Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 615623, 9 pages
doi:10.1155/2010/615623

Research Article

Approximate Minimum Bit Error Rate Equalization for Fading Channels

Lorant Kovacs,[1] Janos Levendovszky,[2] Andras Olah,[3] and Gergely Treplan[3]

[1] *Kecskemet College, Faculty of Mechanical Engineering and Automation, Izsaki ut 10, 6000 Budapest, Hungary*
[2] *Department of Telecommunications, Budapest University of Technology and Economics, Magyar tudosok korutja 2, 1117 Budapest, Hungary*
[3] *Faculty of Information Technology, Pazmany Peter Catholic University, Prater utca 50/a, 1083 Budapest, Hungary*

Correspondence should be addressed to Andras Olah, olah@itk.ppke.hu

Received 1 October 2009; Revised 14 April 2010; Accepted 8 July 2010

Academic Editor: Tim Davidson

Copyright © 2010 Lorant Kovacs et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel channel equalizer algorithm is introduced for wireless communication systems to combat channel distortions resulting from multipath propagation. The novel algorithm is based on minimizing the bit error rate (BER) using a fast approximation of its gradient with respect to the equalizer coefficients. This approximation is obtained by estimating the exponential summation in the gradient with only some carefully chosen dominant terms. The paper derives an algorithm to calculate these dominant terms in real-time. Summing only these dominant terms provides a highly accurate approximation of the true gradient. Combined with a fast adaptive channel state estimator, the new equalization algorithm yields better performance than the traditional zero forcing (ZF) or minimum mean square error (MMSE) equalizers. The performance of the new method is tested by simulations performed on standard wireless channels. From the performance analysis one can infer that the new equalizer is capable of efficient channel equalization and maintaining a relatively low bit error probability in the case of channels corrupted by frequency selectivity. Hence, the new algorithm can contribute to ensuring QoS communication over highly distorted channels.

1. Introduction

Broadband radio channels are susceptible to selective fading due to multipath propagation. In this case, the differences among the propagation delays on the paths may amount to a significant fraction of a symbol interval. Hence, frequency selective fading may yield severe performance degradation. As a result, efficient channel equalization techniques prove to be instrumental to combat intersymbol interference (ISI) in order to avoid large scale drops in system performance.

Since the effect of interferences are especially crucial in wireless communication systems, fast channel equalizer algorithms have to be developed which are simple enough to run on the currently available hardware architectures. This paper aims at developing a low complexity channel equalizer algorithm by directly minimizing the BER instead of minimizing the mean-square error or the peak distortion [1, 2]. Unfortunately, the direct minimization of BER with respect to the equalizer coefficients is of exponential complexity due to the large summation when expressing BER as a function of the equalizer coefficients. Thus, we develop a new bound on BER on which basis the equalizer coefficients can be optimized by a fast algorithms. It is also pointed out that the new bound on BER is sharper than the previously known ones listed in [3].

The first attempts to derive an equalizer based on the minimum BER strategy can be found in the work of Shimbo and Celebiler [4] and Shamas and Yao [5]. The optimal equalizer coefficients were only sought by exhaustive search, thus real-time adaptivity was not guaranteed. In recent years, some new results have been developed for minimum BER equalization. In [6], a low-complexity adaptive algorithm is proposed for 2- or 4-state modulation systems but the convergence is rather slow, while in [7, 8] near minimum BER equalization is carried out by radial basis function neural networks which considerably increases the equalizer

complexity. On the other hand, very complex equalizer schemes have been proposed for DS-CDMA systems in [9–11]. Paper [12] investigates the minimization of BER in MIMO systems with linear equalizers based on QPSK modulation. In this model the information sequence is corrupted by multiuser interference and additive noise. BER is derived for the case of two transmitters and approximated by a simple formula. The optimum then is calculated by analytical tools. When the number of transmitters is larger than two then the authors use a sequence of cascade filters. It is important to note that this article does not address the problem of ISI. In paper [13], BER is minimized subject to some constraints. The authors prove that the constrained BER cost function has only one global minimum and equalization can be achieved by quadratic programming. However, due to the constraints this detector is only optimal in the case of minimum phase channels. In the paper [14, 15], the equalization is carried out by using the Bayes risk criterion. This is a more general approach than BER minimization, but the choice of loss function is left open, realizing that not every loss function will yield fast convergence or yield low error rate. Furthermore, the loss function is very complex and does not lend itself to simple minimization. Therefore, deriving bound on BER for fast minimization with respect to the equalizer coefficients can still yield powerful and real-time channel equalization. The novel algorithm presented by the paper is demonstrated by BPSK modulation scheme.

The results are given in the following structure:

(i) in Section 2, the communication model is outlined;

(ii) in Section 3, BER is expressed as a function of the equalizer coefficients and a gradient based algorithm is discussed for minimizing BER;

(iii) in Section 4, a novel method is derived to approximate BER by using the dominant terms;

(iv) in Section 5, the new equalizer algorithm will be introduced based on the approximation treated in Section 4;

(v) in Section 6, the performance and convergence properties of the new equalizer algorithms are analyzed numerically.

2. The Model

To describe single-user communication over a fading channel, we use the so-called equivalent discrete time white noise filter model (for further details see [1]).

The corresponding quantities are defined as follows:

(i) $y_k \in \{-1, 1\}$ denotes the transmitted information bit at time instant k being a sequence of identically distributed independent Bernoulli random variables with $P(y_k = 1) = P(y_k = -1) = 0.5$;

(ii) the discrete impulse response of the channel is denoted by h_k, $k = 0, \ldots, M$ where M denotes the span of ISI;

(iii) the noise is denoted by ν_k and is assumed to be a stationary zero mean white Gaussian random sequence with constant spectral density N_0;

(iv) the received sequence is denoted by x_k, which is a linearly distorted and noisy version of the transmitted sequence given as

$$x_k = \sum_{j=0}^{M} h_j y_{k-j} + \nu_k; \tag{1}$$

(v) the equalizer is a linear FIR filter, the output of which is denoted by $\widetilde{y}_k$

$$\widetilde{y}_k = \sum_{i=0}^{J} w_i x_{k-i}, \tag{2}$$

where w_i $i = 0, \ldots, J$ denotes the free parameters of the equalizer which are subject to further optimization;

(vi) the decision is carried out by threshold detection in a symbol-by-symbol fashion:

$$\widehat{y}_k = \operatorname{sgn}\{\widetilde{y}_{k-D}\} = \operatorname{sgn}\left\{\sum_{i=0}^{J} w_i x_{k-D-i}\right\}, \tag{3}$$

where D denotes the delay of the channel. (For the sake of brevity, here we assume $D = 0$, while the more general treatment for $D \neq 0$ will be given in Section 4.3);

(vii) the overall channel impulse response function is determined by the cascade of the channel and the equalizer

$$q_k = \sum_{i=0}^{M} h_i w_{k-i}, \quad k = 0, \ldots, L, \tag{4}$$

where $L = M + J$ denotes the support of the overall impulse response.

Traditional equalization algorithms aimed at minimizing the peak distortion [1, 16] defined as

$$\mathbf{w}_{\text{opt}} : \min_{\mathbf{w}} \sum_{i=1}^{L} |q_i| \tag{5}$$

or the mean-square error [1, 2]

$$\mathbf{w}_{\text{opt}} : \min_{\mathbf{w}} \mathrm{E}\left[\left(y_k - \sum_{j=0}^{J} w_j x_{k-j}\right)^2\right]. \tag{6}$$

Both approaches involve the use of linear stochastic approximation schemes [1, 2, 16], but they fell short of providing efficient equalization as the goal functions did not have any direct relationship with BER.

3. Weight Optimization Subject to Minimizing the BER

Since our approach to equalization is based on minimizing the bit error probability, first we express BER as a function of the equalizer coefficients as given in [4]

$$P_E(\mathbf{w}) = \left(\frac{1}{2^L}\right) \sum_{\mathbf{y} \in \mathcal{Y}} \Phi\left(\frac{\sum_{l=0}^{L} q_l y_l}{\sigma}\right), \tag{7}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, $\sigma^2 = N_0 \sum_{j=0}^{J} w_j^2$, and $\mathcal{Y} = \{\mathbf{y} = (y_0, y_1, \ldots, y_L) \mid y_0 = -1;\ y_i \in \{-1, 1\}, i = 1, \ldots, L\}$. Here, we note that in the paper, for expressing the BER, we will use the standard Gaussian distribution function defined as $\Phi(x) := (1/\sqrt{2\pi}) \int_{-\infty}^{x} \mathrm{e}^{(u^2/2)} du$. The relationships of $\Phi(x)$ with the "Q" function $Q(x)$ and the error function $\mathrm{erf}(x)$ are given as follows: $\Phi(x) = 1 - Q(x)$ and $\Phi(x) = 1/2[1 + \mathrm{erf}(x/\sqrt{2})]$. Substituting (4) into (7), we obtain

$$P_E(\mathbf{w}) = \left(\frac{1}{2^L}\right) \sum_{\mathbf{y} \in \mathcal{Y}} \Phi\left(\frac{\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l}{\sqrt{N_0 \sum_{n=0}^{J} w_n^2}}\right). \tag{8}$$

To find the optimal weights of the equalizer which minimize this error probability, we have to solve the following equation:

$$\mathbf{w}_{\mathrm{opt}} : \mathrm{grad}_{\mathbf{w}} P_E(\mathbf{w}) = \mathbf{0}, \tag{9}$$

where the ith component of the gradient is

$$\begin{aligned}
&\frac{\partial P_E(\mathbf{w})}{\partial w_i} \\
&= \frac{1}{2^L \sqrt{2\pi N_0 \left(\sum_{n=0}^{J} w_n^2\right)^3}} \\
&\quad \times \sum_{y \in \mathcal{Y}} \exp\left(\frac{-\left(\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l\right)^2}{2N_0 \sum_{n=0}^{J} w_n^2}\right) \\
&\quad \cdot \left[\left(\sum_{n=0}^{J} w_n^2\right) \cdot \left(\sum_{l=i}^{M+i} h_{l-i} y_l\right) - w_i \left(\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l\right)\right].
\end{aligned} \tag{10}$$

The weights can be optimized by gradient descent, yielding the following equalization algorithm:

$$w_i(k+1) = w_i(k) - \gamma \frac{\partial P_E(\mathbf{w}(k))}{\partial w_i}. \tag{11}$$

Here, $\mathbf{w}(k)$ is the value of the weight vector at the kth iteration. One must note, that the gradient search with fixed step size in general will not guarantee the convergence to the global minimum. However, as our simulations have demonstrated, over standard wireless channels the algorithm in most cases reached the global optimum.

In the forthcoming discussion, the procedure given by formula (11) is termed as true gradient search (TGS). Unfortunately performing TGS is computationally prohibitive, because of the summation over an exponentially large number of vectors in expression (10). Furthermore, this summation in TGS must be calculated in each step of algorithm (11). Thus, TGS can only be applied in practice if the support of the overall impulse response defined in (4) is very limited. Otherwise, near-optimal algorithms must be sought which lend themselves to real-time implementations. To ease this complexity a new bound is derived on BER.

4. New Lower Bound on BER by Using the Dominant Terms

In this section, we derive a new approximation on BER. The purpose of developing this approximation is to estimate BER with an expression which is a computationally simple function of the equalizer weights. This paves the way towards real-time channel equalization.

In order to derive a bound on BER, one can note that function $\Phi(\cdot)$ tends rapidly to zero for negative arguments. As a result, the terms in the summation can differ in several magnitudes. This gives rise to the idea of collecting only the dominant terms to provide a lower bound on BER. This lower bound has been commonly used in other domains, such as reliability analysis and referred to as Li-Silvester bound [17], where the tiresome calculation of an expected value over a large state space is approximated by only using the dominant terms in the summation.

In our case, this bound can be obtained as follows. We may look upon (8) as an expected value, given as follows:

$$\begin{aligned}
G(\mathbf{w}) &= \frac{1}{2^L} \sum_{\mathbf{y} \in \mathcal{Y}} \Phi\left(\frac{\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l}{\sqrt{N_0 \sum_{n=0}^{J} w_n^2}}\right) \\
&= \frac{1}{2^L} \sum_{\mathbf{y} \in \mathcal{Y}} \Phi\left(\frac{\sum_{l=0}^{L} q_l y_l}{\sigma}\right) \\
&= \frac{1}{2^L} \sum_{\mathbf{y} \in \mathcal{Y}} \Phi\left(\frac{\mathbf{q}^T \mathbf{y}}{\sigma}\right) = \frac{1}{2^L} \sum_{\mathbf{y} \in \mathcal{Y}} G(\mathbf{w}, \mathbf{y}) = \mathrm{E}_{\mathbf{y}}[G(\mathbf{w}, \mathbf{y})].
\end{aligned} \tag{12}$$

Introducing the following notations, one can obtain

$$\mathbf{g}(\mathbf{w}) = \mathrm{grad}_{\mathbf{w}} G(\mathbf{w}) = \mathrm{grad}_{\mathbf{w}} \mathrm{E}_{\mathbf{y}}[G(\mathbf{w}, \mathbf{y})] = \mathrm{E}_{\mathbf{y}}[\mathbf{g}(\mathbf{w}, \mathbf{y})]. \tag{13}$$

Let us then separate the set $\mathcal{Y}$ into two disjoint subsets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ ($\mathcal{Y}_1 \bigcup \mathcal{Y}_2 = \mathcal{Y}$, $\mathcal{Y}_1 \bigcap \mathcal{Y}_2 = 0$). Let the number of elements in $\mathcal{Y}_1$ be K, containing the first K vectors belonging to the first K largest values of $G(\mathbf{w}, \mathbf{y})$, for which

$$\begin{aligned}
&G(\mathbf{w}, \mathbf{y}_1) > G(\mathbf{w}, \mathbf{y}_2) > \cdots > G(\mathbf{w}, \mathbf{y}_K), \\
&G(\mathbf{w}, \mathbf{y}_i) > G(\mathbf{w}, \mathbf{y}), \quad \forall i = 1, \ldots, K;\ \forall \mathbf{y} \in \mathcal{Y}_2
\end{aligned} \tag{14}$$

From the properties of $\Phi(\cdot)$ it follows that $0 \le G(\mathbf{w}, \mathbf{y}) \le 1$. This gives rise to the following bounds on BER:

$$\begin{aligned}
&\frac{1}{2^L}\left(\sum_{\forall \mathbf{y}_i \in \mathcal{Y}_1} G(\mathbf{w}, \mathbf{y}_i) + \sum_{\forall \mathbf{y}_i \in \mathcal{Y}_2} 0\right) \\
&\qquad < P_E(\mathbf{w}) < \frac{1}{2^L}\left(\sum_{\forall \mathbf{y}_i \in \mathcal{Y}_1} G(\mathbf{w}, \mathbf{y}_i) + \sum_{\forall \mathbf{y}_i \in \mathcal{Y}_2} 1\right)
\end{aligned} \tag{15}$$

yielding

$$\frac{1}{2^L} \sum_{\forall \mathbf{y}_i \in \mathcal{Y}_1} G(\mathbf{w}, \mathbf{y}_i) < P_E(\mathbf{w}) < \frac{1}{2^L}\left(\sum_{\forall \mathbf{y}_i \in \mathcal{Y}_1} G(\mathbf{w}, \mathbf{y}_i) + |\mathcal{Y}_2|\right), \tag{16}$$

where $|\mathcal{Y}_2|$ denotes the cardinality of $\mathcal{Y}_2$.

Remark. In this case, $\mathbf{y}$ is subject to uniform distribution, thus the left hand side of (15) can be very tight if K is chosen reasonably high. At the same time, the upper bound tends to be loose as the terms G in $\mathcal{Y}_2$ have rather inaccurate upperbounds. In Section 4.2, we will evaluate the tightness of the bound based on the $(K+1)$th dominant sample.

The most "harmful" sequence, denoted by $\mathbf{y}_1$ in $\mathcal{Y}$ (i.e., the sequences beginning with -1) is $\mathbf{y}_1 = [-1, \operatorname{sgn}\{q_1\}, \operatorname{sgn}\{q_2\}, \ldots, \operatorname{sgn}\{q_L\}]$, since $\mathbf{q}^T\mathbf{y}_1 = -1 + \sum_{k=1}^{L} |q_k| = -1 + \mathrm{PD}(\mathbf{q})$, which indicates $\mathbf{y}_1$ to be the absolute dominant term in the summation equation (12).

In order to determine the dominant terms that form the set $\mathcal{Y}_1$, let us introduce the following notation: let $\mathbf{i} \in \mathcal{Z}^L$ be an index array pointing to the different elements of $\mathbf{q}$, where

$$\begin{aligned} i_1 &= \arg\min_j \left| q_j \right| \quad j \neq 0, \\ i_k &= \arg\min_j \left| q_j \right| \quad j \neq 0,\ i_1, i_2, \ldots, i_{k-1}. \end{aligned} \tag{17}$$

Note that i_j points to the jth smallest element of $\mathbf{q}$ in absolute value. The extension of index array i_j for $D \neq 0$ will be given in Section 4.3. The second dominant term $\mathbf{y}_2$ can be deduced from $\mathbf{y}_1$ by changing the sign of the component $\mathbf{y}_1(i_1)$, because in this case $\mathbf{q}^T\mathbf{y}_2 = -1 + \sum_{k=1, k\neq i_1}^{L} |q_k| - |q_{i_1}| = -1 + \mathrm{PD}(\mathbf{q}) - |q_{i_1}|$, where $|q_{i_1}|$ is the smallest possible value for decreasing the PD.

Applying the same reasoning, the first K largest terms can be given as follows:

(1) $\mathbf{y}_1 = [-1, \operatorname{sgn}\{q_1\}, \operatorname{sgn}\{q_2\}, \ldots, \operatorname{sgn}\{q_L\}]$; $\mathcal{C}_1 = \{\}$;

(2) $\mathbf{y}_2 = \mathbf{y}_1$ and change the sign of the component $\mathbf{y}_2(i_1)$; $\mathcal{C}_2 = \{i_1\}$;

(3) $\mathbf{y}_3 = \mathbf{y}_1$ and change the sign of the component of $\mathbf{y}_3(i_2)$; $\mathcal{C}_3 = \{i_2\}$;

(4) FOR $k = 4$ TO K find the index set $\mathcal{C}_k = \{j_1, \ldots, j_N\}$ for which $|q_{j_1}| + |q_{j_2}| + |q_{j_N}|$ is minimal, but $\mathcal{C}_k \neq \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{k-1}$ and $1 \leq N \leq L$;

(5) Form the set $\mathcal{Y}_1 = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K\}$ to be used in the lower bound given in (16).

It is easy to see that the case of $K = 1$ (when $\mathcal{Y}_1 = \{\mathbf{y}_1\}$) results in a cost function which has minimum value over the same coefficient vector as has the peak distortion. Increasing the value of K, the lower bound in (16) tends to the exact P_E and finally the case of $K = 2^L$ and $\mathcal{Y}_1 = \mathcal{Y}$ results in the exact minimum of BER.

We generally can derive an algorithm which identifies the dominant terms for any arbitrary $K = 2^M$ where $M < L$. The following procedure results in the first 4 dominant terms, which seems practically to be a good compromise between $K = 1$ (yielding the PD criterion) and $K = 2^L$ (for further details, see Section 4.2):

(1) $\mathbf{y}_1 = [-1, \operatorname{sgn}\{q_1\}, \operatorname{sgn}\{q_2\}, \ldots, \operatorname{sgn}\{q_L\}]$; $\mathcal{C}_1 = \{\}$;

(2) $\mathbf{y}_2 = \mathbf{y}_1$ and change the sign of the component $\mathbf{y}_2(i_1)$; $\mathcal{C}_2 = \{i_1\}$;

(3) $\mathbf{y}_3 = \mathbf{y}_1$ and change the sign of the component $\mathbf{y}_3(i_2)$; $\mathcal{C}_3 = \{i_2\}$;

(4) If $|q_{i_1}| + |q_{i_2}| \leq |q_{i_3}|$, then $\mathbf{y}_4 = \mathbf{y}_1$ and change the sign of the components $\mathbf{y}_4(i_1)$ and $\mathbf{y}_4(i_2)$; $\mathcal{C}_4 = \{i_1, i_2\}$; ELSE $\mathbf{y}_4 = \mathbf{y}_1$ and change the sign of the component $\mathbf{y}_4(i_3)$; $\mathcal{C}_4 = \{i_3\}$;

(5) Form the set $\mathcal{Y}_1 = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4\}$ to be used in the lower bound (16).

Unfortunately, calculating the set $\mathcal{C}_k$ to find the K largest term is of exponential complexity, as $\mathbf{q}^T\mathbf{y}$ must be calculated and arranged in monotone order for all possible $\mathbf{y}$.

4.1. Optimization of the Bound. The gradient $\mathbf{s}(\mathbf{w})$ of the lower bound in (16) is a truncated version of the gradient of the true BER (10), obtained by carrying out the summation over $\mathcal{Y}_1$ instead of $\mathcal{Y}$:

$$\begin{aligned} &\mathbf{s}(\mathbf{w}) \\ &= \frac{1}{2^L \sqrt{2\pi N_0 \left(\sum_{n=0}^{J} w_n^2\right)^3}} \\ &\quad \times \sum_{\mathbf{y}\in\mathcal{Y}_1} \exp\left(\frac{-\left(\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l\right)^2}{2N_0 \sum_{n=0}^{J} w_n^2} \right) \\ &\quad \cdot \left[\left(\sum_{n=0}^{J} w_n^2\right) \cdot \left(\sum_{l=i}^{M+i} h_{l-i} y_l\right) - w_i \left(\sum_{n=0}^{J} w_n \sum_{l=n}^{M+n} h_{l-n} y_l\right) \right]. \end{aligned} \tag{18}$$

Using the gradient, the following adaptive algorithm can be used for weight optimization

$$w_i(k+1) = w_i(k) - \gamma s_i(\mathbf{w}). \tag{19}$$

Of course, one can use variable step size $\gamma(k)$ in algorithm (19) to improve the speed of convergence. For example, the Armijo rule [15] can be applied to speed up the convergence. However, simulations showed no improvement by applying this rule. Another problem with this method is its high complexity (the gradient has to be evaluated several times). On the other hand, we may introduce a heuristically chosen step size, such as $\gamma(k) = \gamma_0/\sqrt{k}$. The convergence of (19) by using this step size this algorithm is guaranteed by the Kushner-Clark theorem (for more details, see [16]). In the simulation section, the improvement of convergence achieved by variable step size method is also illustrated.

4.2. A Numerical Example for Calculating the New Bound. For the sake of better understanding of the algorithm developed for finding the dominant terms, a numerical example will be given as follows. Let us take the following overall impulse response function

$$\mathbf{q} = [q_0, q_1, \ldots, q_L] = [1, 0.4, 0.08, -0.26, 0.15, -0.06], \tag{20}$$

where $L = 5$. At first, the index array $\mathbf{i}$ will be defined containing the indices of the smallest, second smallest, and so forth. elements of $|\mathbf{q}|$ (omitting q_0), respectively:

$$\mathbf{i} = [5, 2, 4, 3, 11]. \tag{21}$$

The information sequence resulting in maximum distortion, can be calculated as

$$\begin{aligned}\mathbf{y}_1 &= [-1, \operatorname{sgn}\{q_1\}, \operatorname{sgn}\{q_2\}, \ldots, \operatorname{sgn}\{q_L\}] \\ &= [-1, +1, +1, -1, +1, -1],\end{aligned} \tag{22}$$

yielding

$$c_1 = \mathbf{q}^T\mathbf{y}_1 = -1 + 0.04 + 0.08 + 0.26 + 0.15 + 0.06 = -0.05. \tag{23}$$

The information sequence $\mathbf{y}_2$ can be derived from $\mathbf{y}_1$ changing the sign of the element $\mathbf{y}_1(i_1) = \mathbf{y}_1(5)$:

$$\mathbf{y}_2 = [-1, +1, +1, -1, +1, +1], \tag{24}$$

resulting in

$$c_2 = \mathbf{q}^T\mathbf{y}_2 = -1 + 0.04 + 0.08 + 0.26 + 0.15 - 0.06 = -0.17, \tag{25}$$

and $\mathbf{y}_3$ comes from $\mathbf{y}_1$ by changing the sign of the element $\mathbf{y}_1(i_2) = \mathbf{y}_1(2)$

$$\mathbf{y}_3 = [-1, +1, -1, -1, +1, -1], \tag{26}$$

yielding

$$c_3 = \mathbf{q}^T\mathbf{y}_3 = -1 + 0.04 - 0.08 + 0.26 + 0.15 + 0.06 = -0.21. \tag{27}$$

To determine $\mathbf{y}_4$, one can evaluate the sum $|q_1| + |q_5| = 0.14$. Since it is smaller than $|q_{i_3}| = |q_4| = 0.15$, $\mathbf{y}_4$ can be derived from $\mathbf{y}_1$ by changing the sign of the elements $\mathbf{y}_1(2)$ and $\mathbf{y}_1(5)$:

$$\mathbf{y}_4 = [-1, +1, -1, -1, +1, +1], \tag{28}$$

resulting in

$$c_4 = \mathbf{q}^T\mathbf{y}_4 = -1 + 0.04 - 0.08 + 0.26 + 0.15 - 0.06 = -0.33. \tag{29}$$

If one calculates all possible terms up to c_{32} (which equals -1.95), the bit error probability can be calculated using (12) as

$$P_E = \frac{1}{2^5}\sum_{i=1}^{32}\Phi\left(\frac{c_i}{\sigma}\right). \tag{30}$$

In Figure 1, one can see the terms $\Phi(c_i/\sigma)$ versus c_i/σ for two different σ values. In this figure, logarithmic scale on the vertical axis was used, and smaller values than 10^{-14} were omitted. Note that the number of negligible terms depends on the noise level. This is explained by the fact that increasing the noise level, the samples c_i will fall into a region where $\Phi()$ decreases rather fast. The difference between the smallest and largest terms is of 10^{83} magnitudes. Furthermore, one may note that there are no c_i samples on the positive side which proves that channel can indeed be equalized and in this case there are no local minima as proven in [13].

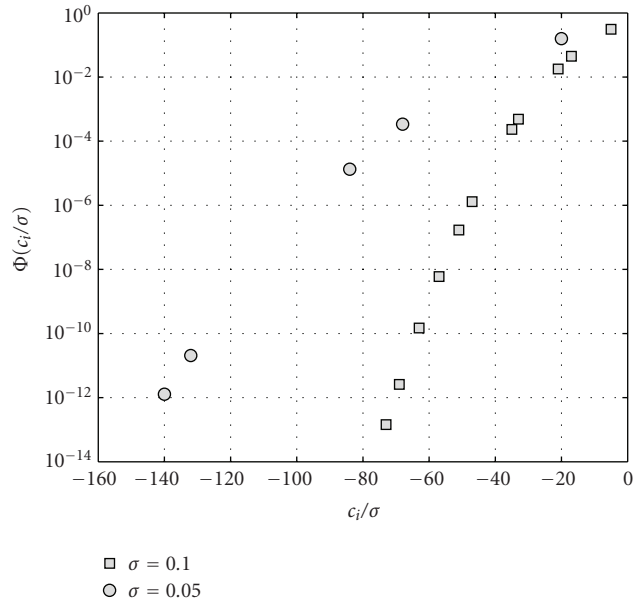

Figure 1: Visualisation of the dominant terms: $\Phi(c_i/\sigma)$ versus c_i/σ versus for two different σ values.

In order to derive a method to identify K, first let us express the bit error probability as a sum of two expressions

$$P_E = \frac{1}{2^L}\sum_{\forall \mathbf{y}_i \in y_1} G(\mathbf{w}, \mathbf{y}_i) + \frac{1}{2^L}\sum_{\forall \mathbf{y}_i \in y_2} G(\mathbf{w}, \mathbf{y}_i), \tag{31}$$

where the second term can be upper bounded by using the $(K+1)$th dominant sample. In this way, one obtains the following bound:

$$P_E < \frac{1}{2^L}\sum_{\forall \mathbf{y}_i \in y_1} G(\mathbf{w}, \mathbf{y}_i) + \frac{2^L - K}{2^L}\sum_{\forall \mathbf{y}_i \in y_2} G(\mathbf{w}, \mathbf{y}_{K+1}). \tag{32}$$

Since the sharpness of (32) depends on the value of the $(K+1)$th dominant sample (it becomes sharp if the value of the $(K+1)$th dominant sample is small), this expression can be used to estimate the number of dominant samples K to be used for giving efficient bound on BER. If the bound using the $(K+1)$th dominant sample drops below a predefined value r then the number of samples K needed to approximate BER can be obtained as follows:

$$rP_E \geq \frac{2^L - K}{2^L} G(\mathbf{w}, \mathbf{y}_{K+1}), \tag{33}$$

where P_E can be substituted by its approximation using the first K dominant samples. Figure 2 analyzes the accuracy of the bound obtained by the dominant samples.

Figure 2 shows two curves belonging to the SNR = 10 and 20 dB, respectively. From this figure, it can be seen that if $r = 0.05$ and SNR = 20 dB then the necessary number of

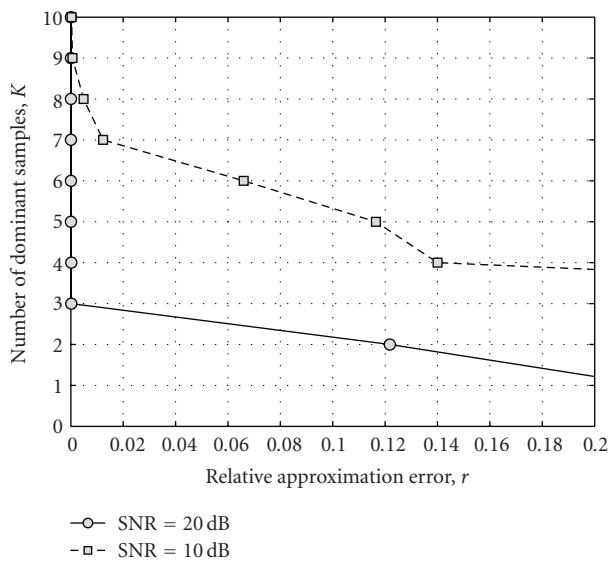

Figure 2: The necessary number of sample as function of the relative approximation error.

samples $K = 3$. This necessary sample number will increase with respect to the decrease of SNR (in the case of SNR = 10 dB the number of samples is seven). This is in line with the reasoning detailed above.

4.3. Handling the Channel Delay. If there is some delay D in the overall channel impulse response function $\mathbf{q}$, a more efficient equalization can be carried out by the decision rule given in (3). The cost function based on the lower bound in (16) has only to be slightly modified in order to handling the delay parameter D. Since in this case (instead of the first element) the Dth element of $\mathbf{q}$ must be set to -1, hence the index array used for calculating the dominant terms have to be changed to

$$\begin{aligned} i_1 &= \arg\min_j \left| q_j \right| \quad j \neq D, \\ i_k &= \arg\min_j \left| q_j \right| \quad j \neq D, i_1, i_2, \ldots, i_{k-1}. \end{aligned} \tag{34}$$

5. Obtaining Channel-State Information

In order to run the proposed algorithm, channel state information is needed (the channel impulse function h_i appears in expression (9)). There are plenty of real time adaptive channel identification algorithms [18] which provide fast and simple channel state information by using a training sequence $\{(y_n, x_n), n = 1, \ldots, N\}$ where $y_n, n = 1, \ldots, N$ is a transmitted binary sequence known at the receiver and $x_n, n = 1, \ldots, N$ is the corresponding received sequence.

We identify the channel estimator with an adaptive FIR filter, the coefficients of which are updated as follows:

$$g_j(n+1) = g_j(n) - \Delta\left(x_n - \sum_{i=0}^{M} g_i y_{n-j}\right) y_{n-j}. \tag{35}$$

This algorithm minimizes the mean square error between the unknown channel impulse response function $h_i, i = 1, \ldots, M$ and the FIR filter coefficients g_i $i = 0, 1, 2, \ldots, M$. Here x_k denotes the received sequence at the output of the channel estimator. Parameters g_i in algorithm (35) converge to the true channel impulse response function h_i in mean square (and in probability) if the degree of the FIR filter is larger than the channel impulse response (overmodeling).

It is noteworthy that the adaptive channel identifier (35) converges rather fast to the true cannel-state because of the narrow eigenvalue-spectrum of the underlying matrices (for further details see [1]). Hence, the combination of identification and equalization can provide real time solutions for low BER communication.

6. Numerical Results

In this section, a detailed performance analysis is given in which the bit error probability achieved by the different equalization methods are compared with each other.

6.1. Channel Characteristics and Channel-State Information. The channel distortion can be modeled by a tapped delay-line model (see Section 2). If the WSSUS (Wide Sense Stationary Uncorrelated Scattering) assumption is made, then the channel coefficient h_i are uncorrelated, and Gaussian distributed. In the project COST 207 [19], several wideband propagation models were proposed for the practical realization of both hardware and software simulators in the context of GSM systems for different classes of environments (an other set of models is ITU-R models for third-generation cellular systems). These models are generally described by power delay profiles from which the discrete time equivalent can be derived.

The simulations were performed on three different discrete channels representing multipath propagation derived from the power delay profiles of the above mentioned models. The corresponding channel characteristics are given by their impulse response as follows: $\mathbf{h}^{(1)} = [1; 0.6; -0.3]^T$, $\mathbf{h}^{(2)} = [1.0000; 0; -0.81; 0.42]^T$, $\mathbf{h}^{(3)} = [1; 0.6; -0.45]^T$.

Note that channel $\mathbf{h}^{(1)}$ has the minimum-phase, while $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(3)}$ have the nonminimum-phase property. The equalization of nonminimum-phase channels is difficult, because these channels have zeros outside the unit circle, and hence, the inverse of the channel has poles outside the unit circle.

6.2. Performance Analysis. In this section we numerically investigate the BER with respect to SNR and we also analyze the convergence properties of the equalization algorithms. The abbreviations used in the figures are as follows:

(i) TGS—True Gradient Search (for details see (11));

(ii) LISIx—our algorithm with x dominant terms in the approximation;

(iii) LMS—Least Mean Square algorithm;

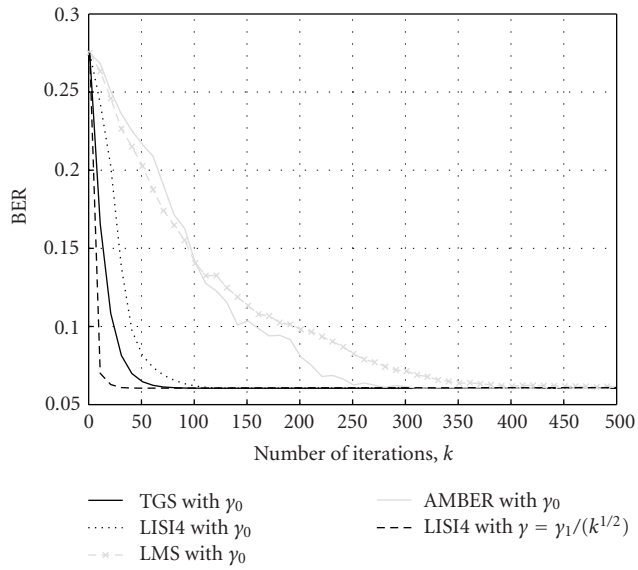

Figure 3: $P_E(\mathbf{w}(k))$ versus k for channel $\mathbf{h}^{(3)}$ in the case of 3 equalizer coefficients, $\gamma_0 = 0.0025$ and $\gamma_1 = 0.025$.

(iv) MMSE—Off-line calculated Minimum Mean Square Error solution

(v) NOEQ—BER without any equalizer;

(vi) AMBER—Adaptive Minimum Bit Error Rate algorithm [6].

As far as the channel-state information is concerned, we assumed no channel-state information to be available at the receiver side, thus channel equalization was preceded by an adaptive channel identifier algorithm given in (35). In all simulations the delay parameter D used in the decision rule (3) was set by exhaustive search. The step size of the gradient-descent-type algorithms was set empirically. The experiments show that the attained BER is not too sensitive to the value of the step size, while the convergence speed is highly dependent on this value as described below. Furthermore the value of the step size depends on the SNR as well, since the error surface tends to be more complicated as SNR increases [6].

6.2.1. Convergence Analysis. Figure 3 demonstrates the convergence properties of the equalization algorithms. One can see, that the TGS algorithm converges rather fast; however, in each step the exponential summation have to be calculated. Algorithm LISI4 exhibits similarly fast convergence but in each step it only needs to evaluate the function $\Phi(\cdot)$ for only some dominant arguments, which results in a considerable decrease in complexity.

Figure 3 also demonstrates that algorithm LISI4 with fix $\gamma = \gamma_0$ step size will yield slower convergence than TGS. However, the convergence speed can be increased by modifying the step size in each step according to the rule $\gamma = \gamma_0/\sqrt{k}$ [16], which indicates that investigation on the step size can also improve the convergence.

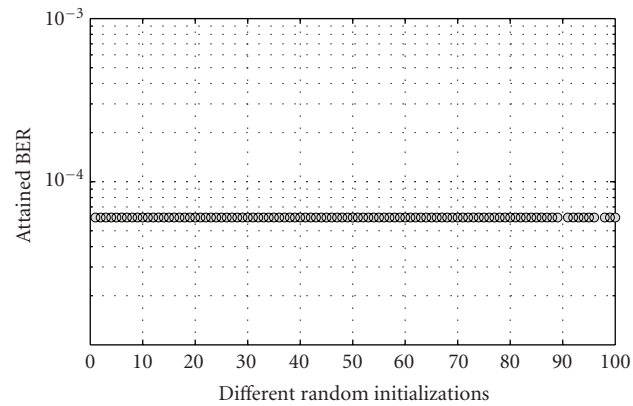

Figure 4: Final BER for 100 different runs of TGS in the case of randomly chosen initial state, channel $\mathbf{h}^{(1)}$ and SNR = 24 dB.

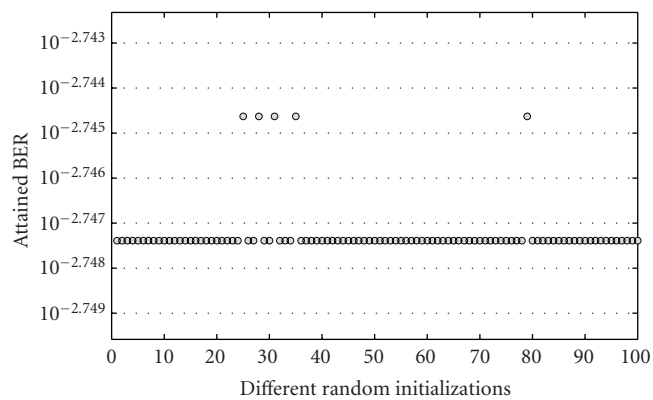

Figure 5: Final BER for 100 different runs of TGS in the case of randomly chosen initial state, channel $\mathbf{h}^{(2)}$ and SNR = 24 dB.

It is proven in [13] that in the case of nonminimum-phase channels the minimum BER error surface is convex, and hence it has only one global minimum, but in the case of nonminimum-phase channels there are local minima [6] in which gradient descent type algorithms can get stuck. In order to detect the chance to getting stuck into local minima for the channel models used in our experiments, the TGS algorithm was randomly initialized with 100 different values, and the attained BER after convergence was depicted in Figures 4, 5, and 6. These figures demonstrate that in the case of minimum-phase channel there is almost no chance of converging to local minima, while in the case of nonminimum phase channels a 5–10 percent of convergence to local minima has been detected. The problem of getting stuck into local minima can be minimized by "good" initialization, for example, iterating the equalizer weights the minimum BER algorithm can be started from an initial weight vector obtained from the MMSE solution.

6.2.2. BER versus SNR. In Figures 7–9, the BER versus SNR achieved by the classical and by the new algorithm are depicted. One can note a sharp improvement in performance achieved by the new algorithm derived from the minimum BER strategy, especially in the case of nonminimum-phase

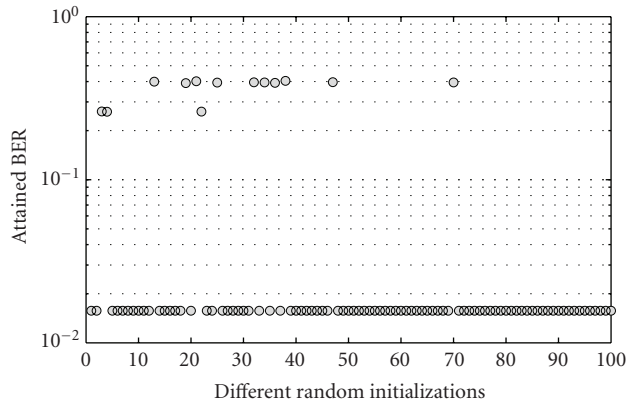

Figure 6: Final BER for 100 different runs of TGS in the case of randomly chosen initial state, channel $\mathbf{h}^{(3)}$ and SNR = 24 dB.

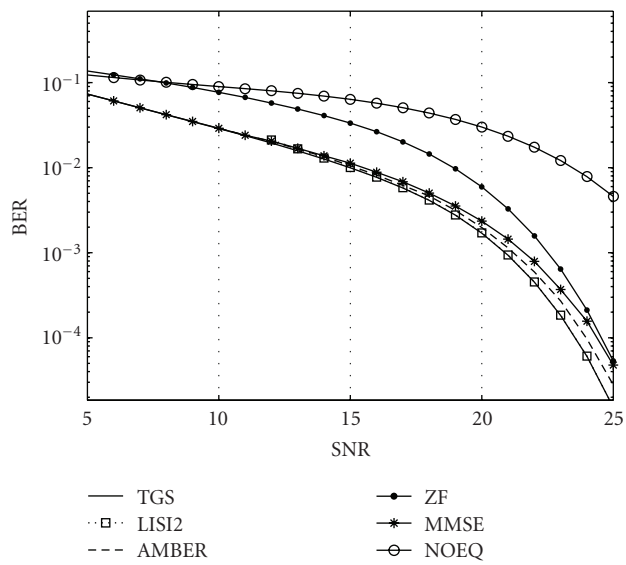

Figure 7: BER versus SNR performance of the different algorithms for channel $\mathbf{h}^{(1)}$ in the case of 3 equalizer coefficients, $D = 0$.

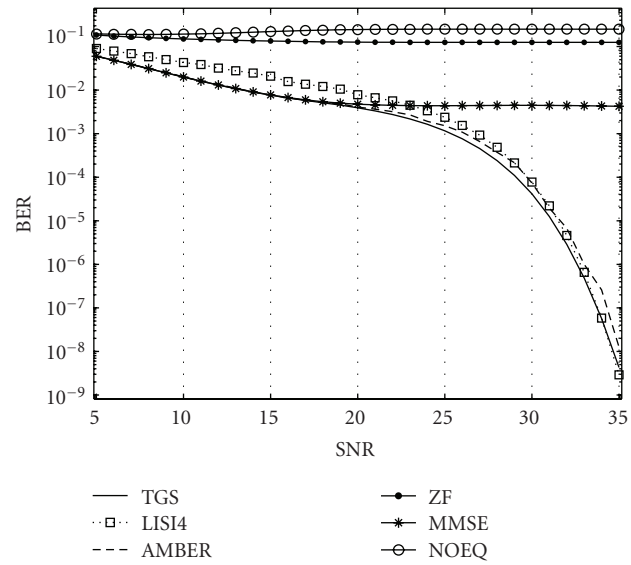

Figure 8: BER versus SNR performance of the different algorithms for channel $\mathbf{h}^{(2)}$ in the case of 3 equalizer coefficients, $D = 3$.

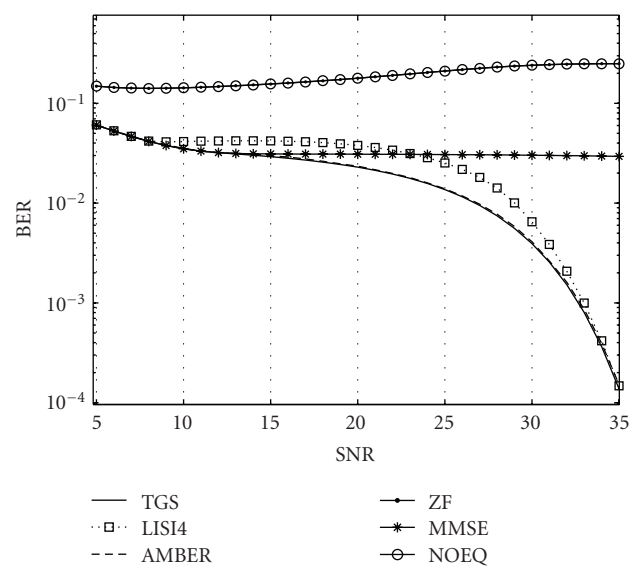

Figure 9: BER versus SNR performance of the different algorithms for channel $\mathbf{h}^{(3)}$ in the case of 6 equalizer coefficients, $D = 2$.

channels (such as $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(3)}$, see Figures 8 and 9) and good SNR circumstances. The LISI algorithm using the dominant terms performs very close to the exact minimum BER solution in the case of good SNRs, since increasing the SNR decreases the number of dominant terms, hence the bound will be sharper. On the other hand, its advantage is its low complexity against TGS (11). The AMBER algorithm introduced in performs very close to the TGS, but converges much slower than the TGS and LISI methods.

7. Conclusions

In this paper, a novel channel equalizer algorithm has been developed based on approximating the BER by dominant terms. Due to the simplicity of this approximation, a fast equalization algorithm can be obtained, the performance of which falls close to optimum. Since this approximation needs channel state information, the equalizer is preceded by an adaptive channel identifier. The combined convergence of channel identification and the new bound-based equalization is still much faster than the convergence of other algorithms (e.g., LMS, AMBER [6]). The operational complexity of the new algorithm is also smaller than TGS (for details see (11)). The new method yielded better performance than the traditional ZF and MMSE equalizer algorithms on standard wireless channels. These benefits make the new algorithm suitable for real time applications.

References

[1] J. Proakis, *Digital Communications*, McGraw–Hill, New York, NY, USA, 4th edition, 2001.

[2] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[3] F. E. Glave, "An upper bound on the probability of error due to intersymbol interference for correlated digital signals," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 356–363, 1972.

[4] S. O. Shimbo and M. Celebiler, "The probability of error due to intersymbol interference and gaussian noise in digital communication systems," *IEEE Transactions on Communications*, vol. 19, no. 2, pp. 113–119, 1971.

[5] E. Shamash and K. Yao, "On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion," in *Proceedings of the 10th International Conference on Communications (ICC '74)*, pp. 1–5, Minneapolis, Minn, USA, 1974.

[6] C.-C. Yeh and J. R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Transactions on Communications*, vol. 48, no. 7, pp. 1226–1235, 2000.

[7] S. Chen, B. Mulgrew, and L. Hanzo, "Least bit error rate adaptive nonlinear equalisers for binary signalling," *IEE Proceedings: Communications*, vol. 150, no. 1, pp. 29–36, 2003.

[8] P. ChandraKumar, P. Saratchandran, and N. Sundararajan, "Minimal radial basis function neural networks for nonlinear channel equalisation," *IEE Vision Image and Signal Processing*, vol. 147, no. 5, pp. 428–435, 2000.

[9] S. Chen, A. K. Samingan, B. Mulgrew, and L. Hanzo, "Adaptive minimum-BER linear multiuser detection," *IEEE Transactions on Signal Processing*, vol. 49, no. 6, pp. 1240–1247, 2001.

[10] A. Dua, U. B. Desai, and R. K. Mallik, "Minimum probability of error-based methods for adaptive multiuser detection in multipath DS-CDMA channels," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 939–948, 2004.

[11] A. Hjorungnes and M. Debbah, "Minimum BER FIR receiver filters for DS-CDMA systems," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 4, pp. 2287–2291, St. Louis, Mo, USA, December 2005.

[12] D. Gesbert, "Robust linear MIMO receivers: a minimum error-rate approach," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2863–2871, 2003.

[13] X. F. Wang, W.-S. Lu, and A. Antoniou, "Constrained minimum-BER multiuser detection," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2903–2909, 2000.

[14] J. Gunther and T. Moon, "Minimum bayes risk adaptive linear equalizers," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4788–4799, 2009.

[15] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, 2nd edition, 1999.

[16] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New York, NY, USA, 4th edition, 2001.

[17] V. O. K. Li and J. A. Silvester, "Performance analysis of networks with unreliable components," *IEEE Transactions on Communications*, vol. 32, no. 10, pp. 1105–1110, 1984.

[18] R. Steele and L. Hanzo, *Mobile Radio Communications*, Wiley-IEEE Press, New York, NY, USA, 2nd edition, 1999.

[19] M. Failli, "Digital land mobile radio communications COST 207," Tech. Rep., European Commission, 1989.