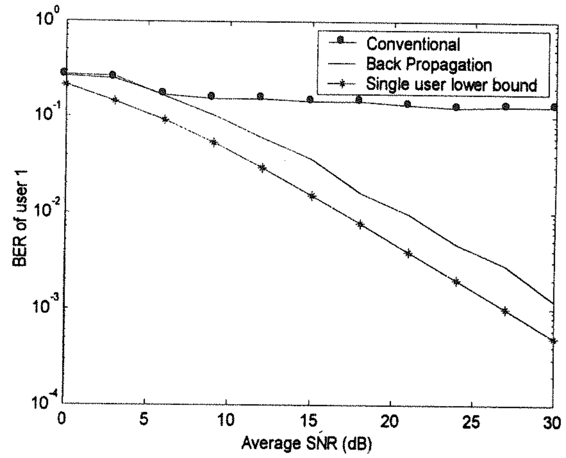

Figure (13) Performance of neural network in synchronous one path fading channel, $E_i / E_1 = 0\,dB$.

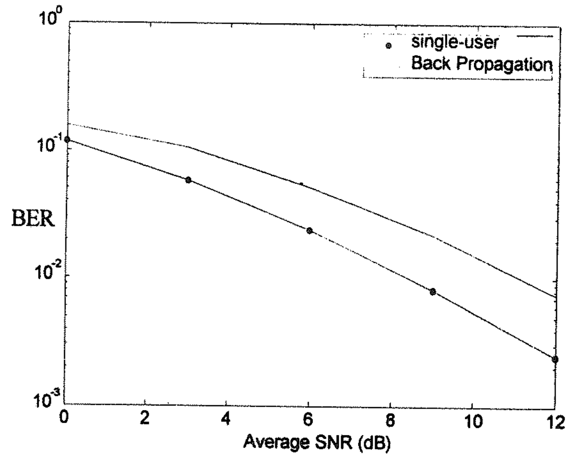

Figure (14) Performance of neural network for three-user system in 2-path fading channel, $E_i / E_1 = 0\,dB$.

References

[1] W. C. Y. Lee, "Overview of cellular CDMA," IEEE Trans. on Vehicular Technology, vt-40, no. 2, pp. 291-302, May 1991.

[2] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," IEEE Trans. on Information Theory, vol. IT-32, no. 1, pp. 85-96, Jan. 1986.

[3] B. Aazhang, B. P. Paris, and G. C. Orsak, "Neural networks for multiuser detection in code division multiple access communications," IEEE Trans. on Communication, vol. 40, no. 7, pp. 1212-1222, July 1992.

[4] U. Mitra, and H. V. Poor, "Neural network techniques for adaptive multiuser demodulation," IEEE Journal on Selected Areas in Communications, vol. 12, No. 9, pp. 1460-1470, Dec. 1994.

[5] G. I. Kechriotis, and E. S. Manolakos, "Hopfield neural network implementation of the optimal CDMA multiuser detector", IEEE Trans. on Neural Networks, vol. 7, no. 1, Jan. 1996.

[6] L. Fausett, "Fundamentals of Neural Networks", Prentice-Hall. 1994.

[7] R. Lupas, S. Verdu, "Linear multi-user detectors for synchronous code-division multiple access channels," IEEE Trans. Information Theory, vol. 35, no. 1, pp. 123-136, Jan. 1989.

[8] M. K. Varansi, B. Aazhang, "Near-optimum detection in synchronous code-division multiple-access systems" IEEE Trans. Communication, vol. 39, no. 5, pp. 725-735, May 1991.

[9] S. Y. Kung, and J. S. Taur, "Decision based neural networks with signal/image classification applications," IEEE Trans. on Neural Networks, vol. 6, no. 1, pp. 170-181, Feb. 1995.

[10] S. Y. Kung, Neural networks, Prentice-Hall, 1993.

[11] B. H. Juang, and S. Katagiri, "Discriminative learning for minimum error classification," IEEE Trans. on Signal Processing, vol. 40, no. 12, pp. 3043-3054, Dec. 1992.

[12] S. Katagiri, B. H. Juang, and C. H. Lee, "Pattern recognition using a family of design algorithms based upon the generalized probabilistic descent method," Proceeding of the IEEE, vol. 86, no. 11, pp. 2345-2373, Nov. 1998.

[13] V. Nedeljkovic, "A novel multilayer neural networks training algorithm that minimizes the probability of classification error", IEEE Trans. on Neural Networks, vol.4, no. 4, July 1993.

[14] P. Zhou, and J. Austin, "Learning criteria for training neural network classifiers", Neural Computing & Application, pp.334-342, 1998.

[15] J. G. Proakis, "Digital Communications," NewYork: McGraw-Hill, 1995.

[16] C. Valadon, and R. Tafazolli, "A fast adaptive multiuser detector for DS-CDMA communications based on an artificial neural network architecture," in Proceeding IEEE ISSSTA, 1998, pp. 873-877.

[17] P. Y. Peng, "An efficient multiuser detection algorithm for direct sequence code division multiple access by using neural networks," in Proceeding IEEE ICC. 2001, pp. 36-40.

[18] M. G. Shayesteh, and H. Amindavar, "DBNN, FDNN, discriminative learning, and back propagation neural networks in DS/CDMA systems," in Proceeding. IEEE IJCNN 2000.

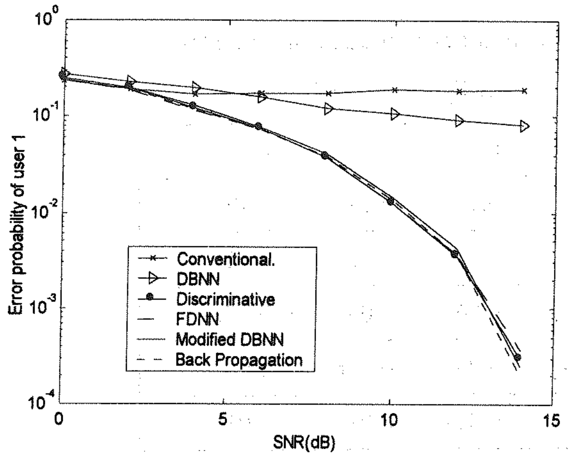

Figure (7) Performance of different nets.

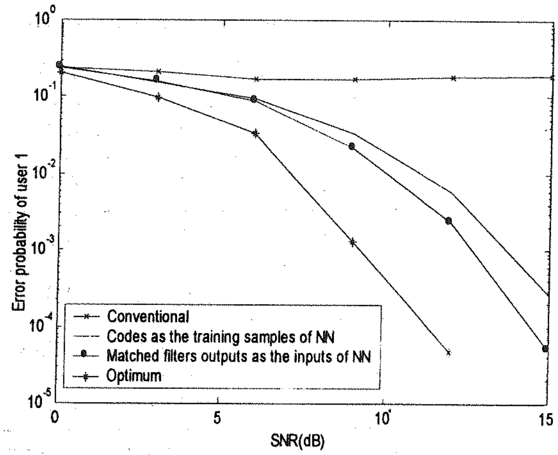

Figure (10) The effect of kind of training samples in neural network performance, $E_i / E_1 = 0dB$.

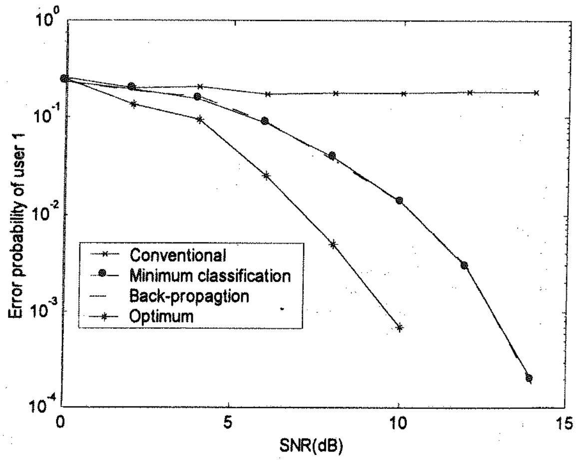

Figure (8) Performance of different criterions in neural network, $E_i / E_1 = 3dB$.

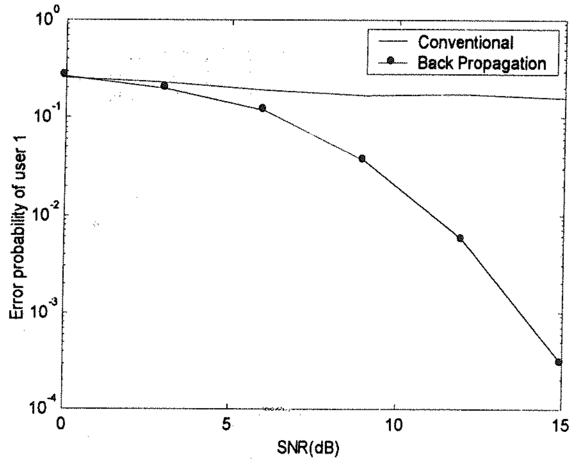

Figure (11) Performance of neural network in asynchronous channel, $E_i / E_1 = 3dB$.

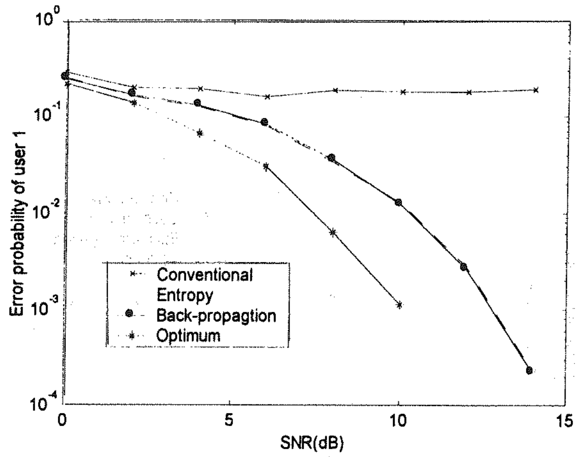

Figure (9) Performance of different criterions in neural network, $E_i / E_1 = 3dB$.

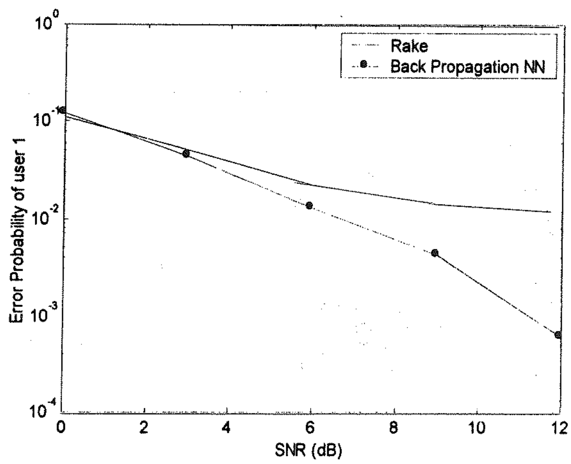

Figure (12) Performance of neural network in two path channel.

Table (1) Error rate of user 1 versus the number of hidden layer nodes.

SNR (dB)	H=4	H=6	H=8	H=10	H=12	H=15
0	0.21	0.22	0.23	0.21	0.27	0.24
2	0.19	0.19	0.19	0.18	0.17	0.17
4	0.12	0.11	0.11	0.115	0.12	0.12
6	0.071	0.07	0.068	0.066	0.065	0.075
8	0.038	0.034	0.034	0.03	0.038	0.03
10	0.012	0.011	0.0118	0.011	0.01	0.014
12	0.0023	0.0022	0.0021	0.0024	0.0023	0.002
14	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002

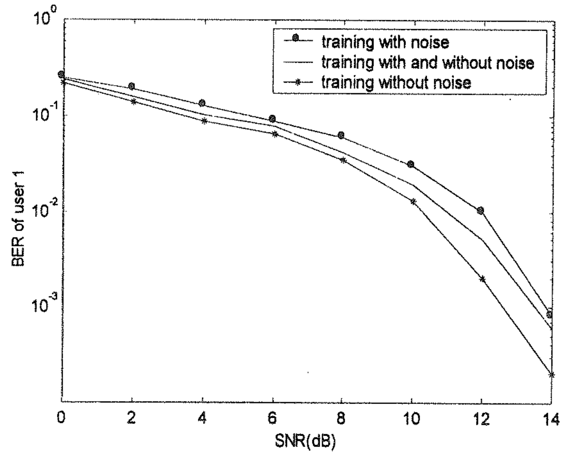

Figure (3) The effect of training of the network with and without noise, $SNR = E_1 / \sigma^2$.

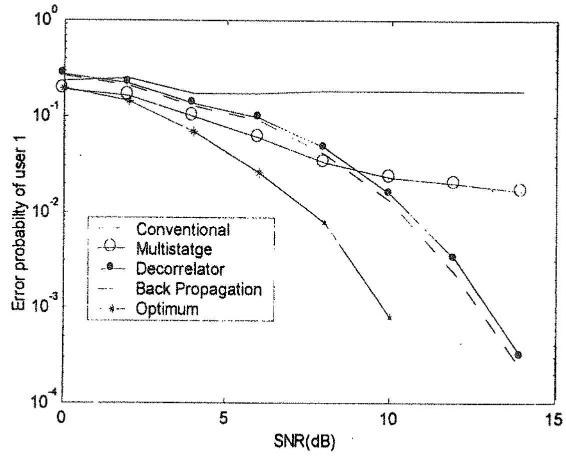

Figure (5) Performance of different detectors, $E_i / E_1 = 3\,dB$.

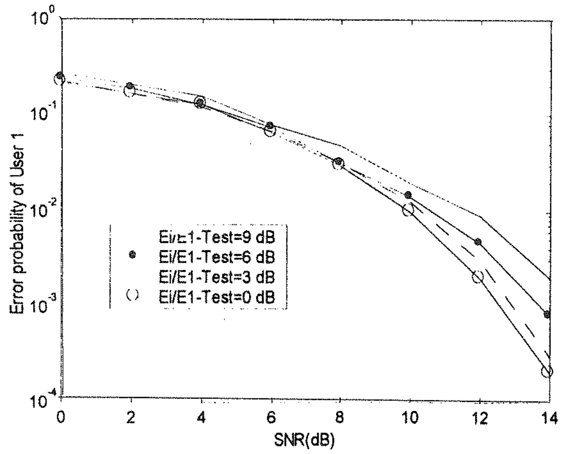

Figure (4) Performance of NN for different near-far ratios of test samples. The net was trained with $E_i / E_1 = 3dB$.

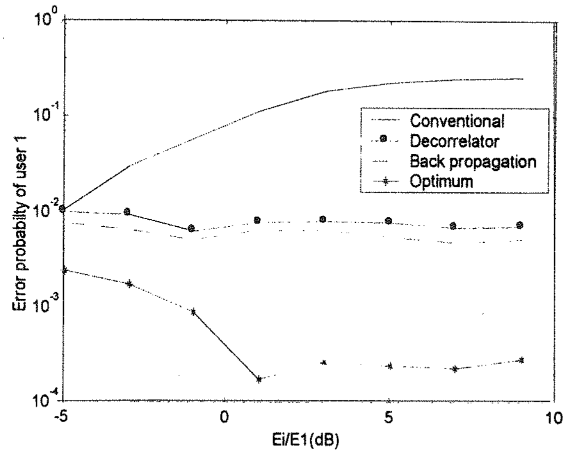

Figure (6)The near-far effect, *SNR*=11 *dB*.

necessary to train the network with 2^K possible samples in synchronous AWGN and one path fading channels. Our results lead to a superior improvement over the previous research in terms of the required training symbols, training time and the number of nodes, consequently, reducing the receiver complexity. We also observed that the back propagation neural network outperforms the conventional, the decorrelator, and the multistage detectors and it is resistant to the near-far effect.

We compared the performance of DBNN, FDNN, and discriminative learning neural network with the back propagation net. The results showed that DBNN has the worst performance among the mentioned neural networks. We also proposed modified DBNN that outperforms DBNN and it has performance equal to the back propagation network. Considering the theory of fuzzy decision and discriminative learning methods, we observed that they have the same ideas. The simulation results show that they have performance comparable to the back propagation net but they require more training epochs.

We observed that the neural networks with minimum classification and cross entropy criterions achieve the same performance as the back propagation method, although they have shown a better performance in the other applications.

We also examined the performance of the back propagation neural network in multipath fading channels. It was shown that it outperforms the Rake receiver and can track the channel variations.

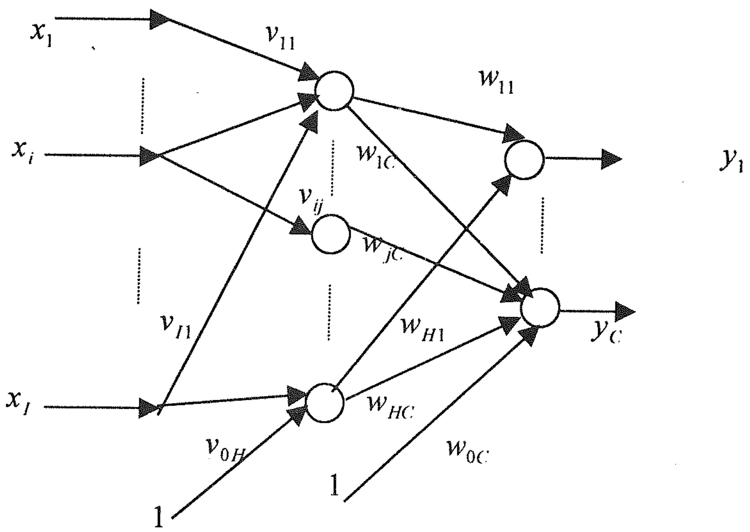

Figure (1) The structure of a typical three layer perceptron neural network.

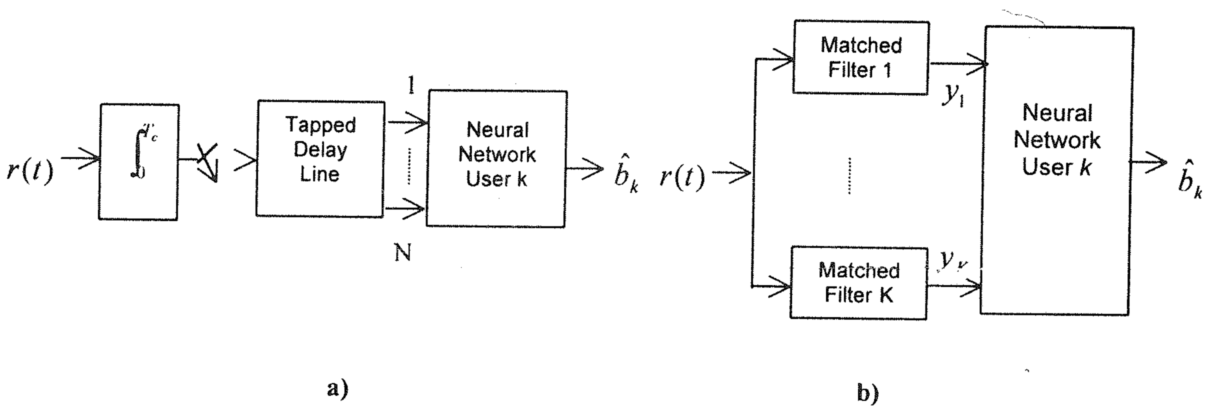

Figure (2) Two kinds of processing of the received signal.

13-In another experiment, the network was trained by the spreading codes of all users. When the desired user's code or its negative are applied into the network, the target (t) value will be 1 and −1, respectively and when the other users' codes or their negatives are fed into the net, the target will be 0 where the BP learning method is used for training. Figure 10 shows that applying these kinds of training patterns gives worse performance than the two configurations explained before, i.e., the outputs of chip matched and code matched filters as training samples. But this has the advantage of requiring $2K$ training samples that are much less than 2^K especially for large K.

C-Figure 11 shows the performance of BP net in asynchronous AWGN channel. The net was trained with specific delays of users. It is seen that NN outperforms the conventional detector.

D-We have done our experiments in multipath fading channel as well. Three kinds of configurations are fed into the neural network. The outputs of chip matched filters, the outputs of KL code matched filters, and the outputs of K Rake receivers. The results are as follows:

14-Figure 12 shows the performance of BP net in a two-path system, where no fading is assumed. The delays of users are kept fixed during the transmission, and each path has unique amplitude. It is seen that BP neural net outperforms the Rake receiver. Also the outputs of chip matched filters and KL code matched filters have the same performance when applied to the net.

15-The effect of fading is considered in Fig 13. One synchronous path is assumed and the fading coefficients remain constant during several symbol intervals (packet length) and they are assumed to be independent in each packet. A lower bound for the error probability can be found as the BER of a single user which is obtained analytically as $BER_{SU} = \frac{1}{2}[1 - \sqrt{\frac{\bar{\gamma}}{\bar{\gamma}+1}}]$ [15] where $\bar{\gamma}$ is the average SNR defined as $\bar{\gamma} = \frac{1}{2}\frac{E_1\eta^2}{\sigma^2}$, and $2\eta^2$ is the power of channel coefficients defined before. It is observed that the neural network outperforms the conventional Rake receiver and can track the variations of the channel adaptively. Its performance is close to that of single user case. We also found that the net trained for synchronous AWGN channel can also be used for one path fading channel in the test mode. The reason is that the channel coefficients can be interpreted as the users' energies that change with the time, so we refer back to part 2 previously discussed. Hence the results of part A also hold true for this case.

Figure 14 shows the performance of neural network for a 3-user system in a two-path fading channel. The single user bound is computed as $BER_{SU} = \left(\frac{1-\lambda}{2}\right)^2(2+\lambda)$ [15] where $\lambda = \sqrt{\frac{\bar{\gamma}}{\bar{\gamma}+1}}$. The net is trained with the different channel coefficients. NN has better performance than the Rake receiver. We also found that the network trained for the two-path and no fading channel can be used for the fading case in the test mode, again refer back to part 2. The simulation results show that if the outputs of the Rake receivers are applied to the network, there is no performance improvement over the Rake receiver.

E-The results show that neural network can be used as a detector in CDMA systems by proper parameter selection. The computational complexity of NN is in the training mode that can be organized in parallel. Of course the hardware implementation of neural network especially for large number of users should be considered.

4-Conclusions

In this paper we analyzed the performance of three layer perceptron neural network using back propagation training algorithm as multiuser detector of DS/CDMA signals in AWGN and fading channels. We observed that it suffices to train the net without noise and with a high ratio of energy of the interfering users to the desired user's energy. As a result it is only

500 samples for 2-user case, here we only need 4 training samples for the same performance, thus 500/4 savings in the number of samples. In the last published research [16], [17] 800 training samples are used for 4-user system and further in [16] the algorithm named RLS-BP is used to reduce the number of samples to 150 where the computational complexity increases. While for the same problem we only require $2^4=16$ training samples without any increase in computational complexity. We also trained the network for 10 epochs. There is no report of the number of epochs in the above papers. Considering the number of training samples, it seems that more than 10 epochs are required to train the nets to achieve the same performance. Therefore the results obtained in our experiments lead to a superior performance improvement in terms of the receiver complexity.

4-Figure 5 compares the performance of different detectors. We observe that the BP neural network outperforms the conventional, the decorrelator, and the multistage detectorsThis is justified by noting the fact that neural network is a nonlinear detector, and with a proper parameter tuning and training algorithm it can approach the optimum detector performance.

5-Figure 6 shows the BER of the network versus the relative energies of the users. It is seen that the neural network is resistant to the near-far effect and it has better performance than the conventional and the decorrelator detectors.

6-Next we examine the effect of the number of nodes in the hidden layer on the performance. The number of hidden layer nodes is changed from 4 to $2^4=16$. Table 1 shows the error rate for different number of nodes. It is seen that increasing the number of nodes does not significantly improve the error rate, while in [3] the number of neurons should grow exponentially with the number of users to achieve the same performance. However, this still is an open problem.

7-We applied the outputs of the decorrelator detector into the network, the results show that the same performance as the two mentioned configurations is obtained.

B-We applied the different criterions explained in Section 2 and compared their performance:

8-It is seen from Fig. 7 that the proposed modified DBNN outperforms the classic DBNN and it has a similar performance as the back propagation neural net. Although DBNN achieves better result than the BP net according to [9], [10], but here it has the worst performance among the nets. This is because DBNN is suitable for clearly separable classes, while in CDMA there is a large amount of overlap between the desired user's binary bits due to noise and MAI. In the modified DBNN we apply weight updating even for the correct classification to separate the classes more, thus we obtain better results.

9-We compared the performance of FDNN, discriminative learning, and back propagation networks. We observed that for the epoch numbers (number of training iterations) less than 100, BP has better performance than FDNN and discriminative learning neural net. We performed such a simulation for system with 6 and 7 users in [18] and obtained similar results. For the epoch number 150, the three nets have the same performance (Fig. 7). This means that FDNN and discriminative learning net need more epochs for training than the BP and modified DBNN to achieve the same performance. However, all nets work better than the conventional detector.

10-The performance of neural net using minimum classification criterion is compared with the BP training method. We observe from Fig. 8 that the two methods have the same performance.

11-The simulation results show that the neural network with the cross entropy criterion has the same performance as the back propagation method (Fig. 9).

12-The above results show that although the mentioned networks perform better than BP in the other applications, but in CDMA system they have the same performance as the BP method, thus they are application dependent. This can be justified by saying that BP net gives the minimum classification error and it can also be interpreted as the cross entropy methodology in CDMA.

samples are obtained for each bit of data. These samples are fed into the tapped delay line that converts the received serial signal into a parallel form. The outputs of the tapped delay line are then fed into the input layer of neural net (Fig. 2.a). The sequence of the received sampled signal has the following model:

$$r[j] = S^T W b[j] + n(j) \tag{63}$$

where S is a matrix consisting of the spreading codes of users and $n(j)$ is a Gaussian noise vector with the covariance matrix of $\sigma^2 I$. This configuration can be used in the mobile station, where the net should have access to the spreading codes of all users in the training phase while in the test mode only the desired user's code is required. The network can be trained in the base station.

b-$r(t)$ is passed through K filters matched to the spreading codes. Then the outputs of the filters are fed into the net (Fig. 2.b), hence the number of samples for each bit is K, the number of users. This configuration is suitable for the base station because of having an access to all users' signature codes required in both the training and test modes.

It is obvious that the two configurations will achieve the same results when applied to the network. This is because all the information contained at the outputs of code matched filters (Y) is also contained at the outputs of chip matched filters (r), that is, Y is a linear combination of r:

$$Y = S^T r. \tag{64}$$

We concentrate more on 4-user system with Gold codes of length 7. These specifications simulate examples where the cross correlation is large, due to the high bandwidth efficiency, defined as K/N. The length 7 codes are chosen as

$$S = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & -1 \end{bmatrix} \Rightarrow$$

$$R = \begin{bmatrix} 1 & -0.14 & 0.43 & 0.43 \\ 0.14 & 1 & -0.14 & 0.43 \\ 0.43 & -0.14 & 1 & -0.14 \\ 0.43 & 0.43 & -0.14 & 1 \end{bmatrix} \tag{65}$$

where R is the normalized correlation matrix. We observe that there is a high correlation among the users. We compute the BER of the first user that has higher correlation values, i.e., the worst case is considered. In simulations we denote SNR (Signal to Noise Ratio) as $SNR = 10\log(E_1/\sigma^2)$.

Our experiments consist of several parts as follows:

A-At first we analyze the performance of BP network in synchronous AWGN channel. For the cases of under consideration we obtain the following results:

1-The network was trained under the following cases: with noise, without noise, and with and without noise. The simulation results (Fig. 3) show that when the net is trained without noise, we achieve better performance in the test phase. Obviously, when the net is trained without noise, it learns the characteristics of the system more clearly. Hence, for the rest of this work we train the networks without noise.

2-Our observations show that when the neural network is trained with a specific ratio of users' energies, that is, $E_i / E_1 = C$, we obtain better results in the test mode for all energy ratios less than C, but the performance decreases for $E_i / E_1 > C$. This is illustrated in Fig. 4. When the network is trained by a high value of interference, we have interference rejection capability in the test mode. Therefore it is better to train the net with a high certain energy ratio to obtain better performance for the range of energy ratios less than that.

3-Thus from the above two species we may conclude that the network can be trained without noise and with a specific energy values of the users. In synchronous transmission because K users have 2^K different combinations, it suffices to train the network with this number of samples. In [3] the net was trained with minimizing (52), the term $\frac{\partial u(y(x,W))}{\partial W}$ appears which contains Dirac delta function. Thus, a modified criterion is introduced by replacing $u(y(x,W))$ by a continuous function $g(y(x,W))$:

$$J_c(W) = P_1 - E\{t(x)\,g(y(x,W)) \mid W\}. \qquad (53)$$

Therefore the weight updating procedure becomes

$$W(new) = W(old) + \mu t(x)\frac{\partial g(y(x,W))}{\partial W}. \qquad (54)$$

where $g(y)$ is binary sigmoid as $g(y) = \frac{1}{1+e^{-y}} \Rightarrow g'(y) = g(y)(1-g(y))$, and bipolar sigmoid is used as activation function of hidden and output layers. The weight updates are computed as

$$\Delta w_{j1} = \mu\delta z_j \qquad \Delta w_{01} = \mu\delta \qquad \delta = t\,g'(y)f'(yin) \qquad (55)$$

$$\Delta v_{ij} = \mu\delta_j x_i \qquad \Delta v_{0j} = \mu\delta_j$$

$$\delta_j = \delta in_j f'(zin_j) \qquad \delta in_j = w_{j1}\delta. \qquad (56)$$

2-7-Cross Entropy Criterion

Here, the objective function is a cross entropy [14] based on a probabilistic model. In this approach it is of interest to model the probabilities of class memberships conditioned on the input data, that is, $p(i \mid x)\ i = 1,\ldots,C$, and to determine unknown model parameters via the maximum likelihood estimation. Specifically, let the kth output, y_k, of a net represents the probability of observing the target values of a kth class pattern, that is, $p(i = k \mid x \in c_k) = y_k$. The likelihood of observing M independent training data is given by

$$\prod_{m=1}^{M} p(i = k \mid x(m) \in c_k) = \prod_{m=1}^{M} y_k(m). \qquad (57)$$

To use the above for supervised training, we can rewrite it to make the target variable $t = [t_1,\ldots,t_C]$ occur explicitly in the likelihood function

$$\prod_{m=1}^{M}\prod_{i=1}^{C} [y_i(m)]^{t_i(m)} \qquad (58)$$

where $t_i = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$. It is more conveniently to minimize the negative logarithm of (58)

$$L = \sum_{m=1}^{M}\sum_{i=1}^{C} t_i(m)\ln y_i(m). \qquad (59)$$

The above is cross entropy function. Since the outputs of the network represent the probabilities, that is, $y_i \in [0,1]$ and $\sum y_i = 1$, the softmax activation function is appropriate for use in the output units:

$$y_k = \frac{\exp(\sum_j w_{jk} z_j)}{\sum_k \exp(\sum_j w_{jk} z_j)}. \qquad (60)$$

We use bipolar sigmoid as the activation function of the hidden layer nodes. The minimization of (59) leads to the following weight updating rules:

$$\Delta w_{jk} = \mu\delta_k z_j \qquad \Delta w_{0k} = \mu\delta_k \qquad \delta_k = t_k - y_k \qquad (61)$$

$$\Delta v_{ij} = \mu\delta_j x_i \qquad \Delta v_{0j} = \mu\delta_j$$

$$\delta_j = \delta in_j f'(zin_j) \qquad \delta in_j = \sum_{k=1}^{C}\delta_k w_{jk} \qquad (62)$$

In CDMA application, two nodes are used in the output layer to represent bit +1 or -1 of the desired user. This criterion has better performance than BP in sonar, radar, and satellite images.

3-Simulations and Results

In this section we analyze the performance of the aforementioned neural nets as multiuser detectors in CDMA system. Monte Carlo simulation is carried out to compute the BER (Bit Error Rate). We use two kinds of processing for the received signal in order to apply into the neural net:

a-The received signal $r(t)$ is passed through the chip matched filter and sampled at the end of every chip interval with the chip rate T_c^{-1}, therefore N (code length)

Reinforced Learning:

$$W_l\,(new) = W_l\,(old) + \mu\varphi'(d_l)\frac{\partial y(x,W_l)}{\partial W_l} \qquad (43)$$

Antireinforced Learning:

$$W_h\,(new) = W_h\,(old) - \mu\varphi'(d_l)\frac{\partial y(x,W_h)}{\partial W_h} \qquad (44)$$

We choose linear activation function in the output layer ($y_k = yin_k$), and bipolar sigmoid function for the hidden layer. We derived the weights update as

$$\Delta w_{jk} = \pm\mu\delta z_j \qquad \Delta w_{0k} = \pm\mu\delta \qquad \delta = \varphi'(d_l) \qquad (45)$$

$$\Delta v_{ij} = \pm\mu\delta_j x_i \quad \Delta v_{0j} = \pm\mu\delta_j$$
$$\delta_j = \delta in_j f'(zin_j) \quad \delta in_j = \sum_{k=1}^{C} \delta\, w_{jk}\,. \qquad (46)$$

We expect to obtain better results in FDNN rather than DBNN in CDMA application because each bit +1 or −1 of the desired user overlaps with the bits of the other users as well as noise.

2-5-Discriminative Learning

The minimum mean squared error criterion does not generally lead to minimum error probability. Thus another cost function is used in lieu of MMSE criterion. Discriminative learning [11], [12] proposes a fundamental technique for designing a classifier that achieves minimum classification error. In this method all classes are lumped into one network with several outputs each one corresponding to one class.

At first a misclassification measure, d_l, similar to (41) is defined for the corresponding outputs of the network. Then a cost function, $\varphi(d_l)$, like as in FDNN, eq. (42), is applied. The net is trained to minimize $\varphi(d_l)$:

$$W(new) = W(old) - \mu\frac{\partial\varphi(d_l)}{\partial W}\,. \qquad (47)$$

It seems that there is a close relationship between the FDNN and this method. The ideas are the same but the difference is that FDNN uses OCON (One-Class-in-One-Network), while discriminative learning method employs ACON (All-Class-in-One-Network). FDNN emphasizes the merit of soft decision based on the sigmoid loss function. We will compare their performance in Section 3. We use linear activation function in the output layer, i.e., $y_k = yin_k$, and bipolar activation function in the hidden layer. We derived the weight changes as follows:

$$\Delta w_{jk} = \pm\mu\delta z_j \qquad \Delta w_{0k} = \pm\mu\delta \qquad k=l,\ h$$
$$\delta = \varphi'(d_l) \qquad (48)$$

$$\Delta v_{ij} = \mu\delta_j x_i \quad \Delta v_{0j} = \mu\delta_j$$
$$\delta_j = \delta in_j f'(zin_j) \quad \delta in_j = \delta(w_{jl} - w_{jh}) \qquad (49)$$

where the signs + and - correspond to the l and h outputs, respectively. This method works better than the BP for speech signals. In CDMA application we use two nodes in the output layer.

2-6-Minimum Classification Criterion (Bayesian back propagation)

This algorithm [13] minimizes the probability of misclassification. For the case of two classes (bits +1 and −1 constitute the two classes in CDMA), the probability of error can be written as

$$J_B(W) = \int_{\Omega_2(w)} p(c_1 \mid x)p_X(x)dx + \int_{\Omega_1(w)} p(c_2 \mid x)p_X(x)dx \qquad (50)$$

where $p(c_i \mid x)$ is the conditional density function of class i given x, and $\Omega_1(W)$ and $\Omega_2(W)$ are the distinct regions of the two classes as

$$\Omega_1(W) = \{x \in R^d : y(x,W) \geq 0\}$$
$$\Omega_2(W) = \{x \in R^d : y(x,W) \leq 0\} \qquad (51)$$

where $y(x,W)$ is the output of the net (one node in output layer) . Equation (50) can be written as

$$J_B(W) = P_1 - E\{t(x)u(y(x,W)) \mid W\} \qquad (52)$$

where P_1 is the priority probability of class 1 (1/2 for bit +1 or −1), $t(x)$ is the target (± 1), and $u(x)$ is the step function. In

$$\Delta v_{ij} = \mu \delta_j x_i \qquad \Delta v_{0j} = \mu \delta_j$$

$$\delta_j = \delta in_j f'(zin_j) \qquad \delta in_j = \sum_{k=1}^{C} \delta_k w_{jk} \,. \quad (33)$$

Here, for CDMA we use one node in the output layer, and bipolar sigmoid as activation function:

$$f(u) = \frac{1-e^{-u}}{1+e^{-u}} \Rightarrow f'(u) = (1-f(u))(1+f(u)) \,. \quad (34)$$

Depending on the sign of the network output, the received signal will be classified to ± 1.

2-2-DBNN (Decision Based Neural Network)

In DBNN [9], [10] one net is assigned to one class only. DBNN is suitable for clearly separable patterns. It has better performance than BP method in pattern recognition applications.

Suppose that $S = \{x^{(1)},...,x^{(M)}\}$ is a set of training patterns, each corresponding to one of the C classes. Each class is modeled by a net, and $y(x,W_j)$ shows the net output (discriminative function) where W_j denotes the weight vector for the jth class. Suppose that the given training pattern belongs to the class l, and the winning class for the pattern is the hth class, i.e.,

$$y(x,W_h) > y(x,W_j) \quad \forall h \neq j \quad (35)$$

When $h = l$, then the pattern x is already correctly classified and no update is needed. When $h \neq l$, that is, x is misclassified, then the following weight updating is performed:

Reinforced learning:

$$W_l\,(new) = W_l\,(old) + \mu \frac{\partial y(x,W_l)}{\partial W_l} \quad (36)$$

Antireinforced learning:

$$W_h\,(new) = W_h\,(old) - \mu \frac{\partial y(x,W_h)}{\partial W_h} \quad (37)$$

and the weights of the other classes remain unchanged. We derived weight changes as

$$\Delta w_{jk} = \pm \mu \delta_k z_j \qquad \Delta w_{0k} = \pm \mu \delta_k \qquad \delta_k = f'(yin_k) \quad (38)$$

$$\Delta v_{ij} = \pm \mu \delta_j x_i \qquad \Delta v_{0j} = \pm \mu \delta_j$$

$$\delta_j = \delta in_j f'(zin_j) \qquad \delta in_j = \sum_{k=1}^{C} \delta_k w_{jk} \quad (39)$$

where the signs + and – correspond to the correct and incorrect classes, respectively. In CDMA application, two classes corresponding to bits +1 or −1 of the desired user are used.

2-3-Proposed Modified DBNN

In our proposed method, the reinforced learning is applied even in the case of correct classification, i.e., the weights of the correct class are changed according to eq. (36) and the antireinforced learning is applied for the incorrect class. In this case the weights of the correct class are moved in the positive direction of gradient of discriminative function, $y(x,W_l)$, and the weights of the incorrect class are moved toward the negative gradient of discriminative function of that class, $y(x,W_h)$. This makes the classes become farther and more separable even in the case of correct classification. Simulation results show a superior performance improvement of the proposed method over the classic DBNN.

2-4-FDNN (Fuzzy Decision Neural Network)

When the patterns are not clearly separable, FDNN [9], [10] is more suitable. Suppose that the class h denotes the leading challenger among all the classes excluding the correct class l, i.e.,

$$h = \arg\max_{h \neq l} y(x,W_h) \,. \quad (40)$$

For a training pattern, a measure of misclassification can be introduced as:

$$d_l = -y(x,W_l) + y(x,W_h) \,. \quad (41)$$

FDNN imposes a penalty function which is a function of misclassification. For example:

$$\varphi(d_l) = \frac{1}{1+e^{-d_l/\xi}} \quad (42)$$

where ξ is a constant. The weight updating for the correct or incorrect classes is:

$$R = \begin{bmatrix} R(0) & \cdots & R(2P) \\ \vdots & \ddots & \vdots \\ R(-2P) & \cdots & R(0) \end{bmatrix}$$

$$R(i) = \begin{bmatrix} R_{1,1}(i) & \cdots & R_{1,K}(i) \\ \vdots & \ddots & \vdots \\ R_{K,1}(i) & \cdots & R_{K,K}(i) \end{bmatrix} \quad (25)$$

where

$$R_{k,k'}(i) = \left[R_{k,k',l,l'}(i)\right]_{L\times L}$$
$$R_{k,k',l,l'}(i) = \int_{-\infty}^{+\infty} s_k(t - t_{k,l}) s_{k'}(t + iT - t_{k',l'}) dt \quad (26)$$

where $R_{k,k',l,l'}(i)$ denotes the correlation between the kth user's lth multipath component and k'th user's l'th multipath component at iT, and n is Gaussian noise vector with the similar notation as Y:

$$n_{k,l}(i) = \int_{iT+t_{k,l}}^{(i+1)T+t_{k,l}} n(t) s_k(t - iT - t_{k,l}) dt \quad (27)$$

The conventional single user detector is the Rake receiver that combines the outputs of matched filters using the Maximum Ratio Combining (MRC) method.

$$Yr_k = \mathrm{Re}\left\{\sum_{l=1}^{L} \alpha^*_{k,l} Y_{k,l}\right\} \quad (28)$$

In this paper we apply neural networks as suboptimum multiuser detectors to classify the received signal to 1 or –1 for the desired user.

2-Description of the Neural Networks Used in Experiments

In this section, we first describe the back propagation (BP) method. Since our goal is to improve the performance of BP neural network, subsequently we explain the networks and criterions that have better performance than the BP in radar, sonar, speech, and pattern recognition applications. We apply them to CDMA system and compare the results with BP.

Figure 1 shows the structure of a three layer perceptron neural network with one hidden layer. The parameters of network are defined as:

- The numbers of nodes in the input, hidden, and output layers are I, H, and C, respectively.
- x_i : the ith unit of input layer
- v_{ij} : weight between the ith input unit and the jth unit of hidden layer
- v_{0j} : bias weight
- w_{jk} : weight between the jth unit of hidden layer and the kth output
- w_{0k} : bias weights
- zin_j : the jth input unit of hidden layer
- z_j : the jth output of hidden layer
- yin_k : the kth input of output layer
- y_k : the kth unit of output layer

$$zin_j = \sum_{i=1}^{I} x_i v_{ij} + v_{0j} \qquad z_j = f(zin_j)$$
$$yin_k = \sum_{j=1}^{H} z_j w_{jk} + w_{0k} \qquad y_k = f(yin_k) \quad (29)$$

- $f(.)$: activation function

2-1-Minimum Mean Square Error (Back Propagation) Criterion

In this common criterion, the optimization of network parameters is obtained by minimization of the sum of square differences between the targets (desired outputs) and the outputs of net [6]:

$$E = \frac{1}{2M} \sum_{m=1}^{M} \sum_{i=1}^{C} [t_i(m) - y_i(m)]^2 \quad (30)$$

where M is the number of training patterns, C is the number of outputs, $t_i(m)$ is the ith component of the mth target (± 1 in CDMA), and $y_i(m)$ is the ith output of the network for the mth input pattern. The weight updating is obtained according to the following rule:

$$W(new) = W(old) - \mu \frac{\partial E}{\partial W} \quad (31)$$

where W is the weights of the net (containing v and w) and μ is the learning rate. The weight change rules are as follows;

$$\Delta w_{jk} = \mu \delta_k z_j \qquad \Delta w_{0k} = \mu \delta_k$$
$$\delta_k = (t_k - y_k) f'(yin_k) \quad (32)$$

(asynchronous) or 2^K (synchronous) possible vectors in order to find the optimum b. Its computational complexity grows exponentially with the number of users and makes it impractical for a realistic environment. As a result, suboptimum multiuser detectors with less complexity and near optimum performance have received considerable attention.

The decorrelator [7] and the multistage [8] are two detectors widely used in CDMA systems for comparative analysis. The decorrelator detector applies R^{-1} to decouple data:

$$Z = R^{-1}Y = Wb + n' \tag{16}$$

where n' is Gaussian noise with covariance matrix $\sigma^2 R^{-1}$. This detector does not need to estimate the received energy of signal and it is resistant to the near-far effect. A disadvantage of the decorrelator detector is noise enhancement.

The multistage detector utilizes previously made decisions of other users to cancel interference present in the signal of the desired user. The (m+1) stage uses decisions of the mth stage as

$$\hat{b}(m+1) = sign\left[Y - (RW - W)\hat{b}(m)\right]. \tag{17}$$

The performance of this detector depends on the initial data estimate that can be the outputs of the conventional or decorrelator detectors. It is suitable for those users whose energies are small relative to the energies of the interfering users. It requires the knowledge of user's energies.

1-2-Fading Channel

It is assumed that the channel of each user is frequency selective Rayleigh fading [15] with L independent paths. We also assume that fading is slow so that the channel parameters remain constant during several symbol intervals. The baseband impulse response of channel for user k is

$$h_k(t) = \sum_{l=1}^{L} c_{k,l} e^{-j\phi_{k,l}} \delta(t - t_{k,l}) \tag{18}$$

where $c_{k,l}$ is the amplitude of the lth path of the kth user channel which has Rayleigh distribution as $p(c_{k,l}) = \frac{c_{k,l}}{\eta^2}\exp(-c_{k,l}^2/2\eta^2)$ with power $E(c_{k,l}^2) = 2\eta^2$, $\phi_{k,l}$ and $t_{k,l}$ are the analogous phase and delay that have uniform distributions in the intervals $[0,2\pi)$ and $[0,T)$, respectively. The received signal can be obtained as

$$\begin{aligned} r(t) &= \sum_{i=-P}^{P}\sum_{k=1}^{K}\sqrt{E_k(i)}\, b_k(i) s_k(t - iT) * h_k(t) + n(t) \\ &= \sum_{i=-P}^{P}\sum_{l=1}^{L}\sum_{k=1}^{K}\sqrt{E_k(i)}\, b_k(i) c_{k,l} e^{-j\varphi_{k,l}} s_k(t - iT - t_{k,l}) + n(t) \\ &= S(t,B) + n(t) \end{aligned} \tag{19}$$

where the symbol $*$ denotes convolution.

The output of the filter matched to the lth path of user k at time iT is obtained as

$$y_{k,l}(i) = \int_{iT+t_{k,l}}^{(i+1)T+t_{k,l}} r(t) s_k(t - iT - t_{k,l})\, dt . \tag{20}$$

The outputs of matched filters can be expressed in the matrix form as

$$Y = RAWb + n . \tag{21}$$

where

$$\begin{aligned} &Y = \left(Y(-P),\ldots,Y(i),\ldots,Y(P)\right) \quad Y(i) = \left(Y_1(i),\ldots,Y_K(i)\right) \\ &Y_k(i) = \left(Y_{k,1}(i),\ldots,Y_{k,L}(i)\right). \end{aligned} \tag{22}$$

In the above equation, $Y(i)$ is the output all K users matched filters, $Y_k(i)$ is the outputs of the kth user's matched filters at time iT, and

$$\begin{aligned} &b = [b(-P),\ldots,b(i),\ldots,b(P)]^T \\ &b(i) = [b_1(i)1^T,\ldots,b_K(i)1^T]^T. \end{aligned} \tag{23}$$

Here $1 = (1,\ldots,1)^L$ is an L vector with all elements equal to 1. A is the matrix of received signal coefficients for all K users defined as

$$\begin{aligned} &A = diag\left(A(-P),\ldots,A(i),\ldots,A(P)\right) \\ &A(i) = diag\left(\alpha_1(i)I,\ldots,\alpha_K(i)I\right) \\ &\alpha_k(i) = \left[\alpha_{k,1}(i),\ldots,\alpha_{k,L}(i)\right]^T \end{aligned} \tag{24}$$

where I is an $L \times L$ identity matrix, and $\alpha_{k,l} = c_{k,l}e^{-j\varphi_{k,l}}$, and R is the correlation matrix as transmitted bits (packet length) in each transmission, and $n(t)$ represents white Gaussian noise with two-sided spectral density equal to σ^2.

The conventional single user detector consists of a bank of filters matched to the spreading codes of users. Based on the sign of the outputs, an estimate of the transmitted bits is obtained. The matrix form of the outputs of matched filters can be expressed as

$$Y = RWb + n \tag{4}$$

where Y, R, W, b, and n are defined as:

$$Y = [Y^T(-P),...,Y^T(P)]^T \qquad Y(i) = [Y_1(i),...,Y_K(i)]^T$$
$$Y_k(i) = \int_{iT+\tau_k}^{(i+1)T+\tau_k} r(t)s_k(t - iT - \tau_k)dt \tag{5}$$

$$b = [b^T(-P),...,b^T(P)]^T \qquad b(i) = [b_1(i),...,b_K(i)]^T \tag{6}$$

$$W = diag\{W(-P),...,W(P)\}$$
$$W(i) = diag\left(\sqrt{E_1(i)},...,\sqrt{E_K(i)}\right) \tag{7}$$

$$n = [n^T(-P),...,n^T(P)]^T \qquad n(i) = [n_1(i),...,n_K(i)]^T$$
$$n_k(i) = \int_{iT+\tau_k}^{(i+1)T+\tau_k} n(t)s_k(t - iT - \tau_k)dt \tag{8}$$

$$R = \begin{bmatrix} R(0) & R(-1) & 0 & \cdots & 0 \\ R(1) & R(0) & R(-1) & & \vdots \\ 0 & R(1) & R(0) & & 0 \\ \vdots & & \ddots & \ddots & R(-1) \\ 0 & \cdots & 0 & R(1) & R(0) \end{bmatrix}_{(2P+1)K\times(2P+1)K}$$
$$R(1) = R(-1)^T \tag{9}$$

$$R(i) = \{R_{jk}(i)\}_{K\times K} \qquad i = 0,\pm1\,.$$
$$R_{jk}(i) = \int_{-\infty}^{+\infty} s_j(t-\tau_j)\,s_k(t+iT-\tau_k)\,dt \tag{10}$$

In equations (4-10), $Y_k(i)$ is the output of matched filter of user k at time iT, W is a diagonal matrix containing users energy, R is the correlation matrix, $R_{jk}(i)$ is the correlation between spreading codes of users j and k at time iT, and n is Gaussian noise vector with covariance matrix of $\sigma^2 R$.

B-Synchronous transmission

In this case $\tau_k = 0 \ \forall k$, and it is seen from (3) that the received signal at time iT depends on the transmitted bits of that time, so that we can drop the index i:

$$r(t) = \sum_{k=1}^{K} \sqrt{E_k}\, b_k(t) s_k(t) + n(t) = S(t,b) + n(t)\,. \tag{11}$$

The outputs of matched filters can be written as in (4) with the following definitions

$$Y = (Y_1,...,Y_K)^T \qquad b = (b_1,...,b_K)^T \qquad n = (n_1,...,n_K)^T \tag{12}$$

$$W = diag(\sqrt{E_1},...,\sqrt{E_K})$$
$$R = \{R_{jk}\}_{K\times K} \qquad R_{jk} = \int_0^T s_j(t)s_k(t)dt\,. \tag{13}$$

The conventional detector is optimum in the presence of only one user and white Gaussian noise. However, in the presence of other users, this detector suffers from two problems. First, as the number of users increases, it can become multiple access interference (MAI) limited. Another disadvantage is the near-far problem, which refers to the phenomenon of high power interferes completely destroying communications from low power transmitters. One solution is multiuser detection where the receiver exploits the knowledge of interfering signals [2]. The optimum multiuser detector maximizes the joint probability density function (maximum likelihood). It selects data sequence b to minimize

$$\Lambda(b) = \int [r(t) - S(t,b)]^2 dt\,. \tag{14}$$

The minimization of above equation results in the following log-likelihood detection

$$\hat{b}_{OMD} = \arg\left\{\max_{b\in\{-1,1\}} \left(2Y^T b - b^T WRWb\right)\right\}. \tag{15}$$

The optimum detector must evaluate the above metrics over all $2^{K(2P+1)}$

decision boundaries formed by the optimal receiver in CDMA can be estimated by NN [3], [4].

The first paper that considered the application of NN in CDMA systems is due to Aazhang *et al.* [3]. They show by applying a complicated training method called "assisted back propagation", where the number of neurons grows exponentially with the number of nodes, the performance of multilayer perceptron is close to that of the optimum receiver in both synchronous and asynchronous Gaussian channels. Although the simulation results show that back propagation (BP) learning rule outperforms the conventional detector, it is still an open problem. The receiver proposed in [4] uses Radial Basis Function (RBF) net where its complexity in terms of centers (neurons, nodes) grows exponentially with the number of users, and, becomes too complex under the multipath environment. In [5], it is shown that the energy function of Hopfield recurrent neural network is identical to the likelihood function encountered in multiuser detection. However, the desired optimal solutions are not guaranteed because a combinatorial optimization problem always involves a large number of local minimums. Also the number of connections grows with the square number of users and the number of neurons increases for asynchronous transmission. However, in these papers multipath fading channels are not considered.

In this paper, we analyze and examine the performance of three layer perceptron neural net using back propagation [6] algorithm for multiuser detection of DS/CDMA signals in AWGN (Additive White Gaussian Noise) and multipath fading channels. The results show superior improvement over the previous research in terms of complexity. We compare the performance of BP network with the decorrelator [7] and multistage [8] detectors widely used for comparative analysis. Since our goal is to improve the performance of BP net, we consider different neural networks. We apply Decision Based Neural Network (DBNN) [9], [10], Fuzzy Decision Neural Network (FDNN) [9], [10], discriminative learning [11],[12], minimum classification [13], and cross entropy [14] nets that have better performance than BP in radar, sonar, pattern recognition, and speech processing applications. We also propose modified DBNN that outperforms DBNN.

This paper is organized as follows. Section 1 describes the model of DS/CDMA system in AWGN and fading channels. In Section 2, we explain different neural networks used in our experiments. In Section 3 simulations and results are discussed. Finally Section 4 presents the conclusion.

1-The Model of DS/CDMA System

The system model consists of K independent simultaneous users. The kth user's transmitted signal assuming BPSK data modulation is of the form

$$r_k(t) = \sum_i \sqrt{E_k(i)}\, b_k(i)\, s_k(t - iT) \qquad (1)$$

where $E_k(i)$ is the energy of the kth user at time iT, $1/T$ is the data rate, $b_k(i) \in \{\pm 1\}$ is the data bit of user k during the ith interval, and $s_k(t)$ is the spreading (signature) waveform of duration T and normalized power which is composed of a spreading sequence of N chips (code length) as

$$s_k(t) = \sum_{n=0}^{N-1} a_n^k(t)\, p(t - nT_c) \qquad (2)$$

where $a_n^k \in (-1,1)$ is the spreading sequence, $p(t)$ is the rectangular waveform of duration T_c, and $T = NT_c$. We obtain the receiver input and output in AWGN and fading channels.

1-1-AWGN channel

We consider asynchronous and synchronous transmissions.

A-Asynchronous transmission

The baseband received signal is obtained as:

$$r(t) = \sum_{i=-P}^{P} \sum_{k=1}^{K} \sqrt{E_k(i)}\, b_k(i)\, s_k(t - iT - \tau_k) + n(t) = S(t,b) + n(t) \qquad (3)$$

where τ_k is the kth user time delay which is in the interval $[0,T)$, $2P+1$ is the number of

Multiuser Detection of DS/CDMA Signals Using Neural Networks

M. G. Shayesteh
Ph.D. Candidate and Lecturer
Department of Electrical Engineering,
Amir kabir University of Technology
Faculty of Engineering, Urmia University

H. Amindavar
Assistant Professor
Department of Electrical Engineering,
Amir kabir University of Technology

Abstract

Multiple access interference and near-far effect cause the performance of the conventional single user detector in DS/CDMA systems to degrade. Due to high complexity of the optimum multiuser detector, suboptimal multiuser detectors with less complexity and reasonable performance have received considerable attention. In this paper, we analyze and examine the performance of multilayer perceptron neural networks using back-propagation algorithm as multiuser detectors of CDMA signals in AWGN and multipath fading channels. Our results show significant improvement over the previous research. We compare the performance of neural network with the other detectors used in CDMA system. We also apply different neural networks and criterions such as the decision based, fuzzy decision, discriminative learning, minimum classification, and cross entropy neural nets and compare their performance. Further, we propose modified decision based network which improves the performance of decision based network.

Keywords

Multiuser detection, DS/CDMA, neural network (NN), MAI (Multiple Access Interference).

Introduction

DS/CDMA (Direct Sequence Code Division Multiple Access) is considered as the third generation of cellular mobile, indoor wireless and personal communication systems. CDMA offers attractive features such as frequency reuse, soft handoff, increased capacity, and multipath combating [1]. In a CDMA system, several users simultaneously transmit information over a common channel using pre-assigned codes.

The conventional single user detector consists of a bank of filters matched to the spreading codes and then deciding on the sign of the outputs. This detector suffers from two problems. First, multiple access interference (MAI) produced by the other co-channel users is a significant limitation to the capacity of this detector. The second problem is the near-far effect which occurs when the relative received power of interfering signals becomes larger.

A potential solution is multiuser detection [2] which exploits the information of signals of interfering users. The optimum multiuser detector evaluates a log-likelihood function over the set of all possible information sequences. It achieves low error probability at the expense of high computational complexity that increases exponentially with the number of users. So this method is extremely complex for a realistic number of users. Consequently, there has been considerable research into suboptimal detectors. These detectors achieve significant performance gains over the conventional detector without the exponential increase in the receiver complexity.

Several factors motivate us to apply neural networks (NN) as multiuser detectors. They are adaptive and computationally efficient. Also the cyclo-stationary structure of MAI and nonlinear