Adaptive Minimum Bit Error Rate Beamforming

S. Chen, N.N. Ahmad and L. Hanzo

Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

Abstract

An adaptive beamforming technique is proposed based on directly minimizing the bit error rate. It is demonstrated that this minimum bit error rate (MBER) approach utilizes the antenna array elements more intelligently, than the standard minimum mean square error (MMSE) approach. Consequently, MBER beamforming is capable of providing significant performance gains in terms of a reduced bit error rate over MMSE beamforming. A block-data adaptive implementation of the MBER beamforming solution is developed based on the Parzen window estimate of probability density function. Furthermore, a sample-by-sample adaptive implementation is considered, and a stochastic gradient algorithm, referred to as the least bit error rate, is derived. The proposed adaptive MBER beamforming technique provides an extension to the existing work [18],[19] for adaptive MBER equalization and multiuser detection.

Keywords: Smart antenna, adaptive beamforming, mean square error, bit error rate, probability density function

1 Introduction

The ever-increasing demand for mobile communication capacity has motivated the employment of space division multiple access for the sake of improving the achievable spectral efficiency. A particular approach that has shown real promise in achieving substantial capacity enhancements is the use of adaptive antenna arrays [1]–[10]. Adaptive beamforming is capable of separating signals transmitted on the same carrier frequency, provided that they are separated in the spatial domain. A beamformer appropriately combines the signals received by the different elements of an antenna array to form a single output. Classically, this has been achieved by minimizing the mean square error (MSE) between the desired output and the actual array output. This principle has its roots in the traditional beamforming employed in sonar and radar systems. Adaptive implementation of the minimum MSE (MMSE) beamforming solution can be realized using temporal reference techniques [2]–[4],[11]–[14]. Specifically, block based beamformer weight adaptation can be achieved for example using the sample matrix inversion (SMI) algorithm [11],[12], while sample-by-sample adaptation can be carried out using the least mean square (LMS) algorithm [13],[14].

For a communication system, it is the achievable bit error rate (BER), not the MSE performance, that really matters. Ideally, the system design should be based directly on minimizing the BER, rather than the MSE. For applications to single-user channel equalization and multiuser detection, it has been shown that the MMSE solution can in certain situations be distinctly inferior in comparison to the minimum BER (MBER) solution, and several adaptive implementations of the MBER solution have been studied in the literature [15]–[19]. This contribution derives a novel adaptive beamforming technique based on directly minimizing the system's BER, rather than the MSE. For the sake of notational simplicity and for highlighting the basic concepts, the modulation scheme is assumed to be binary phase shift keying (BPSK), and the channel is assumed to be non-dispersive with additive Gaussian noise, which does not induce any intersymbol interference (ISI). Furthermore, narrow-band beamforming is considered in this paper. It is demonstrated that the MBER solution utilizes the array weights more intelligently than the MMSE approach. The MBER beamforming appears to be "smarter" than the MMSE solution, since it directly optimizes the system's BER performance, rather than minimizing the MSE, where the latter strategy often turns out to be deficient.

An adaptive implementation of the MBER beamforming technique is investigated in this paper. The classic Parzen window or kernel density estimation technique [20]–[22] is adopted for approximating the probability density function (p.d.f.) of the beamformer's output and a block-data adaptive MBER algorithm is developed, which iteratively minimizes the estimated BER of the beamformer by adjusting the beamformer weights using a simplified conjugate gradient optimization method [23],[19]. Our simulation study shows that this block-data adaptive MBER algorithm converges rapidly, and the length of the data block required for achieving an accurate approximation of the MBER solution is reasonably small. Sample-by-sample adaptation is also considered and an adaptive stochastic gradient MBER algorithm, referred to as the least bit error rate (LBER) techniques, is derived. This LBER algorithm involves more approximations compared with the one given in [18],[19] for equalization and multiuser detection, but it has a significantly lower computational complexity, which is comparable to that of the simple least mean square (LMS) algorithm. Simulation results suggest that the proposed simplified LBER algorithm has a similar performance to the full-complexity LBER algorithm of [18],[19] in terms of its convergence rate and steady-state BER misadjustment.

2 System Model

It is assumed that the system supports M users (signal sources), and each user transmits a BPSK modulated signal on the same carrier frequency of $\omega = 2\pi f$. Let k denote the bit instance. Then the baseband signal of user i is formulated as:

$$m_i(k) = A_i b_i(k), \;\; 1 \le i \le M, \tag{1}$$

where $b_i(k)$ takes value from the set $\{\pm 1\}$ with equal probability and A_i^2 denotes the signal power of user i. Without loss of generality, source 1 is assumed to be the desired user and the rest of the sources are the interfering users. The linear antenna array considered consists of L uniformly spaced elements, and the signals received by the L-element antenna array are given by:

$$x_l(k) = \sum_{i=1}^{M} m_i(k) \exp\left(j\omega t_l(\theta_i)\right) + n_l(k) = \bar{x}_l(k) + n_l(k), \;\; 1 \le l \le L, \tag{2}$$

where $t_l(\theta_i)$ is the relative time delay at array element l for source i, θ_i is the direction of arrival for source i, and $n_l(k)$ is the complex-valued white Gaussian noise having a zero mean and a variance of $E[|n_l(k)|^2] = 2\sigma_n^2$. The desired user's signal to noise ratio is defined as SNR$= A_1^2/2\sigma_n^2$, and the desired signal to interference ratio with respect to user i is defined as SIR$_i = A_1^2/A_i^2$, for $i = 2, \cdots, M$. In vectorial form, the array input $\mathbf{x}(k) = [x_1(k)\; x_2(k) \cdots x_L(k)]^T$ can be expressed as:

$$\mathbf{x}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k) = \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k), \tag{3}$$

where $\mathbf{n}(k) = [n_1(k)\; n_2(k) \cdots n_L(k)]^T$ has a covariance matrix of $E[\mathbf{n}(k)\mathbf{n}^H(k)] = 2\sigma_n^2 \mathbf{I}_L$ with $\mathbf{I}_L$ representing the $L \times L$ identity matrix, the system matrix $\mathbf{P}$ is given by:

$$\mathbf{P} = [A_1\mathbf{s}_1\; A_2\mathbf{s}_2 \cdots A_M\mathbf{s}_M], \tag{4}$$

the steering vector for source i is formulated as:

$$\mathbf{s}_i = [\exp(j\omega t_1(\theta_i))\;\; \exp(j\omega t_2(\theta_i)) \cdots \exp(j\omega t_L(\theta_i))]^T \tag{5}$$

and the transmitted bit vector is $\mathbf{b}(k) = [b_1(k)\; b_2(k) \cdots b_M(k)]^T$.

The beamformer's output is given by:

$$y(k;\mathbf{w}) = \mathbf{w}^H\mathbf{x}(k) = \mathbf{w}^H\bar{\mathbf{x}}(k) + \mathbf{w}^H\mathbf{n}(k) = \bar{y}(k;\mathbf{w}) + e(k;\mathbf{w}), \tag{6}$$

where $\mathbf{w} = [w_1\, w_2 \cdots w_L]^T$ is the complex-valued beamformer weight vector, and $e(k;\mathbf{w})$ is Gaussian distributed having a zero mean and a variance of $E[|e(k;\mathbf{w})|^2] = 2\sigma_n^2\mathbf{w}^H\mathbf{w}$. The estimate of the transmitted bit $b_1(k)$ is given by:

$$\hat{b}_1(k) = \begin{cases} +1, & y_R(k;\mathbf{w}) > 0, \\ -1, & y_R(k;\mathbf{w}) \le 0, \end{cases} \tag{7}$$

where $y_R(k;\mathbf{w}) = \Re[y(k;\mathbf{w})]$ denotes the real part of $y(k;\mathbf{w})$. Classically, the beamformer's weight vector is determined by minimizing the MSE term of $E[|b_1(k) - y(k;\mathbf{w})|^2]$, which leads to the following MMSE solution:

$$\mathbf{w}_{\mathrm{MMSE}} = \left(\mathbf{P}\mathbf{P}^H + 2\sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{p}_1, \tag{8}$$

with $\mathbf{p}_1$ being the first column of $\mathbf{P}$. Although the system matrix $\mathbf{P}$ is generally unknown, the MMSE solution can be readily realized using the block-data based adaptive SMI algorithm [11],[12]. The MMSE solution can also be implemented using the stochastic gradient algorithm known also as the LMS algorithm.

The Discrete Fourier Transform (DFT) of the beamformer weights, which is also referred to as the beam pattern is given by:

$$F(\theta) = \sum_{l=1}^{L} w_l \exp\left(-j\omega_l(\theta)\right), \tag{9}$$

which describes the response of the beamformer to the source arriving at angle θ. In traditional beamforming, the magnitude of $F(\theta)$ is used for characterizing the performance of a beamformer. Using the amplitude response alone, however, can be misleading, since both the magnitude and phase of $F(\theta)$ should be used together for characterizing the beamformer. Ultimately, it is the p.d.f. of the beamformer's output, which fully characterizes the true performance of the beamformer.

3 MBER Beamforming Solution

Denote the $N_b = 2^M$ number of possible transmitted bit sequences of $\mathbf{b}(k)$ as $\mathbf{b}_q$, $1 \le q \le N_b$. Further denote the first element of $\mathbf{b}_q$, corresponding to the desired user, as $b_{q,1}$. The array input signal $\bar{\mathbf{x}}(k)$ one takes values from the signal set defined as:

$$\mathcal{X} \stackrel{\triangle}{=} \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q,\ 1 \le q \le N_b\}. \tag{10}$$

This set can be partitioned into two subsets depending on the specific value of $b_1(k)$, as follows:

$$\mathcal{X}^{(\pm)} \stackrel{\triangle}{=} \{\bar{\mathbf{x}}_q^{(\pm)} \in \mathcal{X} :\ b_1(k) = \pm 1\}. \tag{11}$$

Similarly, the beamformer's output $\bar{y}(k;\mathbf{w})$ takes values from the scalar set $\mathcal{Y}(\mathbf{w}) \stackrel{\triangle}{=} \{\bar{y}_q(\mathbf{w}) = \mathbf{w}^H\bar{\mathbf{x}}_q,\ 1 \le q \le N_b\}$. Thus, the real part of the beamformer's output $\bar{y}_R(k;\mathbf{w})$ can only take values from the set:

$$\mathcal{Y}_R(\mathbf{w}) \stackrel{\triangle}{=} \{\bar{y}_{R,q}(\mathbf{w}) = \Re[\bar{y}_q(\mathbf{w})],\ \ 1 \le q \le N_b\}, \tag{12}$$

which can be divided into two subsets conditioned on $b_1(k)$ as follows:

$$\mathcal{Y}_R^{(\pm)}(\mathbf{w}) \stackrel{\triangle}{=} \{\bar{y}_{R,q}^{(\pm)}(\mathbf{w}) \in \mathcal{Y}_R(\mathbf{w}) :\ b_1(k) = \pm 1\}. \tag{13}$$

Note that the term beamforming here in fact refers to linear beamforming. An implicit assumption is that $\mathcal{X}^{(+)}$ and $\mathcal{X}^{(-)}$ are linearly separable, that is, there exists a weight vector $\mathbf{w}$ such that the two scalar sets $\mathcal{Y}_R^{(+)}(\mathbf{w})$ and $\mathcal{Y}_R^{(-)}(\mathbf{w})$ are completely separable by a linear decision boundary. Otherwise nonlinear beamforming is required, a situation that is similar to non-linear single-user equalization and non-linear multiuser detection [24]–[26].

It can be readily shown that the conditional p.d.f. of $y_R(k;\mathbf{w})$ given $b_1(k) = +1$ is:

$$p(y_R;\mathbf{w}|+1) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} \frac{1}{\sqrt{2\pi\sigma_n^2 \mathbf{w}^H\mathbf{w}}} \exp\left(-\frac{\left(y_R - \bar{y}_{R,q}^{(+)}(\mathbf{w})\right)^2}{2\sigma_n^2 \mathbf{w}^H\mathbf{w}}\right) \tag{14}$$

where $\bar{y}_{R,q}^{(+)}(\mathbf{w}) \in \mathcal{Y}_R^{(+)}(\mathbf{w})$ and $N_{sb} = N_b/2$ is the number of the points in $\mathcal{Y}_R^{(+)}(\mathbf{w})$. Thus, it can be shown that the BER of the beamformer associated with the weight vector $\mathbf{w}$ is given by:

$$P_E(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q\left(g_{q,+}(\mathbf{w})\right), \tag{15}$$

where

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \tag{16}$$

and

$$g_{q,+}(\mathbf{w}) = \frac{+\bar{y}_{R,q}^{(+)}(\mathbf{w})}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} = \frac{\mathrm{sgn}(b_{q,1})\bar{y}_{R,q}^{(+)}(\mathbf{w})}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} = \frac{\mathrm{sgn}(b_{q,1})\Re[\mathbf{w}^H\bar{\mathbf{x}}_q^{(+)}]}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}. \tag{17}$$

Similarly, the BER can be calculated using $\mathcal{Y}_R^{(-)}(\mathbf{w})$. The MBER beamforming solution is then defined as:

$$\mathbf{w}_{\mathrm{MBER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w}). \tag{18}$$

The gradient of $P_E(\mathbf{w})$ (15) with respect to $\mathbf{w}$ can be shown to be:

$$\nabla P_E(\mathbf{w}) = \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}(\mathbf{w})\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right) \mathrm{sgn}(b_{q,1}) \left(\frac{\bar{y}_{R,q}^{(+)}(\mathbf{w})\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \bar{\mathbf{x}}_q^{(+)}\right). \tag{19}$$

Given the gradient of (19), the optimization problem (18) can be solved by iteratively using a gradient-based optimization algorithm. Since the BER is invariant to a positive scaling of $\mathbf{w}$, it is computationally advantageous to normalize $\mathbf{w}$ to a unit-length after every iteration, so that the gradient can be simplified to:

$$\nabla P_E(\mathbf{w}) = \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}(\mathbf{w})\right)^2}{2\sigma_n^2}\right) \mathrm{sgn}(b_{q,1}) \left(\bar{y}_{R,q}^{(+)}(\mathbf{w})\mathbf{w} - \bar{\mathbf{x}}_q^{(+)}\right). \tag{20}$$

The following simplified conjugate gradient algorithm [23],[19] provides an efficient means of finding a MBER solution:

Initialization. Choose a step size of $\mu > 0$ and a termination scalar of $\beta > 0$ (Typically, β can be set to the machine accuracy.); given $\mathbf{w}(1)$ and $\mathbf{d}(1) = -\nabla P_E(\mathbf{w}(1))$; set the iteration index to $\iota = 1$.

Loop. If $\|\nabla P_E(\mathbf{w}(\iota))\| = \sqrt{(\nabla P_E(\mathbf{w}(\iota)))^H \nabla P_E(\mathbf{w}(\iota))} < \beta$: goto *Stop*. Else,

$$\mathbf{w}(\iota+1) = \mathbf{w}(\iota) + \mu\mathbf{d}(\iota)$$

$$\mathbf{w}(\iota+1) = \frac{\mathbf{w}(\iota+1)}{\|\mathbf{w}(\iota+1)\|}$$

$$\phi_\iota = \frac{\|\nabla P_E(\mathbf{w}(\iota+1))\|^2}{\|\nabla P_E(\mathbf{w}(\iota))\|^2}$$
$$\mathbf{d}(\iota+1) = \phi_\iota \mathbf{d}(\iota) - \nabla P_E(\mathbf{w}(\iota+1))$$

$\iota = \iota + 1$, goto *Loop*.

Stop. $\mathbf{w}(\iota)$ is the solution.

The step size μ controls the rate of convergence. Typically, a much larger value of μ can be used compared to the steepest-descent gradient algorithm. Occasionally the search direction $\mathbf{d}$ in the above conjugate gradient algorithm may no longer be a good approximation to the conjugate gradient direction or may even point to the 'uphill' direction, when the iteration index becomes high. A standard measure to prevent this situation from happening is to periodically reset $\mathbf{d}$ to the negative gradient [23]. With a reseting of every iteration, this algorithm reduces to the steepest-descent gradient algorithm.

Unlike the MMSE solution (8), there exists no closed-form MBER solution. In theory, there is no guarantee that the above conjugate gradient algorithm can always find a global minimum point of the BER surface $P_E(\mathbf{w})$. In practice, we have found that the algorithm works well and we have never observed any occurrence of the algorithm being trapped at some local minimum solution. This is likely to be a consequence of the specific shape of the BER surface. Since the BER is invariant to a positive scaling of $\mathbf{w}$, i.e. the size of $\mathbf{w}$ does not matter (except zero size), the BER surface has an infinitely long valley, and any point at the bottom of this valley is a true global MBER solution. For an illustration, see [19]. Once a weight vector $\mathbf{w}$ is near the edge of this infinitely long valley, convergence to the bottom is extremely fast, since the slope or gradient is high.

4 Adaptive MBER Beamforming

The p.d.f. of $y_R(k;\mathbf{w})$ can be shown to be explicitly given by:

$$p(y_R;\mathbf{w}) = \frac{1}{N_b\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_b}\exp\left(-\frac{\left(y_R - \bar{y}_{R,q}(\mathbf{w})\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right) \tag{21}$$

and the BER can alternatively be expressed as:

$$P_E(\mathbf{w}) = \frac{1}{N_b}\sum_{q=1}^{N_b} Q\left(g_q(\mathbf{w})\right), \tag{22}$$

where

$$g_q(\mathbf{w}) = \frac{\operatorname{sgn}(b_{q,1})\bar{y}_{R,q}(\mathbf{w})}{\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \tag{23}$$

and $\bar{y}_{R,q}(\mathbf{w}) \in \mathcal{Y}_R(\mathbf{w})$. In reality, the p.d.f. of $y_R(k;\mathbf{w})$ is unknown. Hence, we will adopt the temporal reference technique for supporting the adaptive implementation of the MBER beamforming algorithm.

4.1 A Block-Data Based Gradient Adaptive MBER Algorithm

A widely used approach of approximating a p.d.f. is known as the kernel density or Parzen window based estimate [20]–[22]. Given a block of K training samples $\{\mathbf{x}(k), b_1(k)\}$, a kernel density estimate of the p.d.f. (21) is readily given by:

$$\hat{p}(y_R;\mathbf{w}) = \frac{1}{K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{k=1}^{K}\exp\left(-\frac{(y_R - y_R(k;\mathbf{w}))^2}{2\rho_n^2\mathbf{w}^H\mathbf{w}}\right), \tag{24}$$

where the kernel width ρ_n is related to the standard deviation σ_n of the channel noise. From the standard results [20]–[22], it is known that the Parzen window estimate (24) is not overly sensitive to the value of ρ_n, and appropriate values for ρ_n lie in a range of values between some lower and upper bounds. From this estimated p.d.f., the estimated BER is given by:

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K}\sum_{k=1}^{K} Q\left(\hat{g}_k(\mathbf{w})\right) \tag{25}$$

with

$$\hat{g}_k(\mathbf{w}) = \frac{\operatorname{sgn}(b_1(k))y_R(k;\mathbf{w})}{\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}}. \tag{26}$$

The gradient of $\hat{P}_E(\mathbf{w})$ is formulated as:

$$\nabla\hat{P}_E(\mathbf{w}) = \frac{1}{2K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{k=1}^{K}\exp\left(-\frac{y_R^2(k;\mathbf{w})}{2\rho_n^2\mathbf{w}^H\mathbf{w}}\right)\operatorname{sgn}(b_1(k))\left(\frac{y_R(k;\mathbf{w})\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \mathbf{x}(k)\right). \tag{27}$$

Upon substituting $\nabla P_E(\mathbf{w})$ by $\nabla\hat{P}_E(\mathbf{w})$ in the conjugate gradient updating mechanism, a block-data based adaptive algorithm is obtained. The step size μ and the kernel width ρ_n are two algorithmic parameters that have to be set appropriately.

4.2 A Stochastic Gradient Based Adaptive MBER Algorithm

In the kernel density estimate (24), a variable width of $\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}$ is used, which depends on the beamformer's weight vector. If an approximation is invoked by using a constant width of ρ_n in a kernel density estimate, the associated computational complexity can be considerably reduced. Formally, this leads to using the kernel density estimate of:

$$\tilde{p}(y_R;\mathbf{w}) = \frac{1}{K\sqrt{2\pi}\rho_n}\sum_{k=1}^{K}\exp\left(-\frac{(y_R - y_R(k;\mathbf{w}))^2}{2\rho_n^2}\right) \tag{28}$$

as an approximation to the true density given by (21), and to using

$$\tilde{P}_E(\mathbf{w}) = \frac{1}{K}\sum_{k=1}^{K} Q\left(\tilde{g}_k(\mathbf{w})\right), \tag{29}$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\operatorname{sgn}(b_1(k))y_R(k;\mathbf{w})}{\rho_n} \tag{30}$$

as the BER estimate. This approximation is valid, provided that the width ρ_n is chosen appropriately. In particular, the appropriate values for ρ in (28) are generally different from those used in (24). Like the kernel density estimate (24), the p.d.f. estimate (28) is not overly sensitive to the value of ρ_n. The gradient of $\tilde{P}_E(\mathbf{w})$ has a much simpler form:

$$\nabla\tilde{P}_E(\mathbf{w}) = \frac{1}{2K\sqrt{2\pi}\rho_n}\sum_{k=1}^{K}\exp\left(-\frac{y_R^2(k;\mathbf{w})}{2\rho_n^2}\right)\operatorname{sgn}(b_1(k))\mathbf{x}(k). \tag{31}$$

In order to derive a sample-by-sample adaptive algorithm, adopt a similar single-sample estimate of $p(y_R;\mathbf{w})$ as used in [18], namely:

$$\tilde{p}(y_R,k;\mathbf{w}) = \frac{1}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{(y_R - y_R(k;\mathbf{w}))^2}{2\rho_n^2}\right). \tag{32}$$

Using the instantaneous stochastic gradient of:

$$\nabla\tilde{P}_E(\mathbf{w},k) = \frac{\operatorname{sgn}(b_1(k))}{2\sqrt{2\pi}\rho_n}\exp\left(-\frac{y_R^2(k;\mathbf{w})}{2\rho_n^2}\right)\mathbf{x}(k) \tag{33}$$

gives rise to a stochastic gradient adaptive algorithm, which we referred to as the LBER algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\operatorname{sgn}(b_1(k))}{2\sqrt{2\pi}\rho_n} \exp\left(-\frac{y_R^2(k;\mathbf{w}(k))}{2\rho_n^2}\right) \mathbf{x}(k). \tag{34}$$

The adaptive gain μ and the kernel width ρ_n are the two algorithmic parameters that have to be set appropriately to ensure a fast convergence rate and small steady-state BER misadjustment. The computational complexity of this LBER algorithm is compared with that of the LMS algorithm in Table 1. Note that the LBER algorithm (34) has considerably lower complexity than the algorithm given in [18],[19], because it involves more simplifications. Previous empirical results derived in a multiuser detection context [27] have shown that this simplified LBER algorithm appears to have a similar convergence speed to the full LBER algorithm of [18],[19], even though it involves a more coarse approximation and exhibits a lower computational complexity.

5 Simulation Study

The example used in our computer simulation study consisted of five signal sources and a two-element antenna array. The antenna array element spacing was $\lambda/2$ with λ being the wavelength. The direction of arrival for source 1, the desired user, was $15°$, and the directions of arrival for the interfering sources $2, 3, 4, 5$ were $-30°$, $60°$, $-70°$ and $80°$, respectively. Fig. 1 compares the BER performance of the MBER solution with that of the MMSE solution under two different conditions: **(a)** the desired user and all the four interfering sources have equal power, and **(b)** all the four interfering sources have 6 dB higher power than the desired user. The BER curves in Fig. 1 were computed using the theoretical BER expression (15) with the MMSE solution given by (8) and the MBER solution obtained by numerical optimization based on the simplified conjugate gradient algorithm portrayed in Section 3. For this example, the superior performance of the MBER beamforming technique over the MMSE benchmarker scheme becomes evident. The results shown in Fig. 1 also indicate that the MBER solution is robust to the near-far effect. This is further confirmed by the result shown in Fig. 2, which was obtained under the following condition: SNR$=10$ dB and SIR$_i = 24$ dB for $i = 3, 4, 5$, were fixed, while SIR$_2$ was varied.

Fig. 3 compares the beam pattern of the MBER beamformer to that of the MMSE beamformer under the condition of SNR$=10$ dB and SIR$_i = 0$ dB for $i = 2, 3, 4, 5$, where $F(\theta)$ has been normalized. Note that the MMSE beamformer appears to have a better amplitude response than the MBER beamformer. Specifically, at the four angles corresponding to the four interfering sources indicated by the circles in Fig. 3, the MMSE beamformer exhibits higher-attenuation magnitude responses at $-70°$, $60°$ and $80°$, and only a slightly inferior magnitude response at $-30°$, compared to the MBER solution. If the amplitude response alone would constitute the ultimate performance criterion of a beamformer, the MMSE beamformer would appear to be more beneficial. However, considering the magnitude response alone can be misleading. At the four angles of the four interfering sources, the phase response of the MBER beamformer is significantly closer to $\pm\frac{\pi}{2}$ than that of the MMSE solution, while maintaining a $0°$ phase for the desired user at the angle of $15°$. Thus, the MBER solution has a significantly better response in terms of $\Re[F(\theta)]$, and this results in an augmented ability to “cancel” interfering signals. Explicitly, the two beamformers optimize the beamformer weights very differently.

It is seen that the operation of the MMSE beamformer appears to break down under the conditions given in Fig. 1 (b) and exhibits a high BER floor. A first attempt of interpreting this phenomenon is made by examining the beam patterns of the two beamformers given SNR$=15$ dB and SIR$_i = -6$ dB for $i = 2, 3, 4, 5$. These two beam patterns, not shown here for reasons of space economy, are found to be similar to those shown in Fig. 3. Thus, these beam patterns cannot explicitly justify, why the operation of the MMSE solution should break down, while the MBER solution remains capable of providing an adequate performance. The beam pattern generated is not directly related to the system's BER performance, and the conditional p.d.f. (14) is the best indicator of a beamformer's BER performance. The conditional p.d.f.s of the two beamformers, under the conditions of maintaining SNR$=15$ dB and SIR$_i=-6$ dB for $i=2,3,4,5$, are illustrated in Fig. 4. In Fig. 4, the beamformer's weight vector has been normalized to a unit length, so that the BER is mainly determined by the minimum distance of the subset $\mathcal{Y}_R^{(+)}(\mathbf{w})$ from the decision threshold of $y_R=0$. There are $N_{sb}=16$ points altogether due to 5 users, calculated based on the formula $N_{sb}=2^M/2$ where M is the number of users. By examining Fig. 4, it becomes clear why the MMSE beamformer has a high BER floor of around $\frac{1.5}{16}$. For the MMSE solution, $\mathcal{Y}_R^{(-)}(\mathbf{w}_{\mathrm{MMSE}})$ and $\mathcal{Y}_R^{(+)}(\mathbf{w}_{\mathrm{MMSE}})$ are linearly non-separable. One of the $N_{sb}=16$ points in $\mathcal{Y}_R^{(+)}(\mathbf{w}_{\mathrm{MMSE}})$ is on the wrong side of the decision boundary of $y_R=0$ and another point is right on $y_R=0$. By comparison, it can be seen from Fig. 4 that the MBER solution successfully separates $\mathcal{Y}_R^{(-)}(\mathbf{w}_{\mathrm{MBER}})$ and $\mathcal{Y}_R^{(+)}(\mathbf{w}_{\mathrm{MBER}})$.

Let us now study the performance of the block-data based gradient adaptive MBER algorithm employing the conjugate gradient updating mechanism. Fig. 5 illustrates the convergence rate of the algorithm under the conditions of SNR$=10$ dB, SIR$_2=$SIR$_5=-6$ dB, SIR$_3=$SIR$_4=0$ dB, and a block size of $K=200$, using two different initial weight vectors, namely (**a**) $\mathbf{w}(0)=[0.1+j0.0\ 0.1+j0.0]^T$ and (**b**) $\mathbf{w}(0)=\mathbf{w}_{\mathrm{MMSE}}$. From Fig. 5, it can be seen that this block-data based adaptive algorithm converges rapidly. The effect of the block size K on the performance of the block-data based adaptive MBER algorithm was also investigated, and it is seen that in conjunction with a short block length of $K=50$, the BER performance of the block-data based adaptive MBER solution degrades only slightly at high SNRs, compared to the theoretical MBER solution. When the block length increases to $K=100$, the block-data based adaptive MBER solution closely matches the performance of the theoretical MBER solution. Space constraints preclude a graphical illustration. The performance of the stochastic gradient based adaptive MBER algorithm portrayed in Section 4.2 is investigated next. Fig. 6 shows the learning curves of the LBER algorithm using two different initial weight vectors, under the conditions of SNR$=10$ dB, SIR$_2=$SIR$_5=-6$ dB and SIR$_3=$SIR$_4=0$ dB. It can be seen that this stochastic gradient algorithm converges reasonably fast, while maintaining a low steady-state BER misadjustment.

To explicitly investigate the influence of weight-initialisation $\mathbf{w}(0)$ on the LBER algorithm, 20 uniformly distributed random initial weight $\mathbf{w}(0)$ were used, under the conditions of SNR$=10$ dB, SIR$_2=$SIR$_5=-6$ dB and SIR$_3=$SIR$_4=0$ dB. Fig. 7 depicts the learning curves obtained. Note that a variable adaptive step-size was used, where over each randomly initialised run, for the first 25 samples we had $\mu=0.2$, which was reduced to $\mu=0.02$ afterward. The need for an adaptive step-size may be explained as follows. When $\mathbf{w}$ is far from the MBER solution, the gradient of the BER surface can be very flat and hence a large adaptive step-size is needed to move away from these flat regions. By contrast, as mentioned at the end of Section 3, once $\mathbf{w}$ is near the edge of the 'infinitely long MBER valley', convergence to the bottom is extremely fast and a small step-size μ is required to avoid 'over-shooting' the MBER solution. It is also worth pointing out that the MMSE solution is not necessarily a 'favourable' initial condition for the LBER algorithm. For the example simulated here, the algorithm appeared to converge well when started from $\mathbf{w}(0)=\mathbf{w}_{\mathrm{MMSE}}$, as can be seen from Fig. 6 (b). However, as confirmed in many other investigations not included here owing to lack of space, the MMSE solution often constitutes an undesirable initial condition, which results in slow convergence and a relatively high steady-state BER misadjustment. To provide a rule of thumb, if the interference scenario encountered is a hostile one owing to having a low angular separation of the interferers, for example, the MMSE weights would be far from the MBER solution and hence they would result in slow convergence, potentially arriving at a local optimum.

6 Conclusions

An adaptive MBER beamforming technique has been developed. It has been shown that the MBER beamformer exploits the system's resources more intelligently than the standard MMSE beamformer and, consequently, can achieve a better performance in terms of a lower BER. Simulation results also suggest that the MBER solution is robust to the near-far effect. Adaptive implementation of the MBER beamformer has also been addressed. A block-data based conjugate gradient adaptive MBER algorithm has been shown to converge rapidly, while requiring a reasonably small block size for accurately approximating the theoretical MBER solution. A stochastic gradient adaptive MBER algorithm, namely the LBER technique, has also been derived. The results obtained in this study have demonstrated the potential of the adaptive MBER beamforming approach. However, several important areas still warrant further research. These include considering hostile fading channels, dispersive wideband channels that induce ISI, and wideband beamforming. Our current research is also considering the extension of the adaptive MBER beamformer to other modulation schemes.

References

[1] J.H. Winters, J. Salz and R.D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Communications*, Vol.42, No.2, pp.1740–1751, February/March/April 1994.

[2] M.C. Wells, "Increasing the capacity of GSM cellular radio using adaptive antennas," *IEE Proc. Communications*, Vol.143, No.5, pp.304–310, 1996.

[3] J. Litva and T. K.Y. Lo, *Digital Beamforming in Wireless Communications*. London: Artech House, 1996.

[4] L. C. Godara, "Applications of antenna arrays to mobile communications, Part I: Performance improvement, feasibility, and system considerations," *Proc. IEEE*, Vol.85, No.7, pp.1031–1060, 1997.

[5] A.J. Paulraj and C.B. Papadias,"Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, Vol.14, No.6, pp.49–83, 1997.

[6] J.H. Winters, "Smart antennas for wireless systems," *IEEE Personal Communications*, Vol.5, No.1, pp.23–27, 1998.

[7] R. Kohno, "Spatial and temporal communication theory using adaptive antenna array," *IEEE Personal Communications*, Vol.5, No.1, pp.28–35, 1998.

[8] P. Petrus, R.B. Ertel and J.H. Reed, "Capacity enhancement using adaptive arrays in an AMPS system," *IEEE Trans. Vehicular Technology*, Vol.47, No.3, pp.717–727, 1998.

[9] G.V. Tsoulos, "Smart antennas for mobile communication systems: benefits and challenges," *IEE Electronics and Communications J.*, Vol.11, No.2, pp.84–94, 1999.

[10] J.S. Blogh and L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking – Smart Antennas and Adaptive Modulation*. Chichester: John Wiley, 2002.

[11] I.S. Reed, J.D. Mallett and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, Vol.AES-10, pp.853–863, 1974.

[12] M.W. Ganz, R.L. Moses and S.L. Wilson, "Convergence of the SMI and the diagonally loaded SMI algorithms with weak interference (adaptive array)," *IEEE Trans. Antennas and Propagation*, Vol.38, No.3, pp.394–399, 1990.

[13] B. Widrow, P.E. Mantey, L.J. Griffiths and B.B. Goode, "Adaptive antenna systems," *Proc. IEEE*, Vol.55, pp.2143-2159, 1967.

[14] L.J. Griffiths, "A simple adaptive algorithm for real-time processing in antenna arrays," *Proc. IEEE*, Vol.57, pp.1696–1704, 1969.

[15] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, pp.316–322, 1998.

[16] I.N. Psaromiligkos, S.N. Batalama and D.A. Pados, "On adaptive minimum probability of error linear filter receivers for DS-CDMA channels," *IEEE Trans. Communications*, Vol.47, No.7, pp.1092–1102, 1999.

[17] C.C. Yeh and J.R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Communications*, Vol.48, No.7, pp.1226–1235, 2000.

[18] B. Mulgrew and S. Chen, "Adaptive minimum-BER decision feedback equalisers for binary signalling," *Signal Processing*, Vol.81, No.7, pp.1479–1489, 2001.

[19] S. Chen, A.K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels," *IEEE Trans. Signal Processing*, Vol.49, No.6, pp.1240–1247, 2001.

[20] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, Vol.33, pp.1066–1076, 1962.

[21] B.W. Silverman, *Density Estimation*. London: Chapman Hall, 1996.

[22] A.W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997.

[23] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley, 1993.

[24] S. Chen, B. Mulgrew and P.M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Trans. Neural Networks*, Vol.4, No.4, pp.570–579, 1993.

[25] S. Chen, B. Mulgrew and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, Vol.41, No.9, pp.2918–2927, 1993.

[26] S. Chen, A.K. Samingan and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Networks*, Vol.12, No.3, pp.604-611, 2001.

[27] S. Chen, "Least bit error rate adaptive multiuser detection," In: L.P. Wang, eds. *Soft Computing in Communications*. Berlin: Springer Verlag, 2003, pp.389–408.

Table 1: Comparison of computational complexity per weight update, where L is the dimension of the weight vector.

	multiplications	additions	$\exp(\bullet)$ evaluation
LBER	$4 \times L + 4$	$4 \times L - 1$	1
LMS	$8 \times L + 2$	$8 \times L - 1$	—

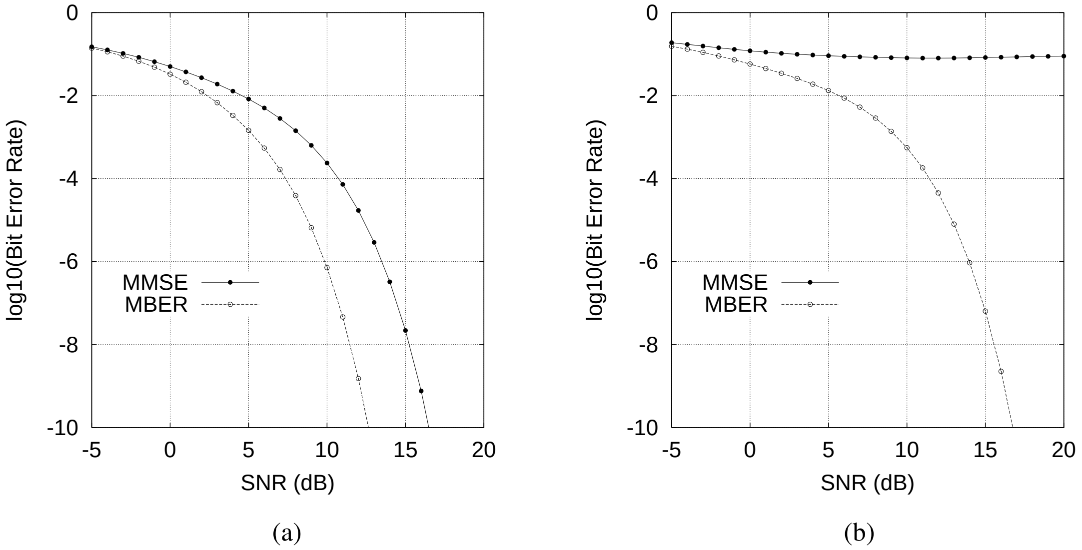


Figure 1: Comparison of the bit error rate performance of the MMSE and MBER beamformers: (a) $\text{SIR}_i = 0$ dB for $i = 2, 3, 4, 5$, and (b) $\text{SIR}_i = -6$ dB for $i = 2, 3, 4, 5$.

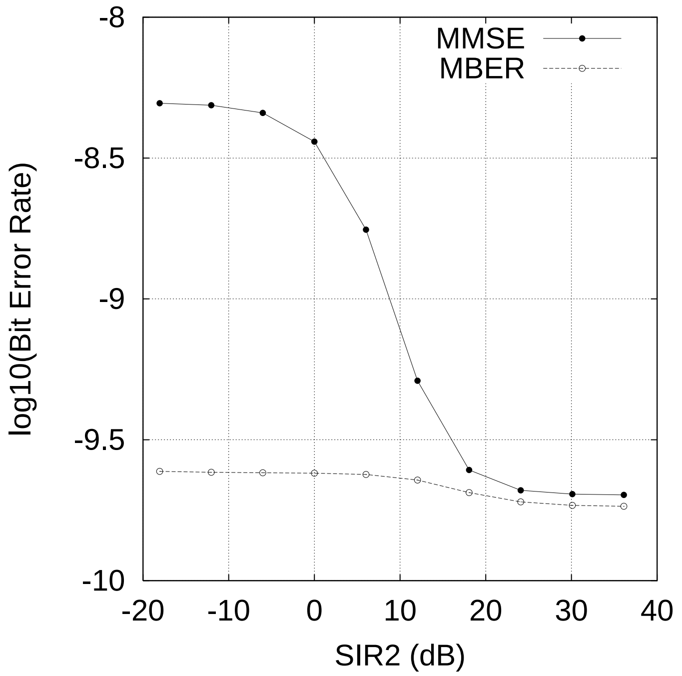


Figure 2: Influence of the near-far effect on bit error rate performance of the MMSE and MBER beamformers for SNR$= 10$ dB and $\text{SIR}_i = 24$ dB for $i = 3, 4, 5$.

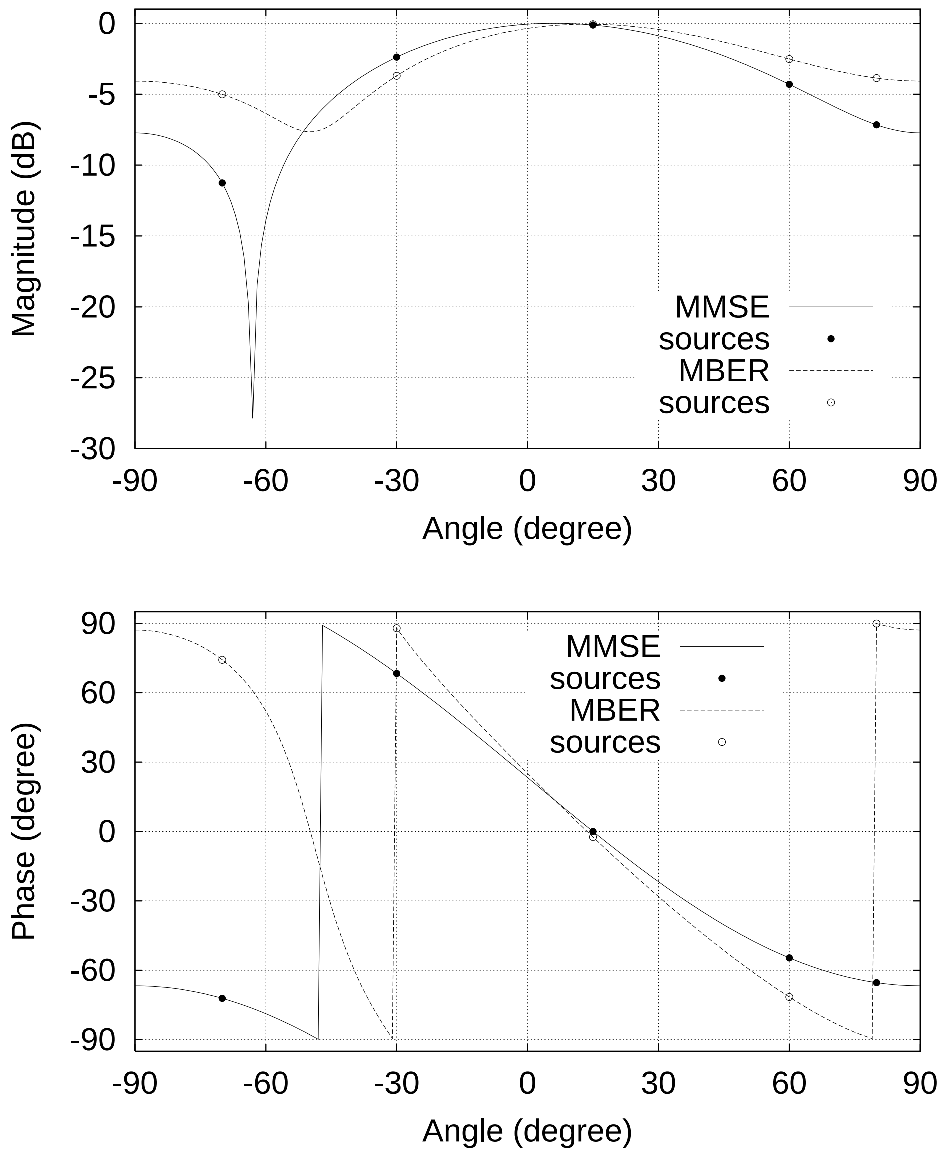
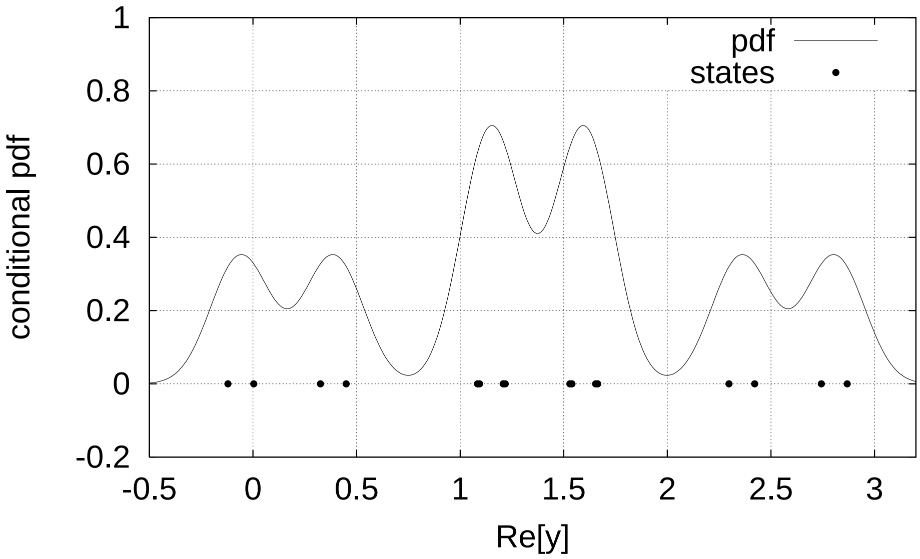
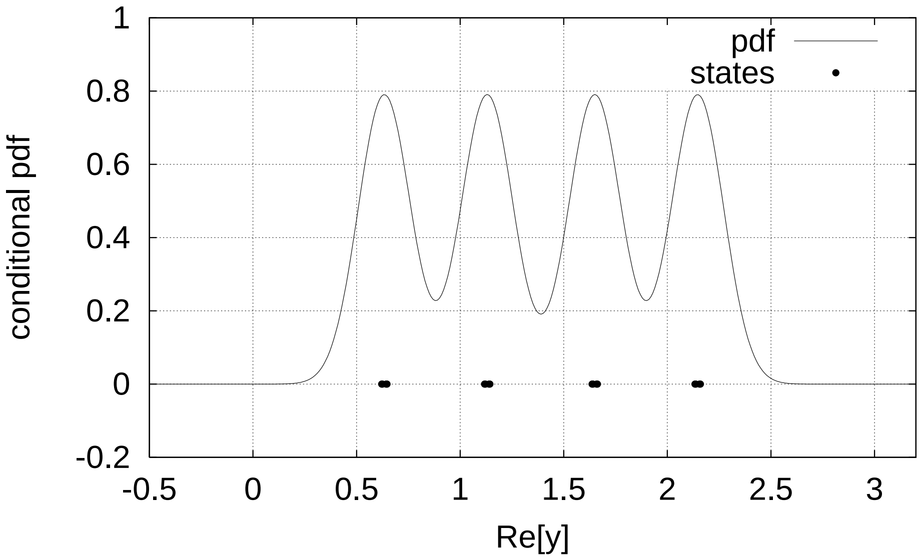


Figure 3: Comparison of MMSE and MBER beam patterns for SNR$=10$ dB and SIR$_i=0$ dB for $i=2,3,4,5$.



(a) MMSE



(b) MBER

Figure 4: Conditional probability density functions given $b_1(k) = +1$ and subsets $\mathcal{Y}_R^{(+)}(\mathbf{w})$ of the MMSE and MBER beamformers for SNR$= 15$ dB and SIR$_i = -6$ dB for $i = 2, 3, 4, 5$.

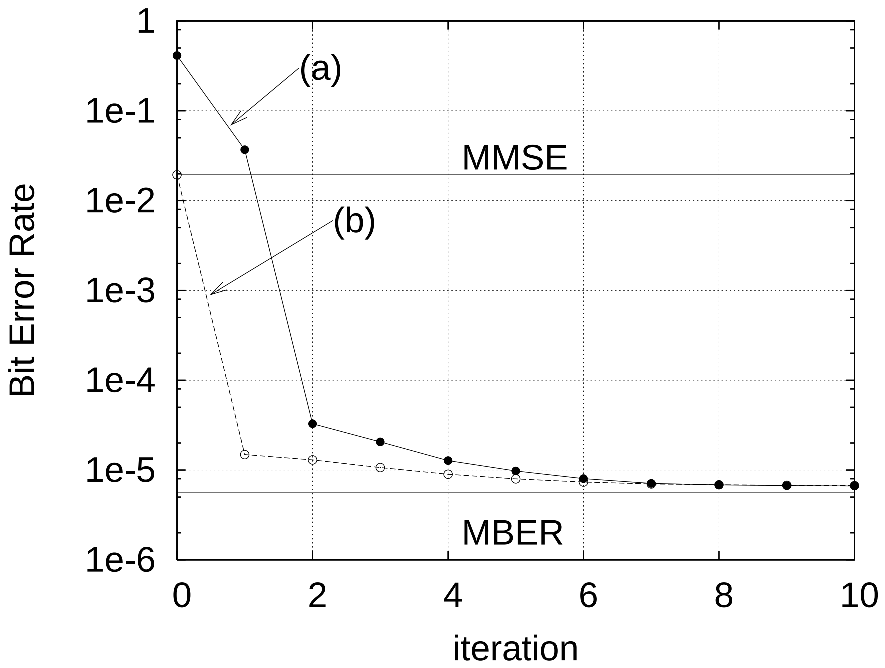
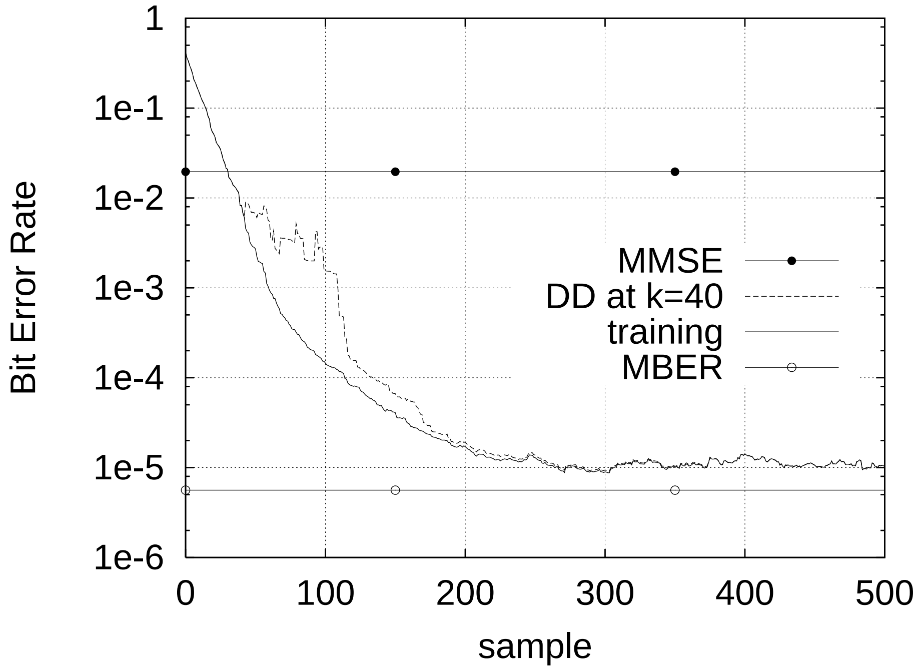
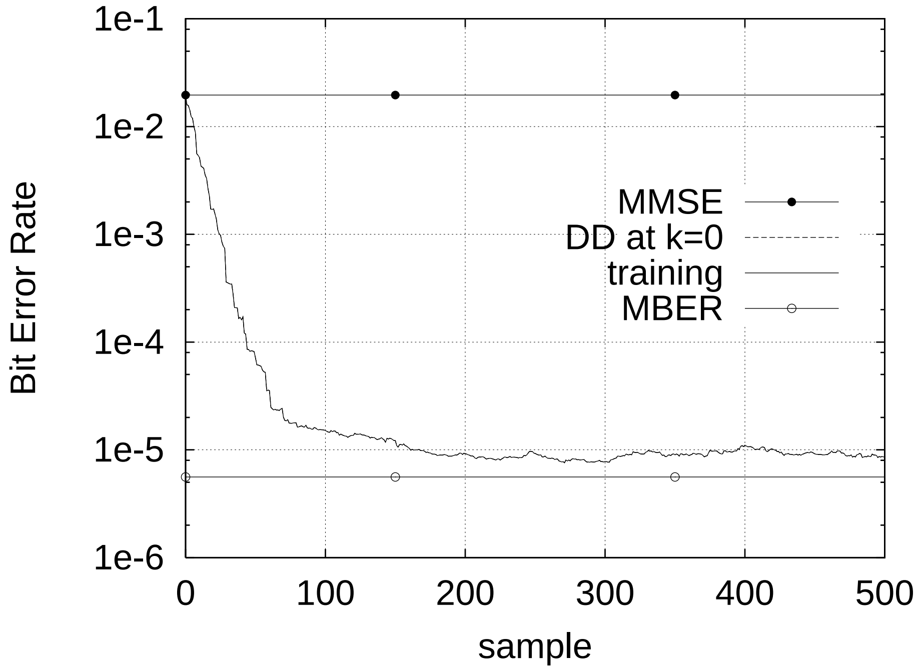


Figure 5: Convergence rate of the block-data based gradient adaptive MBER algorithm of Section 4.1 for a block size of $K = 200$, under the conditions SNR= 10 dB, SIR_2 =SIR_5 $= -6$ dB, and SIR_3 =SIR_4 $= 0$ dB. (a): $\mathbf{w}(0) = [0.1 + j0.0\ 0.1 + j0.0]^T$, $\mu = 0.8$ and $\rho_n^2 = 2\sigma_n^2 = 0.1$; (b): $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, $\mu = 0.5$ and $\rho_n^2 = 2\sigma_n^2 = 0.1$.



(a) $\mathbf{w}(0) = [0.1 + j0.0 \; 0.1 + j0.0]^T$, $\mu = 0.03$ and $\rho_n^2 = 4\sigma_n^2 = 0.2$



(b) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, $\mu = 0.02$ and $\rho_n^2 = 4\sigma_n^2 = 0.2$

Figure 6: Learning curves of the stochastic gradient adaptive MBER algorithm of Section 4.2 averaged over 20 runs. SNR= 10 dB, $\mathrm{SIR}_2 = \mathrm{SIR}_5 = -6$ dB, and $\mathrm{SIR}_3 = \mathrm{SIR}_4 = 0$ dB, DD denotes decision-directed adaptation in which $b_1(k)$ is substituted by its estimate $\hat{b}_1(k)$. Note that in graph (b) decision-directed and training curves are indistinguishable.

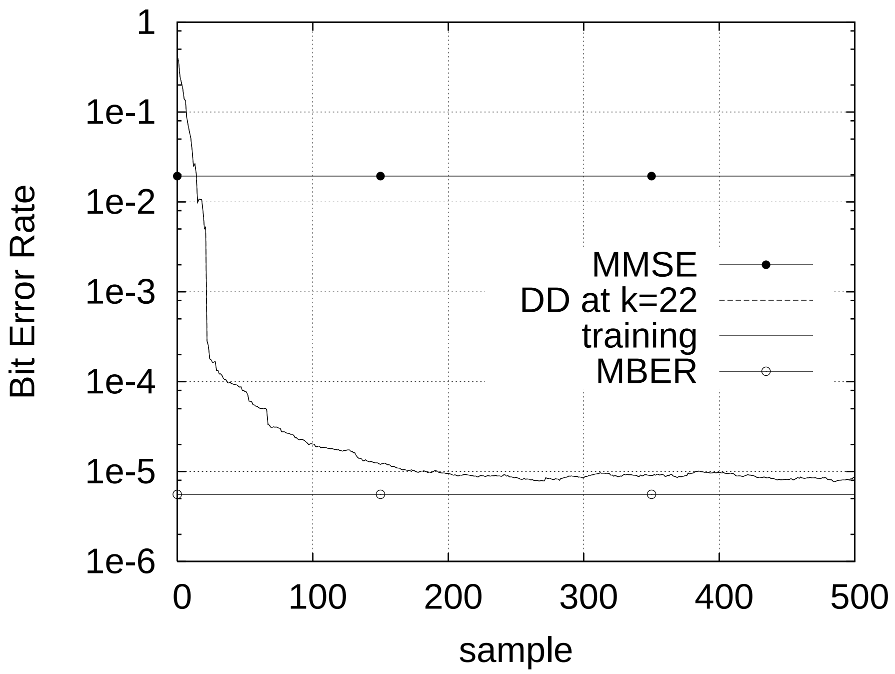


Figure 7: Learning curves of the stochastic gradient adaptive MBER algorithm of Section 4.2 averaged over 20 randomly chosen uniformly distributed initial weight values $\mathbf{w}(0)$. SNR= 10 dB, SIR$_2$ =SIR$_5$ $= -6$ dB, and SIR$_3$ =SIR$_4$ $= 0$ dB, DD denotes decision-directed adaptation in which $b_1(k)$ is substituted by its estimate $\hat{b}_1(k)$. Note that the performance curves based on DD and explicit training are indistinguishable. Furthermore, we had $\rho_n^2 = 4\sigma_n^2 = 0.2$ and $\mu = 0.2$ for the first 25 samples, reducing the step-size to $\mu = 0.02$ afterwards.