

# Maximum Likelihood Estimation of Constellation Vectors for Blind Separation of Co-Channel BPSK Signals and Its Performance Analysis

Anand Kannan and V. U. Reddy, *Senior Member, IEEE*

**Abstract**—In this paper, we present a method for blind separation of co-channel BPSK signals arriving at an antenna array. This method consists of two parts: the maximum likelihood constellation estimation and assignment. We show that at high SNR, the maximum likelihood constellation estimation is well approximated by the smallest distance clustering algorithm, which we proposed earlier on heuristic grounds. We observe that both these methods for estimating the constellation vectors perform very well at high SNR and nearly attain Cramér–Rao bounds. Using this fact and noting that the assignment algorithm causes negligible error at high SNR, we derive upper bounds on the probability of bit error for the above method at high SNR. These upper bounds fall very rapidly with increasing SNR, showing that our constellation estimation-assignment approach is very efficient. Simulation results are given to demonstrate the usefulness of the bounds.

## I. INTRODUCTION

CAPACITY improvement of wireless communication systems is a very important area of current research. The goal is to increase the number of users supported by the system per unit bandwidth allotted. One important way of achieving this improvement is to use multiple antennas backed by intelligent signal processing [2]. In this paper, we focus on a specific uplink capacity improvement scheme and analyze its performance.

Recently, we have proposed a method [1] for blind separation of co-channel users' signals arriving at an antenna array. This method is based on clustering and assignment. The clustering algorithm in [1] is self starting, uses each data vector only once and works very well at moderately high SNR. However, its degradation with increasing level of noise is not graceful. In this paper, we present the maximum likelihood method of estimating the constellation points that, together with the assignment algorithm of [1], gives a blind separation method. We then show that at high SNR, with suitable initialization, it is approximated well by the clustering algorithm given in [1]. We will refer to the clustering algorithm in [1] as *smallest distance clustering* from now on. Using

Manuscript received July 26, 1995; revised December 20, 1996. The associate editor coordinating the review of this paper and approving it for publication was Dr. Petar M. Djuric.

A. Kannan was with Electrical Communication Engineering Department, Indian Institute of Science, Bangalore, India. He is now with Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA.

V. U. Reddy is with Electrical Communication Engineering Department, Indian Institute of Science, Bangalore, India.

Publisher Item Identifier S 1053-587X(97)04952-0.

the fact that these algorithms nearly attain the Cramér–Rao (CR) bounds at high SNR's, we derive upper bounds on the total probability of bit error. Simulation results agree with the predictions.

## II. MAXIMUM LIKELIHOOD ESTIMATION OF CONSTELLATION VECTORS

From the data model of [1] and [3], the output of the antenna array at each instant can be written as

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{w}(n) \quad (1)$$

where

$\mathbf{x}(n)$  complex  $m \times 1$  output of the  $m$ -element antenna array,

$\mathbf{s}(n)$   $d \times 1$  vector of the bits of the  $d$  cochannel users,

$\mathbf{A}$  unknown channel matrix,

$\mathbf{w}(n)$  complex white Gaussian noise with real and imaginary parts of each component being independent, each with variance  $\sigma^2$ .

If each user emits  $\pm 1$  with the same probabilities, then  $\mathbf{s}(n)$  takes one of  $2^d$  possible values  $\mathbf{s}_k, k = 1, 2, \dots, 2^d$ . Suppose  $\mathbf{s}(n) = \mathbf{s}_k$ . Then

$$f(\mathbf{x}|\mathbf{M}, \mathbf{s}(n) = \mathbf{s}_k) = \frac{1}{(2\pi\sigma^2)^m} e^{-\|\mathbf{x} - \boldsymbol{\mu}_k\|^2/2\sigma^2} \quad (2)$$

with

$$\mathbf{M} = [\boldsymbol{\mu}_1 \quad \boldsymbol{\mu}_2 \quad \cdots \quad \boldsymbol{\mu}_{2^d}] = \mathbf{A}\mathbf{S}_0$$

where

$$\mathbf{S}_0 = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots \quad \mathbf{s}_{2^d}]$$

is a  $d \times 2^d$  matrix with  $\pm 1$  elements containing all the  $2^d$  possible columns. Since  $P(\mathbf{s}(n) = \mathbf{s}_k) = 1/2^d$ , we get (with  $L = 2^d$ )

$$f(\mathbf{x}|\mathbf{M}) = \frac{1}{L(2\pi\sigma^2)^m} \sum_{k=1}^L e^{-\|\mathbf{x} - \boldsymbol{\mu}_k\|^2/2\sigma^2}. \quad (3)$$

Note that  $\boldsymbol{\mu}_k$  is an  $m \times 1$  complex vector. If we have  $N$  snapshots, then the log likelihood function is

$$\mathcal{L} = - \sum_{n=1}^N \log f(\mathbf{x}(n)|\mathbf{M}). \quad (4)$$

If we assume that  $\sigma^2$  is known, the ML estimates for  $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_L$ , based on the data  $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)$ , can be obtained by solving

$$\frac{\partial \mathcal{L}}{\partial \hat{\boldsymbol{\mu}}_i} = \mathbf{0}, \quad i = 1, \dots, L. \quad (5)$$

It can be easily verified<sup>1</sup> that this results in the following set of equations:

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{n=1}^N \mathbf{x}(n) g_i(\mathbf{x}(n) | \hat{\mathbf{M}})}{\sum_{n=1}^N g_i(\mathbf{x}(n) | \hat{\mathbf{M}})}, \quad i = 1, 2, \dots, L \quad (6)$$

where

$$g_i(\mathbf{x} | \hat{\mathbf{M}}) = \frac{e^{-\|\mathbf{x} - \hat{\boldsymbol{\mu}}_i\|^2 / 2\sigma^2}}{\sum_{k=1}^L e^{-\|\mathbf{x} - \hat{\boldsymbol{\mu}}_k\|^2 / 2\sigma^2}}. \quad (7)$$

Observe that  $g_i$  is dependent on all the  $\hat{\boldsymbol{\mu}}$ 's. In addition, note that  $\hat{\mathbf{M}} = [\hat{\boldsymbol{\mu}}_1 \ \hat{\boldsymbol{\mu}}_2 \ \dots \ \hat{\boldsymbol{\mu}}_L]$ .

Equation (6) resembles a fixed-point equation. An iterative form for the solution of these equations is given by

$$\hat{\boldsymbol{\mu}}_i(r+1) = \frac{\sum_{n=1}^N \mathbf{x}(n) g_i(\mathbf{x}(n) | \hat{\mathbf{M}}(r))}{\sum_{n=1}^N g_i(\mathbf{x}(n) | \hat{\mathbf{M}}(r))}, \quad i = 1, 2, \dots, L \quad (8)$$

where  $r$  denotes the iteration number. The statistical properties of the solution will have a dependence on the statistics of  $\mathbf{x}(n)$ , particularly on the noise variance  $\sigma^2$ . The numerical solution to this equation has been discussed in the literature [6], but it does not deal with certain issues of interest to us.

For  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_L \in \mathcal{C}^m$ , let

$$E_i = \{\mathbf{x} : \|\mathbf{x} - \boldsymbol{\mu}_i\| < \|\mathbf{x} - \boldsymbol{\mu}_j\|, j \neq i\}.$$

That is,  $E_i$  is the set of all points in  $\mathcal{C}^m$  closer to  $\boldsymbol{\mu}_i$  than to other  $\boldsymbol{\mu}$ 's. The  $E_i$ 's are disjoint sets that are also the Bayesian decision regions for the classification of the  $\mathbf{x}$  vectors [10]. Let  $\chi_i(\mathbf{x} | \mathbf{M})$  denote the indicator function of  $E_i$ .

$$\chi_i(\mathbf{x} | \mathbf{M}) = \begin{cases} 1 & \text{for } \mathbf{x} \in E_i \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

As  $\sigma \rightarrow 0$  (i.e., at high SNR), for  $\|\mathbf{x} - \boldsymbol{\mu}_i\| < \|\mathbf{x} - \boldsymbol{\mu}_j\|$

$$e^{-\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 / 2\sigma^2} \gg e^{-\|\mathbf{x} - \boldsymbol{\mu}_j\|^2 / 2\sigma^2} \quad \forall j \neq i. \quad (10)$$

Thus

$$g_i(\mathbf{x} | \mathbf{M}) \simeq \begin{cases} \frac{e^{-\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 / 2\sigma^2}}{e^{-\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 / 2\sigma^2}} = 1 & \text{if } \mathbf{x} \in E_i \\ \frac{e^{-\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 / 2\sigma^2}}{e^{-\|\mathbf{x} - \boldsymbol{\mu}_j\|^2 / 2\sigma^2}} \simeq 0 & \text{otherwise.} \end{cases} \quad (11)$$

<sup>1</sup>For  $\boldsymbol{\lambda} \in \mathcal{C}^m$  and  $h(\boldsymbol{\lambda})$  real,  $\partial h / \partial \boldsymbol{\lambda} = (\partial h / \partial \lambda_1, \partial h / \partial \lambda_2, \dots, \partial h / \partial \lambda_m)$ . The  $\lambda_i$ 's are complex and the differentiation is according to [12, Th. 1].

Hence

$$g_i(\mathbf{x} | \mathbf{M}) \simeq \chi_i(\mathbf{x} | \mathbf{M}). \quad (12)$$

Using the above approximation in conjunction with (6), we get

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{n=1}^N \mathbf{x}(n) \chi_i(\mathbf{x}(n) | \hat{\mathbf{M}})}{\sum_{n=1}^N \chi_i(\mathbf{x}(n) | \hat{\mathbf{M}})}. \quad (13)$$

The denominator is the number  $n^{(i)}$  of  $\mathbf{x}$  vectors that fall in  $\hat{E}_i$ , where

$$\hat{E}_i = \{\mathbf{x} : \|\mathbf{x} - \hat{\boldsymbol{\mu}}_i\| < \|\mathbf{x} - \hat{\boldsymbol{\mu}}_j\|, j \neq i\}.$$

Hence,  $\hat{\boldsymbol{\mu}}_i$  is the average of the  $\mathbf{x}(n)$ 's that fall in  $\hat{E}_i$ . This shows that at high SNR, the ML solution induces a partition of the data, and each of the  $\hat{\boldsymbol{\mu}}_i$ 's is the mean of the data vectors in a particular  $\hat{E}_i$ . Although this relation gives us some insight into the nature of the solution to (6) at high SNR, it does not tell us how to obtain the solution.

Some general properties of ML solutions hold for this problem as well. It has been shown [5] that the estimates of the  $\boldsymbol{\mu}_i$ 's are asymptotically consistent. That is, the set of  $\hat{\boldsymbol{\mu}}_i$ 's converges to the set of  $\boldsymbol{\mu}_i$ 's almost surely as the data size tends to infinity.

The numerical solution of (6) as a fixed-point iteration requires good starting estimates. We can get these starting estimates by the SD clustering algorithm. Other methods for starting have also been suggested [7]. Although the iterative method (8) suffers from wrong convergence problems at low SNR, these problems reduce substantially as the SNR increases.

Once the  $\boldsymbol{\mu}_i$ 's are estimated, the transmitted bits are determined using the assignment algorithm as described in [1]. The set of bits corresponding to a particular instant will be the  $d$ -tuple corresponding to the  $\boldsymbol{\mu}$  nearest the snapshot vector at that instant.

#### A. Approximation at High SNR

The ML estimate is dependent on the number of sample vectors  $\mathbf{x}(n)$ . This can be explicitly incorporated into the notation of (6) to get

$$\hat{\boldsymbol{\mu}}_i^{(N)} = \frac{\sum_{n=1}^N \mathbf{x}(n) g_i(\mathbf{x}(n) | \hat{\mathbf{M}}^{(N)})}{\sum_{n=1}^N g_i(\mathbf{x}(n) | \hat{\mathbf{M}}^{(N)}), \quad i = 1, 2, \dots, L \quad (14)$$

where

$$\hat{\mathbf{M}}^{(N)} = [\hat{\boldsymbol{\mu}}_1^{(N)} \ \hat{\boldsymbol{\mu}}_2^{(N)} \ \dots \ \hat{\boldsymbol{\mu}}_L^{(N)}].$$

For the case of  $(N - 1)$  sample vectors, we get

$$\hat{\boldsymbol{\mu}}_i^{(N-1)} = \frac{\sum_{n=1}^{N-1} \mathbf{x}(n)g_i(\mathbf{x}(n)|\hat{\mathbf{M}}^{(N-1)})}{\sum_{n=1}^{N-1} g_i(\mathbf{x}(n)|\hat{\mathbf{M}}^{(N-1)}), \quad i = 1, 2, \dots, L. \tag{15}$$

One can start the fixed-point iteration for (14) with the solution of (15) as the initial guess. For large  $N$ ,  $\hat{\boldsymbol{\mu}}_i^{(N-1)}$  will be a good starting estimate, and hence, we can perform the iteration [c.f. (8)] only once. We then obtain

$$\begin{aligned} \hat{\boldsymbol{\mu}}_i^{(N)} &= \frac{\sum_{n=1}^{N-1} \mathbf{x}(n)g_i(\mathbf{x}(n)|\hat{\mathbf{M}}^{(N-1)}) + \mathbf{x}(N)g_i(\mathbf{x}(N)|\hat{\mathbf{M}}^{(N-1)})}{\sum_{n=1}^{N-1} g_i(\mathbf{x}(n)|\hat{\mathbf{M}}^{(N-1)}) + g_i(\mathbf{x}(N)|\hat{\mathbf{M}}^{(N-1)})} \\ i &= 1, 2, \dots, L. \end{aligned} \tag{16}$$

From (12), (13), and (15)

$$\sum_{n=1}^N \mathbf{x}(n)g_i(\mathbf{x}(n)|\hat{\mathbf{M}}^{(N-1)}) = n^{(i)}\hat{\boldsymbol{\mu}}_i^{(N-1)}$$

and  $n^{(i)}$  is the number of  $\mathbf{x}(n)$  that belong to  $\hat{E}_i^{(N-1)}$ , where the superscript  $(N - 1)$  indicates that the set is based on  $\hat{\mathbf{M}}^{(N-1)}$ . Hence

$$\hat{\boldsymbol{\mu}}_i^{(N)} = \begin{cases} \frac{n^{(i)}\hat{\boldsymbol{\mu}}_i^{(N-1)}}{n^{(i)}\hat{\boldsymbol{\mu}}_i^{(N-1)} + \mathbf{x}(N)} = \hat{\boldsymbol{\mu}}_i^{(N-1)} & \mathbf{x}(N) \notin \hat{E}_i^{(N-1)} \\ \frac{n^{(i)}\hat{\boldsymbol{\mu}}_i^{(N-1)} + \mathbf{x}(N)}{n^{(i)} + 1} & \text{otherwise.} \end{cases} \tag{17}$$

This is exactly the “then” part of step 2 of the SD clustering algorithm [1]. It should be noted that the “else” part of the algorithm is used for suitable initialization and correction of possible earlier errors in allotting  $\mathbf{x}(n)$  vectors to different clusters. With increasing  $r$ , the “then” part of the loop is executed with increasingly greater frequency. Thus, the SD clustering algorithm [1] is a good recursive approximation to the ML estimation at high SNR.

The Cramér–Rao bound for the estimate of a parameter occurring in a density function gives the smallest possible variance of any unbiased estimate for that parameter. The covariance matrix of the estimates is lower bounded by the inverse of the Fisher information matrix  $\mathbf{J}$ , which can be approximated using (10) and (12) [2]

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} & \cdots & \mathbf{J}_{1L} \\ \mathbf{J}_{21} & \mathbf{J}_{22} & \cdots & \mathbf{J}_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{J}_{L1} & \mathbf{J}_{L2} & \cdots & \mathbf{J}_{LL} \end{pmatrix} \approx \mathbf{I}_{mL \times mL} / 2L\sigma^2$$

where each of these matrices

$$\mathbf{J}_{ij} = E \left[ \left( \frac{\partial \log f}{\partial \boldsymbol{\mu}_i} \right)^H \left( \frac{\partial \log f}{\partial \boldsymbol{\mu}_j} \right) \right].$$

Here, the superscript  $H$  denotes conjugate transpose. Thus, the smallest variance obtainable for each component of the estimate of  $\boldsymbol{\mu}_i$  is  $2\sigma^2 L/N$ . As the simulation results show, both the SD clustering and ML estimation attain variances close to the CR bound for high SNR.

### III. UPPER BOUND ON THE BIT ERROR PROBABILITY

With the results developed above, we can now derive expressions for upper bounds for probability of error at high SNR for the scheme presented in the previous section. As the error introduced by the assignment algorithm is negligible, all the errors are contributed by the estimation step. In the argument given below, we upper bound the probability of snapshot error for reasonably large  $N$  and high SNR’s. The total probability of bit error for all the users is, at most,  $d$  times this value.

The estimates of  $\boldsymbol{\mu}_i$ ’s induce a partition of  $\mathcal{C}^m$  into regions  $\hat{E}_1, \hat{E}_2, \dots, \hat{E}_L$ , where

$$\hat{E}_i = \{\mathbf{x}: \|\mathbf{x} - \hat{\boldsymbol{\mu}}_i\| < \|\mathbf{x} - \hat{\boldsymbol{\mu}}_j\|; j \neq i\}.$$

These  $\hat{E}_i$ ’s are estimates of the Bayesian decision regions  $E_i$ . If the problem had not been one of blind separation, the regions  $E_i$  would have been known exactly. Assuming equally probable bits for the  $d$  users, we have

$$P(\text{error}) = \frac{1}{L} \sum_{i=1}^L P(\text{error}|i) \tag{18}$$

where  $P(\text{error}|i)$  is the probability that the noiseless part of  $\mathbf{x}$  is the  $i$ th constellation vector  $\boldsymbol{\mu}_i$ , but the Bayesian classification rule classifies it erroneously. Since

$$P(\text{error}|i) = P(\mathbf{x} \in E_i^c|i) = 1 - P(\mathbf{x} \in E_i|i)$$

we can express (18) as

$$P(\text{error}) = \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in E_i^c|i) = 1 - \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in E_i|i). \tag{19}$$

Let

$$F_{ij} = \{\mathbf{x}: \|\mathbf{x} - \boldsymbol{\mu}_i\| < \|\mathbf{x} - \boldsymbol{\mu}_j\|\}.$$

Then

$$E_i^c = \bigcup_{j \neq i} F_{ij}.$$

Therefore

$$P(\text{error}|i) \leq \sum_{j \neq i} P(\mathbf{x} \in F_{ij}|i).$$

We can show that [8], [9]

$$P(\mathbf{x} \in F_{ij}|i) = Q\left(\frac{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|}{2\sigma}\right) \tag{20}$$

where

$$Q(\beta) = \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} e^{-z^2/2} dz$$

is the area under the tail of the Gaussian density function. Using this,  $P(\text{error})$  can be bounded as

$$P(\text{error}) \leq \frac{1}{L} \sum_{i=1}^L \sum_{j \neq i} Q\left(\frac{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|}{2\sigma}\right). \quad (21)$$

However, the problem here is one of blind separation. Therefore, the  $\boldsymbol{\mu}_i$ 's are not known and need to be estimated. Assuming that the estimation technique is good and attains the CR bound very nearly, the additional error because of "blindness" can now be obtained.

In the ML estimation, the  $\hat{\boldsymbol{\mu}}_i$ 's can be assumed distributed Gaussian with mean  $\boldsymbol{\mu}_i$  and variance per component  $2\sigma^2 L/N$ . The probability of error in the "blind separation" case can be expressed as

$$P_{bs}(\text{error}) = \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in \hat{E}_i^c | i) = 1 - \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in \hat{E}_i | i). \quad (22)$$

Combining (19) and (22)

$$\begin{aligned} P_{bs}(\text{error}) - P(\text{error}) &= \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in E_i | i) - \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in \hat{E}_i | i). \end{aligned} \quad (23)$$

Since

$$\begin{aligned} P(\mathbf{x} \in E_i | i) + P(\mathbf{x} \in \hat{E}_i - E_i | i) \\ = P(\mathbf{x} \in \hat{E}_i | i) + P(\mathbf{x} \in E_i - \hat{E}_i | i) \end{aligned} \quad (24)$$

it follows that

$$P_{bs}(\text{error}) - P(\text{error}) \leq \frac{1}{L} \sum_{i=1}^L P(\mathbf{x} \in E_i - \hat{E}_i | i). \quad (25)$$

Note that the expression on the RHS of (25) is again difficult to evaluate but can be approximately upperbounded by a sum of pairwise error terms.

Let

$$\hat{F}_{ij} = \{\mathbf{x} : \|\mathbf{x} - \hat{\boldsymbol{\mu}}_i\| < \|\mathbf{x} - \hat{\boldsymbol{\mu}}_j\|\}.$$

Then

$$E_i - \hat{E}_i \subseteq \bigcup_{j \neq i} (F_{ij} - \hat{F}_{ij}).$$

Hence

$$P(\mathbf{x} \in E_i - \hat{E}_i | i) \leq \sum_{j \neq i} P(\mathbf{x} \in F_{ij} - \hat{F}_{ij} | i). \quad (26)$$

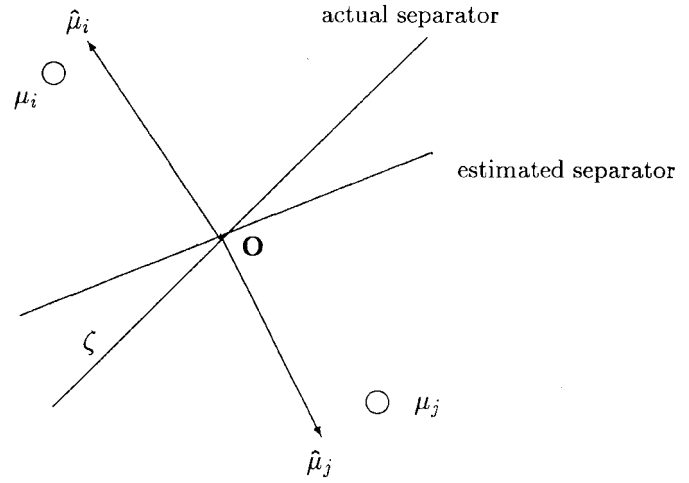


Fig. 1. Hyperplanes separating  $F_{ij}$  and  $F_{ji}$ .

The separator between the regions  $F_{ij}$  and  $F_{ji}$  is the hyperplane orthogonal to  $\boldsymbol{\mu}_i - \boldsymbol{\mu}_j$ . Fig. 1 refers to this as the actual separator. Similarly, the hyperplane that separates  $\hat{F}_{ij}$  and  $\hat{F}_{ji}$  is referred to as the estimated separator in Fig. 1.  $P(\mathbf{x} \in F_{ij} - \hat{F}_{ij} | i)$  is now the probability of that region sandwiched between these two hyperplanes and marked  $\zeta$ . This has to be calculated using the Gaussian density with mean  $\boldsymbol{\mu}_i$  and covariance  $2\sigma^2 I$ . If  $\|\boldsymbol{\mu}_i\|$ ,  $\|\boldsymbol{\mu}_j\|$  and  $\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|$  are  $\gg \sigma$ , which is true at high SNR, this probability will be approximately<sup>2</sup>

$$\begin{aligned} P(\mathbf{x} \in F_{ij} - \hat{F}_{ij} | i) &\leq \left( \begin{array}{l} \text{An upper bound on} \\ \text{the integral of pdf} \\ \text{over } F_{ij} - \hat{F}_{ij} \end{array} \right) \\ &\times \left( \begin{array}{l} \text{Relative probability} \\ \text{of } F_{ij} - \hat{F}_{ij} \text{ in } \mathcal{C}^m. \end{array} \right) \end{aligned} \quad (27)$$

Hence

$$\begin{aligned} P(\mathbf{x} \in F_{ij} - \hat{F}_{ij} | i) &\approx e^{-\|(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)/2 - \boldsymbol{\mu}_i\|^2 / 2\sigma^2} \cdot \frac{\alpha}{2\pi} \\ &\approx e^{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 / 8\sigma^2} \cdot \frac{E\alpha}{2\pi} \end{aligned} \quad (28)$$

where  $\alpha$  is the angle<sup>3</sup> between the vectors  $\boldsymbol{\mu}_i - \boldsymbol{\mu}_j$  and  $\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j$ . The angle  $\alpha$  is a random variable that depends on  $\hat{\boldsymbol{\mu}}_i$  and  $\hat{\boldsymbol{\mu}}_j$ . However, its expected value can be calculated by (29), shown at the bottom of the page. This uses the fact that  $\alpha$  is small compared with  $2\pi$  since the variances of the estimates of the components of  $\boldsymbol{\mu}_i$  are small, and for such  $\alpha$ ,  $\tan \alpha \approx \alpha$ . Hence

$$|E\alpha|^2 \leq \frac{(2m-1)(2\sigma^2)L}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 N}. \quad (30)$$

<sup>2</sup>In the same spirit that  $\int_{-\infty}^{\infty} \exp(-x^2/2) dx \leq \sqrt{2\pi} \exp(-\alpha^2/2)$  for  $\alpha \geq 0$ .

<sup>3</sup>The angle  $\alpha$  between two vectors  $\boldsymbol{\lambda}_1$  and  $\boldsymbol{\lambda}_2 \in \mathcal{C}^m$  is given by  $\cos \alpha = \text{Re}(\boldsymbol{\lambda}_1^H \boldsymbol{\lambda}_2) / \|\boldsymbol{\lambda}_1\| \|\boldsymbol{\lambda}_2\|$ .

$$|E\alpha|^2 \leq E \tan^2 \alpha \approx \frac{\text{sum of variances of components of } (\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j) \perp (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\text{squared length of } (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)} \quad (29)$$

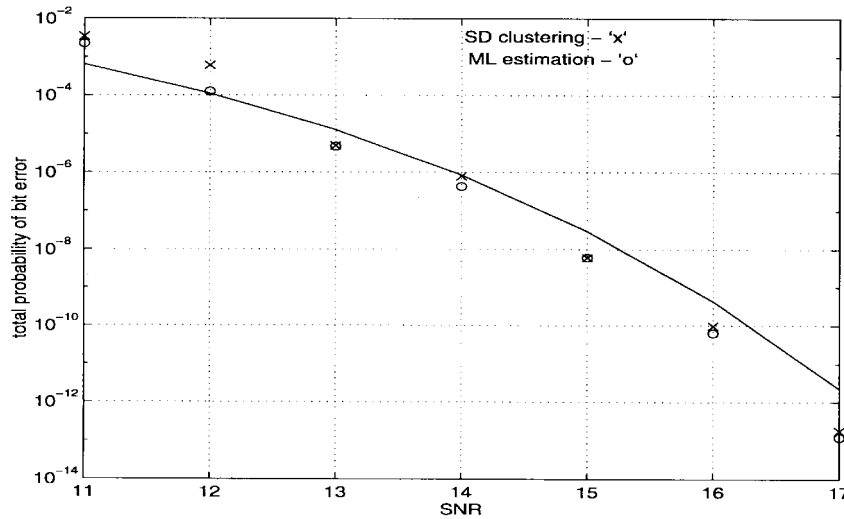


Fig. 2. Simulation results and upper bound (solid line) on total expected bit error rate  $P_B$ .

From (25), (26), and (30)

$$P_{bs}(\text{error}) - P(\text{error}) \leq \frac{1}{L} \sum_{i=1}^L \sum_{j \neq i} e^{-\|\mu_i - \mu_j\|^2 / 8\sigma^2} \frac{\sqrt{(2m-1)L\sigma}}{\pi\sqrt{2N}\|\mu_i - \mu_j\|}. \quad (31)$$

Thus, the probability of snapshot error is upper bounded by

$$P_{bs}(\text{error}) \leq \frac{1}{L} \left( \sum_{i=1}^L \sum_{j \neq i} e^{-\|\mu_i - \mu_j\|^2 / 8\sigma^2} \cdot \frac{\sqrt{(2m-1)L\sigma}}{\pi\sqrt{2N}\|\mu_i - \mu_j\|} + \sum_{i=1}^L \sum_{j \neq i} \cdot Q\left(\frac{\|\mu_i - \mu_j\|}{2\sigma}\right) \right). \quad (32)$$

Since the total expected bit error rate  $P_B$  is at most  $d$  times the probability of snapshot error, we have

$$P_B \leq \frac{d}{L} \left( \sum_{i=1}^L \sum_{j \neq i} e^{-\|\mu_i - \mu_j\|^2 / 8\sigma^2} \frac{\sqrt{(2m-1)L\sigma}}{\pi\sqrt{2N}\|\mu_i - \mu_j\|} + \sum_{i=1}^L \sum_{j \neq i} Q\left(\frac{\|\mu_i - \mu_j\|}{2\sigma}\right) \right). \quad (33)$$

Since SD clustering algorithm performs very much like ML estimation at high SNR, this upper bound holds for the method of [1] as well at high SNR's.

One can replace the ML estimation by any other procedure that nearly attains the CR bounds for the estimation error in the  $\mu_i$ 's, and the above upper bound on the probability of bit error will still hold for that procedure.

#### IV. SIMULATION RESULTS

In our simulations, we used a uniform linear array (ULA) of five elements spaced  $\lambda/2$  apart and assumed three users located at  $-10^\circ, 3^\circ$  and  $10^\circ$ , respectively, from the array

broadside, which emit co-channel BPSK signals. The signal powers arriving at the array were assumed to be 1. The noise variance  $\sigma^2$  was varied to give the desired SNR [which we define as  $20\log(1/\sigma)$ ] over the required range. In the ML estimation case, we initialized the iteration (8) with the output of the SD clustering algorithm [1].

Fig. 2 shows the performance of the SD clustering and the ML estimation methods as compared with the theoretical upper bound given by (33). It is seen that as SNR increases, the total expected bit error rate for both SD clustering assignment and ML estimation assignment methods fall rapidly to values below the upper bound. In addition, they perform nearly the same at higher SNR's. The upper bound falls rapidly with the SNR, dropping nearly an order of magnitude with every decibel increase in the SNR. In the simulations, we used 200 snapshots ( $N = 200$ ) in every trial and averaged the results over 500 trials.

We note from Fig. 2 that violations of the upper bound occur at lower SNR's. This is because the theoretical limit has been derived assuming good behavior of the estimators, which is true only at high SNR's. Further, the SD clustering algorithm gives large errors because of its sensitivity to outliers at these lower SNR's. On the other hand, the ML method does not converge to correct values at lower SNR because of lack of suitable starting values. However, with better initialization, the ML method gives more accurate estimates. This may be seen from the results of the last two rows of Table I, which show the sample variance of  $\hat{\mu}$  vectors for the ML estimation method initialized with the output of the SD clustering and the same method initialized with the actual  $\mu$  vectors. The sample variance is given by

$$\text{Sample variance} = \frac{1}{L} \sum_{j=1}^L \frac{1}{N-1} \sum_{i=1}^N \|\hat{\mu}_j(i) - \bar{\mu}_j\|^2 \quad (34)$$

where

$$\bar{\mu}_j = \frac{1}{N} \sum_{k=1}^N \hat{\mu}_j(k)$$

TABLE I  
SAMPLE VARIANCE OF  $\hat{\mu}$  VECTORS AT DIFFERENT VALUES OF SNR ( $N = 200$ )

SNR	11	12	13	14	15	16	17
CR bound	0.0318	0.0252	0.0200	0.0159	0.0126	0.0100	0.0080
SD clustering [1]	0.0556	0.0275	0.0208	0.0165	0.0131	0.0103	0.0084
ML estimation	0.0498	0.0262	0.0208	0.0165	0.0131	0.0103	0.0084
ML with true initialization	0.0327	0.0262	0.0208	0.0165	0.0131	0.0103	0.0084

TABLE II  
SAMPLE VARIANCE OF  $\hat{\mu}$  VECTORS FOR  
DIFFERENT VALUES OF  $N$  AT SNR = 12 dB

$N$	100	200	500	1000
CR bound	0.0505	0.0252	0.0101	0.0050
SD clustering [1]	0.0565	0.0274	0.0109	0.0051
ML estimation	0.0550	0.0262	0.0100	0.0051
ML with true initialization	0.0550	0.0262	0.0100	0.0051

and  $\mathcal{N}$  denotes the number of trials. The CR bound on the variance ( $2\sigma^2 Lm/N$ ) is also given in the table for comparison. Observe that the CR bound is very nearly attained by the ML estimation and SD clustering methods, at high SNR.

Table II gives the sample variance of the  $\hat{\mu}$  at SNR = 12 dB for different values of  $N$ . It can be seen that the performance of the ML estimation algorithm in the lower SNR range improves with increasing  $N$ . On the other hand, the SD clustering shows slower improvement with increasing  $N$ . This is because bad initialization matters less for the convergence of ML iterations (8) when more data is present. However, SD clustering is a single-shot algorithm, and improvement with increasing  $N$  is only because of more averaging done to get the cluster vectors.

## V. CONCLUSION

In this paper, we have presented a method for blind separation of BPSK signals arriving at an antenna array, explored its relationship to an earlier method, and derived upper bounds on the bit error probability assuming high SNR. This upper bound falls rapidly with increasing SNR. The methods proposed here can be used to increase the uplink capacity of a narrowband digital cellular communication system. The methods consist of two parts: constellation estimation and assignment. The constellation estimation part of the methods can be easily extended to other digital modulation schemes. However, the assignment part is specific to BPSK since it exploits the structure of the BPSK signals. Reference [4] presents the assignment algorithm for M-ary PSK signals.

## REFERENCES

- [1] K. Anand, G. Mathew, and V. U. Reddy, "Blind separation of multiple co-channel BPSK signals arriving at an antenna array," *IEEE Signal Processing Lett.*, pp. 176–178, Sept. 1995.
- [2] K. Anand, "Methods for blind separation of co-channel bpsk signals arriving at an antenna array and their performance analysis," M.Sc.(Engg.) thesis, Indian Inst. Sci., Bangalore, July 1995.
- [3] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Lett.*, vol. 1, pp. 29–31, Feb. 1994.
- [4] B. V. Chandra and V. U. Reddy, "Blind separation of multiple co-channel M-ary PSK signals arriving at an antenna array," submitted for publication.

- [5] R. Redner, "Note on the consistency of the maximum likelihood estimate for non-identifiable distributions," *Ann. Statist.*, vol. 9, pp. 225–228, 1981.
- [6] B. C. Peters, Jr. and H. F. Walker, "An iterative procedure for obtaining maximum likelihood estimates of the parameters for a mixture of normal distributions," *SIAM J. Appl. Math.*, vol. 35, pp. 362–378, 1978.
- [7] B. C. Ng, S. Talwar, and A. Paulraj, "An unconditional maximum likelihood approach for blind estimation of digital signals," submitted for publication.
- [8] J. M. Wozencraft and I. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [9] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1989.
- [10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [11] N. Seshadri, AT&T Bell Labs, personal communication.
- [12] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elec. Eng.*, Pts. F and H, vol. 130, no. 1, p. 11, Feb. 1983.



**Anand Kannan** was born in Kumbakonam, India, in 1972. He received the B.Tech. degree from the Indian Institute of Technology, Chennai, in 1993 and the M.Sc.(Engg.) degree from the Indian Institute of Science, Bangalore, in 1995. He is currently working toward the Ph.D degree at Purdue University, West Lafayette, IN.

He is a research intern at ArrayComm, Inc., San Jose, CA, for the summer of 1997. His primary research interests are in adaptive signal and array processing with applications to wireless communication.

**V. U. Reddy** (SM'82) received the B.E. and M.Tech. degrees in electronics and communication engineering from Osmania University, Hyderabad, India, and the Indian Institute of Technology (IIT), Kharagpur, India, in 1962 and 1963, respectively, and the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, in 1971.

He was an Assistant Professor at IIT, Madras, from 1972 to 1976 and Professor at IIT, Kharagpur, from 1976 to 1979. From 1979 to 1982, he was a Visiting Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA. In April 1982, he joined Osmania University as a Professor and was the Founder-Director of the Research and Training Unit for Navigational Electronics, which was funded by the Department of Electronics, Government of India. Since April 1988, he has been with the Indian Institute of Science, Bangalore, as a Professor of Electrical Communication Engineering (he was the Chair of the Department from March 1992 to April 1995). He held visiting appointments with the Electrical Engineering Department, Stanford University, from 1986 to 1987, March to June 1994, and March to September 1996; he also visited Research Centre Imarat, Hyderabad, from September 1995 to February 1996. He served as a consultant in the signal processing area to several research and development laboratories in Hyderabad and Bangalore. His research interests are in adaptive algorithms, sensor array signal processing, multirate filtering and wavelets, and space-time signal processing in wireless communications.

Dr. Reddy is a Fellow of the Indian Academy of Sciences, the Indian National Academy of Engineering, the Indian National Science Academy, and the Institution of Electronics and Telecommunication Engineers (IETE), India. He received the S.K. Mitra Memorial Award from IETE in 1989 for the best research paper.