Proc. Natl. Sci, Counc. ROC(A)
Vol. 24, No. 1, 2000. pp. 61-72

(Scientific Note)

Adaptive Bayesian Decision Feedback Equalizer for Radio Channel with High-power Amplifier

KUANG-JEN WANG AND SHIUNN-JANG CHERN†

Department of Electrical Engineering,
National Sun Yat-Sen University
Kaohsiung, Taiwan, R.O.C.

(Received September 18, 1998; Accepted July 20, 1999)

ABSTRACT

The optimal equalization solution under the classical symbol-by-symbol decision-making architecture is an inherently nonlinear problem; therefore, some degree of nonlinear decision making ability is desirable in the equalizer structure, even for a linear channel. The Bayesian equalizer has been shown to be an optimum solution for a symbol-by-symbol equalizer in terms of signal detection. In this paper, the performance of the adaptive Bayesian decision feedback equalizer (DFE) with the complex radial basis function/stochastic gradient (CRBF/SG) algorithm used as a non-linear channel (e.g., a radio channel with a high-power amplifier) estimator for channel equalization is investigated. Also, the Bayesian decision making (or classification) of the adaptive Bayesian DFE, for the M-level complex signaling scheme, is implemented using the CRBF network. To evaluate the performance of this adaptive Bayesian DFE, the error rate in symbol detection is evaluated and is shown to be better than that of the conventional least mean square (LMS) DFE and other existing methods.

Key Words: adaptive Bayesian equalizer, high power amplifier, non-linear distortion, complex radial basis function, channel estimator

I. Introduction

Channel equalization is the recovery of a signal distorted in transmission through communication channels having amplitude and delay distortions. It can be viewed as a classification problem where the equalizer is constructed as a decision-making device which reconstructs the transmitted symbol sequence as accurately as possible (Lee *et al*., 1995; Chang and Wang, 1995; Chen *et al*., 1993b, 1995). For signal detection, basically, two categories of equalizers, viz., sequence-estimation and symbol-by-symbol-decision equalizers, are considered. The optimal sequence-estimation equalizer is known to be the maximum likelihood sequence estimator (MLSE) while the optimal symbol-by-symbol equalizer is the Bayesian equalization solution. It has been shown that, for stationary linear channels, the performance of the adaptive Bayesian decision feedback equalizer (DFE) is better than that of the conventional least mean square (LMS) DFE but is inferior to that of the adaptive MLSE (Chen *et al*., 1994b, 1995). However, the adaptive Bayesian DFE has a significant advantage over the adaptive MLSE for rapidly time-varying channels, such as dispersive Mobile radio channels (Chen *et al*., 1995).

The realization of the optimal equalization solution under the classical architecture for making decisions symbol-by-symbol is known to require a nonlinear processing capability. Therefore, some degree of nonlinear decision making ability is desirable in the equalizer structure, even in the case of a linear channel (Gibson *et al*., 1991; Lazzarin *et al*., 1994). On the other hand, nonlinear distortion is now a significant factor hindering further increase in the attainable data rate. Although sources of channel nonlinear characteristics, such as non-linearity in data converters, may be regarded as memory-less, these nonlinear components are connected to or embedded in a linear dynamic network having memory; consequently, the overall channel response is a nonlinear dynamic mapping. That is, the signal received at each sample instant is a nonlinear function of the past values of the transmitted symbols. Due to the

† To whom all correspondence should be addressed.

fact that nonlinear distortion varies with time and from place to place, in general, effective non-linear compensation should be adaptive.

In this paper, we will investigate the performance of the adaptive Bayesian DFE with a non-linear channel estimator for application to a radio channel with a high-power amplifier (HPA) (Chen *et al*., 1994a; Pupolin and Greenstein, 1987). For the radio channel of interest, its input-output response can be totally described by a pair of AM-AM and AM-PM responses. The fundamental description will be addressed in Sec. V.

It is known that in the conventional adaptive Bayesian DFE (Lee *et al*., 1995; Chen *et al*., 1993b), a linear channel estimator is required in order for the linear dispersive channel to estimate the channel output states required for decision making in the Bayesian DFE. Because the channel considered in this paper is non-linear, a non-linear channel estimator is required and is implemented by the complex radial basis function (RBF) network. The operation of RBF networks is based on basis functions, and the functions a RBF network implements are generally non-analytic. However, an RBF network can approximate any analytic function arbitrarily well in a compact domain (Cha and Kassam, 1995). Recently, RBF networks have been extensively used in the area of channel equalization (Chen *et al*., 1991, 1994b; Cha and Kassam, 1995).

In Chen *et al*. (1991, 1993b), an RBF network was employed to implement the Bayesian decision-making function for linear channel equalization based on binary signals (real-valued inputs). Moreover, Chen *et al*. (1994a, 1994b) and Cha and Kassam (1995) extended the real-valued RBF network to obtain the complex-valued RBF network, the so-called complex RBF (CRBF) network. In Chen *et al.* (1994a), the CRBF network with two tracking algorithms, viz., the orthogonal least squared (OLS) and the hybrid clustering (HC) algorithms, was used as a channel estimator to estimate the states of a non-linear channel. On the other hand, in Chen *et al*. (1994b), the CRBF network with a clustering algorithm, which is different from the HC algorithm discussed by Chen *et al*. (1994a), was used to directly estimate the states of a linear dispersive channel and the Bayesian equalizer solution (or Bayesian decision-making function). Also, in Cha and Kassam (1995), the so-called stochastic gradient (SG) training algorithm, which can adopt all the free parameters of RBF networks, was employed for non-linear channel equalization with third-order non-linearity. The SG method has been shown to be well-suited for RBF networks with localized basis functions (e.g., Gaussian function) and to have better performance in terms of the normalized mean-squared-error (NMSE) and symbol-error-rate (SER), compared to the HC algorithm and other existing methods (Cha and Kassam, 1995). It is noted that the SG algorithm is well-suited for on-line adaptive signal processing unlike block-processing algorithms, such as the OLS algorithm.

In this paper, we consider a more complicated non-linear channel, a radio channel with an HPA, where the non-linear channel estimator as well as the Bayesian decision-making function are both implemented using CRBF networks, along with the HC algorithm and SG algorithm, to achieve the optimum solution of the Bayesian DFE. Here, the performance of the adaptive Bayesian DFE in terms of the error rate is investigated. This paper first reviews the conventional Bayesian DFE and discusses the rationale behind it. After that, the non-linear channel estimator based on the CRBF network with the SG training algorithm will be addressed and used to update the channel output states which are required in the Bayesian decision making function for non-linear channel equalization. The performance of the adaptive Bayesian DFE based on the proposed non-linear channel estimator is investigated with 4-ary quadrature amplitude modulation (4-QAM) signals and compared with the non-linear channel estimator with the CRBF/HC algorithm and other conventional methods.

II. Adaptive Bayesian Equalizer with Decision Feedback

In this section, the fundamental idea behind the adaptive Bayesian DFE discussed by Lee *et al*. (1995) and Chen *et al*. (1993b) is introduced. To proceed with the development, the configuration of the basic communication system model with the adaptive Bayesian DFE is illustrated in Fig. 1, where a complex-valued digital sequence $s(k)$ is transmitted through a dispersive complex channel and the channel output is corrupted by added complex-valued noise $e(k)$. Generally, the channel output signal is governed by the following discrete time difference dynamic equation, i.e.,

$$\hat{r}(k) = C(s(k), \cdots s(k - n_a + 1)), \tag{1}$$

where $C(\cdot)$ is a functional of complex-valued digital sequence. For example, if the channel is linear, Eq. (1) can be expressed as

$$\hat{r}(k) = \sum_{i=0}^{n_a-1} a_i s(k-i), \tag{2}$$

where $\{a_i\}$ is the impulse response of the channel and n_a is denoted as the maximum lag of the channel. For the 4-QAM signaling scheme, the constellation of $s(k)$ is given by

$$s(k) = s_R(k) + is_I(k) = \begin{cases} s^{(1)} = 1+i, \\ s^{(2)} = -1+i, \\ s^{(3)} = 1-i, \\ s^{(4)} = -1-i. \end{cases} \tag{3}$$

Here, the real and imaginary parts, $s_R(k)$ and $s_I(k)$, of $s(k)$ are mutually independent sequences. In consequence, the received signal can be represented by

$$r(k) = \hat{r}(k) + e(k). \tag{4}$$

Here, the real and imaginary parts, $e_R(k)$ and $e_I(k)$, of $e(k)$ are assumed to be white Gaussian processes with variance σ_e^2 and to be mutually independent of each other. Also, $e(k)$ and $s(k)$ are assumed to be uncorrelated. In symbol-by-symbol equalization, the task of the equalizer is to reconstruct the transmitted symbols as accurately as possible based on the noisy channel observation $r(k)$.

1. The Optimal Bayesian Solution of Equalizer

In Fig. 1, the equalization process essentially uses the information present in the observed channel output vector, with feed-forward weight order, m, which is defined by

$$\boldsymbol{r}(k) = [r(k) \cdots r(k-m+1)]^T, \tag{5}$$

to produce an estimate, $\hat{s}(k-d)$, of $s(k-d)$, where d denotes the equalizer delay. For a noiseless linear FIR channel as defined in Eq. (2), the filter coefficients of the equalizer may be chosen so that the overall channel-plus-equalizer filter has an impulse

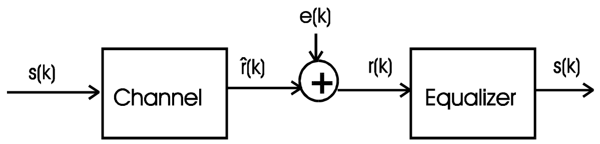


Fig. 1. Baseband discrete-time model of a data transmission system.

response that best approximates a unit impulse at time $k-d$. However, to achieve the optimum solution of the symbol-by-symbol equalizer for a non-linear channel, the channel estimator and the Bayesian solution approach, depicted in Fig. 1, have to be employed. In this subsection, the optimal Bayesian solution for the DFE equalizer is first reviewed.

Since from Eq. (5), the channel output vector involves m terms of delayed version of received signals and recognizing that the order of the channel is n_a, for 4-QAM, we can have $n_s = 4^{n_a+m-1}$ combinations of channel input sequences, i.e.,

$$\boldsymbol{s}(k) = [s(k) \cdots s(k-m-n_a+2)]^T. \tag{6}$$

Correspondingly, the noise-free channel output vector is

$$\hat{\boldsymbol{r}}(k) = [\hat{r}(k) \cdots \hat{r}(k-m+1)]^T, \tag{7}$$

and it has n_s desired corresponding states. Thus, the equalization problem can be treated as a classification problem. The task is to assign regions within the observation space spanned by the noisy vector, $\boldsymbol{r}(k)$ defined in Eq. (5), to represent the input $s(k-d)$ with four possible values, in the case of 4-QAM. To proceed the our discussion, we denote $\boldsymbol{R}_{m,d}$ as the set of noise-free channel output vectors. It can be partitioned into four subsets according to the value of $s(k-d)$, i.e.,

$$\boldsymbol{R}_{m,d} = \bigcup_{1 \le l \le 4} \boldsymbol{R}_{m,d}^{(l)}, \tag{8}$$

where

$$\boldsymbol{R}_{m,d}^{(l)} = \left\{ \hat{\boldsymbol{r}}(k) \middle| s(k-d) = s^{(l)} \right\} \quad 1 \le l \le 4 \tag{9}$$

with the number of states in $\boldsymbol{R}_{m,d}^{(l)}$ being $n_s^{(l)} = n_s/4$ and with each element denoted as $\hat{\boldsymbol{r}}_q^{(l)}$, where $\hat{\boldsymbol{r}}_q^{(l)} \in \boldsymbol{R}_{m,d}^{(l)}$ for $1 \le l \le 4$ and $1 \le q \le n_s^{(l)}$. It is noted that, in the noisy case, $\hat{\boldsymbol{r}}(k)$ can only be estimated based on $\boldsymbol{r}(k)$. Also, for convenience, we denote $\theta_q^{(l)}$ as the a priori probability of $\hat{\boldsymbol{r}}_q^{(l)}$. Under the assumption that all the desired states have the same probability of appearing, $1/n_s$, we have the conditional probability density function (p.d.f.) of $\boldsymbol{r}(k)$, given $s(k-d) = s^{(l)}$, as

$$\begin{aligned} &\Pr(\boldsymbol{r}(k) | s(k-d) = s^{(l)}) \\ &= \eta^{(l)}(\boldsymbol{r}(k)) = \sum_{q=1}^{n_s^{(l)}} \theta_q^{(l)} p_e(\boldsymbol{r}(k) - \hat{\boldsymbol{r}}_q^{(l)}), \; 1 \le l \le 4, \end{aligned} \tag{10}$$

where $p_e(\cdot)$ is the p.d.f. of the noise vector

$\boldsymbol{e}(k) = [e(k)...e(k-m+1)]^T$, and the sum in Eq. (10) is more than $\boldsymbol{r}_q^{(l)} \in \boldsymbol{R}_{m,d}^{(l)}$.

For convenience, we may introduce a complex decision function, $f_B(\boldsymbol{r}(k))$, denoted by

$$\begin{aligned} f_B(\boldsymbol{r}(k)) &= \Big\{\big[\eta^{(1)}(\boldsymbol{r}(k)) + \eta^{(3)}(\boldsymbol{r}(k))\big] \\ &\quad -\big[\eta^{(2)}(\boldsymbol{r}(k)) + \eta^{(4)}(\boldsymbol{r}(k))\big]\Big\} \\ &\quad + i\Big\{\big[\eta^{(1)}(\boldsymbol{r}(k)) + \eta^{(3)}(\boldsymbol{r}(k))\big] \\ &\quad -\big[\eta^{(3)}(\boldsymbol{r}(k)) + \eta^{(4)}(\boldsymbol{r}(k))\big]\Big\} \\ &= (1+i)\eta^{(1)}\boldsymbol{r}(k)) + (-1+i)\eta^{(2)}(\boldsymbol{r}(k)) \\ &\quad + (1-i)\eta^{(3)}(\boldsymbol{r}(k)) + (-1-i)\eta^{(4)}(\boldsymbol{r}(k)). \end{aligned} \tag{11}$$

From Appendix A of Chen *et al.* (1994b), we have an equivalent Bayesian solution, i.e.,

$$\hat{s}(k-d) = \mathrm{sgn}(f_B(r(k))). \tag{12}$$

Here, sgn(·) is the complex signum function and is defined as

$$\mathrm{sgn}(f_B) = \begin{cases} 1+i, & \mathrm{Re}[f_B] \ge 0 \cap \mathrm{Im}[f_B] \ge 0, \\ -1+i, & \mathrm{Re}[f_B] < 0 \cap \mathrm{Im}[f_B] \ge 0, \\ 1-i, & \mathrm{Re}[f_B] \ge 0 \cap \mathrm{Im}[f_B] < 0, \\ -1-i, & \mathrm{Re}[f_B] < 0 \cap \mathrm{Im}[f_B] < 0. \end{cases}$$

Substituting Eq. (10) into Eq. (11), the optimal Bayesian decision function takes the form

$$f_B(\boldsymbol{r}(k)) = \sum_{q=1}^{n_s} h_q p_e(\boldsymbol{r}(k) - \hat{\boldsymbol{r}}_q). \tag{13}$$

where

$$h_q = \begin{cases} \theta_q^{(1)} + i\theta_q^{(1)}, & \hat{\boldsymbol{r}}_q \in \boldsymbol{R}_{m,d}^{(1)}, \\ -\theta_q^{(2)} + i\theta_q^{(2)}, & \hat{\boldsymbol{r}}_q \in \boldsymbol{R}_{m,d}^{(2)}, \\ \theta_q^{(3)} - i\theta_q^{(3)}, & \hat{\boldsymbol{r}}_q \in \boldsymbol{R}_{m,d}^{(3)}, \\ -\theta_q^{(4)} - i\theta_q^{(4)}, & \hat{\boldsymbol{r}}_q \in \boldsymbol{R}_{m,d}^{(4)}, \end{cases}$$

with $\theta_q^{(l)}$ being defined as the a priori probability of $\hat{\boldsymbol{r}}_q^{(l)}$, for $1 \le l \le 4$. The intimate link between the Bayesian decision function and the CRBF network will be discussed in Sec. III.

2. Bayesian Decision Feedback Equalizer

As shown in Fig. 1, the estimated feedback vector of the equalizer with feedback order n is designated by

$$\hat{\boldsymbol{s}}_f(k-d) = \big[\hat{s}(k-d-1)\cdots\hat{s}(k-d-n)\big]^T. \tag{14}$$

The conventional DFE is based on a linear filtering of this equalizer input vector and is defined by

$$DFEEQ(\boldsymbol{r}(k), \hat{\boldsymbol{s}}_f(k-d)) = \boldsymbol{w}^T\boldsymbol{r}(k) + \boldsymbol{b}^T\hat{\boldsymbol{s}}_f(k-d), \tag{15}$$

where $\boldsymbol{w} = [w_1 \cdots w_m]^T$ and $\boldsymbol{b} = [b_1 \cdots b_n]^T$ are the coefficients of the feed-forward and the feedback of the equalizer. It has been shown that it is sufficient to employ a feedback order,

$$n = n_a + m - 2 - d, \tag{16}$$

for the adaptive Bayesian DFE to achieve the desired result (Chen *et al.*, 1993b, 1995). In consequence, the feedback vector $\hat{\boldsymbol{s}}_f(k-d)$ may have $n_f = 4^n$ possible combinations. If we denote $\boldsymbol{s}_{f,i}$, $1 \le i \le n_f$ as the feedback states, then a subset of the channel states $\boldsymbol{R}_{m,d}^{(l)}$ can be partitioned into n_f subsets according to the feedback states, that is,

$$\boldsymbol{R}_{m,d}^{(l)} = \bigcup_{1 \le i \le n_f} \boldsymbol{R}_{m,d,i}^{(l)}, \quad 1 \le l \le 4, \tag{17}$$

where

$$\boldsymbol{R}_{m,d,i}^{(l)} = \left\{\hat{\boldsymbol{r}}(k) \middle| s(k-d) = s^{(l)} \cap \hat{\boldsymbol{s}}(k-d) = \boldsymbol{s}_{f,i}\right\} \quad 1 \le i \le n_f. \tag{18}$$

Now, each subset $\boldsymbol{R}_{m,d,i}^{(l)}$ contains $n_{s,i}^{(l)} = n_s^{(l)}/n_f = 4^d$ states. Under the assumption of a correct feedback vector $\hat{\boldsymbol{s}}_f(k-d)$, the conditional Bayesian decision function, given $\hat{\boldsymbol{s}}_f(k-d) = \boldsymbol{s}_{f,i}$, takes the form (Chen *et al.*, 1994b)

$$f_B(\boldsymbol{r}(k) | \hat{\boldsymbol{s}}_f(k-d) = \boldsymbol{s}_{f,i}) = \sum_{l=1}^{4} \sum_{q=1}^{n_{s,i}^{(l)}} h^{(l)} \exp(-(\boldsymbol{r}(k) - \hat{\boldsymbol{r}}_q^{(l)})^H (\boldsymbol{r}(k) - \hat{\boldsymbol{r}}_q^{(l)}) / 2\sigma_e^2), \tag{19}$$

where the terms in the inner sum are those, $\hat{\boldsymbol{r}}_q^{(l)} \in \boldsymbol{R}_{m,d,i}^{(l)}$. It is noted that as discussed by Chen *et al.*

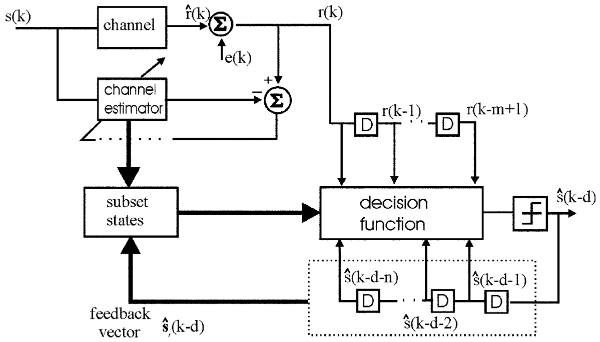


Fig. 2. A schematic diagram of a non-linear communication system with the adaptive Bayesian DFE.

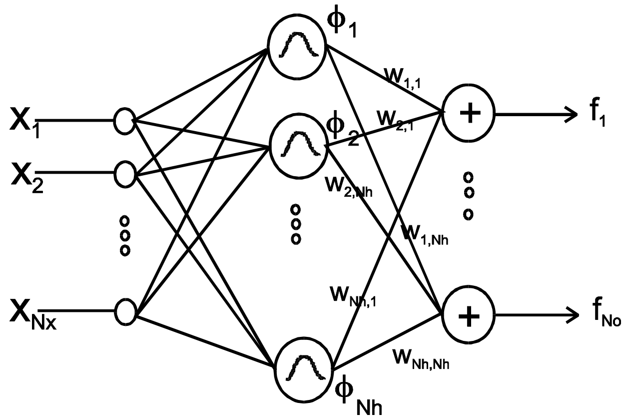


Fig. 3. Schematic diagram of a radial basis function (RBF) network.

(1994a, 1994b), the decision feedback in the Bayesian DFE can be employed not only to improve the performance of the Bayesian feed-forward equalizer, but also to reduce the computation complexity dramatically.

We may conclude that the structure of a Bayesian DFE is specified by the equalizer delay d, feed-forward order m and feedback order n. Here, d specifies the number of states required for computing the Bayesian decision function and, thus, determines the computational complexity. Given d, it can be proved that $m = d + 1$ is sufficient (Chen *et al.*, 1994a). This means that a Bayesian DFE of $m = d + 1$ will have the same performance as that of $m > d + 1$. Substituting this result into Eq. (16) gives rise to the corresponding feedback order $n = n_a - 1$. The overall computation load of the Bayesian DFE consists of three parts, namely, the channel estimator (based on the CRBF network with the training algorithm), calculation of $\boldsymbol{R}^{(l)}_{m,d,i}$ based on the states obtained by the channel estimator, and computation of $f_B(\boldsymbol{r}(k))$ (Chen *et al.*, 1995) (which is implemented by another CRBF network).

III. Complex Radial Basis Function Networks

The RBF network is known to be an extremely powerful type of feed-forward artificial neural network (Mulgrew, 1996; Widrow and Winter, 1988). It is a single-hidden-layer network, which can be used to implement a non-linear mapping on inputs by linearly combining the outputs of hidden nodes, each of which is computed by evaluating the radially symmetric basis function of the distance between the input and a parameter vector, called the center, associated with each hidden node.

A schematic of an RBF network with N_x inputs, N_0 outputs and N_h hidden nodes is depicted in Fig. 3. As shown in Fig. 3, an RBF network contains the following: (1) an input layer of branching nodes, one for each feature component, just as does a multilayer perceptron (MLP) (Chang and Wang, 1995); (2) a hidden layer of neurodes, where each neurode has a special type of activation function centered on the center vector of a cluster or sub-cluster in the feature space; therefore, the function has non-negligible response for input vectors close to its center; (3) an output layer of neurodes that sum the outputs from the hidden neurodes; that is, the output layer uses a linear activation function.

Such a network implements a mapping $\boldsymbol{f}$: $\boldsymbol{R}^{N_x} \rightarrow \boldsymbol{R}^{N_o}$ according to

$$f_j = \sum_{q=1}^{N_h} \phi_q w_{q,j} \quad 1 \le j \le N_o, \tag{20}$$

where $\phi_q = \phi(\| \boldsymbol{x} - \boldsymbol{c}_q \|)$ is a given function, $\|\cdot\|$ denotes the Euclidean norm, $\boldsymbol{c}_q \in \boldsymbol{R}^{N_x}$, $1 \le q \le n_h$ are known as the RBF centers, $\boldsymbol{x} \in \boldsymbol{R}^{N_x}$ is the input vector, and $w_{q,j}$ represents the connection weights. Theoretical investigation and practical results suggest that the choice of the non-linearity $\phi(\cdot)$ is not crucial to the performance of an RBF network. In this paper, the Gaussian function is chosen and is defined by

$$\phi(v) = \exp(-v^2 / \rho^2), \tag{21}$$

where the parameter ρ in Eq. (21) is used to control the spread of the radial basis functions such that its values decrease more slowly or more rapidly as $\boldsymbol{x}$ moves away from the center vector $\boldsymbol{c}_q$ as $\| \boldsymbol{x} - \boldsymbol{c}_q \|$ increases. Also, in Eq. (21), we see that $\phi(v) \rightarrow 0$ as $v \rightarrow \infty$.

So far, we have introduced the RBF network

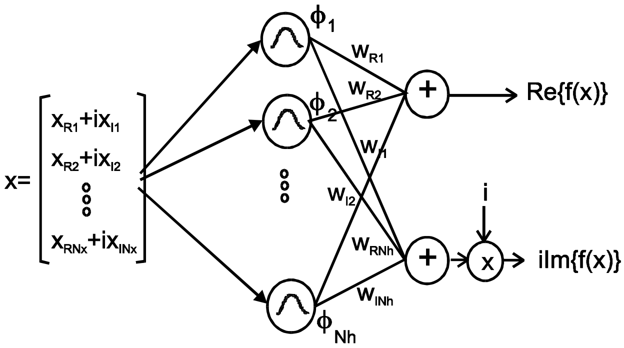


Fig. 4. Schematic diagram of a complex radial basis function network (CRBFN).

with real-valued input and output signals, and real-valued centers. In this paper, complex-valued signals, like 4-QAM modulated signals, are considered in the equalization problem. It is natural to extend the RBF network to obtain the CRBF network. The CRBF network model, since it preserves the radial symmetry of its real counterpart, is a natural extension of the RBF network. Chen *et al*. (1994a) and Cha and Kassam (1995) proposed an essentially identical complex RBF network extension structure independently. Such a network is depicted in Fig. 4.

In a CRBF network, both the input signals and the centers are complex-valued. The Euclidean norm of the difference between the input and each center is now defined in the usual way for complex-valued vectors, and the basis functions remain real-valued. In order to implement complex-valued outputs, we assign two different sets of weights, one for the real part of the network output and the other for the imaginary part. For simplicity, the network output vector is reduced to one dimension; therefore, in this case, the CRBF network operates on a complex-valued input vector $\boldsymbol{x} \in \boldsymbol{C}^{N_x}$ to produce a complex output according to Chen *et al*. (1994a) and Cha and Kassam (1995), i.e.,

$$f(\boldsymbol{x}) = \sum_{q=1}^{N_h} \phi(\|\boldsymbol{x}-\boldsymbol{c}_q\|)(w_{Rq} + iw_{Iq}), \tag{22}$$

where

$$\|\boldsymbol{x}-\boldsymbol{c}_q\|^2 \equiv (\boldsymbol{x}-\boldsymbol{c}_q)^H(\boldsymbol{x}-\boldsymbol{c}_q). \tag{23}$$

Here, the superscript H denotes the complex conjugate transpose, $\phi(\cdot)$ is still a real-valued radial basis function as shown in Eq. (20), and $w_q = w_{Rq} + iw_{Iq}$ and $\boldsymbol{c}_q \in \boldsymbol{C}^{N_x}$, $1 \le q \le N_h$ denote the N_h complex connection weights and N_x-dimensional vector complex-valued centers, respectively. Examination of Eqs. (22) and (23) reveals that the CRBF network treats the real and imaginary parts of an input as if they were two separate real inputs. Therefore, a CRBF network with N_x complex inputs and a complex output can be viewed alternatively as a real RBF network with $2N_x$ real inputs and two real outputs. In this sense, the complex RBF network is a straightforward extension of the real RBF network. It is noted here that to use a CRBF network as a channel estimator, the weights in Eq. (22) have to be updated by the training algorithms.

Next, we will discuss the intimate link between the Bayesian decision-making function of the equalizer and the CRBF network. By comparing Eq. (22) with Eq. (13), we learn that both equations have the same form; the connection weights, w_q, of the CRBF network and the coefficients h_q of $f_B(\cdot)$ are pairs. Also, the centers, $\boldsymbol{c}_q$, of the CRBF network correspond to the channel states, $\hat{\boldsymbol{r}}_q$, and the response of each hidden node resembles a component conditional p.d.f. in Eq. (13). Moreover, because the noise p.d.f. $p_e(\cdot)$ is real, the choice of real non-linearity for hidden nodes is well justified. Finally, by using $\boldsymbol{r}(k)$ as the input vector of the Bayesian DFE and $s(k-d)$ as the desired output, the CRBF network can be used to realize the solution of the Bayesian equalizer. In the next section, we will introduce the adaptive implementation of the Bayesian DFE, which consists of a channel estimator followed by Bayesian decision making, implemented using an independent CRBF network which has a similar structure.

IV. Adaptive Implementation

Basically, in the Bayesian DFE, the channel estimator first updates the connection weights of the CRBF network during a training period; thus the channel estimate may be fixed throughout a transmission period. In Chen *et al*. (1995), the LMS channel estimator was employed to estimate the linear channel model as defined in Eq. (2). Since, in this paper, a nonlinear channel is considered, the CRBF network with the SG training algorithm suggested by Cha and Kassam (1995) is adopted as a channel estimator. On the other hand, the CRBF network can also be used to implement the Bayesian decision functions defined in Eq. (13) and Eq. (19), respectively.

1. Estimating the Non-linear Channel Model Using the CRBF Network

In this section, we will address a CRBF network with the SG training algorithm used as an nonlinear channel estimator to estimate the channel states required for Bayesian decision-making. Note that Chen *et al*. (1995) employed the LMS channel estimator to estimate the channel states using linear channel model. First, for a linear channel model, as defined in Eq. (2), we can define the channel estimate as

$$\hat{\boldsymbol{a}}(k) = \left[\hat{a}_0(k)\cdots\hat{a}_{n_a-1}(k)\right]^T. \tag{24}$$

The LMS algorithm can be applied to the channel estimator to obtain a channel estimate:

$$\varepsilon(k) = r(k) - \hat{\boldsymbol{a}}(k-1)^T\hat{\boldsymbol{s}}_a(k), \tag{25}$$

$$\hat{\boldsymbol{a}}(k) = \hat{\boldsymbol{a}}(k-1) + \alpha_a\varepsilon(k)\hat{\boldsymbol{s}}_a(k), \tag{26}$$

where α_a is the adaptive gain or step size. During data transmission, a decision-directed and a delayed version of Eqs. (25) and (26) are used to track time-varying channels:

$$\varepsilon(k-d) = r(k-d) - \hat{\boldsymbol{a}}^T(k-d-1)\hat{\boldsymbol{s}}_a(k-d), \tag{27}$$

$$\hat{\boldsymbol{a}}(k-d) = \hat{\boldsymbol{a}}(k-d-1) + \alpha_a\varepsilon(k-d)\hat{\boldsymbol{s}}_a(k-d), \tag{28}$$

where $\hat{\boldsymbol{s}}_a(k-d) = [\hat{s}(k-d)\cdots\hat{s}(k-d-n_a+1)]^T$ is the estimated $\boldsymbol{s}_a(k-d)$, consisting of detector symbols. Given the estimated channel model $\hat{a}$, it is straightforward to compute the set of channel states $\boldsymbol{R}_{m,d}$. A noise variance estimator can also be incorporated into the channel estimator.

However, in the case of a nonlinear channel, the linear transversal filter structure with the LMS algorithm channel estimator can not cope with the serious distortion caused by the nonlinear property, inter-symbol interference (ISI) and added noise. That is, the linear estimator structure can not be used to identify the nonlinear channel correctly; the result of the computed channel states of $\boldsymbol{R}_{m,d}$ may not be precise and the performance, in terms of error probability, maybe degraded. Based on the above discussion, it is reasonable to adopt the CRBF network with the SG training algorithm as a non-linear channel estimator to obtain a better channel state estimate for Bayesian decision-making. Furthermore, the same CRBF network structure for channel estimation can also be used for Bayesian decision-making. This is easy and straightforward in terms of hardware implementation.

2. The Stochastic Gradient Training Algorithm

As described earlier, Cha and Kassam (1995) proposed a simple SG training algorithm for updating all the free parameters of a CRBF network simultaneously by using stochastic gradient descent for the error criterion, in the case of third-order non-linearity channel equalization. The SG algorithm takes the instantaneous gradient of the squared error $\|e(k)\|^2$ at the output and moves the network parameters in the direction of opposite that of their respective gradients. Here, as a non-linear channel estimator, the error signal $e(k)$ at time k is defined as $e(k) = f(\boldsymbol{x}(k)) - d(k)$ ($f(\boldsymbol{x}(k))$, a complex output of CRBF for the channel estimator as defined in Eq. (22)), where $d(k)$ denotes the complex-valued desired response. Thus, a network parameter (which can be a weight, a center, or a spread parameter) is adapted at time k according to

$$\lambda(k+1) = \lambda(k) - \alpha\frac{\partial\|e(k)\|^2}{\partial\lambda(k)}, \tag{29}$$

where α controls the speed of adaptation. The SG algorithm has certain advantages over existing methods (Cha and Kassam, 1995). First, all the free network parameters are adapted simultaneously, usually yielding improved overall solutions. This method can also provide greater robustness for poor initial choices of parameters, especially for the centers. Second, the algorithm is well-suited for on-line adaptive signal processing, unlike block-process algorithms, such as the Moody-Darken or the OLS algorithms (Chen *et al*., 1994b). It is also computationally quite feasible. All the parameters in a CRBF network are updated as follows:

$$w_t(k+1) = w_t(k) + \alpha_w e(k)\phi_t(\boldsymbol{s}_a(k)), \tag{30}$$

for $t = 1, 2, \ldots, N_h$, with null initial weights, when the SG algorithm is used to train the channel states, and

$$\sigma_t(k+1) = \sigma_t(k) + \alpha_\sigma\phi_t(\boldsymbol{s}_a(k))[\mathrm{Re}\{w_t(k)\}\mathrm{Re}\{e(k)\} + \mathrm{Im}\{m_t(k)\}\mathrm{Im}\{e(k)\}]\frac{\|\boldsymbol{s}_a(k)-\boldsymbol{c}_t(k)\|^2}{\sigma_t^3(k)}, \tag{31}$$

$$\boldsymbol{c}_t(k+1) = \boldsymbol{c}_t(k) + \alpha_c\phi_t(\boldsymbol{s}_a(k))\left[\frac{\mathrm{Re}\{w_t(k)\}\mathrm{Re}\{e(k)\}\mathrm{Re}\{\boldsymbol{s}_a(k)-\boldsymbol{c}_t(k)\} + i\,\mathrm{Im}\{w_t(k)\}\mathrm{Im}\{e(k)\}\mathrm{Im}\{\boldsymbol{s}_a(k)-\boldsymbol{c}_t(k)\}}{\sigma_t^2(k)}\right], \tag{32}$$

respectively. In our case, the centers are initialized by forming them using the first few symbols, ($\boldsymbol{s}(k)$), in the training mode (Cha and Kassam, 1995), where each center requires three data symbols and each symbol has four possible values as defined in Eq. (3). This is because the training symbols are available at the receiver end. For the purpose of comparison, to estimate the channel output states, we can also apply the so-called HC algorithm (see Appendix B of Chen *et al.* (1994a)) to the CRBF network as a channel estimator to estimate the channel states.

3. The CRBF Network for Bayesian Decision-Making

Comparing Eq. (22) with Eq. (19), we find that the structure of the CRBF network is identical structure to that of the Bayesian DFE. The CRBF network is, therefore, an ideal means of implementing the latter. We can set the number of hidden nodes in the network to match the number of desired states, $N_h = n_s$ (see Sec. III), and group the hidden nodes in accordance with the conditional subsets of the channel states. Whether a CRBF network can realize an optimal equalizer strongly depends on whether the centers can be corrected at the desired states. Here, the CRBF centers, $\boldsymbol{c}_q$, are, in fact, the corresponding states, $\hat{\boldsymbol{r}}_q \in \boldsymbol{R}_{m,d}$, and the feedback vector determines which subsets of hidden nodes are active at each sample k.

As suggested by Chen *et al.* (1993a), the width of the hidden nodes can be selected as $\rho = 2\sigma_e^2$ and the non-linearity of the hidden nodes can be selected as $\phi(x^2;\rho) = \exp(-x^2/\rho)$ (the Gaussian function). The hidden nodes realize conditional p.d.f's in the Bayesian solution. In this case, the connection weights of the CRBF network have fixed values due to the fact that the complex coefficients, $h^{(l)}$, in the Bayesian decision function have equal probability. Thus, the CRBF network precisely implements the Bayesian solution. As described by Chen *et al.* (1995), due to the nature of the Bayesian decision function, small variations in σ_e^2 cause hardly any change in the Bayesian decision boundary. Using the estimated noise variance to set the network, satisfactory accuracy in realizing the Bayesian equalizer can usually be obtained. Adjusting the RBF centers adaptively so that they converge to the channel states is the key in implementing the Bayesian solution. Thus, for Bayesian decision-making, we first apply the CRBF network to implement Eq. (19) (or Eq. (13)) in order to evaluate $f(\boldsymbol{x}(k))$ using the SG algorithm. After evaluating $f(\boldsymbol{x}(k))$, Eq. (12) is employed for final decision-making. As stated earlier, since $h^{(l)}$ in Eq. (19) has equal probability, the weights in Eq. (30) remain constant. We only need to update Eqs. (31) and (32) adaptively.

V. Computer Simulations

To investigate the performance of the proposed method, a communication channel with serious nonlinearity is considered. In this channel model, besides a high power amplifier, there additive white Gaussian noise and a static nonlinear device are included in the channel. We will first describe the general system blocks of this channel.

1. Communication System with the Nonlinear Transmit Amplifier Pupolin's Model

The digital radio link to be studied here is depicted in Fig. 5. The dashed boxes represent operations that may or may not be included in the design. All the time waveforms shown are either base-band or base-band equivalents (complex envelopes of band-pass signals). Similarly, all the frequency responses are either low-pass functions or low-pass equivalents of band-pass functions. Finally, all the sequences shown have complex values. In what follows, we will describe each component shown in Fig. 5.

(1) **Input Data:** The input is a sequence of M-level complex data values, $s(k)$, where k is the time index and successive data are spaced by T seconds. Each $s(k)$ is taken from an M-point letter of the alphabet $\{\gamma_m\}$, where m is the index for the M discrete points. The set consisting of these points in the complex

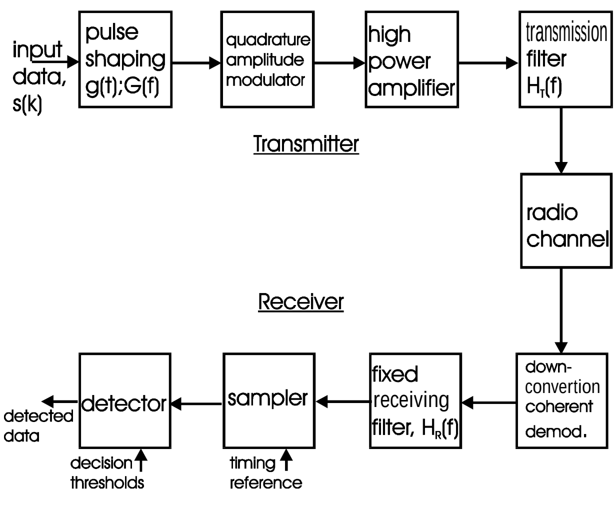


Fig. 5. Baseband equivalent model of a digital radio channel.

plane is called a data constellation.

(2) **Pulse Shaping:** The pulse shaping circuit transforms the input (complex) data stream into a pulse stream, $\sum_k s(k)g(t-kT)$. The choice of $g(t)$ or $G(f)$ (the Fourier transform of $g(t)$) is dictated by the need for spectrally efficient transmission. Accordingly, we will confine our attention to two classes of band-limited pulses, namely, cosine roll-off and root-cosine roll-off pulses:

$$G(f) = TC_{\cos in}(f;\rho) \text{ or } G(f) = T[C_{\cos in}(f;\rho)]^{1/2}, \quad \mathbf{(33)}$$

where $C_{\cos in}(f;\rho)$ is the cosine-roll-off function typically used in digital radio and ρ is the roll-off factor ($0 \le \rho \le 1$).

(3) **Modulator:** The (complex) base-band pulse stream is applied to a quadrature amplitude modulator, the bandpass output of which has a complex envelope, i.e.,

$$w(t) = a\sum_k s(k)g(t-kT). \quad \mathbf{(34)}$$

The constant a is an amplitude-scaling factor whose value determines the high-power amplifier (HPA) input power. If modulation is performed at IF, then up-conversion to RF is conducted prior to high-power amplification.

(4) **High-Power Amplifier:** For the radio channel of interest, wherein the bandwidths are 40 MHz or smaller, any HPA can be characterized as a memory-less band-pass non-linearity. This means that its input-output response can be completely described by a pair of AM-AM and AM-PM responses. For concreteness, we will begin this study with a functional description of z in terms of w that is known to be accurate for traveling wave tubes (TWT's). We will then apply a simple polynomial approximation and identify the input signal range over which this approximation is accurate. The functional description we begin with is

$$z = \underbrace{\frac{2w\sqrt{P_{\max}}}{1+|w|^2}}_{AM-AM}\underbrace{\exp\left\{j\phi_o\frac{2|w|^2}{1+|w|^2}\right\}}_{AM-PM}. \quad \mathbf{(35)}$$

This description assumes, for convenience only, that the maximum possible HPA output power $P_{\max}$ occurs with an input power of 1 Watt. Thus, $|z|^2 = P_{\max}$ when $|w|^2 = 1$, and the phase shift for that maximizing input power is ϕ_o. We will use $\phi_o = \pi/6$, which is a typical value.

(5) **Transmission Filter and Propagation Channel:** We assume that the Federal Communications Commission (FCC) mask for common carrier radio channels sets the limit for the transmitted power spectral density. Whenever the HPA output spectrum does not lie under this mask, an RF transmit filter is assumed to have a ς-th order Butterworth response:

$$|H_T(f)|^2 = \frac{H_0^2}{1+(2f/\mu)^{2\varsigma}}, \quad \mathbf{(36)}$$

where μ is the 3 dB RF bandwidth and H_o^2 is the mid-band filter gain. The line-of-sight radio path has a flat frequency response under normal conditions.

(6) **Down-conversion/Coherent Demodulation**: Our primary assumption about the down-conversion/coherent demodulation is that the carrier is ideal, i.e., is locked to the received carrier and has no phase jitter.

(7) **Fixed Receive Filter:** We assume a fixed receive filter for which, in the absence of non-linearity and selective fading, the end-to-end link response to a unit input datum will be a cosine-rolloff pulse. Thus, $H_R(f)$ is designed such that

$$G(f)H_T(f)H_R(f) = TC_{\cos ine}(f;\rho), \quad \mathbf{(37)}$$

where ρ is the roll-off factor.

(8) **Sampler and Detector:** A sampling in each symbol period is assumed to occur at the peak of the designed cosine-rolloff data pulse. We have not studied sensitivity to timing offsets, but the method of analysis is general enough to accommodate arbitrary timing epochs.

2. The Simulation Channel Model

The simulation channel model and the mathematical equations for each block are shown in Fig.

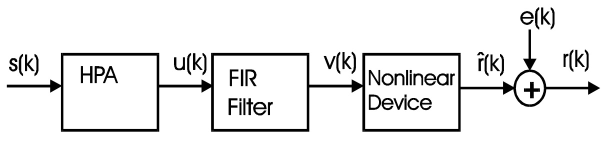


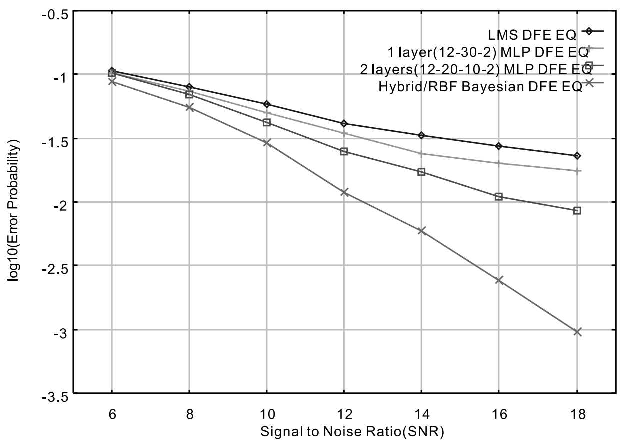
Fig. 6. Baseband equivalent discrete-time model of a nonlinear channel.

6. The input signals are 4-QAM and are transmitted to a high power amplifier working at saturation point:

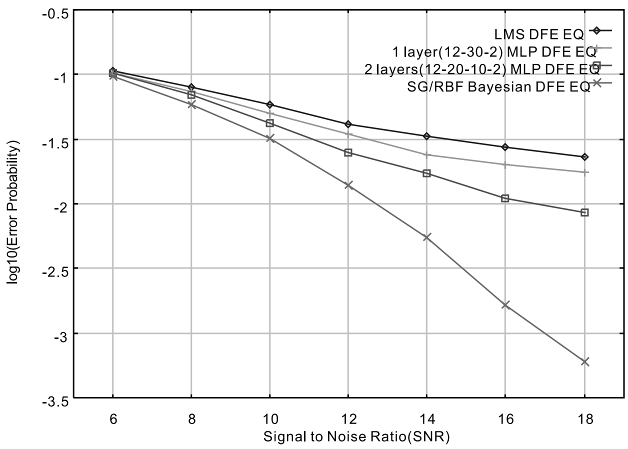
$$u(k) = f_s(s(k)) = \frac{2s(k)}{1+|s(k)|^2}\exp(j\frac{\pi}{3}\frac{|s(k)|^2}{1+|s(k)|^2}). \quad \textbf{(38)}$$

The output of Eq. (38) is then passed through a FIR filter, which represents the memory effect of the transmission filter, propagation channel and receiver filter, i.e.,

$$A(z) = \frac{V(z)}{U(z)} = \sum_{i=0}^{n_a-1} a_i z^{-i} = (0.3725+i0.2172) \cdot(1-(0.35+i0.7)z^{-1}) \cdot(1-(0.5+i)z^{-1}). \quad \textbf{(39)}$$



(a)



(b)

Fig. 7. (a) Comparison of CRBF/SG with other methods in terms of the error rate with varying SNR. **(b)** Comparison of the CRBF/HC with other methods in terms of the error rate with varying SNR.

Finally, a Volterra static nonlinear device is added in order to include the other nonlinear effect:

$$\hat{r}(k) = f_v(v(k)) = v(k) + 0.2v^2(k) - 0.1v^3(k), \quad \textbf{(40)}$$

and

$$e(k) = e_R(k) + je_I(k) \quad \textbf{(41)}$$

is additive complex Gaussian noise. In all cases, the number of symbols used in training depends on the value of SNR, which will correspond to the error rate.

Finally, to document the advantage of the CRBF/SG algorithm, a comparison of the performance, in terms of the symbol error-rate (SER) with different signal-to-noise ratios (SNR) and $d = 2$ (the optimum decision delay for the nonlinear channel), is given in Fig. 7. It is noted that, in our simulation, 1.1×10^4 total symbols were used, where ten percent of symbols were employed in the training mode for equalization. The training data symbols were, in fact, available at the receiver end. Also, in the channel estimator, the initial centers of the CRBF network were formed randomly by applying the 64 sets (with each set having 3 symbols) of symbols, $s(k)$, in the training mode, where each data symbol has four levels as defined in Eq. (3).

From Fig. 7, we learn that the conventional LMS DFE equalizer did not perform well. As by Cha and Kassam (1995), we found that the MLP network equalizers performed better than the one with the LMS DFE due to the fact that the MLP has the nonlinear processing ability. Moreover, the two-hidden-layer MLP outperformed the one having just one layer (MLP).

Next, the channel estimator approach using both the CRBF/SG and CRBF/HC algorithms had the best performance compared to the other approaches. Here, the performance obtained using the CRBF/SG algorithm, as shown in Fig. 7(b), was found to be better than that obtained using the CRBF/HC algorithm as shown in Fig. 7(a), especially, for SNR greater than 14 dB. Moreover, as described earlier, the CRBF/SG algorithm has certain advantages over existing methods. Typically, all free network parameters are adapted simultaneously. This provides greater robustness for poor initial choices of parameters. Also, it is computationally quite feasible although from Figs. 8 and 9, we learn that it may not be able to calculate the channel states concisely using the CRBF/SG and CRBF/HC algorithms as channel estimators. However, if we can devise some techniques which we can use to discard the overlapping states, then the performance can be further improved. This will be investigated in our future research. However, as indicated in Fig. 7, the channel estimator approach did have the best performance compared with the other conventional approaches.

VI. Conclusions

In this paper, we have investigated the performance of the adaptive Bayesian DFE using the non-linear channel estimator approach, along with the CRBF/SG algorithm, for radio channels having HPA. From the simulation results, we have observed that the performance obtained using the CRBF/SG algorithm was better than that obtained using the CRBF/HC algorithm and other conventional methods. This finding agrees quite well with the results obtained by Cha and Kassam (1995) for third-order non-linearity channels. This indicates that the CRBF/SG algorithm is well suited for implementing the Bayesian DFE equalizer, which is the optimum equalizer under the classical symbol-by-symbol decision-making architecture, in a more complicated non-linear channel like the one discussed in this paper. From Figs. 8 and 9, we see that to further improve the performance, the overlapping states should be discarded in the decision-making process. This will be investigated in our future research.

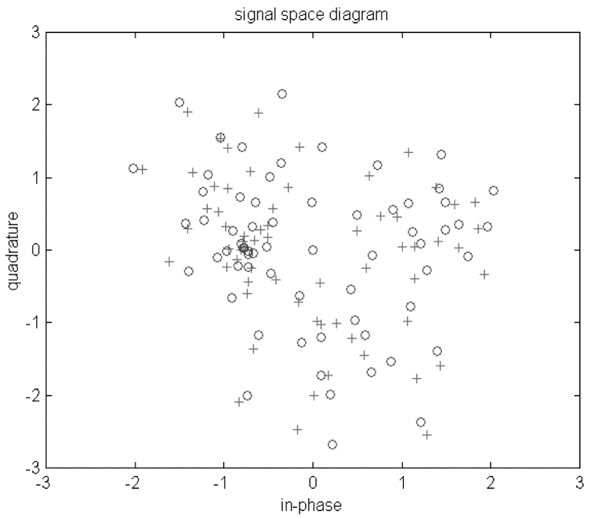


Fig. 8. State constellation. 'o' channel, '+' states identified using the hybrid algorithm with SNR=18.

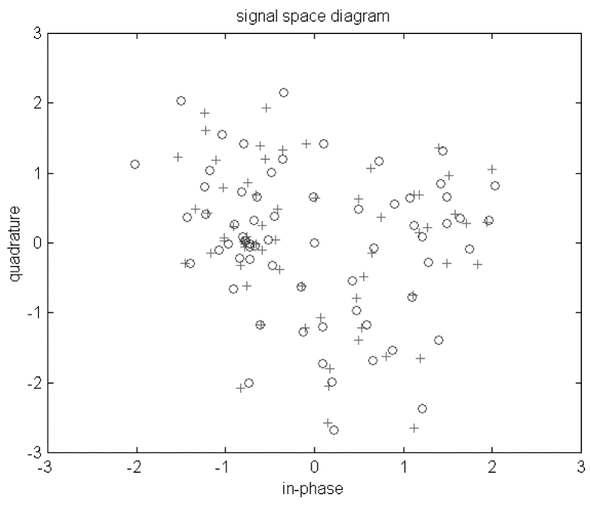


Fig. 9. State constellation. 'o' channel, '+' states identified using the SG algorithm with SNR=18.

References

Cha, I. and S. A. Kassam (1995) Channel equalization using adaptive complex radial basis function networks. *IEEE Journal on Selected Areas in Communications*, **13**(1), 122-131.

Chang, P. R. and B. C. Wang (1995) Adaptive decision feedback equalization for digital satellite channels using multilayer neural networks. *IEEE Journal on Selections Area in Communications*, **13**(2), 316-324.

Chen, S., G. J. Gibson, C. F. N. Cowan, and P. M. Grant (1991) Reconstruction of binary signals using an adaptive radial basis function equalizer. *EURASIP Signal Processing*, **22**(1), 77-93.

Chen, S., B. Mulgrew, and P. M. Grant (1993a) A clustering technique for digital communications channel equalization using radial basis function networks. *IEEE Trans. on Neural Networks*, **4**(4), 570-579.

Chen, S., B. Mulgrew, and S. McLaughlin (1993b) Adaptive Bayesian equalizer with decision feedback. *IEEE Trans. on Signal Processing*, **41**(9), 2918-2927.

Chen, S., S. McLaughlin, and B. Mulgrew (1994a) Complex valued radial basis function networks: network architecture and learning algorithms (part I). *EURASIP Signal Processing Journal*, **36**(1), 19-31.

Chen, S., S. McLaughlin, and B. Mulgrew (1994b) Complex valued radial basis function networks: application to digital communications channel equalization (part II). *EURASIP Signal Processing Journal*, **36**(2), 175-188.

Chen, S., B. Mulgrew, S. McLaughlin, and P. M. Grant (1995) Adaptive Bayesian decision feedback equalizer for dispersive mobile radio channels. *IEEE Trans. on Communications*, **43**(5), 1937-1946.

Gibson, G. J., S. Siu, and C. F. N. Cowan (1991) The application of nonlinear structures to the reconstruction of binary signals. *IEEE Trans. Signal Processing*, **39**(8), 1877-1884.

Lazzarin, G., S. Pupolin, and A. Sarti (1994) Non-linearity compensation in digital radio systems. *IEEE Trans. on Communications*, **42**(2/3/4), 988-999.

Lee, G., S. B. Gelfand, and M. P. Fitz (1995) Bayesian decision feedback techniques for deconvolution. *IEEE Journal on Selected Area in Communications*, **13**(1), 155-166.

Mulgrew, B. (March 1996) Applying radial basis functions. *IEEE Signal Processing Magazine*, 50-65.

Pupolin, S. and L. J. Greenstein (1987) Performance analysis of digital radio links with nonlinear transmit amplifiers. *IEEE Journal on Selected Area in Communications*, **5**(3), 535-546.

Widrow, B. and R. Winter (1988) Neural nets for adaptive filtering and adaptive pattern recognition. *IEEE Compu. Magazine*, **21**, 25-39.

適應性貝氏決定－回授式非線性高功率無線通訊等化器之研究

王光仁　陳巽璋

國立中山大學電機工程學系

摘　要

在通訊系統接收端以符碼(symbol)為主之檢測(decision making)的架構下，貝氏(Bayesian)決定－回授(decision feedback)通道等化器已被證實為最佳等化器。本文將以複數基底函數類神經網路(CRBF network)為架構來執行適應性貝氏決定－回授通道等化器，以探討其在非線性高功率無線通訊頻道之應用。為說明其優點我們將以符碼錯誤率為品質指標，並與其它傳統方法做比較以證實其品質確實優於傳統方法。