]$ denote soft decisions of iteration μ. After soft cancellation of ISI, a soft decision for symbol $a[k]$ is calculated by

$$\hat{a}^{\mathrm{s}}_\mu[k] = \tanh\left(\frac{\rho_{hh}[0]\, y_\mu[k]}{\hat{\sigma}^2_\mu[k]}\right) \tag{3}$$

[1]This means, precursors of significant energy precede the main tap.

and used for ISI cancellation in the next time steps of the current iteration. In (3), $\hat{\sigma}_\mu^2[k]$ denotes the expected variance of the error signal after soft cancellation, which is given by [4]

$$\hat{\sigma}_\mu^2[k] = \sigma_n^2 + \sum_{\kappa=1}^{q_h} \rho_{hh}^2[\kappa] \left(2 - \left(\hat{a}_\mu^{\mathrm{s}}[k-\kappa]\right)^2 - \left(\hat{a}_{\mu-1}^{\mathrm{s}}[k+\kappa]\right)^2\right) \tag{4}$$

where σ_n^2 denotes the variance of $n[k]$. According to, e.g., [5] and [6], the soft decisions $\hat{a}_\mu^{\mathrm{s}}[k]$ minimize the mean-squared error (MMSE) after feedback in the current iteration, if the sum of noise and ISI can be modeled as Gaussian random variable with zero mean.

The principle of using soft decisions for ISI cancellation in a DFE has been first introduced in [3].

After arriving at the position of the last data symbol, a new iteration starts. Equalization is terminated, if soft decisions change only slightly from one iteration to the next, i.e., $\max_{k\in\{0,1,\cdots,L-1\}} |\hat{a}_\mu^{\mathrm{s}}[k] - \hat{a}_{\mu-1}^{\mathrm{s}}[k]| < \epsilon$, with a small constant ϵ, or the iteration number exceeds a prescribed limit $\mu_{\max}$. Hard estimates for the data symbols are given by the sign of the soft decisions of last iteration μ_{stop}

$$\hat{a}[k] = \mathrm{sign}\left(\hat{a}_{\mu_{\mathrm{stop}}}^{\mathrm{s}}[k]\right), \quad k \in \{0, 1, \cdots, L-1\} \triangleq \mathcal{D}. \tag{5}$$

Initialization is done according to

$$\hat{a}_0^{\mathrm{s}}[k] = 0, \quad k \in \mathcal{D}, \tag{6}$$

$$\hat{a}_0^{\mathrm{s}}[k] = a[k], \quad k \in \{-q_h, \cdots, L+q_h-1\} \backslash \mathcal{D} \triangleq \mathcal{T} \tag{7}$$

which takes into account, that symbols of the training instants are known at the receiver side (therefore, (7) is valid for any iteration number), and initial estimates for the data symbols are their maximum-likelihood (ML) *a priori* estimates.

It should be noted that a related scheme has been independently proposed in [7], where feedback of hard decisions or of soft values generated with simplified nonlinearities is considered. However, the feedback strategy proposed in this paper using the hyperbolic tangent function with adaptive slope according to (4) turns out to be crucial for optimum performance, cf. also Section V.

IV. Interpretation as Neural Network

In this section, it is shown that the proposed scheme can be interpreted as a discrete-time Hopfield neural network (HNN) for approximate MLSE, giving rise to some convergence results.

The ML function to be minimized by MLSE is given by

$$f(\tilde{\boldsymbol{a}}) = \sum_{k=0}^{L+q_h-1} \left(r[k] - \sum_{\kappa=0}^{q_h} h[\kappa]\, \tilde{a}[k-\kappa]\right)^2 \tag{8}$$

with $\tilde{\boldsymbol{a}} \triangleq [\tilde{a}[0] \quad \tilde{a}[1] \quad \cdots \quad \tilde{a}[L-1]]^T$, where $\tilde{a}[k] \in \{\pm 1\}$, $k \in \mathcal{D}$, denote the trial symbols of the MLSE. For simplicity of notation, training symbols have also been denoted by $\tilde{a}[k]$, $k \in \mathcal{T}$, in (8). It can be proved by straightforward calculations, cf. [8], that minimization of $f(\cdot)$ is equivalent to that of the energy function

$$\begin{aligned} J(\tilde{\boldsymbol{a}}) &= \frac{1}{2} \sum_{k_1 \in \mathcal{D}} \sum_{k_2 \in \mathcal{D}\backslash\{k_1\}} \tilde{a}[k_1]\, \tilde{a}[k_2]\, \rho_{hh}[k_1 - k_2] \\ &\quad - \sum_{k\in\mathcal{D}} \tilde{a}[k] \left(u[k] - \sum_{\kappa\in\mathcal{T}} \tilde{a}[\kappa]\, \rho_{hh}[k-\kappa]\right) \\ &\triangleq -\frac{1}{2} \tilde{\boldsymbol{a}}^T \boldsymbol{R} \tilde{\boldsymbol{a}} + \tilde{\boldsymbol{a}}^T \boldsymbol{v} \end{aligned} \tag{9}$$

with obvious implicit definitions of matrix $\boldsymbol{R}$ and vector $\boldsymbol{v}$.

It is well known that an HNN can be employed for solution of this kind of discrete optimization problem [9]. It can be shown, that such a serial HNN always converges to a stable state, which corresponds to a local minimum of the energy function in (9) associated with $\boldsymbol{R}$ and $\boldsymbol{v}$, if the connection matrix $\boldsymbol{R}$ is symmetrical and has vanishing diagonal elements [9]. Because the matrix $\boldsymbol{R}$ in the energy function of MLSE satisfies this condition, a local minimum of the ML function can be found with the network, however, not necessarily the global one.

Comparing an HNN for minimization of the ML function to the iterative soft-decision feedback algorithm of Section III, it follows that both schemes are identical except that the $\tanh(\cdot)$-function in (3) has to be replaced by a hard limiter to get a discrete-time HNN. But by soft limitation, and especially with variable slope, the probability of being trapped in an undesired local minimum in the course of the optimization procedure is decreased [9]. Furthermore, using the concept of MMSE feedback, an analytical rule for slope adaptation results, which seems to be well suited to the problem. This is in contrast to related work on the application of neural networks to multiuser detection, e.g., [10]–[12], where an activation function with a fixed, empirically optimized slope has been applied.

V. Numerical Results and Discussion

In contrast to conventional DFE, performance of the proposed iterative soft-decision feedback algorithm has not been able to be analyzed analytically, yet. Therefore, performance has been evaluated by simulations.

In the following, performance of the proposed algorithm is compared to that of conventional DFE, DFSE, and MLSE for a channel with a discrete-time impulse response given by Fig. 2(a) ($q_h = 59$), which has been randomly generated; the corresponding filter autocorrelation sequence is shown in Fig. 2(b). Such an impulse response may be encountered, e.g., in a high-rate transmission over a (mobile) multipath channel with path weights of equal average power, cf. [13].

The bit-error rate (BER) of DFSE can be approximated by [2]

$$\mathrm{BER} \approx Q\left(\sqrt{d_{\min}^2 \frac{E_b}{\mathcal{N}_0}}\right) \tag{10}$$

where $d_{\min}^2$ is defined as normalized minimum squared Euclidean distance, depending on CIR and number of trellis states Z of DFSE. It should be noted that this approximation is quite

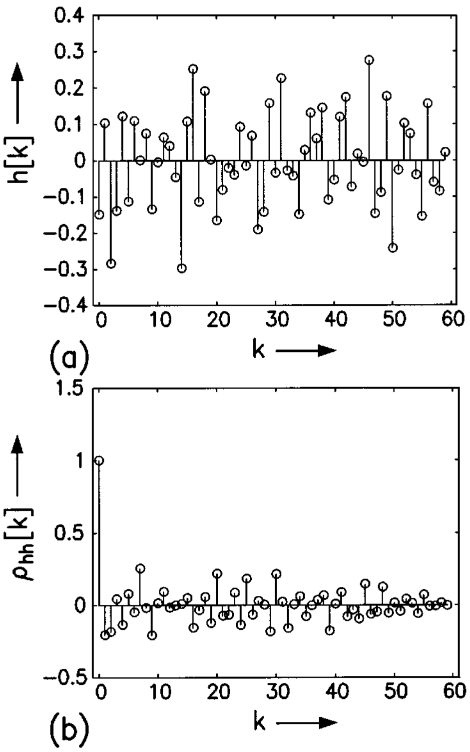

Fig. 2. (a) Example channel impulse response ($q_h = 59$). (b) Corresponding filter autocorrelation sequence for $k \geq 0$ ($\rho_{hh}[-k] = \rho_{hh}[k]$).

optimistic for high state reduction, because it ignores the effects of error propagation. This point is further addressed in Section VI. For $Z = 1$ and $Z = 2^{q_h}$, conventional DFE and MLSE, respectively, result as special cases. In order to achieve optimum performance, the CIR has to be transformed into its minimum-phase equivalent by allpass filtering before equalization, if RSSE is applied, i.e., $Z < 2^{q_h}$. Thus, distance calculations have been performed for the minimum-phase CIR corresponding to $h[k]$ of Fig. 2(a).[2] In contrast to that, performance of the proposed algorithm does not depend on the phase of the CIR, i.e., performance depends only on $\rho_{hh}[k]$, but not on the CIR itself.

For CIRs with an autocorrelation sequence according to Fig. 2(b), $d^2_{\min}(\text{MLSE}) = 2$, i.e., asymptotically, the power efficiency of the AWGN channel is attained by MLSE. According to [1], (10) is also a lower bound for BER in such a case. For DFE, (10) is a lower bound for any CIR.

For simulations, data block length has been chosen to $L = 200$, number of iterations maximally tolerated to $\mu_{\max} = 100$, and $\epsilon = 10^{-3}$. Fig. 3 shows that the proposed algorithm performs less than 0.2 dB worse than MLSE for $\text{BER} = 10^{-5}$. A gain of ≈3 dB results compared to DFE and of at least 1.4 dB compared to DFSE algorithms, which seem to be practically implementable, i.e., are limited to $Z = 1024$ states. In Fig. 4, the expected value of the number of executed iterations $E\{\mu_{\text{stop}}\}$ is shown as a function of $E_b/\mathcal{N}_0$. Obviously, complexity is quite moderate for BERs of practical interest.

Slope adaptation according to (4) plays a crucial role for high performance of the algorithm. Using a fixed slope or a hard limiter instead of the hyperbolic tangent, i.e., an infinite slope, a loss of several decibels may result; even for the optimum fixed choice, which is not known in a practical implementation, a degradation of up to 0.5 dB occurs.

[2]For this, Dijkstra's algorithm [14] has been employed, which can cope with quite long CIRs, provided they are of minimum phase.

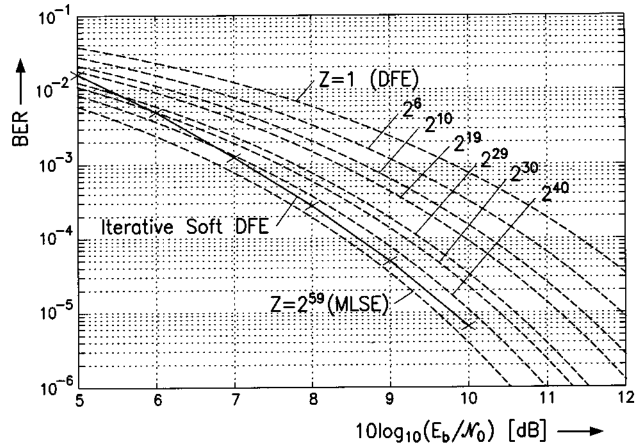

Fig. 3. BER versus $E_b/\mathcal{N}_0$ for iterative soft DFE (solid line, simulation results) and DFSE with different numbers of states Z [dashed lines, approximation according to (10)].

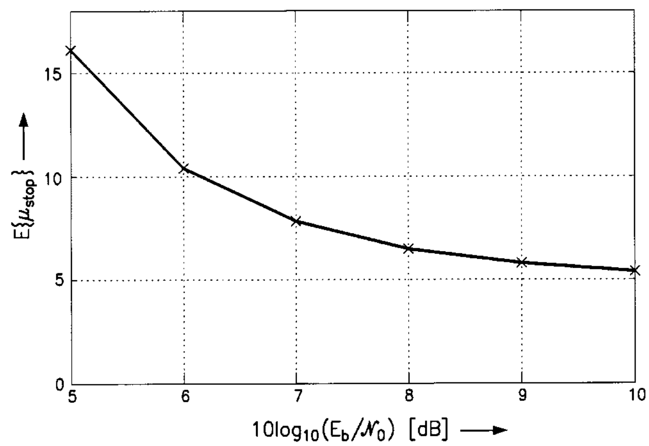

Fig. 4. Expected number of iterations versus $E_b/\mathcal{N}_0$.

By a series of further examples, efficiency of the algorithm for long CIRs with randomly varying coefficients has been confirmed. Here, the event of being trapped in a local minimum was found to be rare. The reason might be that the Gaussian assumption in (3) becomes quite accurate in this case, and that the energy function might possess few undesired local minima. For CIRs of shorter length and/or with well-defined shape, the performance of MLSE cannot be approached and the curves are flattening out at relatively high BERs, which also happens for long random CIRs, but at a markedly lower level. However, also for small to moderate q_h, a substantial gain compared to conventional DFE can be attained.

VI. Effects of Nonideal Channel Estimation

Until now, the availability of ideal channel estimates at the receiver side has been assumed. In this section, the effects of noisy estimates on performance are investigated. For channel

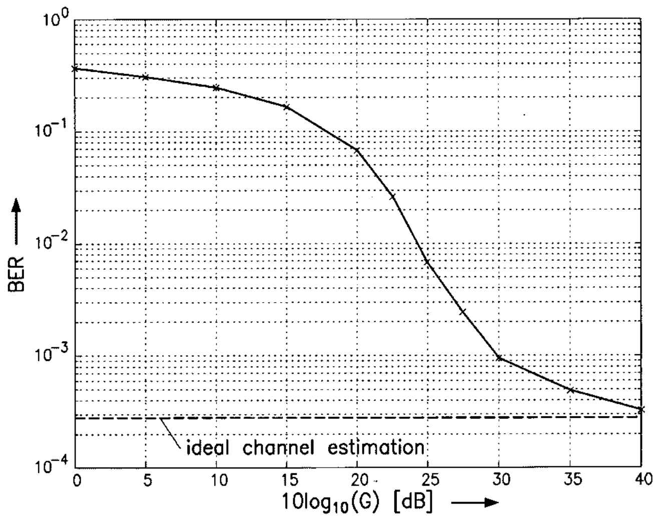

Fig. 5. BER for iterative soft DFE versus noise suppression factor G of channel estimation ($10\ \log_{10}(E_b/\mathcal{N}_0) = 8$ dB).

sounding, ML channel estimation, cf., e.g., [15]–[17], using training sequences designed for optimum noise suppression, is considered. In this case, the estimated channel coefficients are given by

$$\hat{h}[\kappa] = h[\kappa] + e[\kappa], \quad \kappa \in \{0, 1, \cdots, q_h\} \tag{11}$$

where the Gaussian channel estimation errors $e[\kappa]$ are mutually uncorrelated and have equal variance

$$\sigma_e^2 = \frac{\sigma_{n_0}^2}{G}. \tag{12}$$

Here, $\sigma_{n_0}^2$ denotes the variance of $n_0[k]$, cf. Section II, and for the noise suppression factor G

$$G = N - q_h \tag{13}$$

results, where N is the length of the training sequence.

For the following simulation results, noisy channel estimates individually generated for each data block according to (11) have been used for matched filtering and calculation of $\rho_{hh}[\cdot]$, which is required in (2)–(4) of the proposed algorithm. The corresponding BER of iterative soft-decision feedback equalization using noisy channel estimates is shown in Fig. 5 as a function of the noise suppression factor G. Again, the test channel of Section V has been used, and $10\ \log_{10}(E_b/\mathcal{N}_0) = 8$ dB is valid. Obviously, a noise suppression of at least 30 dB is required in order to obtain a BER comparable to that for ideal channel estimation. Thus, a quite long training sequence ($N \approx 1000$) is necessary. As a consequence, also relatively long data blocks should be transmitted in order to limit the rate loss due to training symbols. This is not a major problem, because simulations have indicated, that the performance of the proposed scheme depends only weakly on block length L.

For conventional DFE, it has to be taken into account, that the results presented in Section V are valid for ideal feedback, i.e., of known data symbols. For a realizable DFE using decided symbols in the feedback path, however, error propagation often causes a loss of several decibels compared to an ideal DFE at low to medium signal-to-noise ratios (cf., e.g., simulation results in [18]), especially for long feedback filters (CIRs). Channel estimation errors additionally produce uncancelled ISI and even increase the problem of error propagation. Similar problems also occur for DFSE with only a few number of states.

Thus, the results of Section V are quite optimistic for DFE and DFSE, because channel estimation errors as well as error propagation have been neglected, whereas for the proposed scheme, only ideal channel estimation has been assumed.

For MLSE applied to channels with long impulse responses and DFSE with a high number of states, the effects of channel estimation errors cannot be assessed, because a simulation is impossible. These schemes serve only as benchmarks and cannot be implemented in practice. Also analytical results can be hardly obtained [19]. However, in [20], it has been proved using capacity arguments, that for channels with long memory, it becomes in general increasingly difficult to measure the CIR with sufficient accuracy for a reliable detection, independent of the detection scheme actually used.

VII. Concluding Remarks

A novel iterative soft-decision feedback equalization algorithm has been introduced and analyzed using results from neural network theory. For long CIRs with random coefficients, performance of MLSE can be approached very closely. Although the scheme can be interpreted as a neural network, no training of its coefficients is necessary, as done in most other neural network equalizers, e.g., [21], provided the CIR is known.

For the derivations in this letter, a block-oriented transmission has been assumed. Nevertheless, the algorithm can be also employed for continuous transmission. Here, $\mu_{\max}$ soft DFEs, uniformly spaced by $q_h + 1$ time intervals, have to be run simultaneously, each one using the results of its predecessor. The increase in BER compared to block-oriented transmission with training symbols for termination is only slight.

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, pp. 363–378, May 1972.

[2] A. Duel-Hallen and C. Heegard, “Delayed decision-feedback sequence estimation,” *IEEE Trans. Commun.*, vol. 37, pp. 428–436, May 1989.

[3] D. P. Taylor, “The estimate feedback equalizer: A suboptimum nonlinear receiver,” *IEEE Trans. Commun.*, vol. COM-21, pp. 979–990, Sept. 1973.

[4] R. Müller and J. Huber, “Iterated soft-decision interference cancellation for CDMA,” in *Broadband Wireless Communications*, Luise and Pupolin, Eds. London, U.K.: Springer, 1998, pp. 110–115.

[5] F. Tarköy, “MMSE-optimal feedback and its applications,” in *Proc. ISIT’95*, Whistler, BC, Canada, Sept. 1995, p. 334.

[6] S. Müller, W. Gerstacker, and J. Huber, “Reduced-state soft-output trellis-equalization incorporating soft feedback,” in *Proc. IEEE Global Telecommunication Conf. (GLOBECOM)*, London, U.K., Nov. 1996, pp. 95–100.

[7] E. de Carvalho and D. T. M. Slock, “Burst mode noncausal decision-feedback equalizer based on soft decisions,” in *Proc. Vehicular Technology Conf. (VTC)*, 1998, pp. 414–418.

[8] G. Ungerböck, “Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems,” *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.

[9] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. Stuttgart, Germany: B. G. Teubner, 1993.

[10] G. I. Kechriotis and E. S. Manolakos, “Hopfield neural network implementation of the optimal CDMA multiuser detector,” *IEEE Trans. Neural Networks*, vol. 7, pp. 131–141, Jan. 1996.

[11] T. Nagaosa, T. Miyajima, and T. Hasegawa, “Multiuser detection using a Hopfield network in asynchronous M-ary/SSMA communications,” in *Proc. 4th Int. Symp. Spread Spectrum Techniques and Applications (ISSSTA)*, Mainz, Germany, Sept. 1996, pp. 837–841.

[12] W. G. Teich and M. Seidl, “Code division multiple access communications: Multiuser detection based on a recurrent neural network structure,” in *Proc. 4th Int. Symp. Spread Spectrum Techniques and Applications (ISSSTA)*, Mainz, Germany, Sept. 1996, pp. 979–984.

[13] T. S. Rappaport, *Wireless Communications*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[14] T. Larsson, “A state-space partitioning approach to trellis decoding,” Chalmers Univ. of Technol., Göteborg, Tech. Rep. 222, 1991.

[15] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, “Least sum of squared errors (LSSE) channel estimation,” *IEE Proc.—F*, vol. 138, pp. 371–378, Aug. 1991.

[16] K.-H. Chang and C. N. Georghiades, “Iterative joint sequence and channel estimation for fast time-varying intersymbol interference channels,” in *Proc. Int. Conf. Communications (ICC)*, Seattle, WA, June 1995, pp. 357–361.

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

[18] A. Rony and S. Shamai (Shitz), “Suboptimal detection for intersymbol interference inflicted cable channels,” *Int. J. Electron. Commun. (AEÜ)*, vol. 51, no. 5, pp. 246–254, 1997.

[19] S. Ölçer, “Performance analysis of Viterbi detection of class IV partial response for data transmission over multipair cables,” *IEEE Trans. Commun.*, vol. 39, pp. 1423–1426, Oct. 1991.

[20] R. Gallager and M. Médard, “Bandwidth scaling for fading channels,” in *Proc. Int. Symp. Information Theory (ISIT)*, Ulm, Germany, June/July 1997, p. 471.

[21] G. Kechriotis, E. Zervas, and E. S. Manolakos, “Using recurrent neural networks for adaptive communication channel equalization,” *IEEE Trans. Neural Networks*, vol. 5, pp. 267–278, Mar. 1994.