Compensating for Channel Fading in DS-CDMA Communication Systems Employing ICA Neural Network Detectors

David OVERBYE

DeVry University
One Tower Lane, Suite 1000
Oakbrook Terrace, IL 60181

and

Roland PRIEMER*

University of Illinois at Chicago
Department of Electrical and Computer Engineering
851 South Morgan Street (M/C 154)
Chicago, Illinois 60607-7053
priemer@ece.uic.edu

* Corresponding Author

ABSTRACT

In this paper we examine the impact of channel fading on the bit error rate of a DS-CDMA communication system. The system employs detectors that incorporate neural networks effecting methods of independent component analysis (ICA), subspace estimation of channel noise, and Hopfield type neural networks. The Rayleigh fading channel model is used. When employed in a Rayleigh fading environment, the ICA neural network detectors that give superior performance in a flat fading channel did not retain this superior performance. We then present a new method of compensating for channel fading based on the incorporation of priors in the ICA neural network learning algorithms. When the ICA neural network detectors were compensated using the incorporation of priors, they give significantly better performance than the traditional detectors and the uncompensated ICA detectors.

Keywords: CDMA, Multi-user Detection, Rayleigh Fading, Multipath Detection, Independent Component Analysis, Prior Probability Hebbian Learning, Natural Gradient

1. INTRODUCTION

As DS-CDMA communication systems continue to be developed and deployed for widespread multiuser commercial applications, channel equalization for purposes of interference suppression is an important area of study. One problem that quickly becomes apparent in the use of DS-CDMA systems is the so-called “near/far” problem, a result of a situation in which multiple users are transmitting at the same power, but are located at different distances from the receiver. Under these conditions the use of a multiuser detector (MUD) becomes essential.

A mathematical formulation of the optimal MUD problem was achieved by Verdu [1]. Lupas [2] showed the optimal MUD was near/far resistant and Verdu [3] proved that the optimal MUD was NP-hard in the number of users. This implies that work should concentrate on sub-optimal approaches, such as neural networks.

Independent Component Analysis (ICA) has been developed in recent years to solve the most general type of blind signal separation (BSS) problem [4],[5]. In this paper we extend work that was reported in [6], in which we applied linear and nonlinear neural network approaches to the CDMA MUD problem in flat fading channels, and we obtained very good results. Specifically, here we present a new method for compensating ICA neural network detectors based on prior knowledge of the mixing process, in this case the Rayliegh fading channel model.

We first develop a channel model for a DS-CDMA system in Section 2. Section 3 describes the multiuser detectors, and Section 4 describes how the learning algorithms can be modified to include prior knowledge. The detectors are implemented in Section 5, and the results are presented in Section 6. Finally, conclusions are given in Section 7.

2. CHANNEL MODEL

In a DS-CDMA system the user's transmitted bit information is spread in the frequency domain via the modulation of the data signal with a unique signature waveform code. The transmitted user waveforms are of the form

$$s_k(t) = \sum_{j=0}^{N-1} \beta_k(j)\ \psi_{kl}(t - jT_c)\ ,\ t \in [0, T] \quad (2.1)$$

where $\beta_k(0),\beta_k(1),...,\beta_k(N\text{-}1)$, $\beta_k(i)\in\{+1,-1\}$ is the unique pseudo-random noise sequence assigned to the k^{th} user, also represented as vector $\boldsymbol{\beta}_k = [\beta_k(0)\ \beta_k(1)\ \dots\ \beta_k(N\text{-}1)]^T$, and the chip waveform for the k^{th} user over the l^{th} path is given by $\psi_{kl}(t)=A\cos(\omega_c t+\theta)$, $\theta\in[0,\pi]$.

If we have a synchronous, multi-user spread-spectrum communication system, the received signal can be modeled as

$$r(t) = \sum_{k=1}^{K} A_k e^{j\varphi_k} \sum_{i=-P}^{P} b_k(i) \sum_{l=1}^{L} c_{kl} * s_k(t - iT) + \sigma n(t) \quad (2.2)$$

where K is the number of users, $A_k e^{j\varphi_k}$ is the received amplitude and phase of the carrier relative to the local oscillator at the receiver, $b_k(i)$: $i = 0, \pm1, \dots, \pm P$: $b_k(i)\in\{+1,-1\}$ is the symbol stream of the k^{th} user, 2P+1 is the frame length, $s_k(t)$: $0 \le t \le T$ is the modulation waveform of the k^{th} user as defined in Eq (2.1), L is the number of propagation paths, c_{kl} is the complex channel fading gain of the k^{th} user over the l^{th} path, n(t) is zero-mean Gaussian noise with unity power and σ^2 is the noise power.

We express Eq (2.2) in matrix form by defining the N x L matrix $\mathbf{T}_k = [\psi_{k1}\boldsymbol{\beta}_k\ \psi_{k2}\boldsymbol{\beta}_k \dots \psi_{kL}\boldsymbol{\beta}_k]$. The matrix $\mathbf{T_k}$ represents a symbol bit sent by user k over each of the L multiple propagation paths. We then define the N x KL matrix **S** as $\mathbf{S} = [\mathbf{T}_1 \dots \mathbf{T}_K]$, which modulates a symbol bit sent by all users over each of the L multiple propagation paths. We call the matrix **S** the code spreading matrix.

Also, it is useful to collect the bit streams of all users as the KL vector **b**(i) given by $\mathbf{b}(i) = [b_{11}(i), b_{12}(i),\dots, b_{1L}(i),\ b_{21}(i), b_{22}(i),\dots b_{2L}(i),\ b_{K1}(i), b_{K2}(i),\dots b_{KL}(i)]^T$. These bit streams are assumed to be independent and identically distributed (I.I.D.) random variables with mean value equal to zero. Thus, the product **Sb**(i) is the modulated bit-stream of all users over all paths.

For each user we define the transfer function for each of the L multiple propagation paths. Thus, the vector $\mathbf{h}_k$ given by $\mathbf{h}_k = [h_{k1}\ \dots h_{kL}]^T$ represents each of the L channel transfer functions faced by the k^{th} single user. We then create the KL x KL diagonal matrix $\mathbf{H} = \mathrm{diag}(\mathbf{h}_1,\dots\mathbf{h}_k)$, which represents the channel transfer functions faced by all users over all paths. Finally, define a KL x KL carrier matrix **A**, as $\mathbf{A} = \mathrm{Diag}[\mathbf{A}_1\mathbf{I}_L \dots \mathbf{A}_K\ \mathbf{I}_L]$ where $\mathbf{I}_L$ is an identity matrix of dimension L and $\mathbf{A}_k$ is an L x L diagonal matrix containing the received powers and phases of each path for a single user. Collecting all terms, the received signal becomes

$$\mathbf{r}(i) = \mathbf{SHAb}(i) + \sigma\mathbf{n}(i) \quad (2.3)$$

To generalize this model to allow for an FIR channel of length M, we must reformulate the matrix **H** to include the convolution matrix of each of the paths. In the discrete case, each of the KL channels is modeled as an FIR filter of length M and we can define the M x 1 channel vector as $\mathbf{h}_{kl}=[h_{kl}(0)\ h_{kl}(1) \dots h_{kl}(M\text{-}1)\ ]^T$. We define an M x 2M-1 Toeplitz convolution matrix $\mathbf{H}_{kl}$ for the k^{th} user over the l^{th} channel as

$$\mathbf{H}_{kl} = \begin{bmatrix} h_{kl}(0) & h_{kl}(1) & \cdots & h_{kl}(M\text{-}1) & & & & \\ & h_{kl}(0) & h_{kl}(1) & \cdots & h_{kl}(M\text{-}1) & & & \\ & & \ddots & \ddots & & \cdots & \ddots & \\ & & & & h_{kl}(0) & h_{kl}(1) & \cdots & h_{kl}(M\text{-}1) \end{bmatrix}$$

To account for intersymbol interference we define the matrix **P** as

$$\mathbf{P} = \mathbf{H}\begin{bmatrix} \mathbf{SA} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{SA} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{SA} \end{bmatrix}$$

Then, the received signal becomes

$$\mathbf{r}(i) = \mathbf{Pb}(i) + \sigma\mathbf{n}(i) \quad (2.4)$$

Allowing for intersymbol interference requires that an entire frame of 2P+1 bits sent be processed. To do this, the vector **b**(i) $= [b_{11}(i), b_{12}(i),\dots, b_{1L}(i), \dots, b_{K1}(i), b_{K2}(i),\dots b_{KL}(i)]^T$ can be extended to a KL x (2P+1) matrix **B** such that $\mathbf{B} = [\mathbf{b}(1) \dots \mathbf{b}(2P+1)]$. Finally, define the noise matrix **N** as $\mathbf{N} = [\mathbf{n}(1) \dots \mathbf{n}(2P+1)]$, and define the received signal matrix **R** as

$$\mathbf{R} = \mathbf{PB} + \sigma\mathbf{N} \quad (2.5)$$

When formulated in this way, the multi-user detection problem can be viewed as that of a convolutive mixture, which can be unmixed with methods of Independent Component Analysis.

3. MULTIUSER DETECTORS CONSIDERED

In our investigations four detector types were used. The first was the conventional single-user detector. The second was a one-shot semi-blind LMMSE multiuser detector which, because of its well understood properties, is a useful benchmark. The specific detector used here is an LMMSE detector with an enhancement in the form of a noise estimator based on subspace decomposition of the received signal autocorrelation matrix [7].

The third and forth detector types employed are based on neural networks and are described below.

3.1 Multi-Stage Hopfield Neural Network Detector

The Hopfield neural network can be applied to CDMA MUD by taking advantage of *a-priori* knowledge of the user spreading codes and using these as the stored patterns. The Hopfield net is essentially compensating for code cross-correlation, since no practical code set will be completely orthogonal.

Kechriotis and Manolakos [8] and [9] noted that the optimal MUD formulation is essentially the same formulation as the Hopfield neural network. Varanasi and Aazhang [10] proposed a multistage detector that replicates the bi-directional associative memory principle employed in the Hopfield neural network. An extension of this detector that was developed and explored in [11] will be considered here.

3.2 Independent Component Analysis Detector

Independent Component Analysis (ICA) is a technique whereby the output data from some unknown process is transformed such that the transformed variables are statistically independent to the extent possible. Because nothing needs to be known about the process, ICA is a blind technique. The only constraint on the input signals is that they be assumed independent. Since the user bit streams in a CDMA communications system fit this assumption, ICA is a good technique to explore for multiuser detection.

There are two major approaches to solving the ICA problem, statistical approaches and neural network approaches. In the statistical approaches, higher order cumulants of the observed signal are used to define an objective function that is then minimized, usually in batch mode. Comon [12] explores statistical approaches and presents a unified formulation based on the idea of maximum mutual information (infomax) which was defined by Linsker [13]. Bell [14] derived a self-organizing ICA algorithm that maximizes the transmitted information. Lee [15] extended this work to create a unifying information theoretic framework that proves the infomax, maximum likelihood, and negative entropy approaches all lead to the same learning rule.

In neural network approaches to ICA, the nonlinearity introduced by the neurons replaces the cumulants used in the statistical approaches. In addition, the neural network approaches are suitable to recursive implementation. Jutten [16] proposed such an approach. Karhunen [17] summarizes a number of approaches and algorithms and concludes that neural approaches to ICA have a wide range of applicability and show great promise.

ICA can be formulated by considering a vector **b** of dimension M, where each element of **b** is an independent signal $b_i(t)$, and defining $\mathbf{b}(t)=[b_1(t),...,b_M(t)]^T$. If we then define an output vector **r** as $\mathbf{r}(t)=[r_1(t),...r_N(t)]^T$ such that

$$\mathbf{r}(t) = \mathbf{Pb}(t) \tag{3.2.1}$$

where the matrix **P** is some “mixing” matrix, we will now have an output that no longer consists of independent signals.

In discrete terms we can define $\mathbf{b}(i)=[b_1(i),...,b_M(i)]^T$ and $\mathbf{r}(i)=[r_1(i),...r_N(i)]^T$. If we define $\mathbf{B} = [\mathbf{b}(1)\ \mathbf{b}(2) \ldots \mathbf{b}(2P+1)]$ and $\mathbf{R} = [\mathbf{r}(1)\ \mathbf{r}(2) \ldots \mathbf{r}(2P+1)]$ then

$$\mathbf{R}=\mathbf{PB} \tag{3.2.2}$$

Notice that this is exactly the formulation of Eq(2.5) above. The goal of ICA is to restore **B** from **R,** by finding **W**, the pseudo-inverse of **P**.

The mixing matrix **P** may be linear or nonlinear. In the linear case, often called principal subspace analysis, simple gradient descent is employed and there is no nonlinearity introduced into the solution in the form of a neural activation function.

To solve the non-linear case Hyvärinen and Oja [18] investigated the use of neural networks employing both Hebbian and non-Hebbian learning rules. This work was then extended by Hyvärinen and Pajunen [19] to the point where general existence and uniqueness results were developed for non-linear ICA.

Since each element of **b** is an independent signal $b_i(t)$ we can write the probability distribution of **b,** p(**b**), as

$$p(\mathbf{b}) = \prod_{i=1}^{N} p_i(b_i) \tag{3.2.3}$$

The mutual information of the observed vector **r** is given by the Kullback-Leibler (KL) divergence I(**r**), which is defined as

$$I(\mathbf{r}) = \int p(\mathbf{r}) \log \frac{p(\mathbf{r})}{\prod_{i=1}^{N} p_i(r_i)} d\mathbf{r} = D\left(p(\mathbf{r}) \,\|\, \prod_{i=1}^{N} p_i(r_i) \right) \tag{3.2.4}$$

I(**r**), is always greater than or equal to zero. In the case where I(**r**), is equal to zero, perfect recovery of the signal has been achieved.

The normalized log likelihood of a received signal of length M would be

$$\log L = \sum_{m=1}^{M} \sum_{i=1}^{N} \log p_i(\mathbf{w}_i^T \mathbf{r}(m)) + M \log |\det(\mathbf{W})| \tag{3.2.5}$$

where $\mathbf{w}_i$ is a column of **W**. We can reformulate Eq (3.2.5) by replacing the summation over M with the expectation operator to obtain

$$\frac{1}{M} \log L = E\left\{ \sum_{i=1}^{N} \log p_i(\mathbf{w}_i^T \mathbf{r}) \right\} + \log |\det(\mathbf{W})| \tag{3.2.6}$$

This log likelihood function then becomes the penalty function J(**W**)**.** The optimum **W** is found by maximizing J(**W**) with respect to **W**. This is done via an iterative gradient ascent algorithm expressed as

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k \nabla J(\mathbf{W}_k)\,|_{\mathbf{y}=\mathbf{y}_k} \tag{3.2.7}$$

Because Eq (3.2.6) is expressed in terms of det**W** we need to reformulate it by considering that the gradient of det**W** can be expressed as

$$\frac{\partial}{\partial \mathbf{W}} \det \mathbf{W} = \left(\mathbf{W}^T\right)^{-1} \det \mathbf{W} \tag{3.2.8}$$

and, if **W** is an invertible matrix, then $\mathbf{W}^{-1}$ can be expressed as

$$\mathbf{W}^{-1} = \frac{1}{\det \mathbf{W}} \operatorname{adj}(\mathbf{W}) \tag{3.2.9}$$

Making use of Eq (3.2.8) and Eq (3.2.9) we have

$$\frac{1}{\mathrm{M}}\frac{\partial \log L}{\partial \mathbf{W}} = \left[\mathbf{W}^{\mathrm{T}}\right]^{-1} + E\left\{f(\mathbf{Wr})\mathbf{r}^{\mathrm{T}}\right\} \quad (3.2.10)$$

where $f(\bullet)$ is a monotonic, odd, nonlinear function that is related to the probabilty distributions of the independent sources.

We can eliminate the expectation operator in Eq(3.2.10) when the ascent algorithm is performed iteratively. If we do this, as well as multiply the right hand side of Eq (3.2.10) by $\mathbf{W}^{\mathrm{T}}\mathbf{W}$, we obtain the result that

$$\Delta\mathbf{W} \propto [\mathbf{I}\text{-}f(\mathbf{r})\mathbf{r}^{\mathrm{T}}]\mathbf{W} \quad (3.2.11)$$

which can be used in the gradient assent algorithm given by

$$\mathbf{W}_{q+1} = \mathbf{W}_q + \mu[\mathbf{I}\text{-}f(\mathbf{r})\mathbf{r}^{\mathrm{T}}]\mathbf{W}_q \quad (3.2.12)$$

Since, in general, we do not know the probability distributions of the sources at the receiver, the essence of the blind separation of sources problem becomes one of identifying either an acceptable mapping function $f(\bullet)$ or modifying the basic gradient assent algorithm. The natural gradient method described below does the former and the various other neural learning structures, like the non-linear Hebbian algorithm do the latter.

Some work has been done on the application of ICA to the CDMA detection problem. Joutsensalo [20], [21] and Cristescu [22], for example, have made use of the fast fixed point ICA algorithm based on higher order statistics, specifically fourth-order cumulants. Their work illustrates good results and motivates additional investigation into the use of ICA algorithms for blind detection in CDMA systems. The approach that we took to applying ICA to multiuser detection is unique in two ways. The first is that we used the neural network formulations rather than cumulants. The second is that our detector is a hybrid, making use of classical detector methods for preprocessing. The detectors described in [7] and [11] were used for initial processing, followed by additional processing by ICA. The bit error rate (BER) performance with several learning algorithms was examined. In each algorithm $\mathbf{z}$ is a whitened version of the received signal vector $\mathbf{r}$ and $\mathbf{y} = \mathbf{Wz}$.

The ***principal subspace algorithm*** is a linear algorithm [23] that makes no use of an activation function. Its main advantages are simplicity and ease of implementation. The update function used is

$$\mathbf{W}_{q+1}(\mathrm{i}) = \mathbf{W}_q(\mathrm{i}) + \mu\mathbf{y}(\mathrm{i})\mathbf{z}^{\mathrm{H}}(\mathrm{i}) \quad (3.2.13)$$

This algorithm rarely converged in this application.

The ***modified principal subspace algorithm***, proposed by Karhunen and Oja [24], employs the algorithm

$$\mathbf{W}_{q+1}(\mathrm{i}) = \mathbf{W}_q(\mathrm{i}) + \mu[\mathbf{y}(\mathrm{i})\mathbf{z}^{\mathrm{H}}(\mathrm{i}) - \mathbf{y}(\mathrm{i})\mathbf{y}^{\mathrm{H}}(\mathrm{i})\mathbf{W}_q(\mathrm{i})] \quad (3.2.14)$$

The addition of the term $\mathbf{y}(\mathrm{i})\mathbf{y}^{\mathrm{H}}(\mathrm{i})\mathbf{W}_q(\mathrm{i})$ tends to stabilize the algorithm. As with the unmodified principal subspace algorithm, this is a linear algorithm.

A ***Hebbian learning algorithm*** can, in general, be any algorithm in which the gradient of the weight matrix is proportional to the input multiplied by some function of the input plus some feedback terms, with a learning rule: $\Delta\mathbf{W} \propto f(\mathbf{y})\mathbf{z}^{\mathrm{T}}$ + ... [feedback].

A non-linear Hebbian algorithm has been proposed [18] that employs

$$\mathbf{W}_{q+1}(\mathrm{i}) = \mathbf{W}_q(\mathrm{i}) - \mu f[\mathbf{y}(\mathrm{i})]\mathbf{z}^{\mathrm{H}}(\mathrm{i}) + \alpha[\mathbf{I} - \mathbf{W}_q(\mathrm{i})\mathbf{W}_q^{\mathrm{H}}(\mathrm{i})]\mathbf{W}_q(\mathrm{i}) \quad (3.2.15)$$

where $f(\mathbf{y})$ is the neural activation function. This algorithm was selected for application to the CDMA MUD problem because of its generality and success in its original application of feature extraction in images [25], which is close to that of CDMA MUD.

The ***natural gradient algorithm*** is a variant of classical optimzation based on Newton's method in which the neural activation function is serving as a Riemannian metric tensor that transforms the problem from one of optimization in a Euclidean N-Space to one of optimization on a Riemannian manifold [26]. It employs $\mathbf{W}_{q+1}(\mathrm{i}) = \mathbf{W}_q(\mathrm{i}) - \mu[\mathbf{I} - f\{\mathbf{y}(\mathrm{i})\}\mathbf{y}^{\mathrm{H}}(\mathrm{i})]\,\mathbf{W}_q(\mathrm{i})$.

4. PRIORS IN ICA DETECTION

Recalling the general formulation of the ICA problem presented in Eq (3.2.1), if we define a new variable Ω to represent any prior knowledge, and use Bayes theorem, we can express this prior knowledge as

$$\mathrm{p}(\mathbf{P}, \mathbf{b}|\mathbf{r}, \Omega) \propto \mathrm{p}(\mathbf{r}|\mathbf{P},\mathbf{b}, \Omega)\,\mathrm{p}(\mathbf{P}|\mathbf{b}, \Omega)\,\mathrm{p}(\mathbf{b}|\Omega) \quad (4.1)$$

If the properties of the mixing matrix $\mathbf{P}$ do not depend on the properties of the source signal, as is the case in a CDMA communication system, we can replace the term $\mathrm{p}(\mathbf{P}|\mathbf{b}, \Omega)$ with $\mathrm{p}(\mathbf{P}|\Omega)$ and rewrite Eq (5.1.1) as

$$\mathrm{p}(\mathbf{P}, \mathbf{b}|\mathbf{r}, \Omega) \propto \mathrm{p}(\mathbf{r}|\mathbf{P},\mathbf{b}, \Omega)\,\mathrm{p}(\mathbf{P}|\Omega)\,\mathrm{p}(\mathbf{b}|\Omega) \quad (4.2)$$

Since $\mathbf{P}$, $\mathbf{b}$, and $\mathbf{r}$ are related by Eq (3.1.1) we can express Eq (4.2) as

$$\mathrm{p}(\mathbf{P} \mid \mathbf{r}, \Omega) \propto \mathrm{p}(\mathbf{P} \mid \Omega)\int \mathrm{p}(\mathbf{r} \mid \mathbf{P}, \mathbf{b}, \Omega)\,\mathrm{p}(\mathbf{b} \mid \Omega)\mathrm{d}\mathbf{b} \quad (4.3)$$

Essentially, Eq (4.3) says that the likelihood of a mixing model given the received signal and incorporating any prior information, is proportional to the product of our prior knowledge of the mixing process, the likelihood of the received signal given the mixing process, source signals, and prior information, and our prior knowledge of the sources.

In the case where we have no knowledge of the mixing process, $\mathrm{p}(\mathbf{P}|\,\Omega) = \mathbf{1}$, i.e., a uniform prior. When we have no knowledge of the sources, other than that they are statistically independent, $\mathrm{p}(\mathbf{b}|\Omega)$ is $\mathrm{p}(\mathbf{b})$. Finally, when the mixture is instantaneous and linear, the term $\mathrm{p}(\mathbf{r}|\mathbf{P},\mathbf{b}, \Omega)$ becomes a delta function. The combined result of this ignorance of all priors is the formulation for log L expressed in Eq (3.2.5) and Eq (3.2.6). This then leads to the general formulation of the update algorithm in Eq (3.2.12).

When we have some knowledge of the problem, we can make use of this knowledge to derive an update algorithm that takes this knowledge into account. For example, if we know the probability distributions of the elements P_{ij} of the mixing matrix

P, and know that the elements of **P** are independent, then we can express p(**P**| Ω) as

$$p(\mathbf{P} \mid \Omega) = \prod_{i,j} p(\mathbf{P}_{ij} \mid \Omega) \tag{4.4}$$

The normalized log likelihood of a received signal of length M would then be

$$\log L = \log p(\mathbf{P} \mid \mathbf{r}, \Omega) = \sum_{m=1}^{M} \sum_{i=1}^{N} \log p_i(\mathbf{w}_i^T \mathbf{r}(m)) +$$
$$M \log |\det(\mathbf{W})| + M \sum_{i,j} \log p(\mathbf{P}_{ij} \mid \Omega) \tag{4.4}$$

where $\mathbf{w}_i$ is a column of **W.** We can reformulate Eq (5.1.5) by replacing the summation over M with the expectation operator to obtain

$$\frac{1}{M} \log L = E\left\{ \sum_{i=1}^{N} \log p_i(\mathbf{w}_i^T \mathbf{r}) \right\} +$$
$$\log |\det(\mathbf{W}) + \sum_{i,j} \log p(\mathbf{P}_{ij} \mid \Omega)| \tag{4.6}$$

Recalling that Eq (3.2.10) defined the gradient of log L without the use of prior information, we will concentrate on the gradient for the last term of Eq (4.6) only. The gradient of this term can be expressed as

$$\frac{\partial}{\partial \mathbf{W}_{ij}} \left[\sum_{i,j} \log p(\mathbf{P}_{ij} \mid \Omega) \right] = \frac{\frac{\partial}{\partial \mathbf{W}_{ij}} p(\mathbf{P}_{ij} \mid \Omega)}{p(\mathbf{P}_{ij} \mid \Omega)} \tag{4.7}$$

We now define a new matrix **M**, the prior knowledge matrix, in which each element is calculated using Eq (4.7). The natural gradient algorithm, when expressed to include **M** becomes

$$\mathbf{W}_{q+1}(i) = \mathbf{W}_q(i) - \mu[\mathbf{I} - f\{\mathbf{y}(i)\}\mathbf{y}^H(i) + \mathbf{W}^H(i)\mathbf{M}\mathbf{W}_q(i) \tag{4.8}$$

The modified principle subspace algorithm is

$$\mathbf{W}_{q+1}(i) = \mathbf{W}_q(i) + \mu[\mathbf{M}\ \mathbf{y}(i)\mathbf{z}^H(i) - \mathbf{y}(i)\mathbf{y}^H(i)\mathbf{W}_q(i)] \tag{4.9}$$

and the non-linear Hebbian learning algorithm is

$$\mathbf{W}_{q+1}(i) = \mathbf{W}_q(i) - \mu\mathbf{M}f[\mathbf{y}(i)]\mathbf{z}^H(i) + \alpha[\mathbf{I} - \mathbf{W}_q(i)\mathbf{W}_q^H(i)]\mathbf{W}_q(i) \tag{4.10}$$

4.1 Application to Channel Models

The flat fading channel is perhaps the simplest form of statistical model. In order to incorporate a flat fading channel assumption we assume a form for an element P_{ij} of the mixing matrix as $P_{ij} = a_{ij}$ where a_{ij} represents the channel attenuation. Since the attenuation is between a user and the base station, we express the probability of a_{ij} as $p(a_i \mid \Omega) = (b_2 - b_1)$; $b_1 \le a_i \le b_2$; 0 otherwise, where, b_1 and b_2 are the maximum and minimum fade respectively. Incorporating the assumption that the prior probability is uniform allows us to formulate the prior for P_{ij} as

$$p(P_{ij} \mid \Omega) = \int_0^\infty p(a_i \mid \Omega) da_i = \int_0^\infty \frac{1}{b_2 - b_1} da_i \tag{4.1.1}$$

If we assume that $b_2 = 1$ and $b_1 = 0$, representing a spread between no attenuation and complete loss of signal, then we can express the prior probability as $p(P_{ij} \mid \Omega) = a_i$, and each element of the prior knowledge matrix **M** will be

$$M_{ij} = \frac{\frac{\partial}{\partial W_{ij}} p(P_{ij} \mid \Omega)}{p(P_{ij} \mid \Omega)} = 0 \tag{4.1.2}$$

Because **M** = 0, the problem will reduce to the base algorithms of Section 3. The prior assumption of flat fading adds no information. This is consistent with the well-known result [4] that the maximum a posteriori estimator equals the maximum likelihood estimator when the prior is uniform. When the ICA detectors were examined in a flat fading channel [6] they displayed very good results without any compensation. Our aim here is to formulate a prior knowledge matrix **M** that can be applied to the harsher Rayleigh fading channel.

4.2 Rayleigh Distribution

A common model for wireless channels is Rayleigh fading. The Rayleigh distribution is related to the chi-square distribution. Considering X to be a Gaussian distribution, letting $Y = X_1^2 + X_2^2$, then the pdf of Y is

$$p_Y(y) = \frac{1}{2\sigma^2} e^{-y/2\sigma^2}, \, y \ge 0 \tag{4.2.1}$$

If we define a new statistical variable R as

$$R = \sqrt{X_1^2 + X_2^2} \tag{4.2.2}$$

the pdf of R, the Rayleigh distribution, is

$$p_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, r \ge 0 \tag{4.2.3}$$

In order to formulate the matrix **M** for the Rayleigh channel, we first define a form P_{ij} as $P_{ij} = a_i r_{ij}$ where r_{ij} is the distance between the transmitter and the receiver and a_i is the power. The equivalent form to Eq (4.1.1) is expressed as

$$p(P_{ij} \mid \Omega) = \int_0^\infty p(a_i \mid \Omega) da_i \int_0^\infty \delta(P_{ij} - a_i r_{ij}) p(r_{ij} \mid \Omega) dr_{ij} \tag{4.2.4}$$

where the delta term reflects our confidence in the form of the mixing matrix. Evaluating the integral with respect to r_{ij} we have

$$p(P_{ij}\mid\Omega)=P_{ij}\int_0^\infty \frac{1}{a_i}p(a_i\mid\Omega)p(r_{ij}\mid\Omega)da_i \quad (4.2.5)$$

We use the distribution for a_i, as in Section 4.1. We then assign the Rayleigh distribution to r. Noting that $r_{ij} = P_{ij}/a_i$, the probability then becomes

$$p(P_{ij}\mid\Omega)=P_{ij}\int_0^\infty \frac{1}{a_i(b_2-b_1)}\frac{(P_{ij}/a_i)}{\sigma^2}e^{-q_{ij}^2}da_i \quad (4.2.6)$$

where $q_{ij} = P_{ij}\sigma/\mathrm{sqrt}(2)a_i$. Evaluating the integral in Eq (4.2.6) we obtain

$$p(P_{ij}\mid\Omega)=\frac{P_{ij}\sqrt{\pi}\sqrt{2}\mathrm{erf}(q_{ij})}{2(b_2-b_1)\sigma} \quad (4.2.7)$$

The derivative of Eq (4.2.7) may then be evaluated as

$$\frac{\partial}{\partial P_{ij}}p(P_{ij}\mid\Omega)=2\frac{P_{ij}e^{-q_{ij}^2}}{2a_i^2(b_2-b_1)}-\frac{P_{ij}^3\sigma^2e^{-q_{ij}^2}}{2a_i^4(b_2-b_1)} \quad (4.2.8)$$

Finally, the elements M_{ij}, are expressed as

$$M_{ij}=\frac{\frac{\partial}{\partial P_{ij}}p(P_{ij}\mid\Omega)}{p(P_{ij}\mid\Omega)}=\frac{2\frac{e^{-q_{ij}^2}}{a_i^2}-\frac{P_{ij}^2\sigma^2e^{-q_{ij}^2}}{a_i^4}}{\frac{\sqrt{2\pi}\mathrm{erf}(q_{ij})}{\sigma}} \quad (4.2.9)$$

Notice that in order to implement an update algorithm it is necessary to work with the gradient of **W** rather than **P**. Because we pre-whitened the signal, we are able to use the conjugate transpose of **P** for **W.**

5. IMPLEMENTATION

Explicit calculation of bit error rates for nonlinear multi-user detectors is mathematically intractable. Therefore, the detectors described above were simulated and bit error rate performance analysis was based on this simulation. The various learning algorithms for ICA were implemented independently.

The chip modulation is BPSK using a sinusoidal pulse sampled at a rate of 8 samples per pulse. The use of such a sampled pulse represents a more accurate approximation of physical systems than does the square pulse often employed. The spreading codes were derived from Walsh functions. A spreading gain of eight (N=8) was used. The frame size is 267 (P=133), corresponding to rate set two, which supports a bit rate of 14,400 bits per second in the IS-95 standard.

Three users were modeled (K=3), each transmitting over a Rayleigh fading channel that included two paths per user (L=2). We explored mildly asynchronous scenarios through the use of phase delay in the primary path up to one chip period, but longer asynchronous delays in the primary path are possible. Also, for simplicity, we used the same secondary channel for each user. The channel model allows for any arbitrary number of paths per user and for each path to have its own FIR properties. However the goal was comparative among the various detectors, and it was decided that extensive diversity of channels would yield little additional benefit.

The detectors were simulated in a variety of noise conditions and user power ratios.

The blind detector employing the ICA learning algorithms used the output of the Hopfield-based semi-blind detector to establish an initial starting point. It should be noted however that this type of preprocessing is not necessary. The ICA neural network detectors can be used in both semi-blind applications, where they will be effective in harsh channel conditions and in environments of significant noise and interference, and in blind applications.

6. RESULTS

The bit error rate performance of the ICA detectors in the Rayleigh channel was compared with that of the conventional detector and the Hopfield-based detector described in [11].

The BER results are presented in Chart 1. In Chart 1, for example, we can see that the bit error rate obtained at a power ratio of 1 and noise power of 1, is 0.0707 for the conventional detector, 0.0395 for the multi-stage Hopfield neural network detector, 0.0173 for the uncompensated ICA detector employing the modified principal subspace algorithm, 0.0179 for the uncompensated ICA detector employing the nonlinear Hebbian learning algorithm, and 0.0525 for the uncompensated ICA detector employing the natural gradient algorithm. The results of the compensated detectors are presented in Chart 1 as well (PSO-C, NG-C and NLH-C), where we see that the bit error rate results obtained at a power ratio of 1 and noise power of 1, are 0.0178 for the compensated modified principal subspace algorithm, 0.0179 for the compensated nonlinear Hebbian algorithm, and 0.0513 for the compensated natural gradient algorithm.

These results are also given graphically in Figure 1, with the bit error rates of the compensated ICA detectors plotted as solid lines and the bit error rates of the uncompensated ICA detectors plotted as dotted lines. Figure 2 displays results bit error rate results when the power ratio equals 2.

Upon examining Chart 1 and Figures 1 and 2 we notice that the Rayleigh fading channel has a negative impact on the bit error rate results at low noise power. There is also a degradation in the relative performance of the ICA detectors compared to the conventional and Hopfield neural network-based detectors. If we examine Figure 1, this degradation in relative performance becomes apparent. In the higher-noise

Chart 1 - Bit Error Rates With Cosine Pulse - Rayliegh Channel								
	Power Ratio = 1 (1,1,1)							
Noise Power	CON	MS	NG	NG-C	PSO	PSO-C	NLH	NLH-C
0.25	9.15E-04	0	0.0094	0.0013	0.0012	8.32E-05	5.83E-04	1.00E-05
0.5	0.0161	2.20E-03	8.70E-03	8.01E-03	8.32E-04	0.0010	9.16E-04	9.15E-04
1	0.0707	0.0395	0.0525	0.0513	0.0173	0.0178	0.0179	0.0179
2	0.1677	0.1574	0.1680	0.1610	0.1001	0.1008	0.1000	0.1000
4	0.2916	0.2970	0.2955	0.2965	0.2802	0.2798	0.2772	0.2774
	Power Ratio = 2 (1.0,0.75,0.5)							
	CON	MS	NG	NG-C	PSO	PSO-C	NLH	NLH-C
0.25	0.0260	3.10E-03	1.28E-02	1.85E-02	0.022	0.0065	0.0132	0.0036
0.5	0.0591	0.0231	0.0283	0.0403	0.0193	0.0143	0.0216	0.0131
1	0.1195	0.0963	0.1066	0.1163	0.0666	0.0617	0.0607	0.0574
2	0.2275	0.2280	0.2330	0.2302	0.1818	0.1868	0.1848	0.1815
4	0.3379	0.3457	0.3264	0.3416	0.3402	0.3408	0.3389	0.3359
	Power Ratio = 4 (1.0,0.5,0.25)							
	CON	MS	NG	NG-C	PSO	PSO-C	NLH	NLH-C
0.25	0.1592	0.0381	0.0751	0.0716	0.1714	0.1042	0.0907	0.0375
0.5	0.1564	0.0890	0.1094	0.1088	0.1384	0.1203	0.0865	0.0747
1	0.1881	0.1728	0.1897	0.1914	0.1771	0.1676	0.1528	0.1520
2	0.2855	0.2883	0.2892	0.2891	0.2632	0.2628	0.2593	0.2598
4	0.3756	0.3843	0.3772	0.3779	0.3672	0.3748	0.3667	0.3686
CON = Conventional NE = LMMSE with Noise Estimator MS = Multi-stage MPS = Modified Principal Subspace NLH = Non-linear Hebbian C = Compensated								

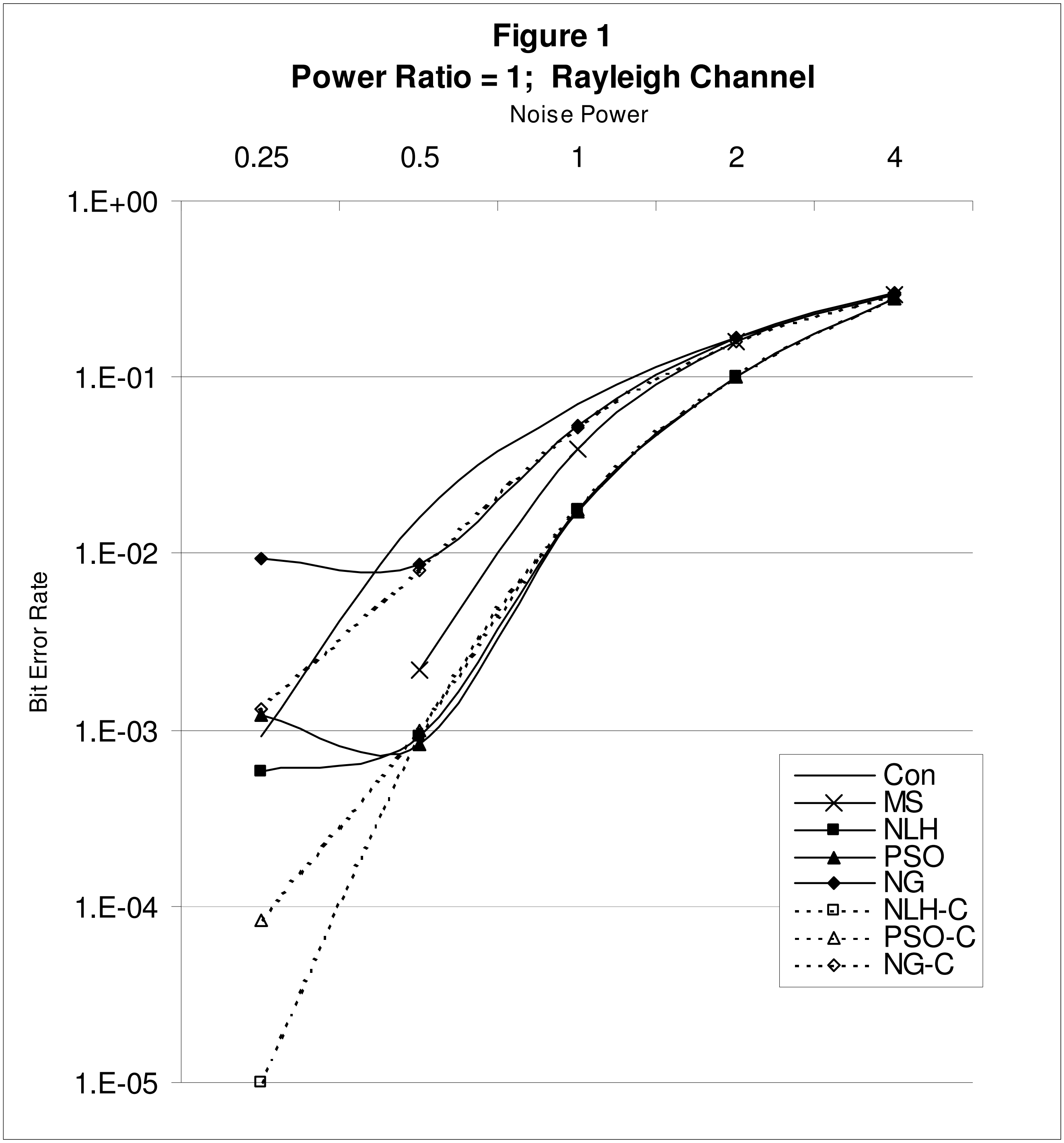
cases, the ICA detectors employing the non-linear Hebbian and the modified principal subspace algorithms retain their relative advantage in bit error rate performance over the two non-ICA detectors. Notice, however, that the ICA detector employing the natural gradient learning algorithm retains a relative advantage only over the conventional detector. Notice also that in the low noise cases, the relative bit error rate performance of the ICA detectors employing all three learning algorithms degrades badly. In the case where the noise power is equal to 0.25, all three ICA detectors lose their bit error rate performance advantage and the ICA detectors employing the natural gradient and principal subspace algorithms also lose their relative advantage even over the conventional detector. Only the ICA detector employing the non-linear Hebbian learning algorithm retains its bit error rate performance advantage over the conventional detector.

When the user power ratio is increased to two, as displayed in Figure 2, little changes. The ICA detectors do retain an advantage over the conventional detector, but this is only because the conventional detector has such poor near/far performance.

When the new compensation scheme, via the use of the prior knowledge matrix **M**, is applied to the ICA learning algorithms, the bit error rate results improve. In Figure 1, notice the significant result that the relative bit error rate performance advantage of the ICA detectors employing all three learning algorithms against both the conventional detector and the multi-stage Hopfield neural network-based detector has been restored under all conditions of noise power. The ICA detectors employing the modified principal subspace algorithm and the non-linear Hebbian learning algorithm display the best bit error rate performance, with the natural gradient learning algorithm performing a bit worse. When the user power ratio is increased to two, similar relative advantages result, as shown in Figure 2.

7. CONCLUSION

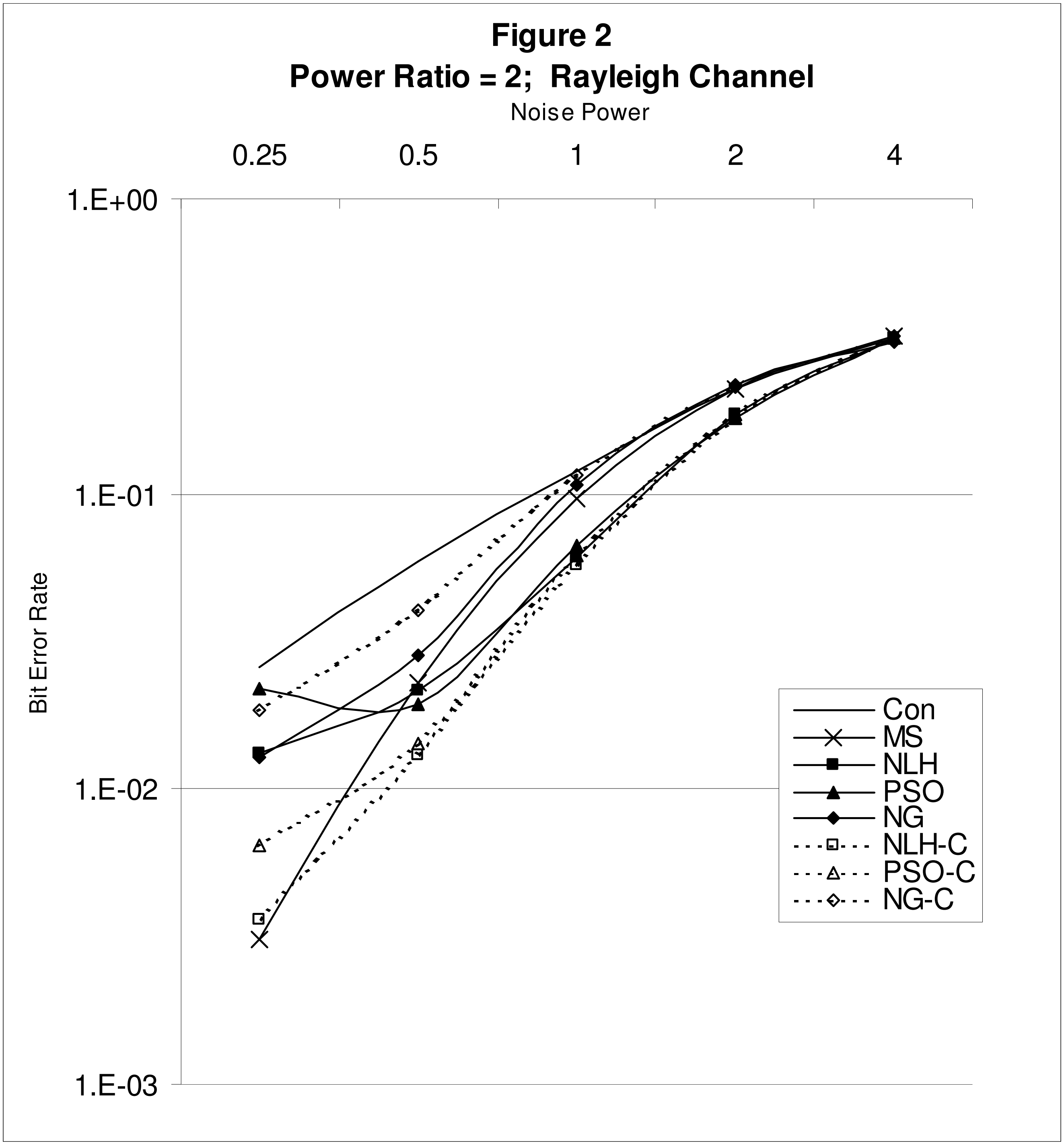
The new method of compensation presented here, through the use of a prior knowledge matrix in the neural network learning algorithm, demonstrates that the ICA neural network detectors are extremely powerful and represent a very attractive way of achieving blind detection of CDMA signals in Rayleigh fading channels. It is worth noting that in this investigation, the neural network detectors yielded superior bit error rate results when operating without knowledge of the user spreading codes and outperformed the Hopfield neural network-based detector,



even though the Hopfield detector was operating with knowledge of the user spreading codes.

When employed in the Rayleigh fading channel, the performance of the uncompensated ICA detectors degraded. This degradation eliminated much of the BER advantage of the ICA detectors. Through the use of a prior knowledge matrix in the learning algorithms, the performance of the ICA detectors was restored.

Opportunities for future research in the application of ICA neural networks to multiuser detection in CDMA systems appear to be very rich. There are additional learning algorithms that have not yet been investigated. There are also a number of other system mixing statistics, such as those related to the pulse shape employed, and those of the spreading codes, that could be incorporated into the prior knowledge matrix.



8. REFERENCES

1. Verdu, S. "Optimum multiuser signal detection," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Illinois – Urbana-Champaign, 1984.
2. Lupas, R. and Verdu, S. "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Transactions on Communications* volume 38, April 1990, Pages 496 – 508.
3. Verdu, S. "Computational complexity of optimum multiuser detection," *Algorithmica*, Vol. 4, No. 3, 1989, Pages 303-312.
4. Hyvärinen, A., Karhunen, J., and Oja, E. *Independent Component Analysis,* John Wiley, 2001.
5. Haykin, S., *Unsupervised Adaptive Filtering: Volume 1 Blind Source Separation,* John Wiley, 2000
6. Overbye, D. and Priemer, R. "Blind multiuser detection for DS-CDMA using an independent component analysis neural network" Proceedings of the Artificial Neural Networks in Engineering (ANNIE) conference, 2002 in

Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Data Mining and Complex Systems, edited by Drs. Dagli, Buczak, Ghosh, Embrechts, Ersoy and Kercel. Pages 111-116. ASME Press.

7. Overbye, D., and Priemer, R. “A semi-blind LMMSE multiuser detector for DS-CDMA using a subspace noise estimator,*” 44th IEEE Midwest Symposium on Circuits and Systems*, 2001.
8. Kechriotis, G.and Manolakos, E. “Hopfield neural network implementation of the optimal CDMA multiuser detector.**”** *IEEE Trans on Neural Networks*, Vol. 7, Jan. 1996, Pages 131-141.
9. Kechriotis, G.; Manolakos, E.S. “A hybrid digital signal processing-neural network CDMA multi-user detection scheme**”** *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Volume 43, Number 2, February 1996, Pages 96 –104.
10. Varanasi, M. K. and Aazhang, B. “Multistage detection in asynchronous code-division multiple access systems”, *IEEE Transactions on Communications* Volume 38, Number 4, April 1990, Pages 509-519.
11. Overbye, D., and Priemer, R. “A Neural Network based multiuser detector for CDMA communications systems,*” Proceedings of 15th International Conference on Systems Engineering*, 2002, Pages 237-243.
12. Comon, P. “Independent component analysis – a new concept?,” *Signal Processing,* volume 36, number 3, 1994, Pages 287-314.
13. Linsker, R, “Self-organization in a perceptual network.” *Computer*, Vol 21, Pages 105-117, 1988.
14. Bell, A.J. and Sejnowski, T.J. 1995. “An information maximisation approach to blind separation and blind deconvolution,” *Neural Computation,* Volume 7, Number 6, 1995, Pages 1129-1159.
15. Lee T-W., Girolami M., Bell A.J., and Sejnowski T.J. ‘A unifying Information-theoretic framework for independent component analysis*,” International Journal on Mathematical and Computer Modeling*, Volume 39, 2000, Pages 1-21.
16. Jutten, C., and Herault, J., “Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture”, *Signal Processing,* volume, 24, July 1991, Pages 1-10.
17. Karhunen J. 1996. “Neural approaches to independent component analysis and source separation,” P*roc 4th European Symposium on Artificial Neural Networks (ESANN '96)*.
18. Hyvärinen, A. and Oja, E. “Independent Component Analysis by General Nonlinear Hebbian-like Learning Rules.” *Signal Processing,* Volume 64, Number 3, 1998, Pages 301-313.
19. Hyvärinen, A. and Pajunen, P. “Nonlinear Independent Component Analysis: Existence and Uniqueness results,” *Neural Networks,* volume 3, number 12, 1999, Pages 429-439.
20. Joutsensalo, J. “Semi-Blind Source Parameter Separation” *Proc. Int. Conf. on Artificial Neural Networks, (ICANN'97)*, Lausanne, Switzerland, October 1997, Pages 577-582.
21. Joutsensalo, J. and P. Pajunen, P., “Blind Symbol Learning Algorithms for CDMA System” *Proc. Int. Joint Conf. on Neural Networks, (IJCNN'98),* Anchorage, Alaska, May 4-9, 1998, volume 1, Pages 152-155.
22. Cristescu, R., Ristaniemi, T., Joutsensalo, J., and Karhunen, J., “Blind Separation of Convolved Mixtures for CDMA Systems.” In *Proc. of the X European Signal Processing Conference (EUSIPCO 2000)*, Tampere, Finland, September 5-8, 2000, Pages 619-622.
23. Pajunen, P., and Karhunen, J., “Least-Squares Methods for Blind Source Separation Based on Nonlinear PCA.” *Int. Journal of Neural Systems*, Volume 8, No. 5/6, 1998, Pages 601-612.
24. Oja, E. and Karhunen, J. “Signal separation by nonlinear Hebbian learning”, in *Computational Intelligence – A dynamic System Perspective,* ed. M. Palaniswami, IEEE Press, 1995.
25. Karhunen, J., Hyvärinen, A., Vigario, R., Hurri, J, and Oja, E., “Applications of Neural Blind Source Separation to Signal and Image Processing.” *In Proc. IEEE 1997 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97)*, April 21 - 24, 1997, Munich, Germany, Pages 131-134.
26. Shun-ichi Amari, “Natural Gradient Works Efficiently in Learning”, *Neural Computation*, Vol. 10, 1998, Pages 251-276.
27. Knuth, K., H., “A Bayesian approach to source separation” *ICA’99 Proceedings*, Aussois, France, Jan. 1999, Pages 283-288.