NONLINEAR SIGNAL PROCESSING FOR ADAPTIVE EQUALISATION AND MULTI-USER DETECTION

Bernard Mulgrew

Department of Electrical Engineering
The University of Edinburgh
King's Buildings, Edinburgh EH9 3EA, UK
Tel: +44 131 650 5580; fax: +44 131 650 6554
Email: B.Mulgrew@ee.ed.ac.uk

ABSTRACT

This paper examines the application of nonlinear signal processing techniques to the development of adaptive equalisers for frequency domain multiple access (FDMA) and multi-user detectors for code division multiple access (CDMA). Current issues are discussed and key problems identified.

1 INTRODUCTION

Adaptive equalisation has a long history [1]. Overviews of the recent application of nonlinear signal processing techniques to the problem are available in [2] and [3]. The closely related topic of multi-user detection (MUD) for code division multiple access is covered by a number of review articles [4-6]. The objective of this paper is to assess the current state of the art in the application of nonlinear signal processing techniques to the design of adaptive equaliser and multi-user detectors.

Care must be exercised in the use of the word ‘‘nonlinear’’. In the communications literature a conventional decision feedback equaliser (DFE) is described as nonlinear because the decisions are based on a nonlinear function of the received signal. Alternatively a conventional DFE can be viewed as a linear function of the received signal and previously detected symbols. In this paper the word nonlinear will be reserved for detectors whose decisions are based on nonlinear combinations of the received signal and previously detected symbols. A similar classification can be applied to MUD schemes.

2 BAYESIAN AND MAXIMUM LIKELIHOOD EQUALISERS

Many digital communications channels are subject to intersymbol interference (ISI). This interference is usually a result of the restricted bandwidth allocated to the channel and/or the presence of multipath distortion in the medium through which the information is transmitted. Many such channels can be characterised by a finite impulse response digital filter and an additive noise source [7]. The appropriate channel model is depicted in Figure 1. The digital data sequence $\{x(k)\}$ is independently identically distributed and drawn with equal probability from a finite alphabet $\{a_p : 1 \le p \le P\}$. It is passed through a linear dispersive channel of finite impulse response (FIR). The observed sequence, $\{y(k)\}$, is formed by adding Gaussian random noise $\{n(k)\}$ to the output of the FIR filter. The relationship between the channel input, $x(k)$, and the channel output, $y(k)$ can be summarised: $y(k) = \Sigma_{i=0}^{N-1} h_i x(k-i) + n(k)$. The transfer function of the FIR filter is $H(z) = \Sigma_{i=0}^{N-1} h_i z^{-i}$ where N is the length of the impulse response.

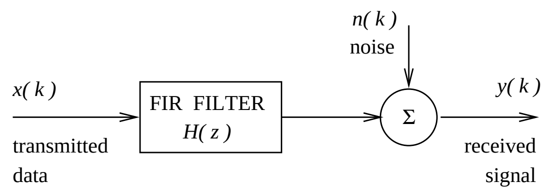


Figure 1: *Digital communications channel*

In many equalisation problems it is convenient to describe the relationship between the transmitted sequence $\{x(k)\}$ and the received sequence $\{y(k)\}$ as a matrix equation. Thus:

$$\mathbf{y}(k) = \mathbf{H}\,\mathbf{x}(k) + \mathbf{n}(k)$$

where $\mathbf{y}(k)$ contains all the available received samples:

$$\mathbf{y}(k) = [\; y(k)\; y(k-1) \cdots y(0)\;]^T$$

$\mathbf{x}(k)$ contains all the transmitted symbols which generated the received samples:

$$\mathbf{x}(k) = [\; x(k)\; x(k-1) \cdots x(0)\; x(-1) \cdots x(-N+1)\;]^T$$

The "impulse response matrix" $\mathbf{H}$ has $k+1$ rows and $k+N$ columns and is Toeplitz:

$$\begin{bmatrix} h_0 & h_1 & h_2 & . & . & h_{N-1} & 0 & . & . & 0 \\ 0 & h_0 & h_1 & . & . & h_{N-2} & h_{N-1} & . & . & 0 \\ . & & & & & . & . & & & . \\ . & & & & & . & . & & & . \\ 0 & 0 & 0 & . & . & 0 & h_0 & . & . & h_{N-1} \end{bmatrix}.$$

and $\mathbf{n}(k)$ contains the Gaussian white noise:

$$\mathbf{n}(k) = [\; n(k)\; n(k-1) \cdots n(0)\;]^T$$

Given a set of observations collected in the vector $\mathbf{y}(k)$, the Bayesian or maximum *a posteriori* (MAP) symbol-by-symbol decision rule is to choose the symbol a_p which is most probable given these observations, i.e.:

$$p = \arg\max_i P_{x|\mathbf{y}}(\; x(k-d) = a_i \mid \mathbf{y}(k)\;) \qquad (1)$$

Thus the detector also provides the probability that its decision are correct, i.e. $P_{x|\mathbf{y}}(x(k-d) = a_p|\mathbf{y}(k))$. The conditional probability of equation (1) can be rewritten as the sum of all the conditional sequence probabilities of the form $P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}(k) = \mathbf{x}_j|\mathbf{y}(k))$ each of which is associated with a particular transmitted sequence collected in vector $\mathbf{x}_j$. All vectors $\mathbf{x}_j$ contain the symbol $x(k-d) = a_i$. Thus:

$$p = \arg\max_i \sum_{\{j:x(k-d)=a_i\}} P_{\mathbf{x}|\mathbf{y}}(\,\mathbf{x}(k) = \mathbf{x}_j|\mathbf{y}(k)\,) \qquad (2)$$

Maximum likelihood sequence estimation (MLSE) involves choosing the sequence of symbols in the vector $\mathbf{x}_l$ which is most probable given the observations contained in $\mathbf{y}(k)$. Thus:

$$l = \arg\max_j P_{\mathbf{x}|\mathbf{y}}(\,\mathbf{x}(k) = \mathbf{x}_j|\mathbf{y}(k)\,) \qquad (3)$$

In common with the MAP symbol-by-symbol detector (MAPSD) the MLSE also provides the probability that its decisions are correct. From the above it is clear that since the probability of the MAP decision is the sum of the individual sequence probabilities, the MAPSD provides a lower bit error rate for a given lag d than the MLSE [8, 9].

Why then is the MLSE so popular? There are a number of reasons: (i) in Gaussian white noise the condition probability calculation of equation (3) is replaced with a simpler minimum Euclidean distance calculation

$$l = \arg\min_j \| \mathbf{y}(k) - \mathbf{H}\mathbf{x}_j \|^2$$

(ii) knowledge of the noise variance σ_n^2 is not required; (iii) the Viterbi algorithm [10] is available to provide an efficient implementation.

Consider a simple example where two observations of the output of a 3-tap channel are available. Thus:

$$\begin{bmatrix} y(1) \\ y(0) \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & 0 \\ 0 & h_0 & h_1 & h_2 \end{bmatrix} \begin{bmatrix} x(1) \\ x(0) \\ x(-1) \\ x(-2) \end{bmatrix} + \begin{bmatrix} n(1) \\ n(0) \end{bmatrix}$$

Figure 2 illustrates the observation space spanned by the vector $\mathbf{y}(1) = [y(1)\ y(0)]^T$. The asterisks and circles represent the observations vectors associated with all the possible input sequences in the absence of noise, i.e.: $\{\mathbf{y}_j : \mathbf{y}_j = \mathbf{H}\,\mathbf{x}_j\}$. The asterisks indicate the observation vectors associated with the symbol $x(0) = -1$ and the circles $x(0) = +1$ (in the absence of noise). The MAP decision boundary is a curve which is a function of the noise variance. As the noise variance approaches zero the MAP decision boundary tends to a piecewise linear function [11]. The classification regions for the MLSE are indicated by the remaining lines. It is clear that portions of the MLSE boundaries can be grouped together to approximate the MAP boundary. It is worth noting that since the MLSE reduces to a minimum distance detector it decision boundaries will not change with signal to noise ratio.

Using Bayes theorem and assuming that the symbols are equi-probable, equation (2) can be rewritten as the sum of conditional probability density functions (pdf's) of the form $p_{\mathbf{y}|\mathbf{x}}(\,\mathbf{y}(k) \mid \mathbf{x}(k) = \mathbf{x}_j\,)$. Thus:

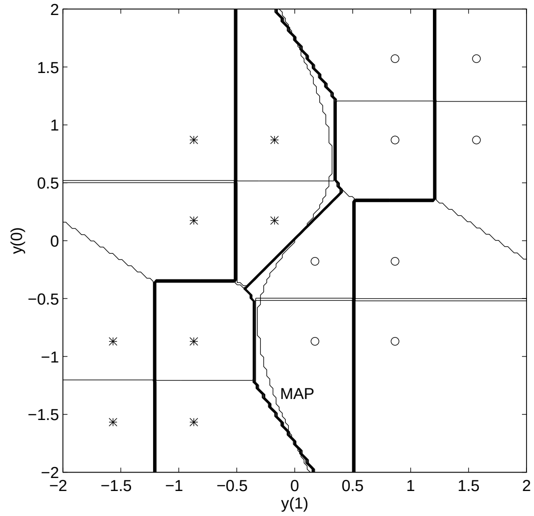


Figure 2: *Observation space for $H(z) = 0.35 + 0.87z^{-1} + 0.35z^{-2}$ showing MAP decision boundary for $d = 1$ and MLSE boundaries.*

$$\begin{aligned} p &= \arg\max_i \sum_{\{j:x(k-d)=a_i\}} p_{\mathbf{y}|\mathbf{x}}(\,\mathbf{y}(k) \mid \mathbf{x}(k) = \mathbf{x}_j\,) \\ &= \arg\max_i \sum_{\{j:x(k-d)=a_i\}} \exp\left(-\frac{\| \mathbf{y}(k) - \mathbf{H}\mathbf{x}_j \|^2}{2\sigma_n^2} \right) \end{aligned} \qquad (4)$$

As σ_n^2 tends to zero one sequence will dominate the summation of equation (4) and the MAPSD and MLSE converge to the same detector. Thus, at high signal to noise ratios (SNR's), their bit error rates will be almost indistinguishable [12, 13]. However, as already indicated, at low SNR the MLSE is inferior to the MAPSD [8, 9]. This theoretical result is supported by the computer simulation of [12].

If the MAPSD is superior, why then has it not gained wider acceptance? The answer is threefold: (i) the MAPSD is computationally more complex than the MLSE - it is worth pointing out that the MAPSD proposed by Hayes et al. [14] can be implemented in a similar manner to the familiar Viterbi algorithm (VA) trellis for MLSE; (ii) the MAPSD requires knowledge of the noise variance σ_n^2 in addition to the channel impulse response required by the MLSE; (iii) the MAPSD requires computation of the conditional pdf's and hence exponential functions rather than the distance calculations associated with MLSE.

In taking a nonlinear signal processing approach to adaptive channel equalisation, the MAPSD plays a pivotal role - it is the optimal solution.

3 ADAPTIVE EQUALISATION

In which application areas are nonlinear equalisers liable to be applied? The answer to this question lies in the complexity. The MAPSD and MLSE have computational complexity which is of $O(P^N)$ where P is the size of the symbol alphabet and N is the length of the channel impulse response sequence. Thus it is unlikely that they can be applied to telephone channels where the impulse response sequence can be tens or hundreds of samples long. A clue comes from the fact that MLSE has seen wide application on mobile radio channels where typically $P = 4$ and $N = 5$. However this application brings its own set of problems. In particular the channel itself usually has an impulse response which is varying slowly with time (fading). Further the SNR is usually low, typically < 20 dB. To permit training of the equaliser, blocks of training data are interspersed with the data. Outside of these training blocks there may be a further requirement for the equaliser to track changes in the channel impulse response using a decision directed mode of operation. The generic structure is illustrated in Figure 3.

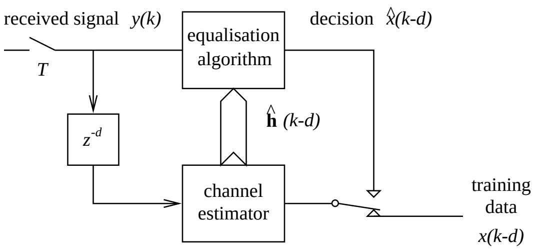


Figure 3: Generic adaptive equaliser for mobile radio applications

During the training period the channel impulse response is estimated. Typically a variant of the least squares (LS) algorithm is used for this purpose because it converges faster than a least mean squares (LMS) algorithm even under these white-input conditions. After training the decisions from the equaliser can be used as a training sequence to allow the channel estimator to track the slowly varying channel. An LMS algorithm is usually used at this stage since there is little difference between the tracking performance of the LMS and recursive LS (RLS) algorithms under white input conditions and the LMS is less complex. Thus there are two processes operating interactively: equalisation and system identification. On a stationary channel the performance of the equaliser will improve with increasing lag d. However increasing the lag in a nonstationary environment will degrade the performance of the channel estimator. A reasonable compromise must be achieved. Typically this lag will be in the region $\{d: 0 \le d < N\}$. At these relatively short lags the performance gain of the MAPSD with respect to the MLSE can become significant.

A number of MAPSD architectures can be accommodated with the generic structure of Figure 3. These are: (i) growing memory; (ii) finite memory; (iii) decision feedback.

Growing memory: Hayes et al. provides a algorithm with complexity proportional of order $O(P^N)$ - this algorithm was given a radial basis function interpretation in [12]. Using Bayes theorem equation (1) can be rewritten in terms of conditional densities of the form $p_{\mathbf{y}|x}(\ \mathbf{y}(k)|x(k-d) = a_i\ )$. To simplify the presentation it is assumed that the equaliser is operating with a lag $d = 0$. In general it can accommodate lags up to N conveniently. Application of the total probability theory and the use of standard probability techniques leads to the following recursion for the probability density of $\mathbf{y}(k)$ conditioned on the last $N - 1$ channel inputs:

$$f_{\mathbf{y}}(k) = ½ f_{n|+}(k)\ \ f_{\mathbf{y}|+}(k-1) + ½ f_{n|-}(k)\ \ f_{\mathbf{y}|-}(k-1) \quad (5)$$

where:

$$f_{\mathbf{y}}(k) = p_{\mathbf{y}(k),N-1}(\ \mathbf{y}(k) \mid x(k), \cdots, x(k-N+2)\ ),$$

$$f_{n|+}(k) = p_n(y(k) - [x(k) \cdots x(k-N+2) + 1]\mathbf{h})$$

and

$$f_{\mathbf{y}|+}(k-1) = p_{\mathbf{y}(k-1),N-1}(\mathbf{y}(k-1) \mid x(k-1), \cdots, x(k-N+1) = +1).$$

$f_{n|-}(k)$ and $f_{\mathbf{y}|-}(k-1)$ are defined in an analogous manner. $p_n(.)$ is the probability density function associated with the noise $n(k)$ and the argument is the difference or error between the scalar channel output $y(k)$ and the scalar noise-free channel output $[x(k)\ x(k-1) \cdots x(k-N+1)]\mathbf{h}$ where $\mathbf{h}$ is the channel impulse response vector, i.e.: $\mathbf{h} = [h_0\ h_1 \cdots h_{N-1}]^T$. Thus for Gaussian noise this density function provides a scalar radial basis function expansion with single input $y(k)$ and a set of centres given by all possible values of $[x(k)\ x(k-1) \cdots x(k-N+1)]\mathbf{h}$. The outputs from this expansion provide the inputs to a recurrent network defined by equation (5). The desired decision function $p_{\mathbf{y}|x}(\mathbf{y}(k)|x(k) = a_i)$ can be obtained from $p_{\mathbf{y}(k),N-1}(\ \mathbf{y}(k) \mid x(k), \cdots, x(k-N+2)\ )$ by repeated application of the total probability theorem e.g.:

$$p_{\mathbf{y}|x}(\mathbf{y}(k)|x(k)) = ½ p_{\mathbf{y}|x}(\mathbf{y}(k)|x(k), x(k-1) = +1) + ½ p_{\mathbf{y}|x}(\mathbf{y}(k)|x(k), x(k-1) = -1)$$

Thus the final layer in the network is a simple linear combiner. The network architecture for a simple case where $N = 3$ is illustrated in Figure 4. The network is trained by estimating the channel impulse response and the noise variance can be estimated as a by product of either the LMS or the LS channel estimation algorithms [15].

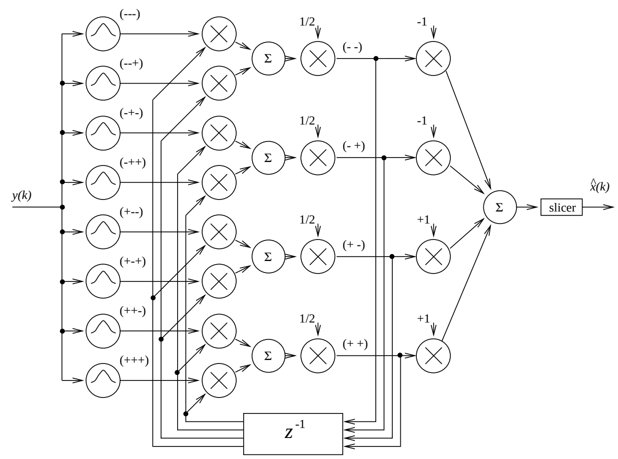


Figure 4: *Recurrent MAPSD for* $N = 3$. *(The notation* $(+\,+\,+)$ *on the layer of basis functions indicates the scalar centre which has been used e.g.* $(+\,+\,+)$ *implies the centre* $[+1, +1, +1]\mathbf{h}$. *The notation* $(+\,-)$ *indicates the values of the channel inputs upon the densities are conditioned e.g.* $(+\,-)$ *indicates densities conditioned on* $x(k) = +1$ *and* $x(k-1) = -1$.*)*

Finite memory: If the MAPSD decisions are based on a sliding window of M observations i.e. on the contents of the M-vector $\mathbf{y}_M(k) = [y(k)\ y(k-1) \cdots y(k-M+1)]^T$ where $M \ll k$, then it is finite memory. A radial basis function implementation of this detector is illustrated in Figure 5. Although the performance of this detector is superior to a FIR equaliser with the same memory it is inferior to the growing memory detector. In common with the FIR equaliser the performance does not improve monotonically with increasing lag d and there does not appear to be a systematic method for choosing the optimum value for d. It can be trained in a similar manner using a estimate of the channel impulse response. However it can also be trained using clustering techniques which facilitates its use on stationary nonlinear channels - training times will be longer than estimating the channel response. A major drawback is that its complexity grows with both N and M i.e. $O(P^{N+M})$ - recently however fuzzy logic implementation has reduced this to $O(P^N)$[16].

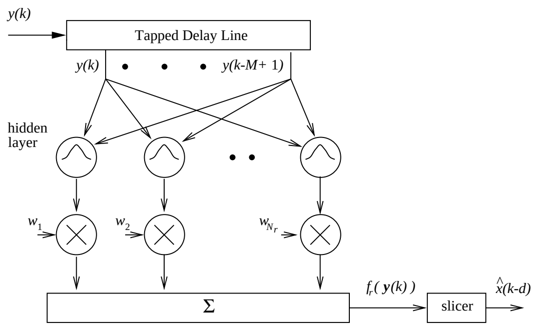


Figure 5: *Finite memory MAPSD - RBF implementation*

Decision feedback: Decision feedback can be viewed as a method of extending the memory of the finite memory detector by using previous decisions in making the current decision. One possible structure is illustrated in Figure 6. The nonlinear decision function can be implemented using a RBF. It is also worth noting that although the decision function is nonlinear the feedback network $\mathbf{H}_2$ is linear. In [17] a structure similar to this was demonstrated to outperform a VA-based MLSE on fading multipath channels at medium to high SNR's. The reasons for this was linked to the robustness of the MAP detector to errors in the channel impulse. It is worth pointing out however that recent advances [18] in the design of MLSE detectors for mobile radio channels may alter this conclusion. In common with the other MAP based detectors, the decision feedback variant is trained using an estimate of the channel response and the noise variance.

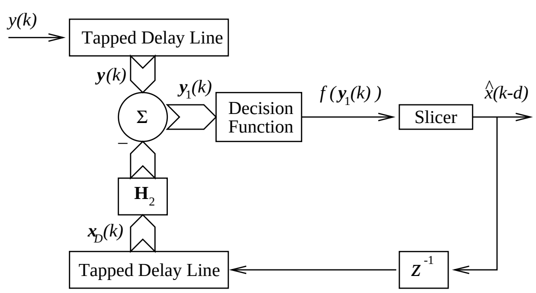


Figure 6: *Decision feedback MAPSD with linear feedback network* $\mathbf{H}_2$

If the MAP/RBF-based detectors are optimal and can be trained quickly, why then should we consider any other nonlinear signal processing architecture for channel equalisation? The answer is again complexity!

Even with the most efficient algorithms complexity grows exponentially i.e. $O(P^N)$. Thus MAPSD and MLSE have tend to be restricted to systems with small symbol alphabets and ISI that extends over only a few symbols . It is pertinent at this stage to point out that as a step towards third generation mobile radio systems (and higher bit rates) the European mobile radio standard GSM incorporates what is known as EDGE (enhanced data rates for GSM evolution). This part of the standard will employ 16-QAM data modulation i.e. $P = 16$ rather than the current $P = 4$. Further, the use of a linear rather than a nonlinear modulation scheme may increase the extent of the ISI. Thus, in the drive to increase bit-rate on a channel, the complexity of the equaliser will grow. The second factor that increases computational load is a requirement on some systems to reject co-channel interference (CCI) and/or adjacent channel interference (ACI). In this sort of environment the complexity grows at an even more alarming rate.

Both CCI and ACI bring with them an additional problem - training data is usually only available for the users own channel. While it is not necessary to equalise the adjacent or co-channel (i.e. detect their symbols) a full MAPSD or MLSE requires that the pdf structure of the component of the received signal associated with CCI and/or ACI be fully characterised e.g. [19]. Characterisation of this pdf structure can be a slow process since it involves unsupervised clustering. In addition the CCI and/or ACI component of the received signal presents itself as coloured noise to system identification algorithms used to estimate the impulse response of the users own channel. In particular the LS algorithm is no longer the best linear unbiased estimate (BLUE) and hence its performance will degrade with respect to the same level of white noise. Thus in the presence of CCI and/or ACI the complexity of the MAPSD structures of Figures 4-6 grow more rapidly than in the single channel case and the convergence performance will be poorer.

There are now a wide variety of nonlinear signal processing techniques which have been applied to the problem of adaptive equalisation of communications channel with ISI and which could also, at least theoretically, accommodate systems with CCI and/or ACI. These include: multi-layer perceptrons (MLP's) [20-24]; Volterra series [25]; polynomial perceptrons [25-27]; radial basis functions [3, 28]; piece-wise linear techniques [29-31]; functional link networks [32-34]; fuzzy filters [16, 35-38]. Here, for reasons of space restrictions, only three will be discussed. All three are drawn from fairly recent journal publication and hence may be representative of state-of-the-art.

Polynomial perceptrons: These date back to [25]. In their simplest forms they are a truncated Volterra expansion followed by memoryless nonlinearity such as a sigmoid function. The motivation behind this structure is the slow convergence (in the series sense) of the Volterra expansion which can lead to very large networks to approximate fairly moderate MAPSD's. The addition of the sigmoid function was demonstrated to reduce the complexity with respect to the Volterra expansion for a given level of performance. However, in adding the sigmoid function, the network is no longer linear-in-the-parameter and hence a LS training strategy is not available. In [26] the functional approximation properties of the polynomial perceptron were proved and the improved performance of fractionally spaced and recurrent variants of the algorithm were demonstrated. In particular the fractionally spaced bilinear perceptron is faster converging and significantly less complex than a multi-level perceptron equaliser. However the convergence speed was too slow for mobile radio applications taking 2000 symbols to converge. Another significant aspect of the results is that the channel used was a fairly realistic simulation of a radio channel with CCI and used a large symbol alphabet i.e. 16-QAM. A gradient lattice polynomial perceptron was introduced in [27] which improves convergence speed and makes converge less dependent on channel characteristics. Thus, while the BER performance of these structure looks extremely promising the open issue is to find faster training methods. In this context it should be recalled that LMS identification of the channel impulse under white input conditions does not converge fast enough for mobile radio applications. Gradient based algorithms do not converge fast enough even if they employ an orthogonalising network. Perhaps what is required is a method to calculate the weights of a polynomial perception from the channel estimate or at least use the channel estimate to provide a good initial condition for the gradient algorithm?

Clustering-based techniques: Supervised clustering [39] provides a fast and efficient method of estimating the conditional means of the vector $\mathbf{y}_M(k)$ associated with the finite memory MASPD. Although it is slower than estimating the channel impulse response in the case of a single linear channel, it will accommodate nonlinear finite memory channels and it is not degraded by the presence of CCI or ACI. The use of supervised clustering was extended in [40] by observing that the conditional mean of $y(k)$ is dependent on the conditional mean of $y(k-1)$ because of the channel memory. This fact is exploited by constructing a Viterbi trellis based on the vector $[y(k)\ y(k-1)]^T$ to provide a MLSE. Further, a Mahalanobis rather than an Euclidean distance is employed - effectively modelling the CCI and/or ACI as coloured noise, correlated over 2 symbols. The other advantage of the Mahalanobis distance is that complexity reduction can be easily accommodated by using a smaller number of clusters than there are local means and modelling the distribution of local means around a cluster as coloured Gaussian noise. Other methods for reducing complexity are discussed in [41] and cluster based blind nonlinear channel estimation is considered in [42].

Conditional distribution learning: Most nonlinear signal processing approaches to channel equalisation invoke a mean squared error (MSE) cost function. However, only when a normalised RBF network is employed, as in [43], will global minimisation of this cost function produce a MAPSD. In general minimisation of the MSE cost function will not produce the MAPSD - the appropriate cost function is probability of a symbol error. An alternative approach, described in [24], is to use an MLP to approximate the conditional probability density associated with a MAPSD. The authors show that by minimising the partial likelihood (PL) cost function the distance between the conditional pdf estimate and the actual conditional pdf is also minimised. (Likelihood refers to the estimate of the parameters of the MLP rather than a MLSE and distance is distance in the Kullback-Leiber sense.) The well-formed nature of the negative log PL cost function reduces the chances of local minima which severely restrict the use of minimum MSE MLP equalisers. Although the convergence rate of the algorithm is only slightly superior to a minimum MSE MLP equaliser the PL-based algorithm is much more robust to abrupt changes in the channel response and hence may be appropriate for equalisers operating in a decision directed mode.

4 MULTI-USER DETECTORS FOR CDMA SYSTEMS

By their very nature code division multiple access (CDMA) spread spectrum systems are subject to co-channel interference. The effect of multi-path distortion on these systems is to reduce the inherent immunity to co-channel interference provided by mutually orthogonal spreading codes. There are now a wide variety of receivers, known as multi-user detectors, that have been proposed to deal with such signalling environments. Several recent review articles are available [4-6]. The development of the adaptive MUD closely parallels that of the adaptive equalisers - particularly equalisers used to combat both ISI and CCI. The optimum MLSE for the problem was defined and analysed in [44]. However the sheer complexity of this receiver has prohibited its use. The focus of this section will be on the application of nonlinear signal processing techniques to the problem. To fit within the page budget the discussion will be limited to what is known as synchronous CMDA - in particular the downlink from the base station to the mobile handset. The discrete-time model of this is illustrated in Figure 7 using symbology from the multi-rate filtering literature. There are Q-users each transmitting at the symbol rate e.g. the ith user transmits a symbol sequence $\{u_i(k)\}$. The sampling rate is increased by the spreading factor L by zero-padding and the resultant sequence is applied to an L-tap FIR code filter $C_i(z)$. The outputs of each of these filter are summed before transmission through a multipath channel with sampled impulse response $H(z)$.

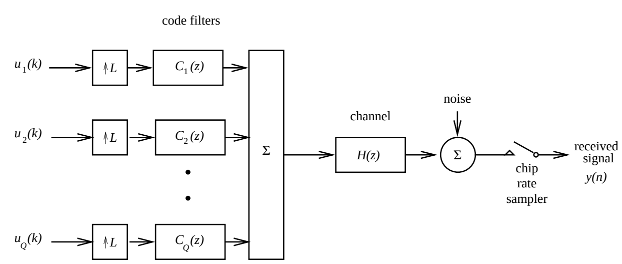
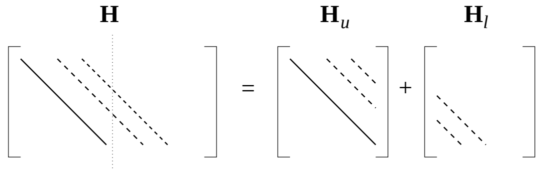


Figure 7: *Discrete-time model of CDMA downlink.*

Assuming that convolution of code filter $C_i(z)$ and channel impulse response $H(z)$ extends over a maximum of $2L$ samples (i.e. 2 symbols), a vector $\mathbf{y}_L(k)$ of L consecutive received samples can be written as:

$$\mathbf{y}(k) = [\ y(kL)\ y(kL-1) \cdots y(kL-L+1)\ ]^T$$
$$= \mathbf{H}\begin{bmatrix} \mathbf{Cu}(k) \\ \mathbf{Cu}(k-1) \end{bmatrix} + \mathbf{n}(k)$$

where $\mathbf{H}$ is the $(L\times 2L)$ channel impulse response matrix which is sparse and can be partitioned in the following manner:



$\mathbf{C}$ is the $(L \times Q)$ code matrix - the columns are the codes - $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2 \cdots \mathbf{c}_Q]$ where $\mathbf{c}_i = [c_{i0}\ c_{i1} \cdots c_{i(L-1)}]^T$ and $C_i(z) = \sum_{j=0}^{L-1} c_{ij} z^{-j}$; Q is the number of users and $Q < L$; $\mathbf{u}(k) = [u_1(k)\ u_2(k) \cdots u_Q(k)]^T$ contains the transmitted symbols from each of the user at time k. Thus we can write:

$$\mathbf{y}_L(k) = \mathbf{H}_u \mathbf{C} \mathbf{u}(k) + \mathbf{H}_l \mathbf{C} \mathbf{u}(k-1) + \mathbf{n}(k)$$

Using this formulation the similarity to the equalisation problem is evident. For a 2-element symbol alphabet, all possible combinations of the user symbol vectors give rise to 2^{2Q} local means of $\mathbf{y}_L(k)$. If we are only interested in one particular symbol from one particular user we have a Bayesian classification problem. In [45] a feedforward MLP was applied to a CDMA system. As ISI was not considered the state equations reduce to: $\mathbf{y}_L(k) = \mathbf{C}\mathbf{u}(k) + \; + \mathbf{n}(k)$. Since there is no memory in the system optimal detection can be achieved by considering $\mathbf{y}_L(k)$ alone i.e. a one-shot detector. The conclusions drawn were similar to experience in the application of MLP's to adaptive equalisation - the performance could be a lot better than a linear detector but training times could be long and unpredictable. Mitra and Poor [46] applied RBF techniques to the same problem. Because of the intimate link between the Bayesian detector and the RBF, the architecture of the former is well defined (unlike the MLP). The training times were better and more predictable than the MLP but still not fast enough for practical application. The complexity grows exponentially with number of users which also restricts application. A finite memory Bayesian detector was considered in [47] to improve performance in ISI. The detector input uses the aggregate vector $[\mathbf{y}_L^T(k)\ \mathbf{y}_L^T(k-1)]^T$ and hence incurs an increase in complexity to 2^{3Q} local means of this vector. The performance gains of this RBF receiver with respect to other detectors are also highlighted. Volterra series detectors have also been explored [48].

Decision feedback: If we are prepared to make hard decisions on all users at $k-1$ then we can assume $\mathbf{u}(k-1)$ is known - hard decisions are indicated as $\hat{\mathbf{u}}(k-1)$. This gives the state equations which defines a MAP detector:

$$\mathbf{y}'_L(k) = \mathbf{y}(k) - \mathbf{H}_l \mathbf{C} \hat{\mathbf{u}}(k-1) = \mathbf{H}_u \mathbf{C} \mathbf{u}(k) + \mathbf{n}(k)$$

The resultant decision feedback (DF) detector is illustrated in Figure 8. The similarity to Figure 6 is striking. The local means of $\mathbf{y}'_L(k)$ are defined by all possible combinations of $\mathbf{H}_u \mathbf{C} \mathbf{u}(k)$ and hence the number of centres is 2^Q - all possible combinations of $\mathbf{u}(k)$. However in a mobile terminal, only the signal from one user may be required. Thus if we are only prepared to make hard decisions on one users at $k-1$ then we can assume only $u_1(k-1)$ is known - call this hard decisions $\hat{u}_1(k-1)$. The the state equations are thus:

$$\mathbf{y}'_L(k) = \mathbf{y}(k) - \mathbf{H}_l \mathbf{C}\ [\hat{u}_1(k-1)\ 0 \cdots 0]^T$$
$$= \mathbf{H}_u \mathbf{C} \mathbf{u}(k) + \mathbf{H}_l \mathbf{C}\ [0\ u_2(k-1) \cdots u_M(k-1)]^T + \mathbf{n}(k)$$

In this case there are 2^{2Q-1} centres - DF only reduces the centres by a factor of 2. Thus in designing a MUD the classification problem is similar to that observed in equalisation but it is in general much more complex because the number of noise free states or local means is much larger. Linear-combiner decision feedback structures for CDMA are considered in [5].

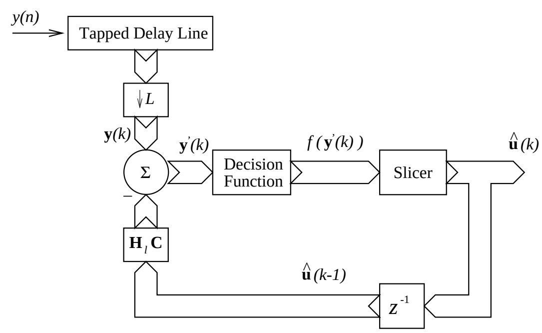


Figure 8: *Bayesian DF MUD*

Decision feedback improves the performance of finite memory MAP detectors as it does in equalisation. Although it reduces the complexity the reductions are not significant enough to make it viable. This has motivated the search for techniques which reduce the complexity further and still maintain some of the performance gains associated with a MAPSD detectors. Thus, for example, in [49] a RBF network is combined with a piecewise linear detector.

Training: As the transmit code filters $\{C_i(z)\}$ are generally known to the receiver, the key element in training is estimating the channel impulse response. If training data was inserted on all user channels simultaneously, identification would be fairly straightforward. However it is common in some systems to label one user as a pilot channel and transmit a simple training sequence continuously on this pilot channel alone. The transmitted power of the pilot is larger than that of the other users and the problem becomes one of system identification in coloured noise. Having estimated the channel impulse response a variety of detectors can be formed.

Hopfield neural networks: The other major nonlinear signal processing technique that has been applied to the problem of constructing a MUD in CDMA is the Hopfield network. In [50] the one-to-one relationship between the maximum-likelihood cost function used in the optimum detector of [44] and the energy cost function, which is minimised by a Hopfield network, is demonstrated. A Hopfield network is a nonlinear dynamical system with multiple inputs and outputs and many feedback paths. For a synchronous CDMA system the inputs are formed by sampling a bank of code matched filters at the symbol rate and the outputs are estimates of the current symbol from the individual user - as illustrated in Figure 9. The Hopfield network is defined by the Q state variables $\{v_i : 1 \le i \le Q\}$. The dynamics are characterised by differential equations of the form:

$$\dot{v}_i = -\frac{v_i}{\tau} - 2 \sum_{\{j:1 \le j \le Q; j \ne i\}} C_{ij}\ \hat{u}_j + 2z_i$$

τ is a time constant; C_{ij} is the element (i, j) of the code correlation matrix $\mathbf{C}^T\mathbf{C}$; $z_i(k)$ is the output of a code matched filter i.e. $z_i(k) = \mathbf{c}_i^T\ \mathbf{y}_L(k)$. The outputs of the network are formed by applying each state variable to a memoryless nonlinearity, i.e.: $\hat{u}_i = \tanh(v_i)$. The summation deliberately prevents direct feedback from the output of a state variable to itself - the restriction that $j \ne i$ is placed on the index i. Without multipath the gains on each of the many feedback paths can be calculated directly from the codes. The detector operates as follows: the outputs from code matched filters are sampled and held at the symbol rate; during the symbol period the network converges to a steady state; in the steady state each output is an estimate of the current symbol from each user. In the steady state the rate of change of all state variables is zero $(\dot{v}_i = 0; \forall i)$ and hence:

$$v_i = 2\tau(\ z_i - \sum_{\{j:1 \le j \le Q; j \ne i\}} C_{ij}\ \hat{u}_j\ )$$

Thus, at equilibrium, the summation in the brackets is equal to the component of the matched filter output due to all users except the *i*th one and the difference between the matched filter output and this summation is the residual component due to the *i*th user alone. Given a reasonable large time constant τ the output $\hat{u}_i$ will be a hard decision - either +1 or −1. This approach is further examined in [51] and [52] and extended to systems with multipath in [53]. The attraction of the approach is that the Hopfield network lends itself to implementation in analogue VLSI with potential gains in cost and power consumption.

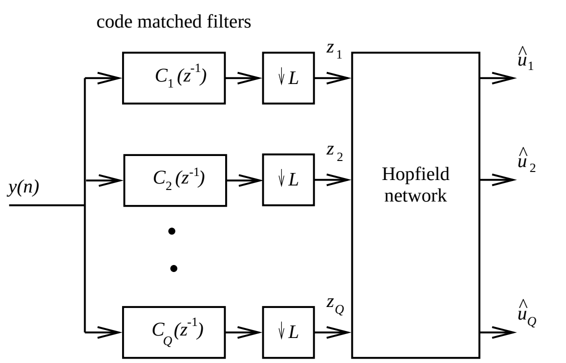


Figure 9: *Hopfield neural network detector for synchronous CDMA*

5 CONCLUSIONS

Direct adaptive equalisation schemes, where a channel estimate is used to contruct a MAPSD or MLSE, are the only techniques which converge and track fast enough for nonstationary environments such as mobile radio system. However the growth in symbol alphabet size and multipath duration increases the complexity of such detectors exponentially. Further, any requirement to combat CCI and ACI both increases complexity and training times. In this light nonlinear signal processing architectures can provide a good compromise between complexity and performance - the key issue and challenge is how to improve dramatically the training times? MUD is a younger and more open topic and a number of linear-combiner type techniques have been applied to it. The key issue with regard to the application of nonlinear signal processing techniques is complexity reduction.

Acknowledgements

Thanks to Rudi Tanner and David Cruickshank for many useful discussions on MAPSD, MLSE and MUD.

References

1. S.U.H. Qureshi, ‘‘Adaptive Equalization’’, *Proceedings IEEE*, vol. 73, no.9, pp. 1349-1387, 1985.
2. J. Cid-Sueiro and A.R. Figueiras-Vidal, ‘‘Channel Equalisation with Neural Networks’’, *Discrete Signal Processing in Telecommunications*, pp. 257-312, Springer, London, 1996.
3. B. Mulgrew, ‘‘Applying Radial Basis Functions’’, *IEEE Signal Processing Magazine*, vol. 13, no. 2, pp. 50-65, March 1996.
4. S. Verdú, ‘‘Adaptive Multiuser Detection’’, *Proceedings ISSSTA 94*, pp. 43-50, Oulu, Finland, July 1994.
5. A. Duel-Hallen, J. Holtzman, and Z. Zvonar, ‘‘Multuser Detection for CDMA’’, *IEEE Personal Communications*, pp. 46-58, Apr 1995.
6. S. Moshavi, ‘‘Multi-User Detection for DS-CDMA Communications’’, *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124-136, Oct 1996.
7. J.G. Proakis, *Digital Communications*, McGraw-Hill, 1984.
8. K. Giridhar, J.J. Shynk, A. Mathur, S. Chari, and R.P. Gooch, ‘‘Nonlinear Techniques for the Joint Estimation of Cochannel Signals’’, *IEEE Trans Communications*, vol. 45, no. 4, pp. 473-484, Apr. 1997.
9. R.A. Iltis, J.J. Shynk, and K. Giridhar, ‘‘Bayesian Algorithms for Blind Equalisation’’, *IEEE Trans Communications*, vol. 42, no. 2/3/4, pp. 1019-1032, Feb/Mar/Apr 1994.
10. G.D. Forney, ‘‘Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference’’, *IEEE Transactions Information Theory*, vol. IT-18, pp. 363-378, May 1972.
11. R.A. Iltis, ‘‘A Randomized Bias Technique for the Importance Sampling Simulation of Bayesian Equalisers’’, *IEEE Trans Communications*, vol. 43, no. 2/3/4, pp. 1107-1115, Feb/Mar/Apr 1995.
12. J. Cid-Sueiro, A. Artes-Rodriguez, and A.R. Figueiras-Vidal, ‘‘Recurrent Radial Basis Function Networks for Optimal Symbol-by-Symbol Equalisation’’, *Signal Processing*, vol. 40, pp. 53-63, 1994.
13. R. Arnott, ‘‘Diversity Combining for Digital Mobile Radio Using Radial Basis Function Networks’’, *Signal Processing*, vol. 63, no. 1, pp. 1-16, Nov 1997.
14. J.F. Hayes, T.M. Cover, and J.B. Riera, ‘‘Optimal Sequence Detection and Optimal Symbol-by-Symbol Detection: Similar Algorithms’’, *IEEE Trans Communications*, vol. 30, no. 1, pp. 152-157, Jan. 1982.
15. B. Mulgrew and C.F.N. Cowan, *Adaptive Filters and Equalisers*, Kluwer Academic Press, 1988. ISBN 0-89838-285-8
16. S.K. Patra and B. Mulgrew,, ‘‘Efficient Architecture for Bayesian Equalisation using Fuzzy Filters’’, *IEEE Trans Circuits & Systems II*,, to appear 1998.
17. S Chen, S McLaughlin, B Mulgrew, and P M Grant, ‘‘Adaptive Bayesian Decision Feedback Equaliser for Dispersive Mobile Radio Channels’’, *IEEE Transactions on Communications*, vol. 43, no. 5, pp. 1937-1946, May 1995.
18. G. Castellini, F. Conti, E. Del Re, and L. Pierucci, ‘‘A Continuously Adaptive MLSE Receiver for Mobile Communications: Algorithm and Performance.’’, *IEEE Trans Communications*, vol. 45, no. 1, pp. 80-89, Jan 1997.

19. S. Chen and B. Mulgrew, ‘‘Overcoming Co-channel Interference using an Adaptive Radial Basis Function Equaliser’’, *Signal Processing*, vol. 28 No. 1, pp. 91-107, 1992.
20. S. Siu, G.J. Gibson, and C.F.N. Cowan, ‘‘Decision Feedback Equalisation using Neural Network Structures and Performance Comparison with the Standard Architecture’’, *IEE Proceedings Part I*, vol. 137, no. 4, pp. 221-225, Aug. 1990.
21. G.J. Gibson and C.F.N. Cowan, ‘‘On the Decision Regions of Multilayer Perceptrons’’, *Proceedings IEEE*, vol. 78, no. 10, pp. 1590-1594, Oct. 1990.
22. M. Meyer and G. Pfeiffer, ‘‘Multilayer Perceptron Based Decision Feedback Equalisers for Channels with Intersymbol Interference’’, *IEE Proceedings*, vol. 140, no. 6, pp. 420-424, 1993.
23. G. Kechriotis, E. Zervas, and E.S. Manolakis, ‘‘Using Reccurrent Neural Networks for Adaptive Communications Channel Equalization’’, *IEEE Trans Neural Networks*, vol. 5, no. 2, pp. 267-278, Mar 1994.
24. T. Adali, X. Liu, and M. K. Sönmez, ‘‘Conditional Distribution Learning with Neural Networks and its Application to Channel Equalization’’, *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 1051-1064, Apr. 1997.
25. S. Chen, G.J. Gibson, and C.F.N. Cowan, ‘‘Adaptive Channel Equalisation Using a Polynomial Perceptron Structure’’, *IEE Proceedings part I.*, vol. 137, no. 5, pp. 257-264, Oct 1990.
26. Z.-J. Xiang, G.-G. Bi, and T. Le-Ngoc, ‘‘Polynomial Perceptrons and Their Applications to Fading Channel Equalization and Co-Channel Interference Suppression’’, *IEEE Transactions on Signal Processing*, vol. 42, no.9, pp. 2470-2480, Sept 1994.
27. Z.-J. Xiang and G.-G. Bi, ‘‘A New Lattice Polynomial Perceptron and Its Applications to Frequency-Selective Channel Equalization and ACI Suppression’’, *IEEE Trans Communications*, vol. 44, no. 7, pp. 761-767, July 1996.
28. S. Chen, G.J. Gibson, C.F.N. Cowan, and P.M. Grant, ‘‘Reconstruction of Binary Signals Using an Adaptive Radial Basis Function Equaliser’’, *Signal Processing*, vol. 22, no. 1, pp. 77-93, 1991.
29. C.P. Callender and C.F.N. Cowan, ‘‘A Comparison of 6 Different Equaliser Techniques for Digital Communications Systems’’, *VII European Signal Processing Conference EUSIPCO 94*, pp. 1524-1527, Edinburgh, Sept 1994.
30. X. Liu and T. Adali, ‘‘Recurrent canonical piecewise neural network and its application to adaptive equalization’’, *Proc. IEEE Int. Conf. on Neural Networks*, pp. 1969-1973, Washington, DC, June 1996.
31. C.Z.W. Hassell Sweatman, B. Mulgrew, and G.J. Gibson, ‘‘Two Algorithms for Neural Network Design and Training with Application to Channel Equalisation’’, *IEEE Trans Neural Networks*, to appear 1998.
32. W.S. Gan, J.J. Soraghan, and T.S. Durrani, ‘‘Application of the Functional Link Technique for Channel Equlaisation’’, *Electronics Letters*, vol. 28, no. 17, pp. 1643-1644, Aug. 1992.
33. J.C. Patra and R.N. Pal, ‘‘A Functional Link Artificial Neural Network for Adaptive Channel Equalisation’’, *Signal Processing*, vol. 43, pp. 181-195, 1995.
34. A. Hussain, J.J. Soraghan, and T.S. Durrani, ‘‘A New Adaptive Functional-Link Neural-Network-Based DFE for Overcoming Co-Channel Interference’’, *IEEE Trans Communications*, vol. 45, no. 11, pp. 1358-1362, Nov 1997.
35. L-X. Wang and J.M. Mendel, ‘‘Fuzzy Adaptive Filters with Application to Nonlinear Channel Equalization’’, *IEEE Trans Fuzzy Systems*, vol. 1, pp. 161-170, Aug 1993.
36. K.Y. Lee, ‘‘Fuzzy Adaptive Decision Feedback Equaliser’’, *Electronics Letters*, vol. 30, no. 10, pp. 749-751, 1994.
37. K.Y. Lee, ‘‘Complex RLS Fuzzy Adaptive Filter and Its Application to Channel Equalisation’’, *Electronics Letters*, vol. 30,, pp. 1572-1574, 15th Sept 1994.
38. K.Y. Lee, ‘‘Complex Fuzzy Adaptive Filter with LMS’’, *IEEE Trans Signal Processing*, vol. 44,, pp. 424-426, Feb 1996.
39. S Chen, B Mulgrew, and P M Grant, ‘‘A Clustering Technique for Digital Communications Channel Equalisation using Radial Basis Function Networks’’, *IEEE Trans. Neural Networks*, vol. 4, no. 4, pp. 570-579, July 1993.
40. S. Theodoridis, C.F.N. Cowan, C.P. Callender, and C.M.S. See, ‘‘Schemes for Equalisation of Communications Channels with Nonlinear Impairments’’, *IEE Proceedings Communications*, vol. 142, no.3, pp. 165-171, June 1995.
41. K. Georgoulakis and S. Theodoridis, ‘‘Efficient Clustering Techniques for Channel Equalisation in Hostile Environments’’, *Signal Processing*, vol. 58, pp. 153-164, 1997.
42. Y.J. Jeng and C.C. Yeh, ‘‘Cluster Based Blind Nonlinear Channel Estimation’’, *IEEE Trans Signal Processing*, vol. 45, no.5, pp. 1161-1172, May 1997.
43. I. Cha and S.A. Kassam, ‘‘Interference Cancellation using Radial Basis Function Networks’’, *Signal Processing*, vol. 47, no. 3, pp. 247-268, Dec 1995.
44. S. Verdú, ‘‘Minimum Probability of Error for Asynchronous Gaussian Multiple-Acess Channnels’’, *IEEE Trans Information Theory*, vol. IT-32, no. 1, pp. 85-96, Jan 1986.
45. B. Aazhang, B.-P. Paris, and G. C. Orsak, ‘‘Neural Network for Multiuser Detection in Code Division Multiple Access Communications’’, *IEEE Trans Communications*, vol. 40, no. 7, pp. 1212-1222, 1992.
46. U. Mitra and H.V. Poor, ‘‘Adaptive Receiver Algorithms for Near-Far Resistant CDMA’’, *IEEE Transactions Communications*, vol. 43, no. 2/3/4, Part III, pp. 1713-1724, Feb/Mar/Apr 1995.
47. D.G.M. Cruickshank, ‘‘Radial Basis Function Receivers for DS-CDMA’’, *Electronics Letters*, vol. 32, no. 3, pp. 188-190, 1st February 1996..
48. R. Tanner and D.G.M. Cruickshank, ‘‘Volterra Based Receivers for DS-CDMA’’, *8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC ’97*, vol. 3, pp. 1166-1170, Sep 1997.
49. R. Tanner, D.G.M. Cruickshank, C.Z.W. Hassell Sweatman, and B. Mulgrew, ‘‘Receivers for Nonlinearly Separable Scenarios in DS-CDMA’’, *Electronics Letters*, vol. 33, no. 25, pp. 2103-2105, Dec 1997.
50. T. Miyajima, T. Hasegawa, and M. Haneishi, ‘‘On the Multiuser Detection Using a Neural Network in Code-Division Muliple-Access Communications’’, *IEICE Trans Communications*, vol. E76-B, no. 8, pp. 961-968, Aug 1993.
51. G.I. Kechriotis and E.S. Manolakis, ‘‘Hopfield Neural Network Implementation of the Optimal CDMA Multiuser Detector’’, *IEEE Trans Neural Networks*, vol. 7, no. 1, pp. 131-141, Jan 1996.
52. G. Kechriotis and E.S. Manolakis, ‘‘A Hybrid Digital Signal Processing-Neural Network CDMA Multiuser Detection Scheme’’, *IEEE Trans Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, no. 2, pp. 96-104, Feb 1996.
53. W.G. Teich and M. Seidi, ‘‘Code Division Multiple Access Communications: Multiuser Detection based on a Recurrent Neural Network Structure’’, *ISSSTA 96*, pp. 979-984, Mainz, Germany, 1996.